



**Investigation of the structure and function of the  
metal transporter CzcB from *Thermus  
thermophilus* and the molecular chaperone trigger  
factor from *Psychrobacter frigidicola* by X-ray  
crystallography**

**PhD Thesis**

**Małgorzata Wrońska**

Supervisor: Dr. J. Gerard Wall, Department of Microbiology

Submitted to the National University of Ireland, Galway

September 2014

---

## Table of Contents

|   |           |
|---|-----------|
| Declaration.....  | VI        |
| Acknowledgements .....  | VII       |
| Abstract.....   | VIII      |
| Abbreviations .....   | IX        |
| List of Figures.....  | XI        |
| List of Tables .....  | XIV       |
| <b>1. INTRODUCTION .....</b>  | <b>1</b>  |
| <b>1.1. Importance of proteins and the structure-function relationship.....</b> | <b>1</b>  |
| <b>1.2. Introduction of proteins of interest in this work.....</b>              | <b>2</b>  |
| <b>1.2.1. Cadmium-zinc resistance protein B (CzrB).....</b>                     | <b>2</b>  |
| <b>1.2.2. Trigger factor (TF).....</b>  | <b>7</b>  |
| <b>1.3. Methods of solving protein structures .....</b>                         | <b>10</b> |
| <b>1.4. Obtaining protein for crystallographic studies .....</b>                | <b>12</b> |
| <b>1.4.1. Expression systems.....</b>   | <b>15</b> |
| <b>1.4.2. <i>E. coli</i> as a host .....</b>                                    | <b>18</b> |
| <b>1.5. Principles of protein crystallisation .....</b>                         | <b>19</b> |
| <b>1.5.1. Crystallisation methods .....</b>                                     | <b>21</b> |
| <b>1.6. Studying membrane proteins.....</b>                                     | <b>23</b> |
| <b>1.6.1. Detergents .....</b>  | <b>24</b> |
| <b>1.6.2. Crystallisation of membrane proteins.....</b>                         | <b>29</b> |
| <b>1.7. Basics of crystallography .....</b>                                     | <b>34</b> |
| <b>1.8. Overview of this project .....</b>                                      | <b>35</b> |
| <b>2. MATERIALS AND METHODS.....</b>  | <b>37</b> |
| <b>2.1. Materials.....</b>  | <b>37</b> |
| <b>2.1.1. Suppliers .....</b>   | <b>37</b> |

---

|            |   |    |
|------------|---|----|
| 2.1.2.     | General Reagents .....                            | 38 |
| 2.1.3.     | PCR Oligonucleotides .....                        | 38 |
| 2.1.4.     | Bacterial Strains .....                           | 39 |
| 2.1.5.     | Plasmids .....                                    | 40 |
| 2.1.6.     | Antibiotics .....                                 | 41 |
| 2.1.7.     | Antibodies .....                                  | 41 |
| 2.1.8.     | Molecular Weight Markers .....                    | 41 |
| 2.2.       | Methods .....                                     | 41 |
| 2.2.1.     | Competent Cell Preparation .....                  | 41 |
| 2.2.1.1.   | Electrocompetent Cells .....                      | 41 |
| 2.2.1.2.   | Chemically Competent Cells .....                  | 42 |
| 2.2.2.     | DNA Manipulation .....                            | 42 |
| 2.2.2.1.   | PCR .....   | 42 |
| 2.2.2.2.   | Plasmid Construction .....                        | 43 |
| 2.2.2.2.1. | Cloning of non-dimerising CzcB .....              | 43 |
| 2.2.2.3.   | Digestion .....                                   | 43 |
| 2.2.2.4.   | Ligation .....                                    | 43 |
| 2.2.2.5.   | Transformation .....                              | 43 |
| 2.2.3.     | DNA extraction .....                              | 44 |
| 2.2.3.1.   | DNA extraction from agarose gels .....            | 44 |
| 2.2.3.2.   | Plasmid extraction from bacterial cells .....     | 44 |
| 2.2.3.3.   | Genomic DNA extraction from bacterial cells ..... | 44 |
| 2.2.4.     | Agarose Gel Electrophoresis .....                 | 44 |
| 2.2.5.     | Protein Expression .....                          | 45 |
| 2.2.5.1.   | Expression of His-tagged TFPf .....               | 45 |
| 2.2.5.2.   | Expression of His-tag free TFPf .....             | 45 |
| 2.2.5.3.   | Expression of CzcB .....                          | 45 |

---

|             |  |    |
|-------------|--|----|
| 2.2.6.      | Protein extraction from bacterial cells.....                           | 46 |
| 2.2.6.1.    | Soluble protein extraction .....                                       | 46 |
| 2.2.6.2.    | Extraction of CzcB using a thermal treatment.....                      | 46 |
| 2.2.6.3.    | Cell fractionation to collect cell membranes for CzcB extraction ..... | 46 |
| 2.2.6.4.    | Extraction of CzcB from membranes.....                                 | 47 |
| 2.2.7.      | SDS PAGE .....   | 47 |
| 2.2.7.1.    | Coomassie staining.....  | 47 |
| 2.2.7.2.    | Western Blotting .....   | 47 |
| 2.2.8.      | Protein purification.....  | 47 |
| 2.2.8.1.    | Purification of TFPf for structural studies .....                      | 48 |
| 2.2.8.1.1.  | Immobilised Metal Affinity Chromatography (IMAC).....                  | 48 |
| 2.2.8.1.2.  | Removal of His-tag with TEV protease.....                              | 48 |
| 2.2.8.1.3.  | Ion-Exchange Chromatography .....                                      | 48 |
| 2.2.8.1.4.  | Size Exclusion Chromatography .....                                    | 48 |
| 2.2.8.2.    | Purification of CzcB for structural studies.....                       | 49 |
| 2.2.8.2.1.  | Immobilised Metal Affinity Chromatography (IMAC).....                  | 49 |
| 2.2.8.2.2.  | Size Exclusion Chromatography .....                                    | 49 |
| 2.2.9.      | Protein analysis .....   | 49 |
| 2.2.9.1.    | Protein concentration and quantification .....                         | 49 |
| 2.2.9.2.    | Circular dichroism (CD) .....  | 50 |
| 2.2.9.3.    | TETRA detector array .....   | 50 |
| 2.2.9.4.    | Thermofluor-based stability assay .....                                | 50 |
| 2.2.10.     | Crystallisation .....  | 51 |
| 2.2.10.1.   | Crystallisation of TFPf .....  | 51 |
| 2.2.10.2.   | Crystallisation of CzcB .....  | 51 |
| 2.2.10.2.1. | Vapour-diffusion .....   | 51 |
| 2.2.10.2.2. | Lipid Cubic Phase (LCP) .....  | 52 |

---

|  |           |
|--|-----------|
| 2.2.10.2.3. Data collection .....  | 52        |
| 2.2.11. Construct design for HTP cloning and expression at Oxford Protein<br>Production Facility (OPPF) .....  | 52        |
| 2.2.11.1. Primer design.....   | 52        |
| 2.2.11.2. High-throughput PCR.....   | 52        |
| 2.2.11.3. Purification of PCR products by magnetic beads .....   | 53        |
| 2.2.11.4. In-fusion ligation .....   | 53        |
| 2.2.11.5. Plasmid purification.....  | 53        |
| 2.2.11.6. Small scale expression screening .....   | 53        |
| 2.2.11.7. Small scale expression analysis .....  | 54        |
| <b>3. RESULTS OF STRUCTURAL STUDIES OF C<sub>zr</sub>B FROM <i>THERMUS</i><br/><i>THERMOPHILUS</i>.....</b>  | <b>55</b> |
| 3.1. Production of C <sub>zr</sub> B for structural studies.....   | 55        |
| 3.1.1. Construct design and expression optimisation .....  | 55        |
| 3.2. Extraction of C <sub>zr</sub> B for structural studies.....   | 57        |
| 3.2.1. Extraction of C <sub>zr</sub> B by thermal treatment .....  | 58        |
| 3.3. Purification of C <sub>zr</sub> B .....   | 60        |
| 3.3.1. Immobilised Metal Affinity Chromatography purification of C <sub>zr</sub> B.....  | 60        |
| 3.4. Initial crystallisation trials of C <sub>zr</sub> B.....  | 63        |
| 3.5. High-throughput approach to C <sub>zr</sub> B crystallisation in Membrane Protein<br>Laboratory (MPL) at Diamond Light Source (DLS), Oxfordshire, UK..... | 66        |
| 3.5.1. TETRA-detector based analysis of protein.....   | 66        |
| 3.5.2. Crystallisation trials at MPL .....   | 69        |
| 3.5.3. Thermofluor-based stability assay .....   | 69        |
| 3.5.4. Cell lysis using the cell disruptor .....   | 72        |
| 3.5.5. Effect of different buffers on dimerisation .....   | 75        |
| 3.5.6. Extraction of C <sub>zr</sub> B from membranes after fractionation .....  | 79        |
| 3.5.7. Crystal analysis by X-ray .....   | 82        |

---

|         |  |            |
|---------|--|------------|
| 3.5.8.  | Purification of CzrB in DDM and DM.....  | 85         |
| 3.5.9.  | Screening of the effect of detergents on CzrB by analytical HPLC.....  | 91         |
| 3.6.    | Construction of CzrB variants and preliminary expression screening at OPPF .   | 93         |
| 3.7.    | Crystallisation of CzrB expressed from pOPINFCzrB .....  | 97         |
| 3.8.    | Crystallisation of CzrB expressed from pOPINEHALO7CzrB .....   | 102        |
| 4.      | <b>RESULTS FROM STRUCTURAL STUDIES OF TRIGGER FACTOR FROM<br/><i>PSYCHROBACTER FRIGIDICOLA</i> .....</b>               | <b>106</b> |
| 4.1.    | Taking crystallographic approach to investigate TFPf.....  | 106        |
| 4.2.    | Expression of full-length TFPf for crystallisation studies.....  | 107        |
| 4.3.    | Extraction and purification of trigger factor for crystallisation .....  | 109        |
| 4.4.    | Purification of full-length TFPf.....  | 110        |
| 4.4.1.  | Immobilised Metal Affinity Chromatography .....  | 110        |
| 4.4.2.  | Ion-exchange chromatography (IEC) .....  | 112        |
| 4.5.    | Preliminary crystallisation trials of trigger factor from <i>P. frigidicola</i> .....                                  | 114        |
| 4.6.    | Optimisation of preliminary hits.....  | 116        |
| 4.6.1.  | Optimisation of needle crystals .....  | 117        |
| 4.6.2.  | Seeding of needle crystals to improve their quality .....  | 119        |
| 4.7.    | Size Exclusion Chromatography (SEC).....   | 120        |
| 4.8.    | Expression, extraction and purification of trigger factor with a removable N-<br>terminal His-tag (HIS-TEV-TFPf) ..... | 124        |
| 4.9.    | Analysis of trigger factor by circular dichroism.....  | 126        |
| 4.10.   | Crystallisation of HIS-TEV-TFPf .....  | 127        |
| 4.10.1. | Stability check on ThermoFluor using SybrOrange.....   | 128        |
| 4.10.2. | Crystallisation of HIS-TEV-TFPf at MPL .....   | 129        |
| 4.11.   | Construction of TF variants and preliminary expression screening at OPPF ...   | 130        |
| 5.      | <b>CONCLUSIONS AND FUTURE WORK.....</b>  | <b>139</b> |

---

## Declaration

I, the PhD candidate, Małgorzata Wrońska, certify that the Thesis is all my own work and that I have not obtained a degree in this University or elsewhere on the basis of any of this work.

Signed: ----- Date: -----

## Acknowledgements

The completion of this project and thesis became reality with the invaluable help from many people to whom I am greatly grateful and want to say a sincere thank you.

I owe special thanks to:

My supervisor Dr. Gerard Wall, for his continuous support, encouragement and patience throughout the project.

The technical staff in Microbiology, who were always ready to provide support in matters big and small: Caroline, Mike, Ann, Katrina and especially Maurice, whose kindness and helpfulness make him a living legend.

All lecturers in Microbiology for encouraging a great working atmosphere in the department.

My fantastic lab-mates: Claire, Iain, Chus and Akshay for their advice and support with everyday lab “turbulances” but also for all the good times and laughs throughout the years.

All postgraduate students in Microbiology, members of the tag rugby team and everyone who, throughout the years, became dear friends, in particular Aileen, Claire, Elizabeth, Iain, Maria, Kate and Paul.

Colleagues that I have met during my travels related to the project, who provided support and direction in progressing this work:

Everyone at MPL, especially Isabel, Mat, Momi, So, Indran, Nien-Jen, Tien, James F. and James B.

Everyone at OPPF, especially Louise, Ray, Heather, Anil

Tewik, Sylvain, Olga and Mohamed at the University of Limerick

My family, especially my parents, my granny and my sister, who have always supported my ventures, decisions and plans.

My friends for always being there for me during good times and bad times, to celebrate or to drown my sorrows, namely Claire, Dave, Maria, Rob, Tracy, Aine, Ewelina, Ania, Didi, Gosia, Filipek and Lala. And also Rafal who already brought few of my computers back from the brink of death.

Finally, I want to thank my husband and best friend Damien Ridge, without whom I would not have arrived at the end of this project in one piece. There are no words to describe how grateful and lucky I am to have you in my life.

## Abstract

### **Investigation of the structure and function of the metal transporter CzrB from *Thermus thermophilus* and the molecular chaperone trigger factor from *Psychrobacter frigidicola* by X-ray crystallography**

**Candidate: Małgorzata Wrońska**

A crystallographic approach was taken to investigate the structure and function of two proteins of interest previously isolated in our research group and identified as having potential application in improving the production of recombinant proteins in *Escherichia coli*. The first, CzrB from the thermophilic bacterium *Thermus thermophilus*, is a membrane-bound protein involved in the transport of metal ions to ensure cell homeostasis. The second target, trigger factor (TF) from the psychrophilic bacterium *Psychrobacter frigidicola* (TFP<sub>f</sub>), is a molecular chaperone that assists the folding of newly synthesised polypeptides in the cell. Here, we present a systematic study of the expression and characterisation of both proteins with a view towards obtaining diffraction-quality crystals for analysis by X-ray crystallography.

Initial investigation based on previously established protocols led to successful expression of both targets in *E. coli*. CzrB was found to occur as a heterogeneous population of largely monomeric and dimeric molecules, however, which is unsuitable for crystallisation. The use of chromatographic size exclusion did not separate monomeric and dimeric CzrB, while the use of a range of maltoside (DDM, UDM, DM,)- and glucoside (OG, NG)-based detergents also failed to commit CzrB to an exclusively dimeric form. Well-shaped pentagonal crystals were obtained with CzrB purified in UDM exhibited a high detergent content, while CzrB purified in DDM formed triangular crystals that diffracted to 7 Å but were identified as the commonly IMAC-co-purified *E. coli* membrane protein AcrB. Construction and screening of a series of new CzrB constructs containing different fusion tags led to identification of two candidate constructs for large-scale production, while a non-dimerising CzrB construct design is also being investigated to reduce heterogeneity of protein preparations.

The expression of C-terminal His-tagged TFP<sub>f</sub> resulted in a high yield of recombinant protein, however the extraction and purification of the protein was characterised by extensive degradation. Optimisation using ion exchange chromatography yielded protein of higher purity and quality, leading to needle-like crystals in a wide range of crystallisation conditions. The morphology did not improve upon pH screening or the inclusion of additives or seeding. Design and generation of a diverse panel of full-length TFP<sub>f</sub> constructs and component domains with a variety of fusion partners and recombinant tags identified differing degradation patterns during expression and extraction, indicating that constructs lacking the C-domain may be less susceptible to degradation. A full-length trigger factor construct that exhibited reduced degradation and is potentially suitable for crystallisation screening was also identified.

Overall, significant progress was made in solubilisation and purification of the two proteins of interest expressed in *E. coli*. While extensive crystallisation investigations did not yield diffraction quality crystals of CzrB or TFP<sub>f</sub>, alternative approaches for future studies have been identified to overcome the bottlenecks of protein heterogeneity and degradation, respectively, in order to allow higher quality crystals to be obtained. The work has also generated important insights into the differences between working with membrane and soluble proteins, and illustrates the limitations of low-throughput versus high-throughput crystallisation approaches.

## Abbreviations

**3C** - Human Rhinovirus 3C protease  
**9.9 MAG** - monoolein, 1-(9Z-octadecenoyl)-rac-glycerol  
**CDF** - cation diffusion facilitator  
**CD** - circular dichroism  
**CMC** - critical micelle concentration  
**CMP** - N-[4-(7-diethylamino-4methyl-3-coumarinyl) phenyl] maleimide  
**Cymal-6** - 6-cyclohexyl-1-hexyl- $\beta$ -D-maltoside  
**CVs** - column volumes  
**DDM** - n-dodecyl- $\beta$ -D-maltopyranoside  
**DG** - n-decyl- $\beta$ -D-glucopyranoside  
**DLS** - dynamic light scattering, also Diamond Light Source  
**DM** - n-decyl-  $\beta$ -D-maltopyranoside  
**GFP** - green fluorescent protein  
**GPCR** - G- protein coupled receptor  
**GST** - Glutathione-S-transferase  
**HIS** - histidine  
**HTG** - n-heptyl- $\beta$ -D-thioglucopyranoside  
**HTP** - high-throughput  
**IMAC** - immobilised metal affinity chromatography  
**IPTG** - isopropyl- $\beta$ -D-thiogalactoside  
**LCP** - lipid cubic phase  
**LDAO** - n-dodecyl-N, N-dimethylamine-N-oxide  
**MBP** - maltose binding protein  
**MW** - molecular weight  
**MWCO** - molecular weight cut-off  
**NG** - n-nonyl- $\beta$ -D-glucopyranoside  
**OD** - optical density  
**OG** - n-octyl- $\beta$ -D-glucopyranoside  
**PAGE** - polyacrylamide gel electrophoresis  
**PCR** - polymerase chain reaction  
**PDB** - Protein Data Bank  
**PDC** - protein detergent complex  
**PEG** - polyethylene glycol  
**PPIase** - peptidylprolyl isomerase  
**RALS** - right angle light scattering  
**RBS** - ribosome-binding site  
**RI** - refractive index  
**SDS** - sodium dodecylsulfate  
**SEC** - size exclusion chromatography  
**SelMet** - seleno-methionine  
**SRP** - signal recognition particle  
**SUMO** - small ubiquitin-like modifier  
**TMD** - transmembrane domain  
**Triton-X-100** -  $\alpha$ -[4-(1,1,3,3-tetramethylbutyl)phenyl]-hydroxy-polyoxy-1,2-ethanediyl  
**TRX** - thioredoxin

**UDM** - n-undecyl-  $\beta$ -D-maltopyranoside

**UV** - ultraviolet light

---

**List of Figures**

|   |    |
|---|----|
| Figure 1. The proposed model of the mechanism of action of <i>T. thermophilus</i> CzrB in <i>E. coli</i> (Cherezov <i>et al.</i> , 2008).....   | 6  |
| Figure 2. Structure of TF from <i>E. coli</i> . .....   | 9  |
| Figure 3. Illustration of trigger factor binding to ribosome through L23. ....  | 10 |
| Figure 4. SEC translocon and YidC insertase system responsible for insertion of membrane proteins into the inner membrane (Xia and Dalbey, 2008).....   | 13 |
| Figure 5. Source of proteins used in structural studies between 1985 and 2010. ....   | 18 |
| Figure 6. Schematic representation of a protein crystallisation phase diagram (Chayen, 2005). .....   | 20 |
| Figure 7. Schematic of the most commonly used crystallisation methods. ....   | 23 |
| Figure 8. Illustration of solubilisation steps of membrane proteins by detergents. ....   | 28 |
| Figure 9. Schematic illustration of membrane protein packing in the three types of crystal lattice.....   | 30 |
| Figure 10. Temperature-composition phase diagram of monoolein, the most common lipid used in LCP crystallisation (Cherezov, 2011).....  | 33 |
| Figure 11. Example of protein complex diffraction pattern, which represents an orderly array of reflections (spots). ....   | 35 |
| Figure 12. The general pipeline involved in solving a protein structure.....  | 55 |
| Figure 13. Immunoblot analysis of CzrB production in autoinducing media ZYP-5052. ....  | 57 |
| Figure 14. Immunoblot analysis of CzrB extraction using thermal treatment in the presence of 1% DDM.....  | 59 |
| Figure 15. SDS-PAGE (A) and Western Blot (B) analysis of CzrB purification in the presence of 0.9% UDM.....   | 62 |
| Figure 16: SDS-PAGE analysis of buffer exchange (A) and concentration of CzrB (B). ....   | 63 |
| Figure 17. Crystals observed in UDM-purified CzrB crystallisation trials carried out at 20°C. ....  | 65 |
| Figure 18. TETRA detector analysis of CzrB in 0.1% UDM. ....  | 68 |
| Figure 19. Summary of the most significant $T_m$ shifts using the ThermoFluor assay. ....   | 71 |
| Figure 20. Coomassie-stained SDS-PAGE analysis of IMAC purification of CzrB extracted using a cell-disruptor and high-temperature treatment. ....   | 73 |
| Figure 21. SEC-based purification of CzrB in the presence of 0.1% UDM. ....   | 74 |
| Figure 22. SDS-PAGE analysis of fractions from SEC purification of CzrB in 50 mM Bis-Tris pH 6.8, 0.1% UDM (A) and analysis of an eluted sample before and after thermal denaturation (B). ....         | 75 |
| Figure 23. SDS-PAGE analysis of the influence of buffer composition on CzrB dimerisation. ....  | 76 |
| Figure 24. Structure of cytoplasmic domain of CzrB with bound zinc.....   | 77 |
| Figure 25. UV chromatogram representing SEC analysis of CzrB purification on SD75.....  | 78 |
| Figure 26. SDS-PAGE analysis of the Superdex75 SEC purification. ....   | 79 |
| Figure 27. SDS-PAGE analysis of comparison of purification of CzrB extracted by thermal extraction from cell extracts (A) and by cell fractionation followed by extraction at low temperature (B). .... | 80 |

---

|   |     |
|---|-----|
| Figure 28. SEC purification analysis of CzrB extracted using a thermal treatment (A) and at 4°C (B). .....  | 81  |
| Figure 29. TETRA detector analysis of CzrB, prepared without thermal treatment, in the presence of 0.1% UDM.....  | 82  |
| Figure 30. Preliminary crystallisation hits observed within 2 days in drops set up by LCP (A) and vapour-diffusion (B and C). .....                     | 83  |
| Figure 31. Examples of crystals grown from UDM-purified CzrB at 20°C during the HTC crystallisation screening .....                                     | 84  |
| Figure 32. An example of a diffraction pattern from UDM-purified CzrB crystals obtained. ....   | 85  |
| Figure 33. CzrB purification in the presence of 0.03 % DDM (A) and 0.26 % DM (B). .....   | 86  |
| Figure 34. SEC analysis of CzrB purified in the presence of DDM (A) and DM (B). .....   | 87  |
| Figure 35. Triangular crystals observed in crystallisation trials set up with CzrB purified in DDM. ....  | 88  |
| Figure 36. Crystals obtained in the control well during the optimisation screen. ....   | 89  |
| Figure 37. MicroMesh loop from MiTeGen with harvested crystal ready for the X-ray analysis inside the beam I04-1. ....                                  | 89  |
| Figure 38. The diffraction pattern obtained from X-ray analysis of crystals grown in vapour diffusion from DDM-purified CzrB.....                       | 90  |
| Figure 39. Determination of monodispersity of CzrB in different detergents using HPLC. .  | 92  |
| Figure 40. Amino acid sequence of CzrB. ....  | 93  |
| Figure 41. Summary of the CzrB expression screen carried out using constructs generated at the OPPF. ....   | 94  |
| Figure 42. SDS-PAGE analysis of preliminary expression screen of CzrB constructs in <i>E. coli</i> . ....   | 95  |
| Figure 43. SDS-PAGE analysis of the efficiency of various detergents in solubilising CzrB. ....   | 98  |
| Figure 44. SDS-PAGE analysis of IMAC purification in the presence of OG (A), DM (B) or NG (C). ....   | 99  |
| Figure 45. SEC analysis of DDM-purified CzrB, with exchange into buffer containing 0.2% NG.....   | 100 |
| Figure 46. SEC analysis of UDM-purified CzrB with exchange into buffer containing 0.2% NG.....  | 101 |
| Figure 47. SDS-PAGE analysis of IMAC purification and His-tag removal from CzrB (A) and SEC purification of CzrB in the presence of 0.09% UDM (B). .... | 103 |
| Figure 48. SDS-PAGE analysis of size exclusion purification of CzrB expressed in pOPINHALO7. ....   | 104 |
| Figure 49. Western blot analysis of yield of TFPf production after up to 16 h induction. ..   | 108 |
| Figure 50. Western blot analysis of TFPf present in expressed and extracted samples.....  | 109 |
| Figure 51. SDS-PAGE Coomassie Blue analysis of the summary of TFPf purification by IMAC. ....   | 111 |
| Figure 52. Western Blot analysis of IMAC-purified TFPf. ....  | 112 |
| Figure 53. Coomassie Blue-stained SDS-PAGE analysis of purification of TFPf by Ion Exchange Chromatography (IEC).....                                   | 113 |
| Figure 54. Result of TFPf initial crystallisation screening. ....   | 115 |
| Figure 55. Needle crystals obtained through optimisation of the initial hits. ....  | 116 |

---

|  |     |
|--|-----|
| Figure 56. Examples of needle crystals obtained at various pH values during optimisation screening of <i>TFPf</i> .....  | 117 |
| Figure 57. Crystals obtained with MDP as precipitant during the additive screen. ....  | 118 |
| Figure 58. SDS-PAGE analysis of purified and concentrated <i>TFPf</i> showing extensive degradation of the protein. ....   | 120 |
| Figure 59. Further purification of trigger factor to increase purity. ....   | 122 |
| Figure 60. SDS-PAGE analysis of (A) <i>TFPf</i> purified by SEC, (B) followed by vivaspin concentration. ....  | 122 |
| Figure 61. Western blot analysis of expression of HIS-TEV- <i>TFPf</i> expressed by autoinduction in ZYP-5052 at 25°C for 16 h (A); expressed and extracted HIS-TEV- <i>TFPf</i> (B). .... | 124 |
| Figure 62. Purification of HIS-TEV- <i>TFPf</i> . SDS-PAGE analysis of Coomassie stained (A) and immunoblotting (B). ....  | 125 |
| Figure 63. CD spectrum of HIS-TEV- <i>TFPf</i> in 10 mM sodium phosphate pH 7.0. ....  | 126 |
| Figure 64. Secondary structure elements (% of amino acids) of TF in different CD spectrum ranges. ....   | 127 |
| Figure 65. Crystals that appeared in the JCSG screen (Hampton Research). ....  | 128 |
| Figure 66. Summary of the melting curve demonstrating a shift in $T_m$ for HIS-TEV- <i>TFPf</i> upon changing buffer, indicating greater protein stability. ....                           | 129 |
| Figure 67. Summary of the <i>TFPf</i> expression screen carried out using constructs generated at OPPF. ....   | 132 |
| Figure 68. SDS-PAGE analysis of preliminary expression screen followed by IMAC purification of <i>TFPf</i> constructs expressed in <i>E. coli</i> Lemo21. ....                             | 133 |
| Figure 69. SDS-PAGE analysis of preliminary expression screen followed by purification by IMAC of <i>TFPf</i> constructs expressed in <i>E. coli</i> Rosetta pLacI. ....                   | 134 |
| Figure 70. Summary of representatives of each construct identified for crystallisation trials. ....  | 136 |
| Figure 71. Immunoblot analysis of expression of <i>TFPf</i> constructs. ....   | 136 |
| Figure 72. Immunoblot analysis of TF full-length constructs expressed in several different vector formats. ....  | 138 |

**List of Tables**

|  |     |
|--|-----|
| Table 1. Detergent groups and representative examples with their characteristics ..... | 26  |
| Table 2. Detergents used during optimising purification of CzcB (Kolaj, 2008).....     | 61  |
| Table 3. Conditions used to investigate the stability of CzcB.....                     | 70  |
| Table 4. Summary of conditions used to set up trials of TFPf.....                      | 123 |

## 1. INTRODUCTION

### 1.1. Importance of proteins and the structure-function relationship

Throughout the 1990s and early 2000s when a major objective of researchers focused on sequencing the genomes of a wide variety of organisms, the focus has moved in more recent years to investigating protein function. Faced with deciphering enormous collections of sequencing data, the challenge remains to turn this information into knowledge about the roles of proteins at a molecular level. Proteins are a versatile class of macromolecules involved in diverse cellular processes in every living cell. By playing critical roles in replicating and transcribing DNA, and in producing, processing and secreting polypeptides, proteins control cell division, metabolic reactions and the flow of information and molecules into and out of the cell. The first step in realising the importance of proteins in living systems is recognising the diversity of their function: complex macromolecules possess distinctive, structure-encoded dynamic properties that are a consequence of complex interatomic interactions and are fully dependent on the molecule's three-dimensional structure.

The modern field of structural biology originated in Cambridge, UK in the 1950's when the first three-dimensional protein structure was determined. It was a structure of myoglobin from sperm whale revealed by English biochemist John Kendrew and Austrian-born molecular biologist Max Perutz. However, the use of crystallisation in biological studies dates back to a century earlier, with the first publication in 1840 on crystallites, which were in fact haemoglobin crystals, observed in blood samples. In the 19<sup>th</sup> Century, crystallisation often occurred accidentally through empirical methods and was used as a means of purification and assessment of a protein's activity by biochemists and physiologists. The development of X-ray crystallography in the 1930's, driven by Bragg and von Laue, changed the face of crystallisation, leading to a rise in advances in crystallisation procedures to produce crystals suitable for structure determination and, in parallel, in crystallographic methods to image the component macromolecules. Gradually, this was followed by more focused efforts in acquiring proteins from their natural sources. Considering that the goal was to develop methods for structure determination, however, proteins targeted were typically those available in

abundance and unproblematic to crystallise. Ultimately, understanding protein function at an atomic level has been revolutionised by the use of high-resolution X-ray crystallography, which has facilitated the understanding of the roles of proteins in areas such as critical cellular processes, in pathogenesis, and as drug targets in the pharmaceutical industry.

## **1.2. Introduction of proteins of interest in this work**

### **1.2.1. Cadmium-zinc resistance protein B (CzrB)**

CzrB, an integral membrane cation transporter, was identified in the host group during a phage display-based screen for chaperone-like molecules encoded in a *Thermus thermophilus* genomic library (Spada *et al.*, 2002). *T. thermophilus* is a Gram-negative rod-shaped, non-motile, thermophilic bacterium which was first isolated from a Japanese thermal spa (Oshima and Imahori, 1974). Since then, *T. thermophilus* has demonstrated considerable biotechnological potential due to its heat resistance, and the ability to exploit its potential was facilitated greatly by completion of its genome sequence in 2004 (Henne *et al.*, 2004). Upon isolation in our group, the *czrB* gene sequence was found to be homologous to cation efflux transporters from a variety of organisms, including *Staphylococcus aureus* (Kuroda *et al.*, 1999, Xiong and Jayaswal, 1998), *Saccharomyces cerevisiae* (Palmiter and Findley, 1995) and *Ralstonia eutropha* (Nies *et al.*, 1989) most of which were termed Czr (for cadmium-zinc resistance) or Czc (for cadmium-zinc-cobalt resistance).

CzrB belongs to the CDF (cation diffusion facilitator) family of membrane transporters that have been identified in eukaryotes, prokaryotes, archaea and plants to date (Montanini *et al.*, 2007). Members of this protein family play a role in the homeostasis of a wide range of bivalent metal cations, including control of intracellular cation accumulation, oxidative stress and cation resistance, regulation of signal transduction cascades, and reciprocally regulating protein synthesis and protein degradation (Bruinsma *et al.*, 2002, Kobayashi *et al.*, 1996, Ellis *et al.*, 2004). The broad spectrum of cations known to be transported by CDF family members, including Zn, Fe, Co, Mn and Cd, allows their participation in diverse cellular processes throughout different compartments of the cell. The majority of CDF family membrane transporters share a common architecture, which is characterised by six

putative transmembrane domains (TMDs), usually highly hydrophobic, and a soluble, cytosolic C-terminal domain (Paulsen and Saier, 1997). Some CDF family proteins exhibit differences in their number of transmembrane helices and size such as Msc2 from *S. cerevisiae* which contains 12 predicted TMDs (Li and Kaplan, 2001), and the human protein ZnT5 with 15 TMDs (Kambe *et al.*, 2002). The amphipathic transmembrane helices I, II, V, VI, which are thought to be involved in metal transport, occur within the most conserved regions of the protein family. Many CDF family members also contain a histidine-rich motif at their N- or C-terminus, which is thought to be cytoplasmic and to potentially act as metal-binding domain (Yang *et al.*, 2009).

Transmembrane protein topology prediction carried out in the group envisaged CzrB to be a membrane protein with a molecular mass of 31,233 and containing six putative membrane-spanning  $\alpha$ -helices, with the four N-terminal spanning domains exhibiting highly hydrophobic characteristics. The *Thermus* CzrB, however, does not contain a histidine-rich region commonly associated with zinc-binding and it also lacks the classic metal-binding motif CXXC (Spada *et al.*, 2002). The *czrB* gene exhibits a high (72%) GC content typical of thermophilic species and its predicted amino acid sequence reveals a high arginine and low serine content, which are also common characteristics of thermostable proteins (Kumar *et al.*, 2000).

Zinc, like most heavy metals, is toxic at high concentrations and yet essential for cellular metabolism. It is also unique among essential metals in having a dual role in structural stability and as a catalytic cofactor in the cell:  $Zn^{2+}$  stabilizes “zinc finger” proteins involved in binding DNA and RNA in eukaryotes (Seneque *et al.*, 2009, Berg and Godwin, 1997), while it is a cofactor of many hydrolytic enzymes like proteases, phosphatases, esterases and deacetylases (Auld, 2001). Due to the opposing activities of many metals *in vivo*, the metal homeostasis machinery is finely tuned to tightly control their intracellular levels to cover all biochemical needs but without reaching antimicrobial levels. According to Outten and O’Halloran (2001), the total zinc content of *E. coli*, known also as the zinc quota, is tightly regulated. The intracellular concentration of free zinc is very low, suggesting that the majority of zinc in the cell exists in a complexed form. Exposure of *E. coli* to excess zinc in the environment triggers transcription of at least four CDF family zinc

transport pumps, including the CzrB homologue ZitB, to ensure these levels are successfully maintained (Outten, 2001).

Previous research in our group revealed that *E. coli* cells expressing *czrB* exhibited significantly higher resistance to  $Zn^{2+}$  and a slightly increased resistance to  $Cd^{2+}$  than control cells (Spada *et al.*, 2002). There was no increase in cobalt resistance, however, suggesting the existence of an independent cobalt efflux system in *T. thermophilus*, as previously reported with *S. aureus* (Xiong and Jayaswal, 1998) and *S. cerevisiae* (Conklin *et al.*, 1994, Palmiter and Findley, 1995), which contain CzrB homologues CzrB and ZRC1, respectively. Furthermore, *czrB* does not appear to be co-regulated with contiguous genes to function as a CzrCB<sub>2</sub>A efflux pump in *T. thermophilus*, as reported in a number of other species (Kuroda *et al.*, 1999, van der Lelie *et al.*, 1997). It is unclear if equivalents of these additional efflux-related genes exist in *T. thermophilus* in a non-contiguous location in the genome or if their absence indicates the presence of a less complex *czrB* system in *T. thermophilus*, possibly due to thermophilic species generally possessing smaller genomes than their mesophilic homologues (Spada *et al.*, 2002).

Additionally, *E. coli* cells containing the full-length *czrB* gene, as well as its constituent cytoplasmic domain ( $\Delta czrB$ ), exhibited a delayed cell lysis phenomenon leading to higher cell densities during production of a recombinant antibody fragment, effectively resulting also in higher yields of the recombinant protein (Spada *et al.*, 2002). As the production of recombinant proteins in *E. coli* often leads to a stress-induced lysis of the host cells, Spada and co-workers (2002) suggested that this cell lysis could arise from overloading of the translocation machinery during expression, leading to aggregation of polypeptides at the membrane and blocking of the metal extrusion system. Nevertheless, the observed positive effect of the presence of the full-length and partial *czrB* genes on host *E. coli* physiology during protein expression provided additional interest in understanding the molecular basis of its activity in its recombinant host.

Although the molecular mechanism for these effects remains unknown, sequence similarities between CzrB and an anti-stress response factor in the  $\sigma^E$ -mediated pathway in *E. coli*, RseA, were noted in subsequent studies of CzrB in our group (Kolaj, 2008). RseA is an anti- $\sigma^E$  factor whose N-terminus (in particular

residues 32-53) binds to  $\sigma^E$  to inhibit gene expression and which is inactivated when a stress signal is communicated through the signal transduction pathway from the periplasm (Campbell *et al.*, 2003). This inactivation leads to activation of the  $\sigma^E$ -mediated stress response (De Las Peñas *et al.*, 1997). Sequence analysis revealed similarities between the C-terminal cytoplasmic tail of CzrB and the similarly cytoplasmic N-terminal domain of RseA, in particular in a conserved RGDTP motif within a region of extensive contact with  $\sigma^E$ . This may indicate a possible activity of CzrB in *E. coli* of delaying cell lysis by mimicking the activity of the native RseA (Kolaj, 2008).

Furthermore, the structure of the cytoplasmic domain of CzrB ( $\Delta$ CzrB) has been solved in this group in an attempt to better understand the structure of the full-length protein and the basis for its delayed cell lysis effect during its co-expression with recombinant proteins in *E. coli* (Cherezov *et al.*, 2008).  $\Delta$ CzrB is a 92-amino-acid extramembranal cytoplasmic domain that is expressed from Met200 in the native *czrB* gene sequence (Hofer *et al.*, 2007). The structure of  $\Delta$ CzrB was determined to a resolution of 1.7 Å in the apo and zinc-bound forms and confirmed that the protein exists as a homodimer, which was consistent with *in vitro* studies including size exclusion chromatography,  $^1\text{H}$  NMR and small-angle X-ray scattering (Cherezov *et al.*, 2008). A hypothesis was proposed based on the structure that the cytoplasmic domain could exist independently in the cell and function as a metallochaperone, thus regulating the zinc-transporting activity of the full-length membrane-bound protein. A structure of the full-length CzrB protein was also proposed, based on molecular replacement using the  $\text{Zn}^{2+}/\text{H}^+$  antiporter YiiP (Lu and Fu, 2007) and concluded that the protein dimer functions as a completely integrated transport device equipped with sensor, actuator and transport parts. The CzrB protomer in its apo form takes the shape of a flattened ellipsoid, while the homodimer is created through hydrogen bonds and hydrophobic contacts, leading to an inverted V-shaped appearance with membrane-anchored protomers splayed apart (Figure 1 (Cherezov *et al.*, 2008)). Upon zinc binding the two splayed-apart monomers snap together along the flattened surface of ellipsoids to bring the arms of the inverted V closer. Each protomer has been associated with four zinc ions; however it is likely that only three are physiologically relevant, whereas the fourth

plays a role in crystallogenesis and crystal packing (Cherezov *et al.*, 2008). In the model proposed by Cherezov, zinc binding triggers a dramatic conformational change in the protein, which facilitates chaperone docking followed by release of a zinc ion for transport through the transmembrane part of protein. Despite the recent availability of the structure of YiiP (Lu and Fu, 2007), a close homologue of CzcB, more structural studies are required to better understand the mechanism of action of this group of important transporters from the CDF family.



**Figure 1. Figure removed due to copyright.**

While most descriptions of CDF proteins to date emphasise the putative role of the cytoplasmic domain in ion transport, Russell and co-workers (2012) reported that MmCDF3 from the marine bacterium *Maricaulis maris*, a 23-kDa homologue of CzcB, interacts selectively with zinc and cadmium despite lacking the C-terminal domain. This would imply that this CDF family member may possess an alternative means of acquiring metal ions from the cytoplasm in order to efflux them from the cell (Russell and Soulimane, 2012).

### 1.2.2. Trigger factor (TF)

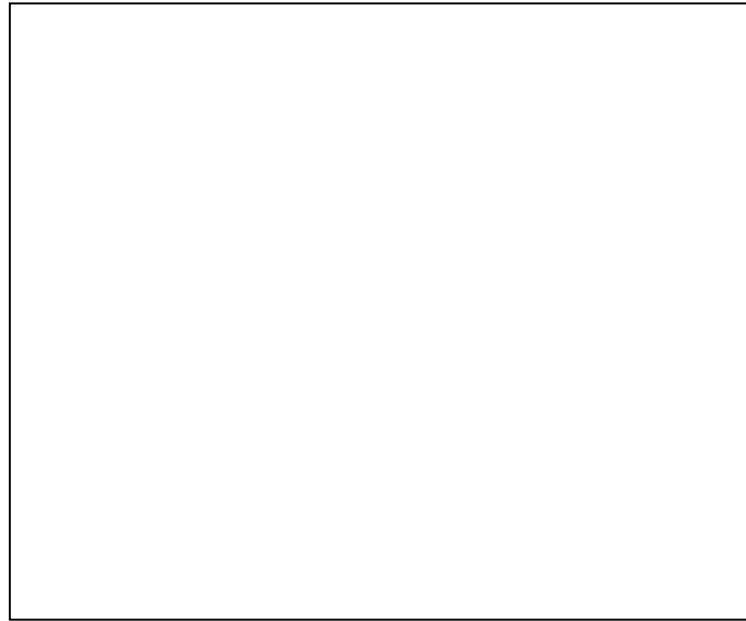
The second protein of interest in the present study is the molecular chaperone trigger factor from *Psychrobacter frigidicola*, a bacterium first isolated from ornithogenic soil collected from an Adéli penguin colony in eastern Antarctica (Bowman *et al.*, 1996). *P. frigidicola* is a cold-adapted, slightly halophilic, Gram-negative, non-motile, coccoid organism with a tendency to occur in pairs or tetrads (Bowman, 2006).

Trigger factor is involved in protein folding in the cell. While many polypeptides fold into well-defined three-dimensional, active states *in vitro*, the vast majority of proteins require the assistance of molecular chaperones to avoid misfolding and aggregation in the challenging and overcrowded cellular environment *in vivo* (Ellis and Minton, 2006). Molecular chaperones assist in *de novo* folding of polypeptides and maintain the pool of existing protein molecules in their native states by counteracting factors such as high protein concentrations and mis- or partially-folded proteins with exposed hydrophobic domains that make proteins susceptible to non-native interactions or aggregation (Maier *et al.*, 2005). Ribosome-associated chaperones are highly conserved among prokaryotes and eukaryotes, and in bacteria include trigger factor (TF), however, the chaperone-assisted folding also include DnaK/DnaJ/GrpE and GroEL/GroES complexes (Hartl, 2002). The first evidence that TF and DnaK act as chaperones came from Deuerling and co-workers in 1999 when they demonstrated that knockout of both genes led to faults in protein folding, resulting in large-scale aggregation of proteins and a reduction in cell viability (Deuerling *et al.*, 1999).

Trigger factor is the first chaperone to come into contact with newly synthesised polypeptides. It provides assistance for proteins to fold into defined three-dimensional structures with functional activity, rather than undergoing aberrant folding and aggregation processes and potentially leading to the formation of toxic species (Robin *et al.*, 2009). First identified in *E. coli* due to its binding of the precursor of outer membrane porin A (proOmpA), TF has also been identified and characterised in *Vibrio cholera* (Ludlam *et al.*, 2004), *Thermatoga maritima* (Martinez-Hackert and Hendrickson, 2007), and several psychrophilic organisms including *Pseudoalteromonas haloplanktis* (Médigue *et al.*, 2005, Piette *et al.*, 2010),

*Psychrobacter arcticus* (Zheng *et al.*, 2007) and *Shewanella livingstonensis* (Kawamoto *et al.*, 2007). Trigger factor may be a particularly important chaperone in cold-adapted species as it has been shown that it is the predominant up-regulated chaperone in *P. haloplanktis*, whereas DnaK and GroEL are down-regulated at 4°C (Piette *et al.*, 2011).

Previous characterisation of trigger factor from *P. frigidicola* (TFPf) in this research group revealed 58% homology between the *P. frigidicola* chaperone and its *E. coli* homologue. Glutaraldehyde cross-linking and fluorescence anisotropy analyses revealed that, unlike its well-characterised *E. coli* homologue which occurs in a monomer-dimer equilibrium, TFPf exists exclusively as a monomer in solution. Additionally, while the *Psychrobacter* TF shares the three putative domains found in the *E. coli* chaperone, TFPf does not exhibit the reduced efficiency of refolding of GAPDH (glyceraldehyde-3-phosphate dehydrogenase) observed at high chaperone concentrations with the *E. coli* molecule (Robin *et al.*, 2009). Furthermore, TFPf does not impede the export of co-expressed recombinant proteins to the periplasm at high TF concentrations, as observed with *E. coli* TF, which indicates potential usefulness of the *Psychrobacter* chaperone as a folding aid in cell-free recombinant protein expression systems (Robin *et al.*, 2009). *E. coli* trigger factor has been demonstrated to adopt a unique elongated structure described as a crouching dragon, with a ‘tail’ formed from the ribosome-binding amino-terminal domain, a ‘head’ represented by a *cis/trans* isomerase (PPIase) domain, ‘arms’ constructed by the carboxy-terminal region positioned between the N-terminal and PPIase domains, and connecting regions building up the ‘back’ of the dragon (Figure 2). The N-terminal ‘tail’ mediates binding of the monomeric TF to the large ribosomal subunit through L23 protein (Kramer *et al.*, 2002, Ferbitz *et al.*, 2004), which is thought to control the progression of freshly synthesised polypeptides through the ribosomal exit tunnel (Baram *et al.*, 2005). The contribution of the PPIase domain to the folding of newly synthesised chains remains unclear, though it may interact with longer polypeptide chains and act as an auxiliary substrate recognition site (Kramer *et al.*, 2002, Merz *et al.*, 2008). The centrally located C-terminal ‘head’ provides the primary binding site for the nascent polypeptide chain (Lakshmipathy *et al.*, 2007).



**Figure 2. Figure removed due to copyright.**

Recent advances in the study of trigger factor revealed its versatile nature and demonstrated that it interacts not only with nascent polypeptide chains as they exit the ribosomal tunnel, but also with polypeptides released into the cytosol and fully-formed proteins upon their assembly into larger complexes (Hartl *et al.*, 2011). It can also reverse premature folding (Hoffmann *et al.*, 2012), bind and protect partially-folded proteins and prevent unproductive interactions that might otherwise lead to stable misfolded states (O'Brien *et al.*, 2012, Mashaghi *et al.*, 2013). A significant milestone in understanding the basis of TF's multi-tasking ability in the cell was provided by a study from Saio and co-workers (2014) who demonstrated through NMR and isotope labelling approaches that TF creates a binding support for PhoA (alkaline phosphatase) containing four individual substrate-binding sites (see Figure 3) that can bind the polypeptide substrate in random order. The binding sites form hydrophobic pockets that shelter hydrophobic residues in the substrate peptide and the flexibility of their architecture allows interaction with varied amino acid sequence motifs, which helps to explain the degree of functional diversity observed in the chaperone (Saio *et al.*, 2014).



**Figure 3. Figure removed due to copyright.**

Conformational similarities between TF from *P. frigidicola* and *E. coli* include the N-terminal domain (TF<sub>Pf</sub> exhibits the conserved, distinctive TF sequence <sup>47</sup>GFRKGNVP which is essential for ribosome binding), and the PPIase catalytic site, in which several characteristic residues are conserved. However, even though the predicted secondary structure of TF<sub>Pf</sub> reveals a very similar organisation to that of TF<sub>Ec</sub>, their reported functional disparities (Robin *et al.*, 2009) imply differences in their tertiary structures. The distinct challenges associated with protein folding in cold-adapted species may have forced TF<sub>Pf</sub> to adapt in order to reduce physicochemical constraints caused by reduced molecular diffusion rates, increased cytoplasmic viscosity, lower biochemical reaction rates, weaker molecular interactions and the increased stability and solubility of toxic metabolites at low temperatures (Piette *et al.*, 2011).

Upregulation of TF was noted in psychrophilic organisms like *Pseudoalteromonas haloplanktis*, *Shewanella livingstonesis* and *Erythronium sibiricum* (Piette *et al.*, 2011). As GroEL is not cold-adapted and therefore becomes inefficient at low temperatures (Piette *et al.*, 2010), this may indicate a more extensive role for TF in psychrophilic than in mesophilic bacteria.

### **1.3. Methods for solving protein structures**

The rapid developments in identification, characterisation and determination of protein structures has led to an exponential increase in the number

of proteins of known structure since 1958, when the structure of myoglobin, the first protein structure, was revealed by Max Perutz and John Kendrew (Kendrew *et al.*, 1958). Since myoglobin, nearly 100,000 macromolecular structures and their assemblies have been solved and made available *via* the Protein Data Bank (PDB). The important depository of macromolecular structures at PDB ensures organised, convenient and free access to complete information about protein structures and the opportunity to share new developments in the area of structural studies (Berman *et al.*, 2000).

Currently, several techniques are employed in structure determination, including X-ray crystallography, nuclear magnetic resonance (NMR) and electron microscopy. The majority of protein structures deposited in the PDB have been determined by X-ray crystallography, which is the most favoured method in structure determination. A smaller portion of deposited structures was unravelled by NMR, while the number of membrane protein structures solved by electron crystallography, though still low, is increasing (PDB). The three techniques should be considered complementary to each other and together provide a more complete and efficient strategy in working out the structure of target proteins. In addition, while it is quite common to use all three techniques to obtain a detailed resolution of a molecule of interest at present, each technique presents a number of limitations. Moreover, there is a growing impact of single particle electron microscopy (EM) as a method of solving low to medium resolution structures. EM is a fast technique, which provides an image of protein in its native state with no upper limit in size of the macromolecule and is especially valuable for flexible and difficult to crystallise proteins.

The undeniable advantage of X-ray crystallography is the possibility to solve structures of molecules of high molecular weight, as well as protein complexes, whereas NMR is limited to molecules of less than 50 kDa in size. However, producing protein crystals that will diffract by X-ray to a resolution that allows structure determination is far from trivial. Moreover, NMR provides a structure of a soluble form of a protein that most likely represents its native state most accurately, while the conformation of protein molecules in crystals might alter due to the interactions caused by molecule packing. In fact, electron crystallography is so far

the only technique to study membrane proteins in their native environment (Stahlberg *et al.*, 2001); it also requires 2D crystals that are one unit cell thick, which can be an alternative route to success with proteins that do not grow sufficiently for X-ray crystallography. An example of a membrane protein that naturally produces 2D crystals is bacteriorhodopsin (Henderson *et al.*, 1990), a light-sensitive proton pump from Archaea, for which a native conformation has been determined by electron crystallography at an atomic resolution of 3.2 Å (Subramaniam and Henderson, 2000). This atomic model of bacteriorhodopsin revealed the shift in the conformation of the cofactor that is responsible for light-driven transport of protons through the membrane.

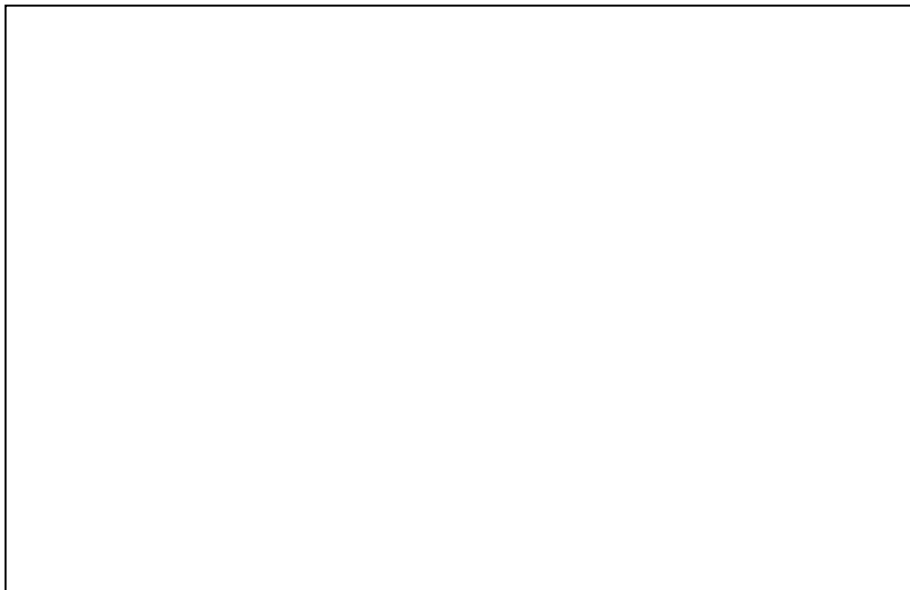
#### **1.4. Obtaining protein for crystallographic studies**

Due to a difficulty in accessing many proteins from their natural sources at sufficient levels for crystallisation, most proteins in structural studies are now produced using recombinant techniques rather than via purification from their native source. The main advantage of recombinant production over purification of proteins from their native sources is the increase in protein yield. In addition, however, purification can be greatly simplified by the incorporation of affinity tags during cloning, modification of proteins to alter properties such as, *e.g.*, their stability is theoretically possible, and, purely from the perspective of crystallography and phasing purposes, labelling with SelMet is much simpler with recombinant expression (Guerrero *et al.*, 2001, Zhao *et al.*, 2007, Walden, 2010). The difficulties associated with recombinant expression tend to arise from the scale of biodiversity in macromolecules, which significantly complicates the task of producing an active target protein in a heterologous host that remains authentic to its native form.

Expressing membrane proteins in recombinant systems is challenging due to a complex process by which they need to be inserted into the membrane bilayer to assume and maintain their native structures. In *E. coli*, membrane proteins are inserted into the inner membrane by the Sec translocase and YidC insertase through the signal recognition particle (SRP) pathway. The pathway requires the presence of the SRP, typically the first TMD of the protein of interest as most membrane proteins lack a cleavable signal sequence (Ulbrandt *et al.*, 1997), and the SRP receptor FtsY at the membrane. Targeting of nascent chains to the membrane

occurs when a hydrophobic helix emerges from the exit tunnel of the ribosome and is recognised as a targeting signal by the SRP (Berndt *et al.*, 2009, Huber *et al.*, 2005), at which point further translation is paused until the TMD segment docks onto the Sec translocon and dissociates from the SRP (Luirink and Sinning, 2004, Dalbey *et al.*, 2011) (Figure 4). Translation subsequently resumes and TMD segments pass through an opening in the SecYEG translocon into the membrane (Park and Rapoport, 2012).

YidC insertase has been shown to insert small proteins into the membrane independently of the Sec translocon (Samuelson *et al.*, 2000). The structure of the periplasmic domain of *E. coli* YidC has been solved by X-ray crystallography and demonstrated to have a super  $\beta$ -sandwich fold (Oliver and Paetzel, 2008, Ravaud *et al.*, 2008), but its function remains unknown. It also remains unclear why some proteins are directed through the YidC instead of the Sec insertion pathway.



**Figure 4. Figure removed due to copyright.**

The stability of a protein, or its ability to maintain its molecular integrity and biological function while subjected to environmental determinants, is a critical determinant of its potential to crystallise uniformly, particularly given the harsh extraction and purification techniques potentially involved in its preparation

(Ericsson *et al.*, 2006). Recombinant production approaches offer potential advantages by facilitating mutagenesis of unstable proteins to improve their conformational stability, such as by introducing additional cysteine residues to create disulfide bridges in the protein (Trivedi *et al.*, 2009). Removing flexible regions in unstable proteins can also be beneficial in crystallisation of some molecules - providing the biological activity of the target is not affected (Hsu *et al.*, 2006). Furthermore, the expression of proteins of interest with added recombinant tags or as fusion proteins frequently improves their stability (Terpe, 2003, Terpe, 2006, Butt *et al.*, 2005), while unstable proteins may also be expressed in the presence of protein-specific cofactors or ligands to increase their stability (Vedadi *et al.*, 2006, Golovanov *et al.*, 2004). Membrane proteins, due to their amphipathic character, can be difficult to manipulate and their stability sometime is low. One approach to increase the stability of membrane proteins is to engineer fusions with T4 lysozyme, which significantly enlarges the hydrophilic surface of the protein, therefore increasing protein stability and allowing for more extensive protein contacts during crystallisation (Grishammer, 2006). Furthermore, membrane protein sequences are thought to evolve much more slowly than globular proteins as residues within TMDs are less exposed, which leads to helical packing being maximised to form stable structures in the absence of the hydrophobic effect (Bowie, 2001, Oberai *et al.*, 2009). Therefore, point mutations are more effective at stabilising membrane proteins than globular proteins (Bowie, 2001, Lau and Bowie, 1997), with the likelihood of finding a stabilising point mutation considerably higher for membrane proteins than water-soluble proteins. Tate and Schertler (2009) substituted each residue of the turkey  $\beta$ 1-adrenergic receptor ( $\beta$ 1-AR) with alanine and measured the effect of each mutation on the thermostability. Eventually, in addition to several standard N- and C-terminal truncations, a combination of six different point mutations was found to stabilise the  $\beta$ 1-AR significantly, with an increase in its melting temperature ( $T_m$ ) by 21°C compared to the wild-type receptor (Serrano-Vega *et al.*, 2008, Tate and Schertler, 2009).

The identification of conditions suitable for the stability of a membrane protein of interest has been a highly labour-intensive, trial-and-error approach (Bill *et al.*, 2011). In a recent advance, however, fusion of green fluorescent protein (GFP)

with membrane proteins allowed their stability to be screened rapidly and easily in a variety of detergents using a fluorescence-based assay (Drew *et al.*, 2008b). This led to the development of identification methods to successfully screen conditions for membrane protein solubilisation, purification and crystallisation with a significant reduction in the time and effort typically required for such a screening procedure (Moraes *et al.*, 2014).

#### 1.4.1. Protein expression systems

Recombinant expression approaches are typically used for the expression and purification of high yields of proteins required for structural investigation. Several bacterial expression systems play an important role in recombinant protein production, amongst these *E. coli* (see 1.4.2), *Lactococcus lactis*, and *Pseudomonas sp.* are the most prominent ones.

*L. lactis* emerged recently as one of the most noteworthy Gram-positive bacterial hosts and finds application within a therapeutic market niche due to being endotoxin free and GRAS (Generally Recognised As Safe), which are desirable characteristics in both biomedicines and food-associated markets (Yeh *et al.*, 2009). It has been successfully used to express membrane proteins (Mierau *et al.*, 2005) and holds promise for extracellular expression due to its Gram-positive cell structure. Moreover, *L. lactis* has just one membrane, which could simplify membrane-targeting, protein processing and the extraction of membrane proteins (Monné *et al.*, 2005), as well as reducing the occurrence of inclusion bodies (Kunji *et al.*, 2003). Kunji and co-workers were the first to successfully express a range of eukaryotic and prokaryotic membrane proteins of different sizes, topologies, quaternary structures and domain organisations in *L. lactis*; the overexpressed proteins were typically inserted into the membrane and assembled into functional complexes without eliciting inclusion body formation (Kunji *et al.*, 2003).

*Pseudomonas* species are known for their rapid growth rates and ability to secrete proteins (Krzeslak *et al.*, 2009). Several *Pseudomonas* species, including *P. fluorescens*, *P. aeruginosa*, and *P. putida*, have been demonstrated to achieve yields of recombinant proteins comparable to or higher than obtained in *E. coli* systems (Chen, 2012). *P. fluorescens* in particular shares features of *E. coli* for hyper-expression of heterologous proteins, with recombinant protein molecules

accumulating to up to 50% of the total protein content, while the species is also capable of reaching high cell densities that favour product accumulation (Huang *et al.*, 2007). *P. fluorescens* also exhibits a lesser oxygen-dependence and no acetate accumulation during production (Huang *et al.*, 2007); therefore its cultivation does not require strictly controlled aeration and glucose levels. Huang and co-workers (2007) described the expression of two insecticidal proteins from *Bacillus thuringiensis* that could be produced at up to 3-4 g/L in *P. fluorescens*, whereas their yields in *E. coli* reached only 100 mg/L. Among other well-established bacterial hosts, *Streptomyces* systems also deserve being mentioned, as this Gram-positive class can secrete large amounts of soluble proteins into the growth medium (Vrancken and Anné, 2009), indicating a high capacity of their secretion systems that could be advantageous in expression and purification of recombinant proteins.

Overall, bacterial expression systems combine simplicity and cost-effectiveness with reasonable yields, at least for proteins lacking extensive post-translational modifications. For more demanding targets, however, including many bio-therapeutics, expression experiments are frequently characterised by the production of insoluble, inactive proteins; poor translation efficiencies due to incompatible codon bias with the native host of the expressed gene; inefficient protein folding and/or processing; and an inability to carry out posttranslational modifications necessary for the structure or activity of the mammalian target (Sahdev *et al.*, 2008). Also, proteins larger than 100 kDa are generally expressed more efficiently in eukaryotic systems such as yeast, insect cells or mammalian cells, while prokaryotic systems cope better with proteins smaller than 30 kDa (Demain and Vaishnav, 2009).

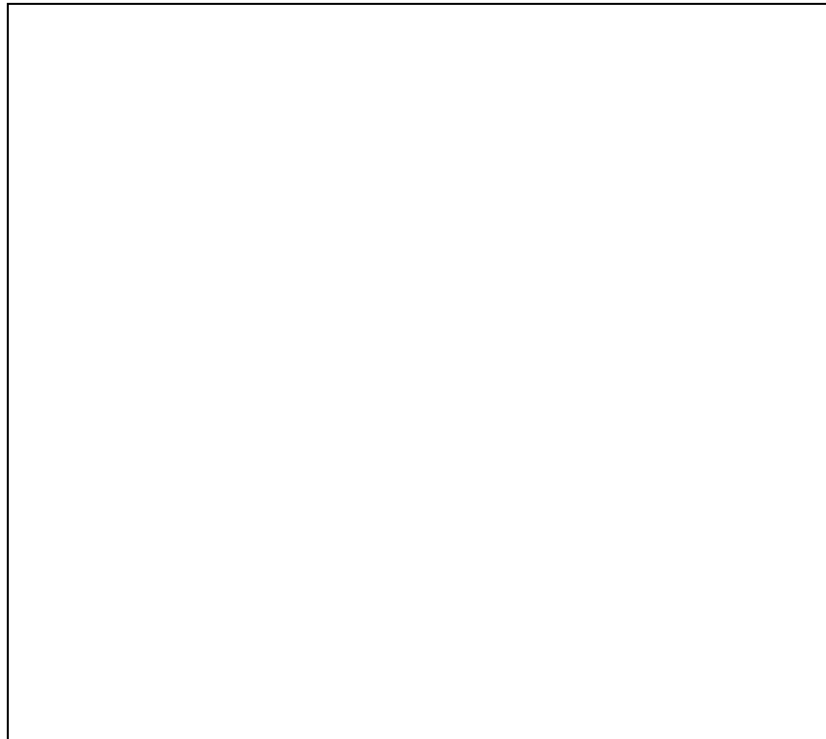
Yeasts, as eukaryotes, combine the simplicity of a unicellular organism with the ability to carry out complex posttranslational modifications. Therefore, yeasts are increasingly used because they are cost effective, can reach high cell densities, are easy to scale-up, and engineered strains deficient in protease and capable of human-like glycosylation pathways are available for improved expression of complex targets (Çelik and Çalık, 2012). Of the numerous yeast expression hosts available, the two most commonly utilised are *S. cerevisiae*, which exhibits the highest capacity for glycosylation amongst yeast expression species (Petranovic and

Nielsen, 2008) and *Pichia pastoris*, which unlike *S. cerevisiae* does not exhibit limited secretion of high molecular weight products (Cregg *et al.*, 2000).

Insect cells are higher eukaryotes than yeast and possess an excellent machinery for folding proteins of mammalian origin and are capable of carrying out complex posttranslational modifications that can rarely be achieved with bacterial systems (Agathos, 2010, Demain and Vaishnav, 2009). The one-vector system adopted for recombinant protein expression in insects is the baculovirus, a natural pathogen of lepidopteran cells and which also infects the fall armyworm (*Spodoptera frugiperda*) larvae (Kost *et al.*, 2005). As well as their cost efficiency, baculovirus-assisted insect cells present other positive features for recombinant protein expression such as their compatibility with post-translational modifications including phosphorylation and glycosylation (Luckow *et al.*, 1993), the high expression levels and high cell densities they can achieve (Taticek and Shuler, 1997, Ikonomou *et al.*, 2003), and their efficient protein folding and disulphide bond formation (Unger and Peleg, 2012). On the other hand, however, protein production in the baculovirus system is slow, time consuming and not as simple as in yeast-based systems (Demain and Vaishnav, 2009).

The most commonly used and reliable system for production of recombinant, glycosylated mammalian proteins is a mammalian cell expression platform. The understandable benefits arise from the ability of mammalian cells to carry out successful protein folding due to the presence of the appropriate chaperonins, as well as assembly and posttranslational modification systems best suited to the, frequently mammalian, target protein. There is a wide selection of cell lines in use for protein production, including baby hamster kidney, mouse myeloma-derived NS0, human retina-derived PerC6, baculovirus-infected human embryonic kidney (HEK293) and cervical cancer HeLa cells (Mohan *et al.*, 2008, Kim *et al.*, 2012). However, over 70% of recombinant therapeutic proteins are obtained from Chinese Hamster Ovary cells (CHO), the gold standard for mammalian cell systems, due to their capability of posttranslational modifications (glycosylation) (Wurm, 2004). While mammalian cell systems are clearly best suited of all heterologous platforms for the expression of mammalian proteins, desirable improvements in such systems in the future might include the development of low-cost cell culture media

and processes, adaptation to high-throughput expression and a greater diversity of DNA vectors for protein production (Li *et al.*, 2010).



**Figure 5. Figure removed due to copyright.**

#### **1.4.2. *E. coli* as a host for protein expression**

Currently, approximately 30% of approved recombinant therapeutics are produced in *E. coli* (Huang *et al.*, 2012, Swartz, 2001) and it continues to be a dominant player in the production of both soluble and membrane recombinant protein for structural studies. The depth of knowledge of its genetics and growth requirements, its short doubling time, its capacity to accommodate high yields of foreign proteins, as well as its low-cost and ease of manipulation, make *E. coli* a remarkable and desirable host for many recombinant expression challenges (Rosano and Ceccarelli, 2014). An enormous diversity of strains, expression vectors, and protein folding and expression enhancement technologies has been developed for use with the bacterium, many designed to ease the high metabolic burden exerted on a cell by high-level expression of a heterologous protein (Bentley *et al.*, 1990). Overexpression of membrane proteins is often problematic in *E. coli* due to toxicity

of the proteins and this step can prove a major bottleneck in structural studies (Wagner *et al.*, 2008, Narayanan *et al.*, 2011). Moreover, membrane proteins often fail to fold properly, leading to the formation of inclusion bodies (IBs) that are difficult to retrieve (Bernaudat *et al.*, 2011, Terpe, 2006). One approach to improving expression of recombinant membrane proteins in *E. coli* BL21(DE3) led to the identification of strains C41 and C43 which are more tolerant to toxic membrane proteins (Miroux and Walker, 1996). Subsequent evaluation of the strains identified a mutation in the *lacUV5* promoter, transforming it into its weaker wild-type counterpart and thus resulting in a lower level of synthesis of the membrane proteins that could be more easily tolerated by the host cells (Miroux and Walker, 1996, Dumon-Seignovert *et al.*, 2004).

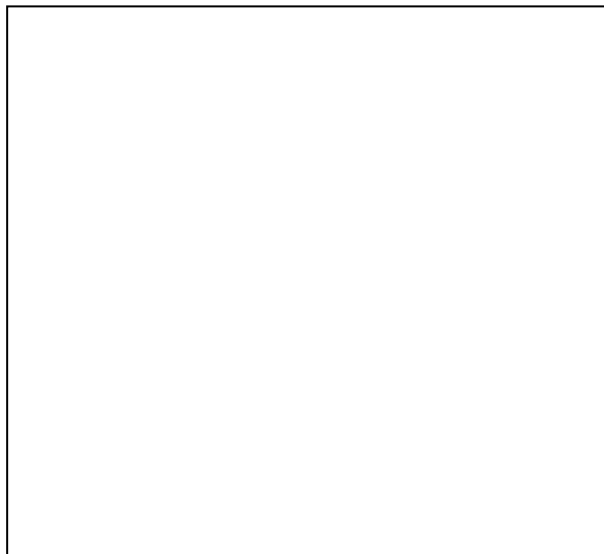
One of the main limitations of *E. coli* as an expression host has long been recognised as its inability to carry out posttranslational modifications, particularly those such as glycosylation that are important for the activities of many biomedically important molecules (Baneyx, 1999). The discovery of the *pgl*-encoded N-linked glycosylation system of *Campylobacter jejuni* and, critically, its adaptation into *E. coli* (Wacker *et al.*, 2002) was therefore a significant advance in the field of protein expression and engineering for research and industrial applications. It has since been demonstrated that *E. coli* cells containing the *pgl* system can produce glycoproteins with the *C. jejuni*-like glycans attached in *E. coli* cells (Lizak *et al.*, 2011), though mimicking exactly the human-like glycan sequence on recombinant, *E. coli*-produced proteins has yet to be achieved.

### **1.5. Principles of protein crystallisation**

Crystallisation is the process, determined by both thermodynamic and kinetic factors, during which identical protein molecules arrange themselves in an orderly manner to form a repetitive three-dimensional array known as a crystal (Rhodes, 2010). It can also be summarised as a transition between states and it consists of two main steps: nucleation and crystal growth. Nucleation takes place in a supersaturated protein solution, which in practice means that protein solubility must be exceeded by a factor of 3 or more (Chernov, 1997). Excessive supersaturation is necessary to overcome the free energy barrier, common during crystal formation, and allow the creation of microscopic clusters of protein which constitute the nuclei

(Bergfors, 2009). The formation of crystallisation nuclei is followed by spontaneous growth surrounding the nuclei, which is characterised by expansion of orderly arrays of protein molecules into crystals. Crystal growth can also be achieved in the absence of the nucleation stage by introducing nucleation clusters, such as crushed crystals, into the metastable phase; this practice is called seeding and is used to control or improve crystal growth in order to obtain diffraction-quality crystals (Bergfors, 2003).

The principles of the crystallisation process can be represented by the phase diagram, which illustrates how crystallisation parameters such as protein concentration, precipitant(s), additive(s), pH, temperature and time influence the different states of protein from soluble, through crystalline to amorphous precipitate (Figure 6 (Chayen, 2005)). A protein will stay in solution up to a certain limiting concentration, above which it will enter a new phase and a new state will appear.



**Figure 6. Figure removed due to copyright.**

Figure 6 shows an example of a typical protein crystallisation phase diagram. It indicates an area of high supersaturation, where protein will precipitate (precipitation zone), an area of moderate supersaturation, where spontaneous nucleation has a chance to occur (nucleation zone or labile zone), an area of lower supersaturation, where crystals are stable and may continue to grow but new

nucleation does not occur (metastable zone), and undersaturated zone where the protein is fully dissolved and will not crystallise. In theory, once spontaneous nucleation takes place, the concentration of the protein will decrease, thus naturally moving the system into the metastable zone, where crystals can continue to grow into a well-ordered and larger shape. Crystal growth will stop upon reaching the level of protein solubility, which means that the protein in solution is in equilibrium with formed crystals (Bergfors, 2009). Too low a supersaturation will result in a slow rate of nucleation while too high a supersaturation will lead to the occurrence of disordered forms of proteins, such as aggregates and precipitates; therefore proteins must be placed in the labile area for nucleation to be observed.

The phase diagram provides a strong theoretical basis for successful crystallisation. Despite this, determining an accurate phase diagram for individual proteins is time-demanding and requires large amounts of product. Therefore, it is more commonly used to optimise existing crystals, the solubility of which is easier to determine, than to identify initial crystallisation conditions (McPherson, 2004).

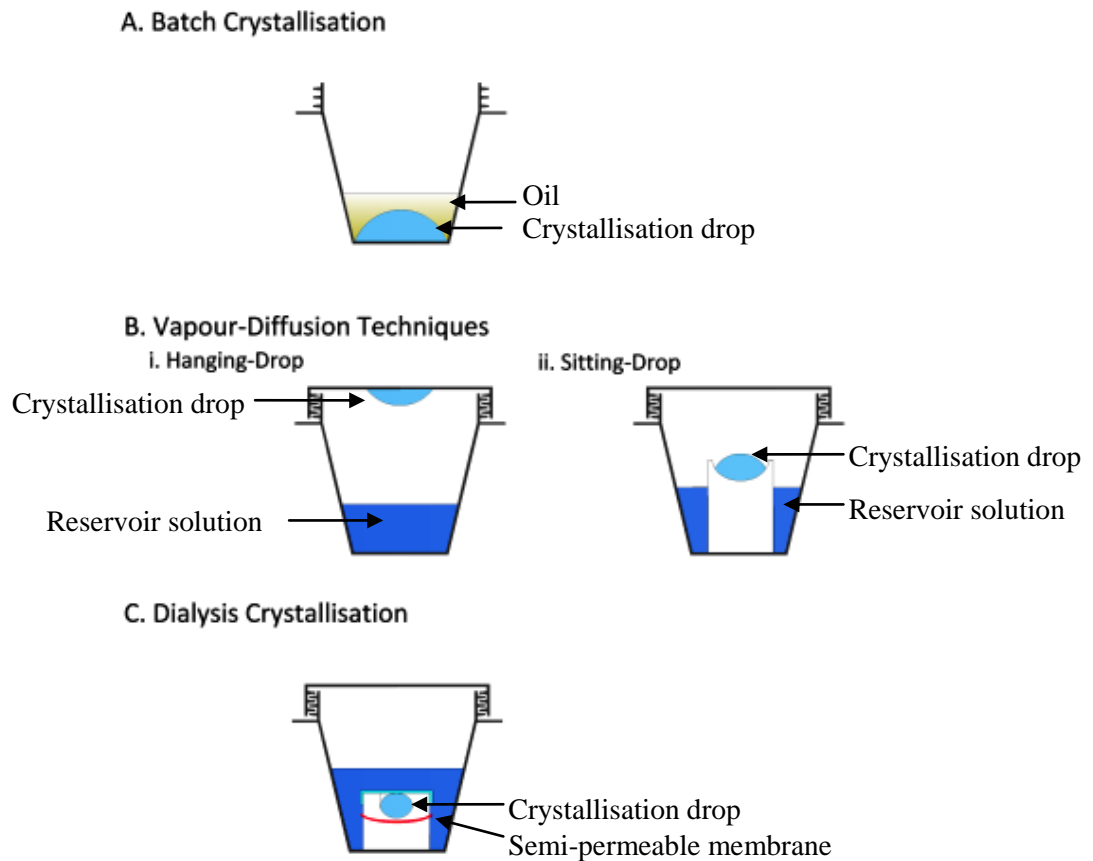
### **1.5.1. Crystallisation methods**

Protein crystallisation experiments can be set up in a number of different ways, all of which combine varying chemicals and parameters to gently push a protein from solution into an ordered crystalline phase. While crystallisation methods – outlined below – differ in their thermodynamic and kinetic components, the main consideration and starting point in all cases should be pure, stable, highly concentrated protein. Therefore, when choosing the initial method to be used in protein crystallisation, factors to be taken into account include the ease or otherwise of setting up trials, screening and harvesting crystals, as well as how many conditions can be screened simultaneously, how much protein is necessary for each method, and, in the high-throughput era, whether the chosen method is readily automated (Luft *et al.*, 2011, Luft *et al.*, 2014).

The main techniques used routinely in crystallography at present (Figure 7) are batch crystallisation, vapour-diffusion and dialysis methods (Chayen and Saridakis, 2008).

In *batch crystallisation*, the target protein is mixed with the precipitant at the final desired concentration and so supersaturation is instant. The crystallisation

drops are incubated under oil to prevent evaporation and preserve the conditions present in the drop (Chayen, 1998, Brumshtein *et al.*, 2008, D'Arcy *et al.*, 1996). *Vapour-diffusion* incorporates hanging- and sitting-drop options and is the most widely used technique in protein crystallisation. In this approach, the crystallisation drop, consisting of a mix of protein and precipitant, is equilibrated against a reservoir containing the crystallising solution at higher or lower concentration. The equilibration-driven crystallisation drop forces a shift in the conditions, allowing the occurrence of supersaturation. Vapour-diffusion accommodates gradual changes in the conditions present in the crystallisation drop and therefore supports the growth of well-ordered and big protein crystals, as well as enabling easy variation of the experimental parameters (Chayen, 1998, Korczynska *et al.*, 2007, Benvenuti and Mangani, 2007). Other advantages of vapour-diffusion are the simplicity of setting up trials, the ease of visualisation of the crystallisation drop, and of harvesting. In addition, upon automation only nanolitre volumes of precious protein samples are needed to set up trials. The principle of the *dialysis* method is that the protein sample is separated from the crystallising agent by a semi-permeable membrane which allows the passage of only small molecules, thus leading to slow mixing between protein molecules and the precipitant solution, and the occurrence of nucleation (Thomas *et al.*, 1989).



**Figure 7. Schematic of the most commonly used crystallisation methods.**

Each drawing represents single well crystallisation trial carried by: Batch crystallisation (A), Vapour-diffusion (B) and Dialysis (C). Description of each method in the text above.

## 1.6. Studying membrane proteins

Membrane proteins account for approximately 30% of proteins encoded in the human (Diehn *et al.*, 2006) as well as bacterial genomes and they are essential components of biological processes such as cell-cell communication, motility, energy generation and metabolism, and cross-membrane transport of ions, water, peptides and drugs (Lluis *et al.*, 2012). The structural complexity of membrane proteins, however, and their amphipathic character, present significant challenges in working with these molecules. Meanwhile solved membrane proteins structures are often important milestones that facilitate understanding of their functional organisation or their mechanism of action at the atomic level. Considering therefore that many membrane proteins are involved in human disease processes and that 60% of available drugs currently target membrane proteins such as GPCRs and ion channels,

this structural knowledge, though difficult to achieve, is invaluable in many fields (Arinaminpathy *et al.*, 2009).

The first structure of a membrane protein was determined in 1985 (Deisenhofer *et al.*, 1985). After years of slow progress subsequent to this breakthrough, the last decade saw the pace of membrane protein research speed up, due largely to the development of new technologies, advances in protein engineering and significant improvements in X-ray diffraction (Bill *et al.*, 2011). Despite this progress, however, the expression, purification and crystallisation of membrane proteins all still pose very considerable bottlenecks not encountered in the study of soluble proteins. This goes some way to explaining the paucity of high-resolution membrane protein structures currently available in protein databases, with only 499 unique membrane proteins of known structure having been reported to date (White, 1998-2014) from a total of ~95,490 protein structures overall .

One of the main difficulties in working with integral membrane proteins is that they naturally occur embedded in phospholipid bilayers in which the bilayer interior offers protection to otherwise unstable, hydrophobic, membrane-spanning regions of the protein. Upon extraction of the membrane protein from its native environment, therefore, the lipid bilayer must be replaced by other solvents capable of mimicking the natural lipid bilayer environment to ensure the protein remains as close as possible to its native state. This typically involves the use of detergents to stabilise the protein.

### **1.6.1. Principles of detergents**

Detergents are small, simple, ubiquitous molecules with amphipathic properties that result in a complex behaviour. They have a tendency to self-associate and to interact with hydrophobic surfaces in other molecules in a concentration-dependent manner, thus allowing them to mimic lipid bilayers (Privé, 2007). Detergents are incorporated throughout vital steps in membrane protein analysis such as protein extraction, purification and crystallisation as they have the potential to maintain such proteins in their folded and active states upon removal from the membrane. Their amphiphilic character is evidently connected to their structure, which consists of a polar head group and hydrophobic tail. The character and stereochemistry of these components determine their specific behaviour and allow

for the division of detergents into three major categories: ionic, non-ionic and zwitterionic (le Maire *et al.*, 2000). *Ionic detergents* contain a hydrophilic head group with a cationic or anionic charge and a hydrophobic chain or steroidal backbone (Helenius *et al.*, 1979). They are very effective at disruption of cellular structures and protein solubilisation but usually lead to some denaturation of proteins (Seddon *et al.*, 2004, Privé, 2007, Moraes *et al.*, 2014). The best known representative of ionic detergents is SDS (sodium dodecylsulfate), which is commonly used for protein denaturation in molecular biology. *Non-ionic detergents* contain an uncharged hydrophilic head group and are generally considered mild and non-denaturing. They are considered the most favourable groups of detergents for solubilisation of membrane proteins as they usually break lipid-lipid and lipid-protein interactions necessary to solubilise membrane proteins, but not protein-protein interactions which would lead to protein denaturation (Seddon *et al.*, 2004, Moraes *et al.*, 2014). Non-ionic detergents include maltosides (n-Dodecyl- $\beta$ -D-maltopyranoside (DDM), n-Decyl- $\beta$ -D-maltopyranoside (DM), n-Undecyl- $\beta$ -D-maltopyranoside (UDM)), glucosides (n-octyl- $\beta$ -D-glucopyranoside (OG), n-Nonyl  $\beta$ -D-glucopyranoside (NG)) and cymals. However, even detergents belonging to this group, which are characterised by short alkyl chains (*e.g.* OG, NG), can inactivate proteins (Tate, 2010, Arnold and Linke, 2008). *Zwitterionic detergents* combine properties of both ionic and non-ionic detergents and have a greater tendency to lead to protein deactivation than non-ionic detergents (Luche *et al.*, 2003, Henningsen *et al.*, 2002). Some representatives of this detergent group, such as LDAO, are nevertheless commonly used in structural biology investigations of membrane proteins (Wiener, 2004, Eshaghi *et al.*, 2005, Raschle *et al.*, 2009).

| Detergent group |                                    | Detergent name                             | Alkyl chain                     | Mr  | CMC (%) | Aggregation number (H <sub>2</sub> O) |
|-----------------|------------------------------------|--|---------------------------------|-----|---------|---------------------------------------|
| Non-ionic       | Alkyl- $\beta$ -D-glucosides       | Octyl- $\beta$ -D-glucoside (OG)           | C <sub>8</sub> H <sub>17</sub>  | 292 | 0.53    | ~27-100                               |
|                 |                                    | Nonyl- $\beta$ -D-glucoside (NG)           | C <sub>9</sub> H <sub>19</sub>  | 306 | 0.2     | ~133                                  |
|                 | Alkyl- $\beta$ -D-thioglucosides   | Nonyl- $\beta$ -D-thioglucoside            | C <sub>9</sub> H <sub>19</sub>  | 322 | 0.093   | -                                     |
|                 | Alkyl- $\beta$ -D-maltosides       | Decyl- $\beta$ -D-maltoside (DM)           | C <sub>10</sub> H <sub>21</sub> | 483 | 0.087   | ~69                                   |
|                 |                                    | Undecyl- $\beta$ -D-maltoside (UDM)        | C <sub>11</sub> H <sub>23</sub> | 497 | 0.029   | ~71*                                  |
|                 |                                    | Dodecyl- $\beta$ -D-maltoside (DDM)        | C <sub>12</sub> H <sub>25</sub> | 511 | 0.0087  | ~78-149                               |
|                 | Alkyl- $\beta$ -D-thiomaltosides   | Nonyl- $\beta$ -D-thiomaltoside (NTM)      | C <sub>9</sub> H <sub>19</sub>  | 484 | 0.15    | -                                     |
| Zwitterionic    | N-Alkyl-N,N'-dimethylamine-N-Oxide | N-Lauryl-N,N'-dimethylamine-N-oxide (LDAO) | C <sub>12</sub> H <sub>25</sub> | 229 | 0.023   | ~76                                   |
| Ionic           | Alkyl Sulphates                    | Sodium Dodecyl Sulphate (SDS)              | C <sub>12</sub> H <sub>25</sub> | 288 | 0.047   | -                                     |

**Table 1. Detergent groups and representative examples with their characteristics**

\*CMC for UDM was measured in 100 mM NaCl, 20 mM HEPES pH 7.5

Detergents are capable of solubilising membrane proteins because they form organised spherical structures called micelles in which the polar groups of the detergent form hydrogen bonds with water molecules and the hydrophobic tail groups aggregate due to hydrophobic interactions (Figure 8). The micelles into which detergents self-associate are small, well-defined assemblies, much simpler than the extended structures characteristic of lipid bilayers. Detergent micelles vary in the number of molecules they contain, ranging from as few as 4 to several hundred (Privé, 2007). This is termed the *aggregation number* and is typical for each detergent. A second important property of detergents is the critical micelle

concentration (CMC), which is the minimum concentration required for micelle formation. In a typical micelle, the hydrophilic head group of the detergent determines the interactions with protein, while the length of the hydrophobic chain influences the CMC and aggregation number of detergent, which in turn will affect solubilisation and crystallisation of membrane protein (Seddon *et al.*, 2004). Mild detergents usually possess a large and uncharged head group and long alkyl chain, whereas a small head group and short alkyl chain characterise harsh detergents. Through creating big micelles, mild detergents are gentle to proteins' native states though they sometimes fail to effectively solubilise membrane protein and are difficult to handle in further applications, including crystallisation, due to the large micelle size (Newby *et al.*, 2009). On the other hand, detergents with shorter hydrophobic chains create smaller micelles and are often more successful in the crystallisation. This may be due to an increase in a protein's accessibility to form crystals but long exposures to these harsh detergents can also lead to denaturation of the target protein (Privé, 2007, Newby *et al.*, 2009).



**Figure 8. Figure removed due to copyright.**

Successful and effective solubilisation of membrane proteins can be achieved by using a detergent concentration much higher than the CMC, termed the critical solubilisation concentration (CSC), which allows complete disruption of the cell membrane (Seddon *et al.*, 2004). While a particular detergent's aggregation number and CMC are difficult to measure precisely and accurately, and their reported values can vary (le Maire *et al.*, 2000), its CSC is empirical as it depends on the detergent, lipid concentration and composition of the membrane system. For example, while the CSC value for DDM typically oscillates between 1-4%, its CMC is 0.009%. On the contrary, the CMC for OG is 0.89% but it is used at concentrations

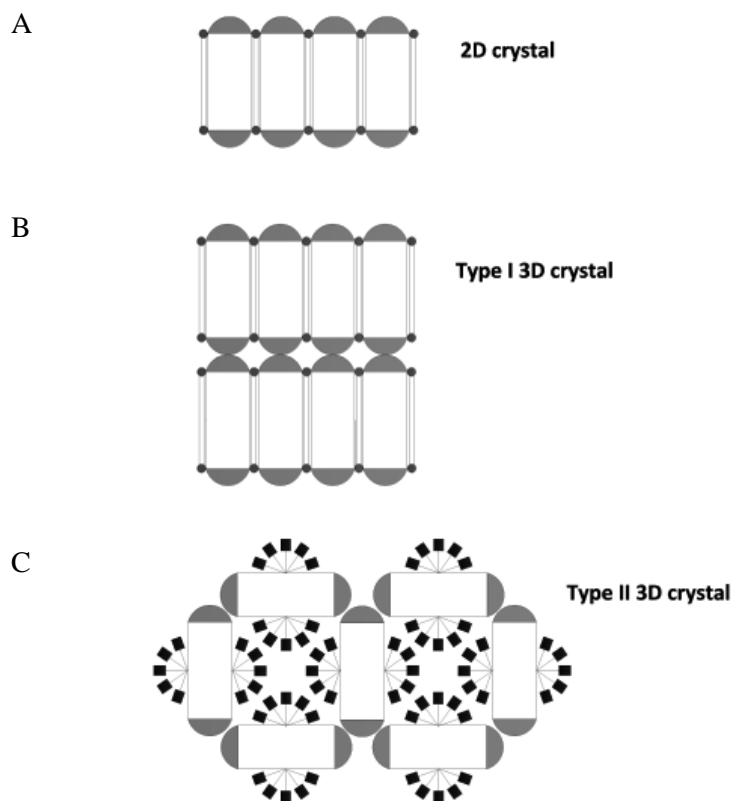
from 1.2-1.5% for membrane solubilisation (Seddon *et al.*, 2004, Arachea *et al.*, 2012). It is evident that OG is more disruptive to membranes and may be more effective at membrane solubilisation due to its shorter alkyl chain but it is also much harsher to the protein, and so DDM is usually the initial detergent utilised to solubilise membranes containing the target protein (Kang *et al.*, 2013).

Solubilised membrane proteins exist in water-soluble, protein-detergent complexes (PDCs) in which hydrophobic regions of the protein are surrounded by the detergent micelle and soluble regions are exposed to solvent. Formation of the crystal lattice of the protein is dependent on complex interactions between the protein, detergent and aqueous buffer - components with remarkably different physical characteristics. Protein-protein interactions form the principal scaffold in the crystal lattice; therefore having a large cytoplasmic domain can be advantageous for membrane proteins in crystallisation trials (Carpenter *et al.*, 2008). The flexible and dynamic nature of detergent micelles surrounding the membrane protein also favour detergents that create smaller micelles around the transmembrane regions of protein, thus supporting stronger contacts between exposed surfaces of the protein and, consequently, leading to well-ordered lattices (Privé, 2007). Unfortunately, small micelles can lead to more rapid destabilisation of the protein, whereas large micelles, while maintaining the protein in a more stable state, can often interfere with the crystallisation process (Gutmann *et al.*, 2007, Stroud, 2011). Therefore, the right choice of detergent to solubilise and crystallise proteins has a huge impact on the outcome of trials and is often the key factor in obtaining diffraction-quality crystals of membrane proteins.

### **1.6.2. Crystallisation of membrane proteins**

Membrane proteins form three crystal lattice variations: 2D crystals, type I 3D crystals and type II 3D crystals (Figure 9 (Iwata, 2003)). 2D crystals are stabilised through hydrophobic interactions of reconstituted membrane and polar interlayer interactions and, while they are not suitable for X-ray crystallography, they find use with electron microscopy (Bill *et al.*, 2011). Type II 3D crystals are the most common for membrane proteins and are usually obtained by *in surfo* methods, where the crystal packing is based on hydrophilic interaction between proteins in a manner similar to soluble protein crystals, which is possible with membrane proteins due to

coverage of the hydrophobic regions by detergent (Hunte *et al.*, 2003). Type II 3D crystals are usually characterised by poor protein-protein interactions and a high solvent content of up to 80%, resulting in reduced data quality (Iwata, 2003). Finally, type I 3D crystals are generated by *in meso* crystallisation methods involving microliposomes and involve protein-detergent-lipid hydrophobic interactions, resulting in planar sheets stacked on top of each other through hydrophobic contacts (Landau and Rosenbusch, 1996).



**Figure 9. Schematic illustration of membrane protein packing in the three types of crystal lattice.**

*2D crystals are formed through hydrophobic interactions of proteins reconstituted into lipid bilayer (A), Type I 3D crystals are formed by 2D sheets stacked on top of each other through hydrophobic contacts, they are typical for “in meso” crystallisation methods (B), Type II 3D crystals are formed through hydrophilic interactions between protein molecules surrounded by detergent micelle, they are typical for “in surfo” crystallisation methods (C). Re-drawn from (Iwata, 2003).*

Several techniques are employed in the crystallisation of membrane proteins; the majority of these *in surfo* techniques employed in the crystallisation of

membrane proteins involve the use of solubilised protein-detergent complexes in vapour-diffusion, microdialysis or microbatch. Alternatively, the *in meso* approach which continues to grow in popularity is based on the formation of a bicontinuous lipidic mesophase and involves partial reconstitution of the solubilised protein into an organised and structured lipid bilayer (Caffrey and Cherezov, 2009).

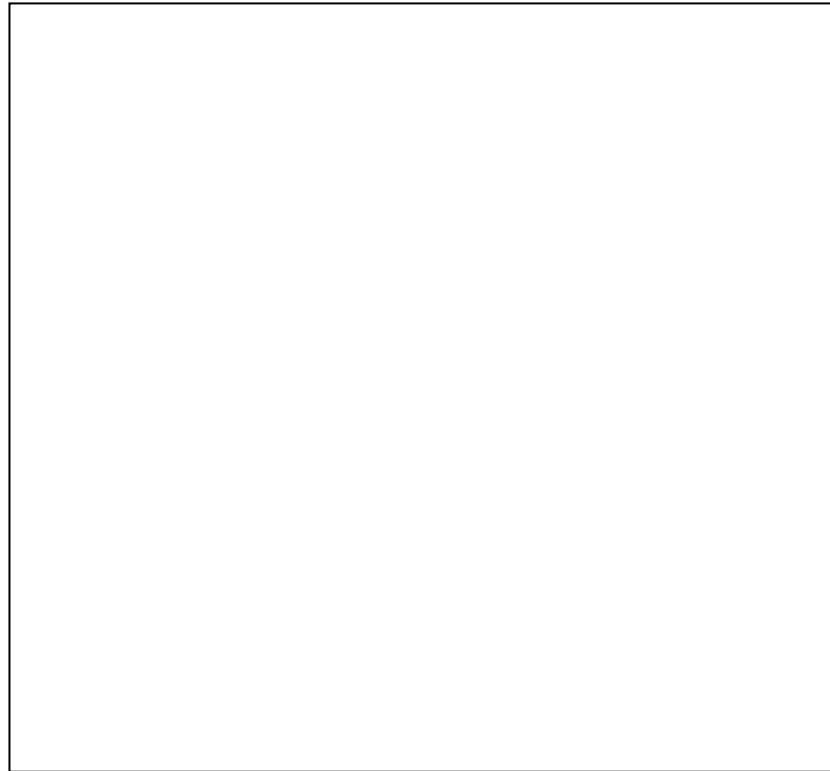
Manipulation of PDCs *in surf* is carried out in the same manner as for soluble proteins, often resulting in poor-quality crystals due to obstruction of protein-protein interactions by detergent (Iwata, 2003). The choice of detergent plays an important role in crystallisation, therefore, because the shape and size of resultant micelles can have a critical influence on the growth of crystals (Garavito and Ferguson-Miller, 2001). Consequently, detergents characterised by small micelles, such as OG or NG, are preferable in crystallisation trials even though their occasional failure to protect the entire hydrophobic surface can result in aggregation of the protein.

The *in meso* lipidic cubic phase (LCP) method has attracted interest as an alternative approach to growing membrane protein crystals due to its success in solving the structures of a human-engineered  $\beta_2$ -adrenergic and the human adenosine A2A G protein-coupled receptors (GPCRs)(Cherezov *et al.*, 2007). Members of the GPCR family transmit signals across the cell membrane in response to a diversity of extracellular stimuli and activate varied intracellular and physiological responses as a result (Lindemann and Hoener, 2005). Previously, this *in meso* method has been successfully applied to solving the structure of bacteriorhodopsin (Luecke *et al.*, 1999) and, subsequently, more than 114 other membrane proteins. However, it is still responsible for only about 10% of all membrane protein structures submitted to PDB (Protein Data Bank) to date (Caffrey and Cherezov, 2009).

The main advantage of LCP is that the membrane protein is removed from the harsh environment of detergent micelles and placed in stabilising, more native-like surroundings in which it should retain its activity and conformation (Cherezov, 2011). Moreover, LCP trials tend to lead to the formation of type I crystals which are characterised by extensive crystal contacts through both hydrophilic and hydrophobic regions (Nollert *et al.*, 2001), resulting in better crystal packing. Structurally, LCP is a three-dimensional, bicontinuous lipid bilayer formed

spontaneously upon energetic mixing of equal volumes of the chosen lipid with an aqueous buffer (the protein-detergent complex) at a temperature of 20°C (Landau and Rosenbusch, 1996). The formation of a mesophase is characterised by the appearance of a transparent and homogenous gel-like layer and this is coincident with reconstitution of the membrane protein in a lipid bilayer. While a powerful approach for studying membrane proteins, LCP is difficult to work with as it is very viscous, difficult to dispense and the mesophase can be difficult to achieve.

Monoolein (9.9 MAG) is the lipid of choice for LCP (Li *et al.*, 2011), although success has also been obtained with related lipids such as monovaccenin (11.7 MAG) for a sensory rhodopsin II-transducer complex (Gordeliy *et al.*, 2002) and 7.7 MAG for the  $\beta$ 2 adrenergic receptor-Gs protein complex (Rasmussen *et al.*, 2007, Cherezov *et al.*, 2007). According to the phase diagram developed for the monoolein/water system (Figure 10 (Cherezov, 2011)), a solid lamellar crystal lipid monoolein ( $L_c$ ) begins to melt at approximately 37°C, forming the so-called fluid isotropic phase ( $L_a$ ). With the addition of water (or other aqueous solvent) to the system, an increase in a number of temperature-dependent lyotropic mesophases is observed. Thus, at high temperatures (80-100°C) the inverted hexagonal phase ( $H_{III}$ ) appears which, upon lowering the temperature, transforms into two different cubic phases: Ia3d phase forms at lower hydration levels than Pn3m phase, which can exist in equilibrium with excess water as a two-phase system, referred to as sponge phase, over a wide range of temperatures. Pn3m, the most desired of cubic mesophases for reconstitution of membrane proteins, is achieved in the monoolein/water system at 20°C at an overall composition of about 40% (wt/wt) water, which has proven highly successful in accommodating the growth of crystals of numerous membrane proteins (Caffrey and Cherezov, 2009).



**Figure 10. Figure removed due to copyright.**

Bicelles are another attractive and powerful lipidic tool for crystallising membrane proteins in a native-like environment. These are bilayer discs that are formed through the mixing of a long-chain phospholipid (*e.g.* DMPC) and an amphiphile (*e.g.* CHAPSO) in an aqueous environment (Ujwal and Bowie, 2011). Proteins reconstituted using bicelles have been demonstrated to be fully functional under physiological conditions and protein–bicelle combinations can be used with almost the same ease as detergent-solubilised membrane proteins, including high-throughput crystallisation approaches (Faham and Bowie, 2002). While the capability of the bicelle concept was demonstrated, like many such membrane protein methods, by crystallisation of bacteriorhodopsin (Faham *et al.*, 2005), the real potential of the method was demonstrated by crystallisation of the human  $\beta$ 2-adrenergic G-protein-coupled receptor in complex with a Fab fragment in a structure that was determined at 3.4 Å/ 3.7 Å resolution (Rasmussen *et al.*, 2007, Cherezov *et al.*, 2007).

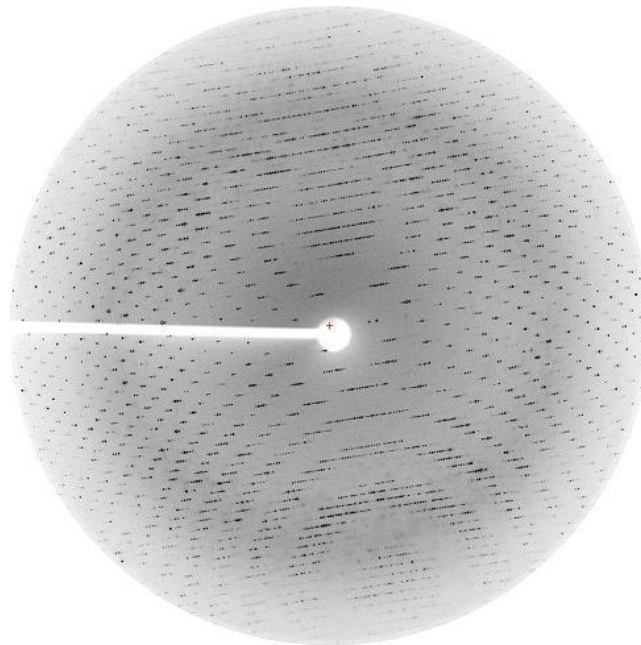
### 1.7. Basics of crystallography

X-radiation, composed of X-rays, was discovered in the 19<sup>th</sup> century by Wilhelm Röntgen while experimenting with an open cathode ray tube and various chemical materials. In fact, the first known X-ray image is of Röntgen's wife's hand. Interestingly, an early application of X-rays was in the shoe-fitting fluoroscope, which became a hit with shoe stores from the 1930s-1950s. "Now, at last, you can be certain that your children's foot health is not being jeopardised by improperly fitting shoes" was the radio commercial tag line. In the 1950s, scientists from UC Berkley published a research study about the dangers of X-ray exposure, putting an end to well-fitting shoes (Mould, 1993, Lewis and Caplan, 1950).

X-rays are electromagnetic waves of the same nature as visible light or radio waves but with very short wavelengths of around 1 Å. This short wavelength is necessary in crystallography due to the interatomic distances of macromolecules of the same order of magnitude. X-rays are emitted by electrons generated by an X-ray tube, which uses a high voltage cathode under vacuum to accelerate electrons to high velocities. The electrons then collide with a metal target, which in crystallography is usually copper. Highly focused X-rays of high intensity are generated in specialised synchrotrons, football stadium-sized machines in which electrons are accelerated to extremely high energy and their direction changed periodically to produce dozens of thin beams directed toward a beamline, where they are used to study the material of interest (Bilderback *et al.*, 2005).

Once crystals of macromolecules are obtained they must be prepared for X-ray data collection by harvesting, cryo-cooling and, in most cases, packing and shipping to one of the 70 synchrotron facilities worldwide. To collect data from a crystal, it is placed in an intense, monochromatic X-ray beam and irradiated with X-rays while being rotated, which results in beam scattering (Snell, 2012). Upon hitting the detector, the scattered X-rays form a diffraction pattern of spots (Figure 11), the strength and angles of which are recorded to produce a regular pattern of reflections. The intensities of these spots can be used to calculate the electron density of the molecules within the crystal, which provides information about the locations of atoms which is used in turn to build a model of the macromolecule (Rhodes, 2010).

Solving a structure typically requires multiple crystals to be evaluated in order to obtain a full set of data, because thousands of diffraction spots need to be collected to solve a protein structure and this usually necessitates the evaluation of large numbers of crystals as they are prone to radiation damage. Data are collected at cryogenic temperatures to minimise radiation damage to the crystals; therefore crystals are dipped into liquid nitrogen during harvesting, while diffraction is carried out in a liquid nitrogen stream to cool the crystal (Garman and Owen, 2005).



**Figure 11. Example of protein complex diffraction pattern, which represents an orderly array of reflections (spots).**

*The position and intensity of each spot are used to determine the molecular structure of molecules.*

## **1.8. Overview of this project**

The study of proteins using X-ray crystallography is an increasingly important field of science due to the essential role of proteins in core biological processes. Expression, purification and crystallisation of heterologous proteins, especially membrane proteins, all pose considerable difficulties and challenges, as outlined above. This PhD work is focused on the recombinant expression and structural analysis of two recombinant proteins: CzrB, a membrane-bound zinc

transporter from *Thermus thermophilus*, and trigger factor, a molecular chaperone from *Psychrobacter frigidicola*. The two target proteins are not related, however, they both piqued the interest in the research group due to their biotechnological application related to improved production of recombinant proteins (see 1.2.1 and 1.2.2). The knowledge of the three-dimensional arrangement of the target proteins will allow the insight into the structural basis for the exhibited biotechnological potential. Moreover, solved structures will lead to better understanding of the roles target proteins play in the cell and their mechanism of action during metal transport (CzrB) and supporting protein folding (TFPf).

A detailed account of the production of high yields of the two target proteins, the preparation of pure and stable products necessary for crystallisation trials, and work towards the growth of diffraction-quality crystals follows.

## **2. MATERIALS AND METHODS**

### **2.1. Materials**

#### **2.1.1. Suppliers**

The following suppliers were used to source materials used throughout the project:

Affymetrix, Wycombe Lane, Wooburn Green, High Wycombe, HP10 0HH, U.K.

BDH Merck Ltd, Glasgow, U.K.

Beckman Coulter Inc., Brea, California, U.S.A.

Bioline Reagents Limited, Unit 16 The Edge Business Centre, Humber Road, London, NW2 6EW, U.K.

Bio-Rad Laboratories, Hemel Hempstead, Hertfordshire, U.K.

Cruinn Diagnostic Ltd., 5b/6b Hume Centre, Park West Industrial Estate, Nangor Road, Dublin 12, Ireland.

Eurofins Genomics, Anzinger Str. 7a, 85560, Ebersberg, Germany.

Expedeon Ltd., Unit 1A Button End, Harston, Cambridgeshire, CB22 7GX, U.K.

Fisher Scientific Ireland, Suite 3, Plaza 212, Blanchardstown Corporate Park, Ballycoolin, Dublin 15, Ireland.

Finnzymes- part of Thermo Fisher Scientific Inc.

GE Healthcare , Little Chalfont, Buckinghamshire HP7 9NA, U.K.

Hampton Research, 34 Journey, Aliso Viejo, CA 92656-3899, U.S.A.

Invitrogen Life Technologies Ltd., Inchinnan Business Park, Paisley, U.K.

Medical Supply Company Ltd., Damastown, Mulhuddart, Dublin 15, Ireland.

Melford Laboratories, Bildeston Road, Chelsworth, Ipswich, Suffolk, IP7 7LE, U.K.

Merck Millipore Ltd, Watford, Hertfordshire, U.K.

Molecular Dimensions Ltd. Unit 6, Goodwin Business Park, Willie Snaith Road, Newmarket, Suffolk CB8 7SQ, U. K.

MyBio Ltd., Kilkenny Research & Innovation Centre, St. Kieran's College, Kilkenny, Ireland.

New England Biolabs Ltd. (NEB), Hitchin, Hertfordshire, SG4 0TY, U.K.

Novagen, 441 Charmany Drive, Madison, Wisconsin 53719, U.S.A.

Qiagen Ltd, Fleming Way, Crawley, West Sussex , RH10 9NQ, U.K.

Rigaku (Europe), Unit B6, Chaucer Business Park, Watery Lane, Kemsing Sevenoaks, Kent TN15 6QY, U.K.

Roche Diagnostics Ltd., East Essex, U.K.

Sartorius Stedim Ireland Ltd., Unit 41, The Business Centre, Stadium Business Park, Ballycoolin Road, Dublin 11, Ireland.

Sarstedt Ltd., Sinnottstown Lane, Drinagh, Wexford, Ireland.

Sigma Aldrich Ltd., Airton Rd., Tallaght, Dublin 24, Ireland.

Sparks Lab Supplies Ltd., Unit J-7, Grants Road, Greenogue Business Park, Rathcoole, Co. Dublin, Ireland.

Thermo Fisher Scientific Inc., 3747 N Meridian Rd, Rockford, IL 61101, U.S.A.

VWR, Orion Business Campus, Northwest Business Park, Blanchardstown Corporate Park, Ballycoolin, Dublin 15, Ireland.

### **2.1.2. General Reagents**

Commonly used buffer components, solvents and general chemicals were sourced from Sigma Aldrich, VWR or Fisher Scientific unless stated otherwise.

### **2.1.3. PCR Oligonucleotides**

All primers were synthesised and purified by HPLC at Eurofins Genomics. Sequences of all primers used in this project are available in the Appendix.

### 2.1.4. Bacterial Strains

The genotypes of the *E. coli* strains used in this work are as follows:

| <i>E. coli</i> strain | Genotype  | Description  |
|-----------------------|---|--|
| TOP10                 | <i>F</i> <i>mcrA</i> Δ ( <i>mrr-hsdRMS-mcrBC</i> ) <i>f80lacZDM15</i> Δ <i>lacX74</i> <i>deoR</i> <i>recA1</i> <i>araD139</i> <i>D</i> ( <i>ara-leu</i> ) 7697 <i>galUgalKrpsL</i> (Str <sup>R</sup> ) <i>endA1</i> <i>nupG</i> | Designed for cloning and plasmid propagation (Invitrogen)  |
| BL21 (DE3)            | <i>F</i> <i>ompThsdS<sub>B</sub></i> ( <i>r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup></i> ) <i>gal dcm</i> λ (DE3)   | Designed for use in high-level heterologous protein production with T7 promoter-based expression systems.  |
| BL21 codon plus RIL*  | <i>F</i> <i>ompThsdS<sub>B</sub></i> ( <i>r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup></i> ) <i>gal dcm</i> λ (DE3) <i>endA</i> Hte [ <i>argUileYleuW</i> Cam <sup>R</sup> ]  | As above, additionally contains plasmid with extra copies of <i>argU ileY leuW</i> tRNA genes.   |
| C41 (DE3)             | BL21 derivative <i>F</i> <i>ompThsdS<sub>B</sub></i> ( <i>r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup></i> ) <i>gal dcm</i> λ (DE3) plus <i>lacUV5</i> promoter mutation.   | Strain is effective in overexpressing toxin proteins.  |
| C43 (DE3)             | Derived from C41  | This strain was derived from C41 (DE3) by selecting for resistance to a different toxic protein and can express a different set of toxic proteins to C41 (DE3).  |
| Lemo21                | <i>fhu</i> [lon] <i>ompT</i> λ (DE3) <i>dcm</i> Δ <i>hsdS</i> /pLemo (Cam <sup>R</sup> )  | BL21 derivative designed for tuneable expression of difficult products by varying levels of lysozyme ( <i>lysY</i> ).  |
| Rosetta pLacI         | <i>F</i> <i>ompThsdS<sub>B</sub></i> ( <i>r<sub>B</sub><sup>-</sup>m<sub>B</sub><sup>-</sup></i> ) <i>gal dcm</i> λ (DE3) pLacIRARE (Cam <sup>R</sup> )   | BL21 derivative designed to enhance expression of proteins that contain codons rarely used in <i>E. coli</i> . Plasmid with extra copies of <i>argU ileY leuW</i> tRNA genes also carries the <i>lac</i> repressor gene. |

## 2.1.5. Plasmids

| Plasmid name            | Description; source  |
|-------------------------|--|
| pACYC184                | <i>E. coli</i> low copy vector with p15A replication origin; Cam <sup>R</sup> , Tc <sup>R</sup> ; (NEB)  |
| p15aratighisPF          | pACYC184 derivative containing arabinose promoter cassette from pBADHisB and <i>tig</i> gene from <i>Psychrobacter frigidicola</i> ; this research group (Robin <i>et al.</i> , 2009)                          |
| pACYCDuet-1             | Vector designed to co-express two target genes; consists of two multiple cloning sites controlled by T7 promoter, p15A replicon, <i>lacI</i> , Cam <sup>R</sup> ; Novagen                                      |
| pACYCDuet-1-HisTEVtigPF | pACYCDuet-1 derivative containing <i>tig</i> gene from <i>P. frigidicola</i> with TEV protease-removable N-terminal histidine tag.   |
| pIVEX2.4d               | Vector designed for high-level expression of His <sub>6</sub> -tagged protein from T7 promoter in the cell-free RTS <i>E. coli</i> system; Amp <sup>R</sup> ; Roche Diagnostics                                |
| pIVEX2.4d/His6TFPf      | pIVEX2.4d derivative for N-terminus His <sub>6</sub> -tagged <i>tig</i> gene.  |
| pET28b+                 | Vector designed for recombinant protein expression from T7 promoter, pBR322 ori, Kan <sup>R</sup> , Novagen  |
| pET28b+CzrBHis6         | pET28b+ derivative containing C-terminal His <sub>6</sub> -tagged <i>czrB</i> gene; this research group  |
| pOPIN Vector Suite      | Versatile suite of expression vectors developed by OPPF used to generate series of expression vectors for multiple hosts consisting of full-length and truncated versions of <i>czrB</i> and <i>tig</i> genes. |

### **2.1.6. Antibiotics**

Antibiotics were used for strain and plasmid selection at the following concentrations unless stated otherwise: ampicillin (100 µg/ml in water), chloramphenicol (34 µg/ml in 100% ethanol), kanamycin (50 µg/ml in water), carbenicillin (50 µg/ml in water).

### **2.1.7. Antibodies**

Monoclonal anti-polyhistidine peroxidase-conjugated antibody produced in mouse (Sigma Aldrich).

### **2.1.8. Molecular Weight Markers**

DNA agarose Gel: Hyperladder I (0.2-10 kb) and Hyperladder IV (0.1-1 kb) (Bioline).

SDS PAGE: PageRuler Plus Prestained Protein Ladder (10-250 kDa) (Thermo Scientific), SeeBlue Plus2 Pre-stained Standard (3-188 kDa) (Invitrogen), Prestained SDS-PAGE Standards Broad Range (Bio-Rad), SigmaMarkers™ (Sigma).

## **2.2. Methods**

### **2.2.1. Competent Cell Preparation**

Competent cells for transformation of recombinant plasmids were prepared based on Sambrook and co-workers (Sambrook *et al.*, 2001).

#### **2.2.1.1. Electrocompetent Cells**

Cells of the desired strain were streaked from glycerol stocks on LB (Luria Bertani) agar plates containing the appropriate antibiotics (if applicable). After overnight growth at 37°C, a single colony from the plate was used to inoculate 10 ml of LB supplemented with the appropriate antibiotics; the culture was grown overnight at 37°C with shaking at 250 rpm. The resulting culture was diluted 100-fold into 500 ml of fresh LB and grown until an OD<sub>600</sub> of 0.5 was reached, after which the culture was placed on ice for 30 min. Cells were then harvested by

centrifugation using a Beckman Avanti J-20 XP and JLA-16.250 rotor at 5,000 g, 4°C for 10 min and washed with 400 ml of 1 mM HEPES and twice with 400 ml of 10% glycerol. After the final centrifugation, the cells were resuspended in 25% glycerol, flash frozen in liquid nitrogen and stored for further use at -80°C.

#### **2.2.1.2. Chemically Competent Cells**

Cells were prepared as above (section 2.2.1.1) until 100-fold dilution into 500 ml of fresh LB, after which they were grown until early log phase was reached and placed on ice for 30 min. Cells were then collected by centrifugation at 5,000 g, 4°C for 10 min, washed with 0.1 mM CaCl<sub>2</sub> and resuspended in ice-cold 0.1 M CaCl<sub>2</sub> prior to 2 h incubation on ice. Cells were finally suspended in 25% glycerol in 0.1 M CaCl<sub>2</sub>, flash frozen in liquid nitrogen and stored at -80°C.

#### **2.2.2. DNA Manipulation**

The following DNA manipulations were carried out according to Sambrook *et al.* (2001), unless otherwise stated.

##### **2.2.2.1. PCR**

Polymerase Chain Reaction (PCR) cycles were carried using a Flexigene thermal cycler (Techne). Annealing temperatures for primers were calculated using an online calculation tool available on the NEB website ([www.neb.com/tools-and-resources/interactive-tools/tm-calculator](http://www.neb.com/tools-and-resources/interactive-tools/tm-calculator)). Typically, amplification reactions were carried out with 0.4 units Phusion High Fidelity DNA polymerase or 0.1-0.8 units Velocity DNA polymerase per reaction (Finnzymes and Bionline, respectively). Overlap PCR reactions were carried out similarly to conventional PCR but with 20 extra cycles at the beginning of the reaction, in the absence of forward and reverse primers, to ensure the formation of adequate amounts of full-length template from fragments to be spliced.

### **2.2.2.2. Plasmid Construction**

#### **2.2.2.2.1. Cloning of non-dimerising CzrB**

The amino acid pairs identified by Cherezov (2008) to create hydrogen bonds between protein monomers in the CzrB dimer are Glu57(N)-Glu52(O) and Val56(N)-Thr54. Glu52 and Thr54 were replaced by Ala using an overlap PCR approach with *czrBNcoIF*, *czrBGluThrToAlaR1*, *czrBGluThrToAlaF2*, *czrBHisHindIIIR* primers (see Appendix) and the pET28b+CzrBHis6 vector as template. The resultant PCR product was digested with *NcoI/HindIII* and cloned into pET28b+, creating the pET28b+CzrBGluThrToAla vector for expression of non-dimerisingCzrB.

#### **2.2.2.3. Digestion**

All restriction enzymes used (*XbaI*, *HindIII*, *NotI*, *PstI*, *EcoRI*, *NcoI*) were purchased from Roche Diagnostics or Fermentas. Digests were carried out at 37°C for 2 to 16 h, unless recommended otherwise by the manufacturer. Digests using FastDigest (Fermentas) enzymes were incubated for up to 25 min.

#### **2.2.2.4. Ligation**

Ligation reactions were carried out using T4 DNA ligase from New England Biolabs. Ligation reactions were typically carried out overnight at 16°C.

#### **2.2.2.5. Transformation**

Transformation of plasmid DNA was carried out by electroporation using an Electroporator 2510 (Eppendorf) or by “heat shock” in a water bath. For electroporation, 1-5 µl of plasmid DNA was added to 50 µl electrocompetent cells and mixed thoroughly. It was then transferred to a pre-chilled cuvette, which was placed into the electroporator and subjected to a potential difference of 1800 V. The cells were transferred to a universal containing 950 µl of LB broth. After culturing for 60 min at 37°C (45 min for plasmids with an ampicillin resistance gene) and shaking at 250 rpm, 100 µl of the mixture was plated on LB agar containing the relevant antibiotics. In the case of heat shock, 1-5 µl of plasmid DNA was added to

50 µl chemically competent cells, mixed thoroughly and placed on ice for 10 min. The mixture was then placed in a water bath at 42°C for 60 sec followed by incubation on ice for 30 min. The mixture was transferred to a universal containing 950 µl of LB broth and incubated for 60 min at 37°C (45 min for plasmids with an ampicillin resistance gene) and 250 rpm, following which 100-200 µl of the culture was plated on LB agar containing the appropriate antibiotics.

### **2.2.3. DNA extraction**

#### **2.2.3.1. DNA extraction from agarose gels**

DNA visualised using a High Performance Ultraviolet Transilluminator was extracted from agarose gels using a scalpel, followed by melting of the gel at 50-60°C. This was followed by purification using Agarose Gel DNA Extraction Kit (Roche) or NucleoSpin Extract II gel extraction kit (Macherey-Nagel), with elution in 35-50 µl of double distilled water.

#### **2.2.3.2. Plasmid extraction from bacterial cells**

Plasmid DNA was isolated throughout the project using ISOLATE Plasmid Mini Kit (Bioline), QiaPrep Spin Miniprep Kit (Qiagen) and NucleoSpin Plasmid Kit (Macherey-Nagel), following the relevant manufacturer's instructions.

#### **2.2.3.3. Genomic DNA extraction from bacterial cells**

Genomic DNA was isolated from bacterial cells using a GeneJet Genomic DNA Purification Kit (Fermentas).

### **2.2.4. Agarose Gel Electrophoresis**

DNA was resolved on 1-2% agarose gels at 90 V for 1 h in an Apollo horizontal electrophoresis tank (Medical Supply Company). Ethidium bromide (Sigma)- or Midori Green (Nippon Genetics)-stained DNA was visualised using a fluorescence G:Box from SynGene using GeneSnap (SynGene).

### **2.2.5. Protein Expression**

#### **2.2.5.1. Expression of His-tagged TFPf**

Expression of His-tagged *TFPf* was carried out by inoculating LB broth containing chloramphenicol (34  $\mu\text{g/ml}$ ) with an overnight culture to a starting  $\text{OD}_{600}$  of 0.05. The culture was grown at 37°C with shaking at 250 rpm. Bacterial growth was monitored using a BioMateUV reader (Thermo Scientific). When an  $\text{OD}_{600}$  of 0.3 was reached, cultures were moved to 25°C in order for bacteria to adjust to the lower temperature, followed by induction of *tig* expression by the addition of 1 mg/ml arabinose at an  $\text{OD}_{600}$  of 0.5. This was followed by 16 h incubation at 25°C. Cells were harvested by centrifugation using Beckman Avanti J-20 XP with a JLA-16,250 rotor at 8,000 *g* for 15 min.

#### **2.2.5.2. Expression of His-tag free TFPf**

Expression of His-tag free *TFPf* was carried out by inoculating ZYP-5052 autoinducing medium containing chloramphenicol (34  $\mu\text{g/ml}$ ) with an overnight culture to a starting  $\text{OD}_{600}$  of 0.05. The culture was grown at 37°C with shaking at 250 rpm until it reached  $\text{OD}_{600} = 1$ , when cultures were moved to 25°C for *tig* expression for 16 h. Cells were harvested by centrifugation using a Beckman Avanti J-20 XP equipped with a JLA-16,250 rotor at 8,000 *g* for 15 min.

#### **2.2.5.3. Expression of CzrB**

Expression of *CzrB* was carried out by inoculating autoinducing media (ZYP-5052, PASM-5052, Power Prime Broth) containing appropriate antibiotics with an overnight culture to a starting  $\text{OD}_{600}$  of 0.05. The culture was grown at 37°C with shaking at 250 rpm until it reached an  $\text{OD}_{600}$  of 0.5-1. Cultures were moved to 25°C for *czrB* expression for 24-48 h, after which cells were harvested by centrifugation as described above.

## **2.2.6. Protein extraction from bacterial cells**

### **2.2.6.1. Soluble protein extraction**

*E. coli* cell pastes were frozen after harvesting, thawed on ice and resuspended in buffer A (50 mM sodium phosphate pH 7.5, 150 mM NaCl) containing protease inhibitor cocktail (Roche Diagnostics), 0.3 kU/  $\mu$ L of rLysozyme (Invitrogen) and 5% CelLytic reagent (Sigma Aldrich). The suspension was incubated for 30 min at room temperature, after which DNase I was added to 50  $\mu$ g/ml and MgCl<sub>2</sub> to 3 mM, followed by 10 min incubation on ice. Supernatant containing soluble protein was collected after 10 min centrifugation at 30,000 *g* and 4°C. Sodium chloride and imidazole were stirred into extracts to concentrations of 500 mM and 20 mM, respectively. Cell extracts were passed through a 0.45  $\mu$ m filter and stored at 4°C for protein purification.

### **2.2.6.2. Extraction of CzcB using a thermal treatment**

After harvest, *E. coli* cells were frozen, followed by thawing on ice. The cell paste was resuspended in 25 ml of buffer (50 mM sodium phosphate pH 7.5, 150 mM NaCl) per g of wet pellet. The addition of 0.3 kU/ml rLysozyme, 50  $\mu$ g/ml DNase I, 3 mM MgCl<sub>2</sub> was followed by sonication with an amplitude of 30% for 45 min in a 30 sec on/off cycle. Afterwards, DDM powder was stirred into the cell extract to a final concentration of 1% and the suspension was incubated at 75°C for 2 h, followed by centrifugation for 30 min at 30,000 *g*.

### **2.2.6.3. Cell fractionation to collect cell membranes for CzcB extraction**

*E. coli* pellets were resuspended in 5 ml per g of wet pellet of PBS. Cells were broken down using a Constant Systems TS Series cell disruptor at 25 k psi. Cell extracts were centrifuged for 10 min at 8,000 *g* to remove unbroken cells and the supernatant was ultracentrifuged (Beckman, Ti45 rotor) at 41 krpm and 4°C for 1 h. Cell membranes were collected and homogenised in 5 ml PBS per pellet using a glass homogeniser (Fisher Scientific). Membranes were used for protein solubilisation in detergent or frozen in liquid nitrogen and stored at -80°C for further use.

#### **2.2.6.4. Extraction of C<sub>zr</sub>B from membranes**

Cell membranes were resuspended in binding buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, 20 mM imidazole) containing 1% DDM at 4°C. The suspension was centrifuged (Beckman, Ti45 rotor) at 41 krpm and 4°C for 1 h and the supernatant containing membrane proteins was collected.

#### **2.2.7. SDS PAGE**

Proteins were separated on 12-16% tris-glycine polyacrylamide gels in a dual mini slab chamber (ATTO) with a discontinuous buffer system, following a standard method (Laemmli, 1970). NuPAGE Novex Bis-Tris gels (12%; Invitrogen) were also used for separation of protein samples.

##### **2.2.7.1. Coomassie staining**

Following polyacrylamide gel electrophoresis, the gel was placed in distilled water and was heated for 1 min in a microwave at high power level. Then, the gel was cooled down by rinsing with distilled water. After removal of the distilled water, 5 ml of InstantBlue (Expedeon) coomassie stain was added and allowed to develop for 30 min up to 2 h.

##### **2.2.7.2. Western Blotting**

Following PAGE, proteins were transferred from gels onto PROTRAN Nitrocellulose Transfer Membrane (Whatman) using a semi-dry transfer apparatus at 60 mA for 45 min (single gel transfer) or 100 mA for 45 min (for two gels). Western blotting was carried out according to Burnette procedure (Burnette, 1981) including blocking with 5% skim milk powder solution (incubation from 1 h to overnight), detection with a murine monoclonal anti-polyHistidine antibody (1 h incubation) and development with TMB.

#### **2.2.8. Protein purification**

Purifications were carried out with the AKTAexpress system (GE Healthcare), peristaltic pump or manually with the use of a syringe and on Econo-Column or Econo-Pac columns (Bio-Rad) using a gravity-flow approach.

### **2.2.8.1. Purification of TFPf for structural studies**

#### **2.2.8.1.1. Immobilised Metal Affinity Chromatography (IMAC)**

IMAC purification was carried out on the AKTExpress system or using manual loading of the relevant column. Cell extracts (obtained as described in 2.2.6.1) were loaded onto a previously equilibrated 1-ml His-Tag HP column (GE Healthcare) at the rate of 1 ml/min. The protein was purified using an imidazole gradient from 20 mM to 500 mM. Fractions containing purified protein were collected and analysed on Coomassie-stained 12% SDS-PAGE gels and by western blotting.

#### **2.2.8.1.2. Removal of His-tag with TEV protease**

IMAC-purified TFPf was concentrated using Amicon Ultra-15 Centrifugal Filter Units (Milipore) with 10 kDa MWCO, followed by buffer exchange into 50 mM Tris-HCl pH 8.0, 0.5 mM EDTA and 1 mM DTT using PD-10 columns (Bio-Rad). A 5  $\mu$ l aliquot of AcTEV<sup>TM</sup> Protease (Novagen) at 10 U/ $\mu$ l was added to the 1 ml fractions of protein sample. His-tag cleavage was allowed to proceed for up to 16 h with gentle agitation at 4°C, followed by “reverse” IMAC on a His-Tag HP column to remove un-cleaved protein and His-tagged TEV protease, whereas the tag-free protein was collected in column flow-through.

#### **2.2.8.1.3. Ion-Exchange Chromatography**

IMAC-purified TFPf was concentrated as above, followed by buffer exchange into 20 mM Tris-HCl pH 8.0 using PD-10 desalting columns (Bio-Rad). Protein samples were applied to an equilibrated Resource<sup>TM</sup> Q ion-exchange column at a flow rate of 1 ml/min. Protein was eluted using a 0-1 M NaCl gradient at the same flow rate.

#### **2.2.8.1.4. Size Exclusion Chromatography**

IMAC-purified TFPf was concentrated as above and applied to a Superdex 200 Hi Load 16/60 column (GE Healthcare) previously equilibrated with the SEC buffer. Purification was performed using an AKTExpress system (GE Healthcare) at

a flow rate of 0.5 ml/min. Fractions containing TFPf were collected and analysed on Coomassie-stained 12% SDS-PAGE and by western blotting.

### **2.2.8.2. Purification of CzrB for structural studies**

#### **2.2.8.2.1. Immobilised Metal Affinity Chromatography (IMAC)**

Extracts from cell membranes containing CzrB were incubated overnight with Superflow Ni-NTA agarose (Qiagen) at 4°C with agitation, at the ratio of 1 ml of resin slurry per 100 ml of membrane extract. The resin was packed onto an Econo-Pac or Econo-Column (Bio-Rad) and washed with approximately 50 column volumes (CV) of binding buffer (50 mM sodium phosphate pH 7.5, 500 mM NaCl, 20 mM imidazole) in the presence of 0.1% UDM, 0.03% DDM, 0.3% DM, 0.26% NG, 1% OG (Anatrace). Washing continued with 20 CVs of buffer containing 50-100 mM imidazole and the protein was eluted with 5-10 CVs of buffer containing 500 mM imidazole. Fractions were analysed on Coomassie-stained SDS-PAGE. Fractions containing purified CzrB were pooled and concentrated using an Ultra-15 Centrifugal Filter Unit (Amicon) with 10, 30 or 50 kDa MWCO. Samples were concentrated for further purification by SEC.

#### **2.2.8.2.2. Size Exclusion Chromatography**

IMAC-purified and concentrated CzrB was applied onto a Superdex 200 HiLoad 10/300 column previously equilibrated with SEC buffer (50 mM Bis-Tris pH 6.8, 150 mM KCl) in the presence of 0.1% UDM, 0.03% DDM, 0.3% DM, 0.26% NG, 1% OG. SEC was carried out on AKTAprime at a flow rate of 0.3 ml/min and elution fractions from the peak detected by UV were analysed on Coomassie-stained SDS-PAGE and by western blotting.

### **2.2.9. Protein analysis**

#### **2.2.9.1. Protein concentration and quantification**

Purified proteins were concentrated using a Vivaspin500 with 30 and 50 kDa MWCO to final concentrations of 18-25 mg/ml (CzrB) and 10-30 mg/ml (TFPf), estimated by absorbance spectroscopy using a NanoDrop (Thermo Scientific) at 280

nm or a Bio-Rad DC<sup>TM</sup> Protein Assay Kit. The molecular extinction co-efficient of the proteins was calculated from their aminoacid sequences using: [<http://web.expasy.org/protparam/>] by ExPASy Bioinformatics Resource.

#### **2.2.9.2. Circular dichroism (CD)**

CD analysis was performed in the group of Professor Tewik Soulimane, University of Limerick using a Chirascan CD spectrometer and Chirascan Pro-Data acquisition software. Baseline spectra and protein spectra for TFPf and CzrB were collected in triplicates at wavelengths 180-280 nm with 4 s time points and 1 nm bandwidth using quartz suprasil (QS) cuvettes (0.1 mm path length). The baseline and protein spectra were separately averaged, and averaged baselines were subtracted from the relevant averaged protein spectra and smoothed using a Savitsky-Golay algorithm (Savitzky and Golay, 1964).

#### **2.2.9.3. TETRA detector array**

The ViscotekTDAmx (Malvern Instruments) system at MPL, Diamond Light Source was used for SEC analysis of the composition of purified CzrB. CzrB was analysed at 1 mg/ml and the analysis was carried out using a Superdex 200 10/300 GL column at 0.5 ml/min flow-rate. The column was equilibrated with the buffer in which the protein was suspended and bovine serum albumin (BSA) was used for calibration of the detectors. The detectors were in the following order: UV, RI (Refractive Index), LS (Light Scattering) and viscometer. Data obtained were analysed using OmniSEC software (Malvern Instruments).

#### **2.2.9.4. Thermofluor-based stability assay**

A Thermofluor-based assay was used to determine protein stability using a fluorescent dye: CMP (Invitrogen) was used for CzrB and Sypro Orange for TFPf to record a  $T_m$  shift. Several detergents (Cymal6, OG, NM, LDAO, DDM, UDM), as well as buffers of various pHs (6.0-8.0) and different salts (NaCl, KCl) were screened in order to observe the shift in  $T_m$ . A protein at a concentration of 1 mg/ml was incubated in the presence of 5% CMP and Sypro Orange in all buffers for 1 h. The assay was carried out using 96-well thin-walled PCR plates placed in the BioRad

DNA Engine Opticon 2 real-time PCR machine and samples were heated from 25-85°C with an excitation wavelength 387 nm and emission wavelength of 463 nm. The melting curves for the proteins were analysed using Opticon Monitor software to calculate thermal shifts.

## **2.2.10. Crystallisation**

### **2.2.10.1. Crystallisation of TFPf**

Crystallisation of TFPf was carried out by vapour-diffusion using sitting-drop 24-well plates (Hampton). Initially, trials at NUI Galway were set up with 1 µl of protein covered with 1 µl of precipitant over a 500 µl precipitant solution reservoir. Protein was used at a concentration of 10-30 mg/ml. Conditions screened included commercial screens JCSG Core Suite I, II, III, IV (Qiagen) and screens prepared *de novo*. Further trials were set up at MPL using a Mosquito (TTP Labtech) crystallisation robot with 100 nl protein and 100 nl precipitant drops over 30 µl reservoir solution. Conditions screened in this set-up included JCSG+ (Qiagen) and Wizard Classic 1, 2, 3, 4 (Rigaku).

### **2.2.10.2. Crystallisation of CzrB**

#### **2.2.10.2.1. Vapour-diffusion**

Crystallisation by vapour-diffusion was carried out using sitting-drop plates. Initially, trials at NUI Galway were set up in 24 well plates (Hampton) with 1 µl of protein covered with 1 µl of precipitant over a 500 µl precipitant solution reservoir. Protein was used at a concentration of 10 mg/ml. Conditions screened included those in Crystal Screen and MemFac Screen commercial screens from Hampton Research, as well as screens prepared *de novo*. Further trials were set up at MPL in 96 well-plates (Greiner CrystalQuick<sup>R</sup>) using a crystallisation robot as described above and at protein concentrations from 12-30 mg/ml. MemGold and MemStart from Molecular Dimensions were used for preliminary screening.

#### **2.2.10.2.2. Lipid Cubic Phase (LCP)**

Monolein lipid was used for formation of a bicontinuous cubic phase. Lipid prepared at MPL was stored at -80°C. LCP was formed using monolein melted at 40°C, mixed with aqueous buffer containing CsrB, purified and concentrated to 20 mg/ml, in 2/3 v/v protein to lipid ratio. Reconstitution was performed at 20°C using two gas-tight removable needle syringes joined by a coupler (Hamilton). The resulting lipid cubic phase mix was dispensed using a Mosquito crystallisation robot onto a glass sandwich plate, covered in a thick adhesive containing a 96-punched-hole spacer, in 35 or 50 nl drops covered with 1000 nl of precipitant.

#### **2.2.10.2.3. Data collection**

Prior to data collection, crystals were harvested and flash-frozen in liquid nitrogen. Diffraction measurements were performed at I04 and I24 beamlines at Diamond Light Source (DLS, Harwell, Oxfordshire, UK).

### **2.2.11. Construct design for HTP cloning and expression at Oxford Protein Production Facility (OPPF)**

#### **2.2.11.1. Primer design**

Primers were designed for In-Fusion cloning for insertion of genes into the pOPIN Vector Suite. Primers were designed to have a  $T_m \geq 68^\circ\text{C}$  and fewer than 45 nucleotides. Primer stocks were 100  $\mu\text{M}$  and they were used at 10  $\mu\text{M}$ .

#### **2.2.11.2. High-throughput PCR**

HTP PCR reactions were set up in 96-well plates using designed primers and KOD Xtreme Hot Start DNA Polymerase (Novagen). An annealing temperature of 68°C was applied throughout the 96-well plate. Aliquots of 5  $\mu\text{l}$  of each PCR product were analysed by electrophoresis in SybrSafe (Invitrogen)-stained 1.25% TBE agarose gels at 100 V for 30 min using the Electro-Fast Stretch 108 system (ABGene).

### **2.2.11.3. Purification of PCR products by magnetic beads**

PCR products were purified using an AMPure Magnetic Bead Purification approach (Beckman). Methylated DNA was broken down by DnpI enzyme (0.5  $\mu$ l/5 units) at 37°C for 60 min prior to the purification procedure. Magnetic beads (90  $\mu$ l) were added into each well of a 96-well plate containing the PCR reactions and mixed thoroughly, followed by incubation for 5 min for maximum recovery. Then 96-well plates were placed on a SPRIplate 96R magnetic platform (Beckman) for 5 min to separate the beads and bound DNA from solution. Magnetic beads were washed twice with 70% ethanol, allowed to dry, and DNA was eluted in 30  $\mu$ l of 10 mM Tris pH 8.0. Recovered PCR products were analysed using agarose gel electrophoresis.

### **2.2.11.4. In-fusion ligation**

Linearised pOPIN vectors with the corresponding inserts were transferred into a plate containing a lyophilised In-fusion reaction mix. Ligations were carried out for 30 min at 42°C in a Veriti thermocycler. Ligated vectors were immediately transformed by “heat shock” into *E. coli* OmniMaxII competent cells and 5 and 25  $\mu$ l aliquots of transformation reactions were plated on 24-well agar plates containing the appropriate antibiotics. Plates were incubated overnight at 37°C.

### **2.2.11.5. Plasmid purification**

Single colonies were picked from transformation plates and grown overnight at 37°C to make glycerol stocks and extract cloned vectors. High-throughput mini plasmid preparation was carried out using Wizard SV96 purification plates on a Bio-Robot 8000, yielding 100  $\mu$ l purified plasmid suitable for transformation into *E. coli* expression strains. Constructed and purified vectors were verified by PCR followed by agarose gel analysis.

### **2.2.11.6. Small scale expression screening**

Constructed and verified pOPIN Suite vectors were transformed into *E. coli* C41, C43, Rosetta pLacI and Lemo21 strains by “heat shock”. Small scale expression was carried out in 24-well deep well plates sealed with a gas-permeable adhesive seal. Expression was carried in PB broth using IPTG induction and auto-

induction in Overnight Express<sup>TM</sup> Instant TB Medium. Each well contained 3 ml of the relevant growth medium, which was inoculated with a 1/20 dilution of an overnight culture. Cultures were grown at 37°C to OD<sub>600</sub> 0.5, after which the temperature was reduced to 20°C and cultures were induced in PB broth by the addition of 1 mM IPTG.

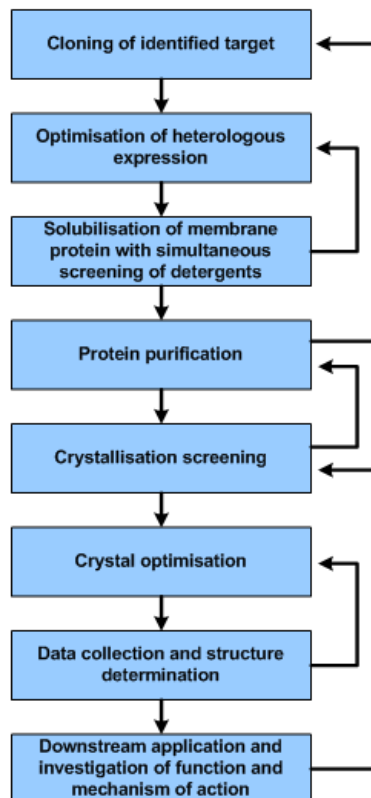
#### **2.2.11.7. Small scale expression analysis**

Following small scale expression described in section 2.2.11.6, cells were harvested by centrifugation at 6000 *g* for 10 min. Pellets were used for protein extraction and HTP IMAC purification. Recombinant proteins were extracted from bacterial cells using NPI-10 lysis buffer supplemented with 1% Tween, 1 mg/ml lysozyme, 3 units/ml Benzonase (and 1% DDM in case of membrane protein constructs) with shaking at 1000 rpm on an orbital shaker for 30 min at 4°C. After centrifugation at 5000 *g* for 30 min at 4°C the supernatant was collected and Ni-NTA magnetic beads were added to each well. The Ni-NTA HTP protein purification was carried out using a Bio-Robot 8000 and results of expression were analysed by Coomassie-stained SDS-PAGE.

### 3. RESULTS OF STRUCTURAL STUDIES OF CzrB FROM *THERMUS THERMOPHILUS*

#### 3.1. Production of CzrB for structural studies

##### 3.1.1. Construct design and expression optimisation



**Figure 12. The general pipeline involved in solving a protein structure.**

*It illustrates typical steps from target selection via protein preparation and crystallisation to data collection and structure determination.*

The ultimate goal of protein crystallography is to make a connection between the protein structure and its cellular function. A general protocol in obtaining high-resolution diffraction data is presented in Figure 12. This starts with target identification and cloning, through optimisation of expression, purification and crystallisation, arriving at data collection. Often one step in this pipeline determines the success of the next; therefore, it is not atypical to move back a number of steps in

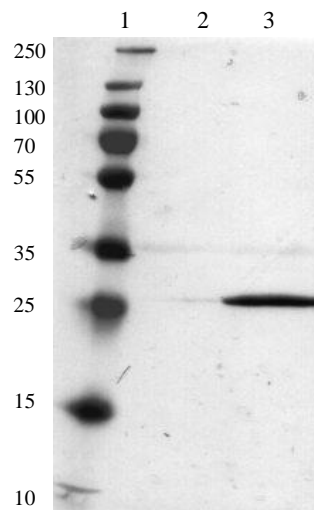
order to improve the success of the overall procedure (Chayen, 2004). Steps as early as construct design and cloning can affect the expression and purification process (Gräslund *et al.*, 2008), while solubilisation and purification procedures can highly influence the crystallisation stage (Kim *et al.*, 2011), so that revisiting earlier steps in the pipeline is frequently necessary to advance progress in obtaining high-quality diffraction data.

Access to high yields of membrane protein for structural studies is often the major bottleneck in structure determination. At the moment *E. coli* is one of the most successful systems used for overexpression of membrane proteins (Vinothkumar *et al.*, 2013). However, production of heterologous proteins in *E. coli* is still far from trivial and can prove toxic to the cell and lead to cell death. The last few years saw developments in the successful expression of membrane proteins in *E. coli* as a well-characterised, inexpensive and easy-to-maintain host (Sørensen and Mortensen, 2005a, Wagner *et al.*, 2008). Developments of particular note include the variety of plasmids (Cronan, 2006, Studier, 1991), mutant strains (Miroux and Walker, 1996) and recombinant fusion partners (Esposito and Chatterjee, 2006) now available for optimisation of expression studies.

Previous work in this group identified optimal conditions for production and purification of CzrB. The full-length *czrB* gene was amplified from *Thermus thermophilus* genomic DNA and cloned with a hexahistidine tag into the pET28b+ vector (Novagen). Two constructs were generated which varied in the position of the histidine tag (N- or C-terminal). Basic expression analysis led to the conclusion that the C-terminal histidine tag construct was expressed better in *E. coli* while investigation of a number of *E. coli* strains (C41, C43, BL21(DE3), BL21-CodonPlus-RIL) and expression temperatures revealed increased cell density and protein yield upon using *E. coli* BL21-CodonPlus-RIL, which contains supplementary tRNA's to better translate rare codons. Despite the fact that CzrB is a thermophilic protein, expression at 46°C and 37°C led to a decreased yield of the protein properly inserted into the cell membrane compared with 25°C (Kolaj, 2008).

In the early stages of this project, the previously optimised conditions described in section 2.2.5.3 were used to obtain protein for structural studies. Thus, CzrB was produced in *E. coli* BL21-CodonPlus-RIL (Stratagene) containing the

pET28b+CzrB/His vector for a 48 h autoinduction period, in ZYP-5052 medium and at 25°C (Figure 13), yielding around 7 mg of protein per 1 L of bacterial culture (Kolaj, 2008). Extra copies of the *argU*, *ileY* and *leuW* tRNA genes present in *E. coli* BL21-CodonPlus-RIL allow more efficient translation of the *T. thermophilus* gene. In addition, pET vectors are derivatives of pBR322 and have been designed to exploit features of the T7 bacteriophage expression system (Pan and Malcolm, 2000, Terpe, 2006). T7 RNA polymerase is several times more efficient than the *E. coli* enzyme at synthesising RNA and its expression is tightly regulated under the control of a T7 promoter (Studier, 1991). Moreover, the use of the autoinduction media ZYP-5052 results in higher density cultures and concomitant higher yields of target protein (Studier, 2005).



**Figure 13. Immunoblot analysis of CzrB production in autoinducing media ZYP-5052.** Lanes; 1: molecular weight marker (values in kDa are presented beside each protein band); 2: whole cell producing CzrB in ZYP5052 at  $OD_{600nm}$  1; 3: whole cell producing CzrB in ZYP5052 after 48 h.

### 3.2. Extraction of CzrB for structural studies

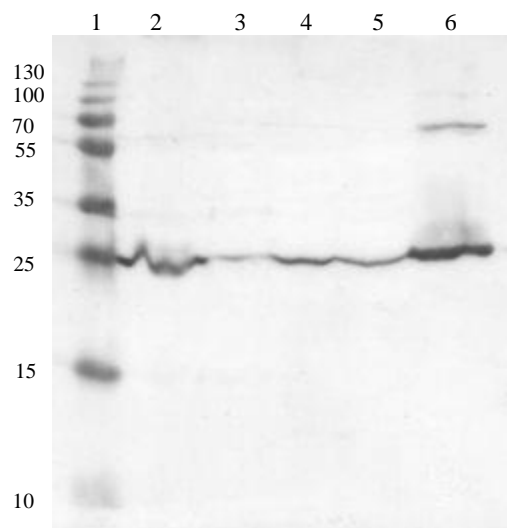
During the extraction of a membrane protein (see 1.6), one of the difficulties is dissolution of the membrane in which the protein is embedded so the molecules subsist in a monodisperse state in solution. Another problem is that membrane proteins are insoluble in aqueous solution; therefore, it is necessary to provide a condition mimicking their natural habitat to maintain them in a soluble

state. Solubilisation of membrane proteins is usually accomplished by the use of a range of detergents consisting of a polar head group and a hydrophobic tail, which tend to form micelles at a certain concentration. The critical micelle concentration (CCM) that is specific for each detergent is a key to ensure the detergent creates micelles in solution that are necessary for the stabilisation of membrane proteins. However, the efficiency of solubilisation often depends on the choice of detergent and is determined by their structure, size and ability to maintain proteins in their native state (Tulumello and Deber, 2012). The detergents that are most capable of solubilisation are often simultaneously harsh on protein molecules and cause protein denaturation, whereas milder detergents are not as efficient in maintaining protein solubility but are less hazardous for the protein's native state. Non-ionic detergents, especially those with longer alkyl chains like DDM and DM, are considered good at solubilising proteins as they break lipid-protein and lipid-lipid interactions while leaving protein-protein interactions untouched (Privé, 2007, Seddon *et al.*, 2004). Moreover, the minimum concentration of detergent that is necessary to break up the cell membrane is called the critical solubilisation concentration (CSC) and this is much higher than a detergent's CMC.

### **3.2.1. Extraction of CzrB by thermal treatment**

In order to release CzrB embedded in the membrane, harvested *E. coli* cells were lysed according to the protocol described in section 2.2.16.2. Membrane dissolution was examined visually after sonication and unbroken cells were separated from broken membranes and cell extract by centrifugation. Kolaj and co-workers (2007) previously optimised the concentration of detergent necessary for the successful extraction of CzrB and 1% DDM proved to be the minimal concentration needed to guarantee efficient solubilisation. DDM is the most commonly used detergent for protein extraction due to its mild character and low CMC (0.0087%), which allows the use of small amounts of this expensive compound to effectively solubilise membrane protein (Privé, 2007, Seddon *et al.*, 2004). Moreover, for the purpose of extraction, the sol-grade detergent (purity greater than 97% as determined by HPLC) can be used; the lower purity of the compound results in lower price.

The thermostable nature of the target protein allowed the solubilisation to be carried out at 75°C, which enhanced elimination of host proteins (Dąbrowski *et al.*, 2002, Lama *et al.*, 2001, Jimbo *et al.*, 2007). Denatured *E. coli* proteins were removed by centrifugation at 30,000 g for 30 min. Therefore, the thermal treatment, which was designed for protein extraction, also served as the first purification step by denaturing host proteins. This approach usually helps to obtain a high purity, homogenous protein preparation in a simple, one-step metal-affinity purification (Sørensen *et al.*, 2003).



**Figure 14. Immunoblot analysis of CzrB extraction using thermal treatment in the presence of 1% DDM.**

*Lanes; 1: molecular weight marker (values in kDa are presented beside each protein band), 2: whole cells producing CzrB, 3: soluble fraction after rLysozyme lysis, 4: soluble fraction after sonication, 5: soluble fraction after solubilisation with 1% DDM at RT, 6: soluble fraction after solubilisation with 1% DDM at 75°C.*

Figure 14 shows the significant increase in the yield of extracted CzrB (lane 6) after the high-temperature incubation in the presence of the detergent compared with the amount of membrane protein extracted post-sonication in 1% DDM at RT (lane 5). The analysis also reveals the presence of a higher molecular weight molecule (lane 6), which is thought to be a dimer of CzrB and is consistent with the solved structure of cytoplasmic domain (Cherezov *et al.*, 2008). Despite

that, the majority of extracted CzrB using the thermal treatment appears to be monomeric.

### **3.3. Purification of CzrB**

#### **3.3.1. Immobilised Metal Affinity Chromatography purification of CzrB**

Immobilised Metal Affinity Chromatography (IMAC) is the most widely used technique in purifying recombinant proteins for structural and functional studies. It is based on the affinity of a recombinant polyhistidine tag, attached during cloning to the protein of interest, for metal ions (i.e. Fe, Co, Cu, Ni, Zn) fixed to a solid resin (Porath *et al.*, 1975, Porath, 1992).  $\text{Zn}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ,  $\text{Co}^{2+}$  are all well-suited for the separation of His-tagged proteins but  $\text{Ni}^{2+}$  is the most commonly used.

To maximize the quantity of purified CzrB for structural studies, extracted CzrB was incubated with stirring with Ni-NTA (Qiagen) resin overnight at 4°C. Despite the thermophilic character of the target protein, all purification steps were carried out at 4°C to improve the stability of the protein in detergent micelles. Membrane proteins are usually stable while embedded in the cell membrane; however, solubilised membrane proteins in detergent micelles are prone to decreased stability (Privé, 2007).

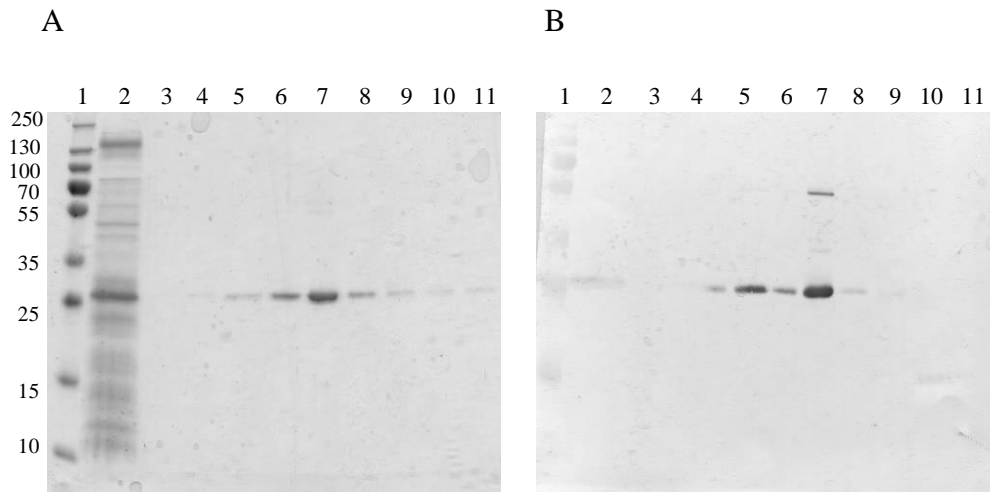
Even though DDM is commonly used to extract membrane proteins, it is preferable to exchange the proteins into a different detergent upon purification, due to the large (~70 kDa) micelles it creates. The optimisation of purification carried out previously in our group by IMAC included purification in the 12 different detergents listed in the Table 2 (Kolaj, 2008).

| Detergent (abbreviation)  | CMC     | Concentration |
|---|---------|---------------|
| n-dodecyl- $\beta$ -D-maltopyranoside (DDM)                             | 0.0087% | 0.05%         |
| n-undecyl- $\beta$ -D-maltopyranoside (UDM)                             | 0.029%  | 0.1%          |
| n-decyl- $\beta$ -D-maltopyranoside (DM)                                | 0.087%  | 0.2%          |
| n-decyl- $\beta$ -D-glucopyranoside (DG)                                | 0.07%   | 0.1%          |
| n-nonyl- $\beta$ -D-glucopyranoside (NG)                                | 0.2%    | 0.45%         |
| n-octyl- $\beta$ -D-glucopyranoside (OG)                                | 0.53%   | 1.1%          |
| n-nonyl- $\beta$ -D-thiomaltopyranoside (NTM)                           | 0.15%   | 0.3%          |
| n-heptyl- $\beta$ -D-thioglucopyranoside (HTG)                          | 0.85%   | 1%            |
| 5-cyclohexyl-1-pentyl- $\beta$ -D-maltoside (Cymal-5)                   | 0.12%   | 0.24%         |
| octaethylene glycol monododecyl ether (C <sub>12</sub> E <sub>8</sub> ) | 0.0048% | 0.01%         |
| n-undecylphosphocholine (FOS 11)*                                       | 0.062%  | 0.125%        |
| n-dodecylphosphocholine (FOS 12)*                                       | 0.047%  | 0.1%          |
| n-dodecyl -N,N-dimethylamine-N-oxide (LDAO*)                            | 0.023%  | 0.1%          |

**Table 2. Detergents used during optimising purification of CzrB (Kolaj, 2008).**

The last three detergents marked with \* are zwitterionic, whereas the rest of detergents in the table are non-ionic. CMC values are given in H<sub>2</sub>O for all detergents, with the exception of C<sub>12</sub>E<sub>8</sub> where values are given in a buffer (0.01 M TES, pH 7.5, 0.05 M NaCl, 0.1 mM CaCl<sub>2</sub>). Concentration refers to the concentration used during purification.

The results of optimising CzrB purification (Kolaj, 2008) revealed a high yield and purity of sample in most of the detergents evaluated, but a reduced yield was observed in case of FOS11 (2,8-Dimethyl-5-Nonylphosphocholine) and C<sub>12</sub>E<sub>8</sub> (Polyoxyethylene(8)dodecyl Ether). In addition, it was evident that the purity of the membrane protein was higher when purified in maltoside-based detergents (DDM, UDM, DM, NTM, Cymal-5), whereas samples purified in glucoside detergents (OG, NG, DG, HTG) showed the presence of an intense product at ~62 kDa. Optimal results were obtained when using undecyl- $\beta$ -maltoside (UDM), another non-ionic detergent, which differs from DDM in the length of the alkyl tail. Therefore, after extraction in DDM as described here, buffer exchange into UDM was carried out during IMAC purification. UDM has a CMC of 0.027%, which is much higher than that of DDM. The fact that UDM creates smaller micelles and consequently smaller PDCs provides a greater chance for protein-protein contacts to occur and so the protein has a higher chance of crystallising.

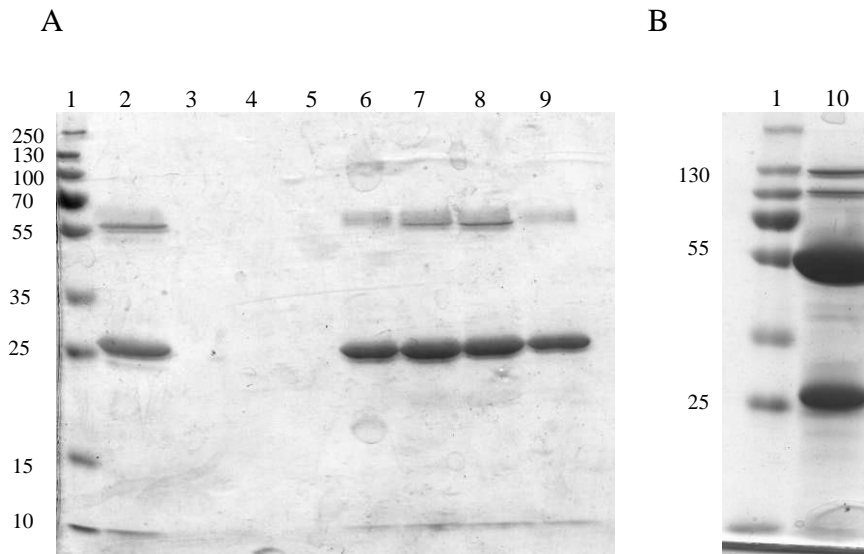


**Figure 15. SDS-PAGE (A) and Western Blot (B) analysis of CzrB purification in the presence of 0.9% UDM.**

*Lanes; 1: molecular weight marker marker (values in kDa are presented beside each protein band), 2: flow-through, 3-6: washes with 20 mM, 40 mM, 60 mM and 100 mM imidazole, respectively, 7-11: elution with buffer containing 500 mM imidazole.*

Figure 15 shows an analysis of CzrB purification in 0.09% UDM (see 2.2.8.2.1). To ensure the solubility of membrane proteins in detergent, their working concentration is kept at 3 x CMC. It was clear that washes with buffers containing 60 and 100 mM imidazole led to a loss of protein, thus requiring the use of a 50 mM imidazole in wash buffer. CzrB was eluted using buffers containing 250-500 mM imidazole to yield a high-purity product (Figure 15A). The structure of the cytosolic domain of CzrB previously solved in our group revealed that the domain exists as a homodimer (Cherezov *et al.*, 2008). Immunodetection of the purified full-length protein in this work consistently revealed a higher molecular weight product (Figure 15B, lane 7) that corresponded with the expected size of a CzrB dimer.

Eluted fractions were pooled, concentrated and applied to a PD-10 desalting column to reduce the concentrations of salt and imidazole, and to remove an excess of detergent (Figure 16A).



**Figure 16: SDS-PAGE analysis of buffer exchange (A) and concentration of CzrB (B).** Lanes; 1: molecular weight marker (values in kDa are presented beside each protein band), 2: purified CzrB, 3-5: PD-10 column flow-through, 6-9: elution of CzrB from PD-10, 10: concentrated CzrB.

CzrB was then re-concentrated to 10 mg/ml using a Vivaspin500 with a 30 kDa molecular weight cut-off (Figure 16B). Concentrating membrane proteins is usually advised to be carried out with a 100 kDa cut-off to ensure the removal of detergent. As the molecular weight of the CzrB monomer is too low, even surrounded by detergent micelle, attempting its concentration with a 100 kDa cut-off membrane resulted in loss of the protein in this study (data not shown). Successfully concentrated CzrB appears on SDS-PAGE largely as a pure mixture of monomer and dimer, with this higher molecular mass molecules present (Figure 16B).

### 3.4. Initial crystallisation trials of CzrB

Growing protein crystals for structure analysis is a major obstacle in solving structures of membrane proteins. While the technique originated as a method for protein purification, many criteria have to be met in order to obtain diffraction-quality crystals, especially when working with membrane proteins. The presence of detergents that are necessary to solubilise membrane proteins also causes difficulties in crystallisation because detergent micelles make the formation of contacts between protein molecules more challenging (Gutmann *et al.*, 2007).

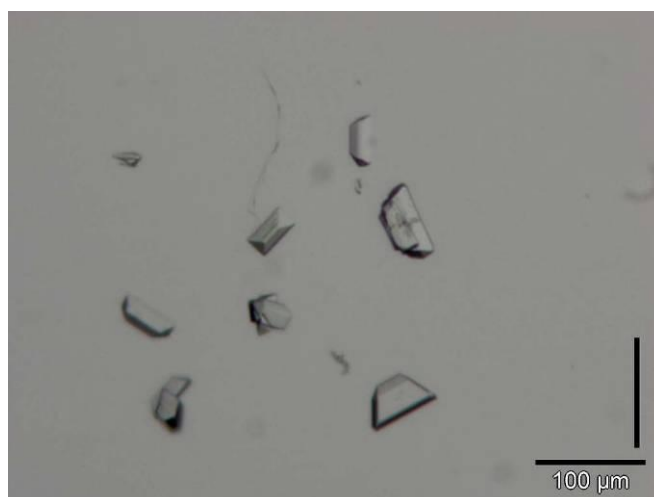
Crystallisation trials were set up with CzrB after simple IMAC purification and concentration due to the high purity of the protein. The high purity of the sample was most likely related to the thermal treatment during extraction of the sample prior the fractionation and led to the elimination of *E. coli* proteins as previously discussed (see 3.2.1).

Crystallisation trials were carried out using vapour-diffusion (see 2.2.10.2.1), which is the most commonly used technique in protein crystallisation (Chayen and Saridakis, 2008). The main appeal of the approach lies in its uncomplicated set up, ease of visualisation and ease of harvesting crystals. The role of crystallisation buffers is to create suitable conditions for crystal formation; buffers vary in their precipitant or other additives, concentrations of constituents and their pH. Initial CzrB crystallisation trials were carried out in 24-well sitting-drop crystallisation plates (Hampton) using conditions previously identified to produce crystals, with buffers prepared *de novo* and commercial screens. Screens customised for membrane proteins typically prefer polyethylene glycol (PEG) as their main precipitant, whereas a wider range of precipitants is utilised in kits designed for crystallisation of soluble proteins.

Conditions that previously produced crystals of CzrB (Kolaj, 2008) were used to design a screen oscillating around those conditions to repeat crystal growth with the protein purified in UDM. The original conditions were: 50 mM NaOAc pH 4.0, 25% (w/v) PEG 1000 or 50 mM Tris-HCl pH 8.0, 25% (w/v) PEG 1500. Therefore, in this work, crystal growth was tested in 50 mM NaOAc, MES, HEPES and Tris-HCl at pH values from 4.0 to 8.0, in the presence of 10-25% PEG 1000 and PEG 1500. These conditions led to the occurrence of either precipitation or phase separation in most wells but no crystals were observed. Addition of 100 mM KCl, which had previously been found to increase the stability of crystals, did not lead to the formation of detectable crystals.

Conditions that yield crystals can differ significantly with different preparations of the same protein, which may have been the reason for the difficulties encountered in reproducing previously obtained results. It is advisable, therefore, to use a single protein prep to optimise crystal formation but this is not always feasible due to low yields of protein typically obtained during purification.

To identify new conditions suitable for crystal formation, a broader screen was performed using the commercial Crystal Screen (50 unique conditions) and MembFac (48 unique conditions) screens (Hampton); the latter screen was designed specifically for preliminary screening of membrane proteins. Sitting-drop vapour-diffusion trials were set up in 24-well plates with CzcB freshly purified in 50 mM Tris pH 8.0, 0.1% UDM buffer and concentrated to 10 mg/ml. Plates were incubated at 20°C. Crystals were observed within 3 weeks in the well containing 0.1 M ADA (N-(2-Acetamido) iminodiacetic Acid) pH 6.5, 1 M  $(\text{NH}_4)_2\text{HPO}_4$ . The crystals were trapezoidal (Figure 17), 50  $\mu\text{m}$  to 75  $\mu\text{m}$  in length, stable and well-shaped, and characterised by very sharp edges and low solvent content.



**Figure 17. Crystals observed in UDM-purified CzcB crystallisation trials carried out at 20°C.**

*Crystals were obtained in the following conditions using MembFac: 0.1 M ADA pH 6.5, 1 M  $(\text{NH}_4)_2\text{HPO}_4$ .*

However, it did not prove possible to reproduce the crystals by re-screening original or similar conditions. Professor Tewik Soulimane confirmed them to be salt crystals upon X-ray diffraction analysis at Swiss Light Source (SLS).

### **3.5. High-throughput approach to CzcB crystallisation in Membrane Protein Laboratory (MPL) at Diamond Light Source (DLS), Oxfordshire, UK**

A collaboration was established between our group and the Membrane Protein Laboratory (MPL) group based at the Diamond Light Source (DLS), Oxfordshire, UK. MPL is a research group and training facility exclusively dedicated to working with membrane proteins and with a proven record of successfully determining structures of membrane proteins (Drew *et al.*, 2008a, Shimamura *et al.*, 2010, Newstead *et al.*, 2011, Hu *et al.*, 2011, Rengachari *et al.*, 2012). MPL, a joint initiative of DLS and Imperial College London, has attracted many talented scientists, led by Professor So Iwata, a world-widely recognised crystallographer. The MPL facility allows researchers to advance their own membrane protein research by accessing the high-throughput technologies available at MPL.

Work carried out in MPL was co-funded by grants from the Society for General Microbiology (SGM) and NUI Galway's Thomas Crawford Hayes Fund (TCH).

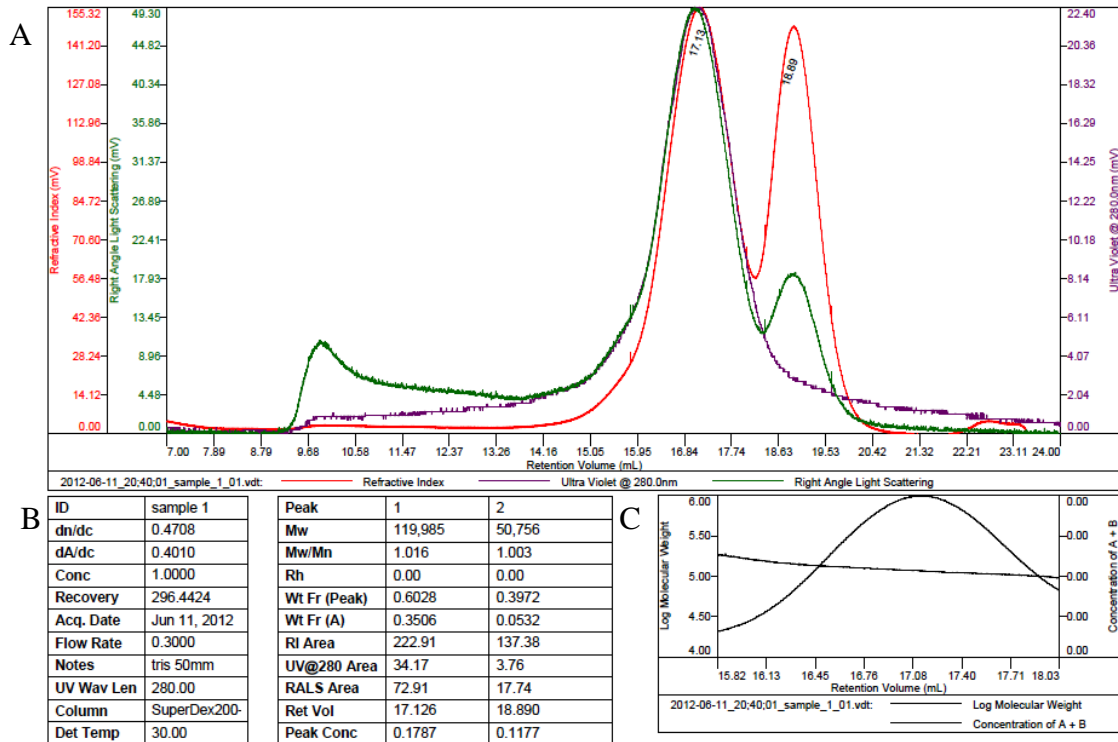
#### **3.5.1. TETRA-detector based analysis of protein**

Membrane proteins constitute roughly 30% of all proteins in cells, carry out many essential cell functions and often are leading candidates for drug targets. Access to high-resolution data on membrane protein structure is a growing need. The past decade saw advances in techniques in cloning, expression, purification and crystallisation of membrane proteins in HTS (high-throughput screening) systems (Gutmann *et al.*, 2007). Membrane protein crystallography is now the most valuable tool in accelerating knowledge of the atomic structure of membrane proteins.

Determining the appropriate conditions for crystal formation is time-consuming and requires large amounts of protein. Automation of the crystallisation set up through the use of crystallisation robots speeds up the process and requires much less protein, therefore allowing screening of many more conditions due to dealing in nanolitre quantities (DeLucas *et al.*, 2003, Stock *et al.*, 2005). Often however, even screening of hundreds of conditions does not identify parameters suitable for crystal growth. In such cases, the focus must be re-directed to the major

variable in protein crystallography – the protein itself. Access to a pure, homogenous and stable protein is the first challenge that must be addressed in the process of structure determination.

Before undertaking a HTP approach to identify appropriate crystallisation conditions for CzrB, UDM-purified CzrB was analysed using a Triple Detector Array (TDA) with additional UV detector (“TETRA”) to determine its quality. The detectors incorporated into TETRA are Refractive Index (RI), Right Angle Light Scattering (RALS), Viscosity Detector and Ultraviolet Photodiode Array (UV-PDA). A single Size-Exclusion Chromatography (SEC) run, coupled with TETRA analysis, can provide very useful details of the concentration, intrinsic viscosity, molecular weight and molecular size of a protein. RI and UV detectors provide details of the concentration of components in the sample, which is especially crucial with membrane proteins in order to avoid excess free detergent in the protein solution. A viscosity detector, which is used to determine the intrinsic viscosity, allows detection of protein-protein interactions. The light scattering detectors determine the size distribution within the protein solution to assess the monodispersity of the sample. The TETRA analysis of CzrB is described in section 2.2.9.3 and the result is presented in the Figure 18.



**Figure 18. TETRA detector analysis of CzrB in 0.1% UDM.**

Chromatogram peaks represent detection with UV (purple), RALS (green), RI (red) (A), table on the left provide information about the run including sample concentration, flow rate and sample, table on the right provides values to calculate molecular weight; [it is calculated by multiplying MW and Wt Fr(A), which indicates the percentage of the MW that is fraction A (protein of interest)] (B), log of MW against NWF (normalised weight fraction) (C).

TETRA detector analysis revealed that CzrB eluted as a broad, but nonetheless fairly well resolved and symmetrical peak, which was detected by UV. The sample also exhibited a slight excess of detergent, which was apparent from RALS detection of a second peak. The presence of the second peak indicates that CzrB in protein-detergent complex (PDC) and UDM elute at different volumes providing separation between the free micelles and PDC's. On the other hand, the MW of the main population of the protein was calculated at around 42 kDa (Figure 18B), which is inconsistent with predicted MW of CzrB (~32 kDa). The shift in the molecular weight from that estimated from the sequence might be due to the presence of a heterogeneous population and weak resolution between two forms of molecules, which was previously shown by SDS-PAGE analysis. Moreover, Figure

18C, a log of MW against normalised fraction, is a measure of reliability of the data, and the slope in the log MW indicates that the MW is not stable across the peak.

### **3.5.2. Crystallisation trials at MPL**

CzrB purified in the presence of 0.09% UDM and analysed on the TETRA detector (Figure 18) was used for crystallisation. In order to remove aggregates that might have been generated during thawing, the protein was centrifuged for 1 h at 18,000 *g* and 4°C. Crystallisation trials were set up as described (see 2.2.10.2.1, 2.2.10.2.2) using a Mosquito crystallisation robot, which guaranteed a very accurate distribution of nanolitres of protein and crystallisation buffers, ensuring precision and intra-experiment reproducibility. The initial HTP screen of crystallisation conditions was set up with MemGold and MemStart commercial crystallisation screens (Molecular Dimension), which are specifically designed for membrane protein crystallisation. These combine the most commonly used conditions that generated diffraction-quality crystals, according to published results in the Protein Data Bank (PDB) (Berman *et al.*, 2000). Both screens contain 96 unique conditions and cover a wide range of buffers, pH, PEGs and additives. Crystallisation plates were incubated at 4°C and 20°C. Almost 400 conditions were screened using only 52 µl of protein (12-30 mg/ml), which consumed approximately 8 times less protein than would have been required to set up the same experiment manually.

### **3.5.3. Thermofluor-based stability assay**

One of the major determinants in successful crystallisation is the stability of the target protein. Solubilisation of membrane proteins often leads to their destabilisation since the formation of a protein-detergent complex is a poor replacement for a lipid bilayer. Determination of optimal buffer composition and identification of possible ligands to enhance the protein stability increases the chances of yielding protein crystals. In order to increase the probability of crystal formation from CzrB, the stability of the protein was investigated using a microscale fluorescent thermal stability assay for membrane proteins. The basis for the assay is

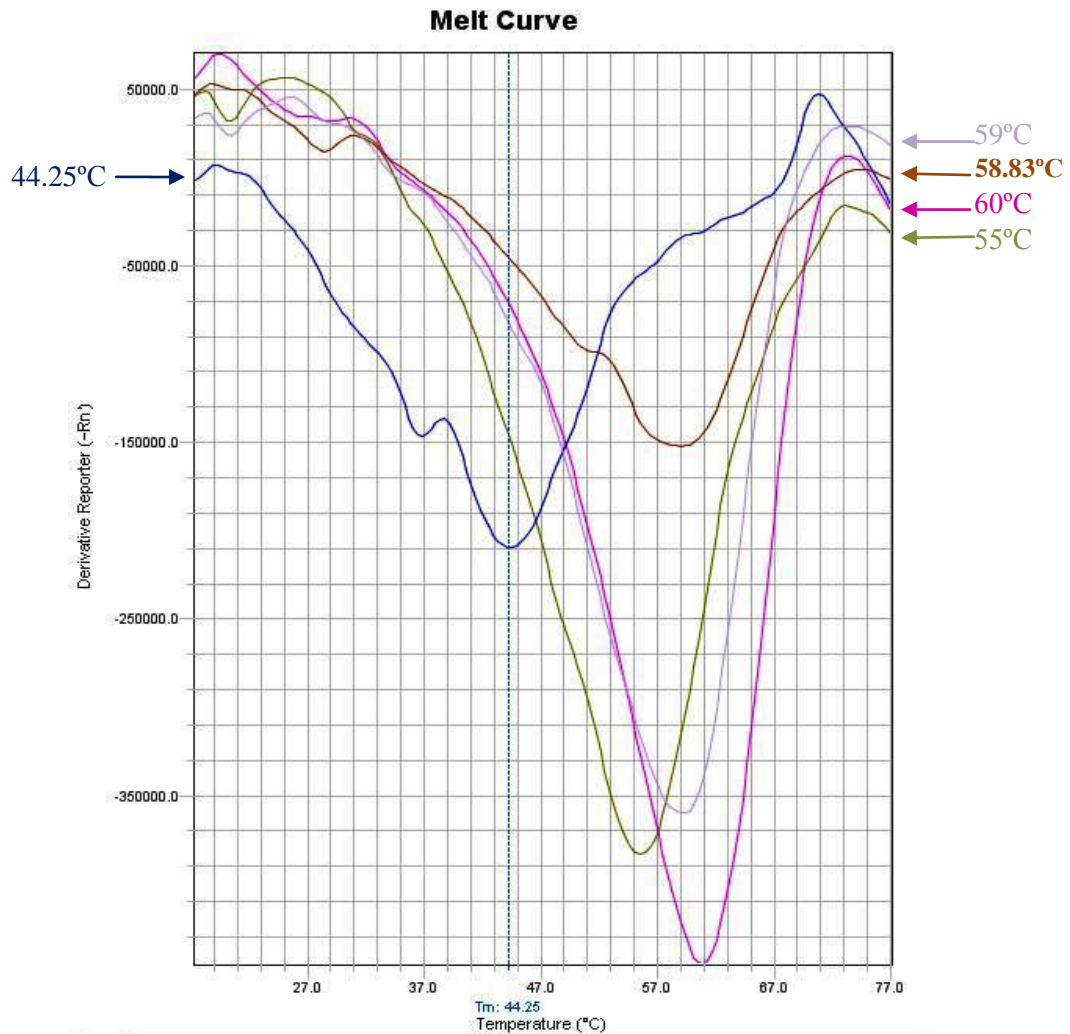
the binding of the fluorescent dye N-[4-(7-diethylamino-4methyl-3-coumarinyl) phenyl] maleimide (CMP) to cysteines in the protein and measurement of its melting temperature ( $T_m$ ), which serves as an indicator of the general integrity of the folded protein state (Alexandrov *et al.*, 2008).

Initially, CzrB purified in 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 0.09-0.1% UDM exhibited a melting temperature of 44°C (Figure 19). The  $T_m$  of the protein was then determined in several detergents (Cymal-6, LDAO, OG, DDM, NM and UDM), in buffers of different pH (6.0-8.0) and in the presence of added salts (NaCl, KCl). CzrB (10 mg/ml) was incubated for 1 h in each buffer prior to addition of CPM dye diluted 1:20 in the appropriate buffer (Table 3).

| Screen    | Buffer      | Buffer composition                             | Results:<br>$T_m$ (°C) |
|-----------|-------------|--|------------------------|
| Detergent | 1           | 50 mM Tris pH 8, 0.09% Cymal 6                 | 47.5                   |
|           | 2           | 50 mM Tris pH 8, 0.07% LDAO                    | 44.7                   |
|           | 3           | 50 mM Tris pH 8, 1% OG                         | 44.68                  |
|           | 4           | 50 mM Tris pH 8, 0.26% DDM                     | 45                     |
|           | 5           | 50 mM Tris pH 8, 0.9% NM                       | 46.7                   |
|           | 6 (control) | 50 mM Tris pH 8, 0.1% UDM                      | 44.25                  |
| Salt      | 7           | 50 mM Tris pH 8, 0.1% UDM, 50 mM NaCl          | Inconclusive           |
|           | 8           | 50 mM Tris pH 8, 0.1% UDM, 100 mM NaCl         | 42                     |
|           | 9           | 50 mM Tris pH 8, 0.1% UDM, 200 mM NaCl         | Inconclusive           |
| pH        | 10          | 50 mM Tris pH 7.5, 0.1% UDM                    | 48.9                   |
|           | 11*         | 50 mM BisTris pH 6.8, 0.1% UDM                 | 58.83                  |
|           | 12*         | 50 mM BisTris pH 6, 0.1% UDM                   | 60                     |
| pH + salt | 13*         | 50 mM BisTris pH 6.8, 0.1% UDM,<br>100 mM NaCl | 55                     |
|           | 14*         | 50 mM BisTris pH 6, 0.1% UDM,<br>100 mM NaCl   | 59                     |
|           | 15          | 50 mM Tris pH 7.5, 0.1% UDM,<br>100 mM NaCl    | 43                     |

**Table 3. Conditions used to investigate the stability of CzrB.**

*Asterisks indicate conditions presented in Figure 19.*



**Figure 19. Summary of the most significant  $T_m$  shifts using the Thermofluor assay.**

*Melt curves demonstrate the shift in  $T_m$  upon buffer change. The control  $T_m$  at 44.25°C was measured in 50 mM Tris pH 8, 0.1% UDM. The buffers leading to increase of  $T_m$  are as follows:  $T_m$ : 59°C- 50 mM BisTris pH 6.8, 0.1% UDM,  $T_m$ : 58.83°C-50 mM BisTris pH 6.8, 0.1% UDM,  $T_m$ : 60°C- 50 mM BisTris 6, 0.1% UDM,  $T_m$ : 55°C- 50 mM BisTris pH 6.8, 0.1% UDM, 100 mM NaCl.*

The Thermofluor assay revealed that the CzrB protein in 50 mM Bis-Tris pH 6.8, 0.1% UDM demonstrated the most significant shift in its melting temperature – from 44°C to 60°C. Generally, a shift in melting temperature when screening different detergents can be substantial, from 24°C in NG to 44°C in DDM in case of Apelin G Protein Coupled Receptor (APJ) reported by Alexandrov (2008). However, a less ample shift of up to several degrees is observed upon the change of pH and up

to 10 degrees shift is observed with increasing concentration of salt (Alexandrov *et al.*, 2008). Therefore, during size exclusion chromatography on a Superdex200 column, the CzrB protein was exchanged into this buffer to improve its stability ahead of setting up crystallisation trials. When trials were set up, crystals formed rapidly but were very temperature sensitive and “melted” under the light of the microscope. These crystals could not be reproduced, even after numerous attempts.

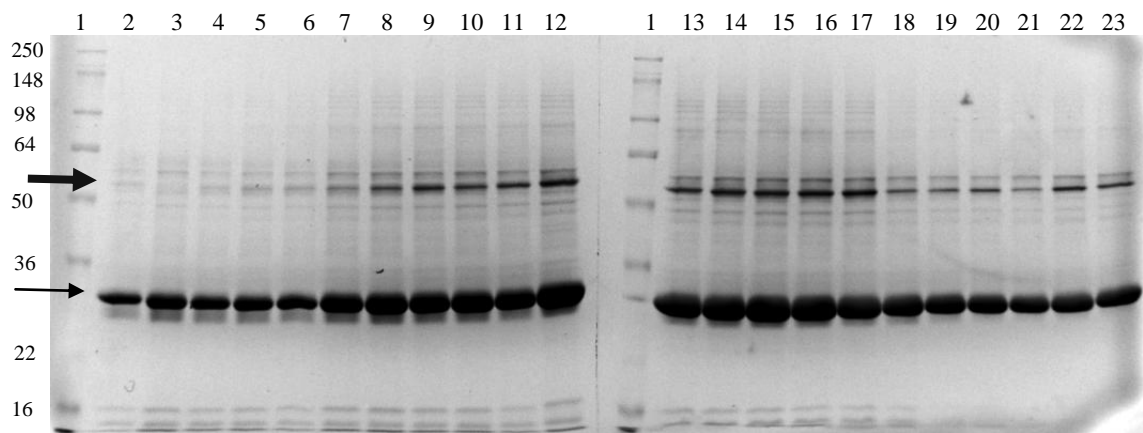
#### **3.5.4. Cell lysis using the cell disruptor**

As discussed previously (see 3.1.1, Figure 12) it is often necessary in protein crystallisation to modify earlier, possibly apparently successful, stages in the procedure in order to achieve a positive outcome later in the process. In this case, due to the lack of crystallisation “hits” obtained, a new method for preparing CzrB was adopted.

At the NUI Galway lab, bacterial cells were ruptured using a combination of freeze-thaw, enzymatic lysis and sonication methods to release the protein from the cell membrane: cells were frozen and then thawed on ice, followed by sonication accompanied by enzymatic lysis using rLysozyme (as described in 2.2.6.2). At MPL, cells were ruptured using a Constant System cell disruptor, which ensures very effective disruption by pressurising cells at up to 25 kpsi and passing them through a small diameter nozzle at high velocity (see 2.2.6.3). Sonication is widely used in medicine, research and industry; however, it is less effective in disrupting of cell membranes than cell disruptors. Moreover, sonication is also associated with the formation of aggregates with properties resembling amyloids (Stathopoulos *et al.*, 2004), a structure commonly adopted by non-disease and disease-associated proteins upon observed destabilisation of a protein’s native state. The occurrence of local high-temperature, mechanical shearing and the presence of free-radicals during sonication can all contribute to disruption of a protein’s native state, resulting in a loss of activity (Carrion-Vazquez *et al.*, 2000, Peng and Li, 2009). Peternel and Komel showed that the extraction of protein from inclusion bodies using sonication can damage their structures, thus decreasing their biological activity, whereas high-pressure homogenisation has very little negative effect on protein biological activity (Peternel and Komel, 2010). Sonication also shears DNA

released from bacterial cells, eliminating the necessity of using enzymes to degrade it, however, it is essential to use DNase enzymes to destroy nucleic acids when using homogeniser in order to reduce the viscosity of the cell extract.

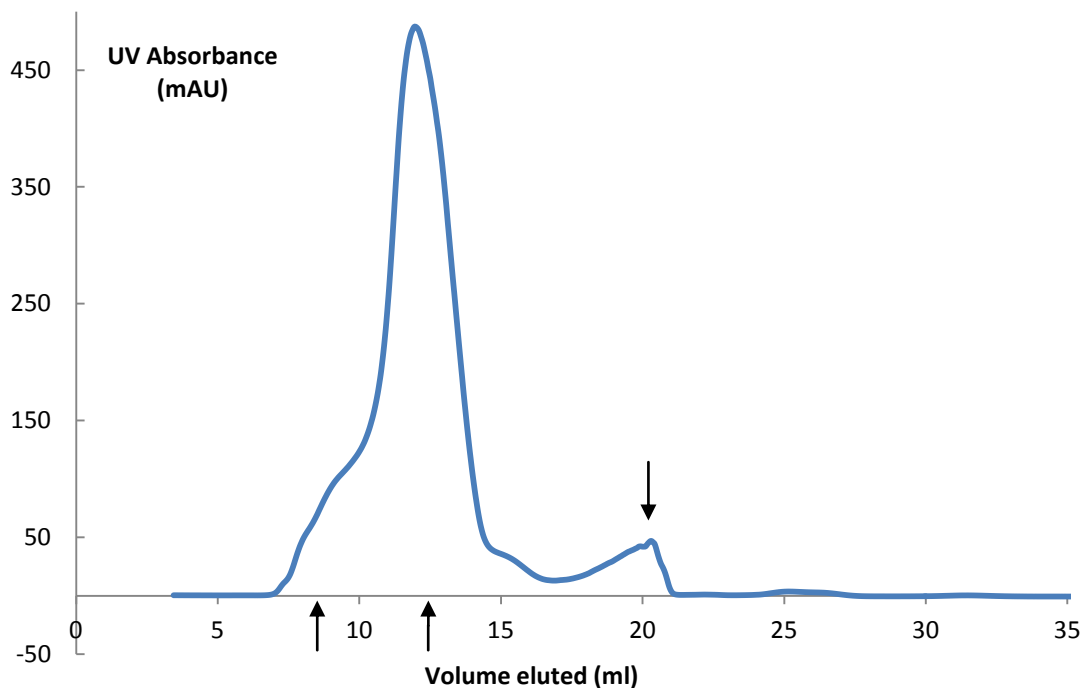
CzrB was solubilised from bacterial membranes formed using a high-pressure cell disruptor by incubating at 75°C for 2 h in the presence of 1% DDM, which was followed by IMAC purification described in 2.2.8.2.1. Figure 20 is a summary of fraction eluted from Ni-NTA Superflow resin.



**Figure 20. Coomassie-stained SDS-PAGE analysis of IMAC purification of CzrB extracted using a cell-disruptor and high-temperature treatment.**

*Lanes: 1: molecular weight markers (values in kDa are presented beside each protein band), 2-23: 1 ml fractions eluted by 250 mM imidazole buffer. The light arrow indicates monomeric CzrB and the thicker arrow a product corresponding to the size of the dimeric protein.*

Analysis of CzrB fractions purified by IMAC showed that the eluted protein is relatively clean and the majority of it appears to be in the monomeric form (Figure 20). However, one of the critical factors in successful crystallisation is obtaining not only pure but also conformationally homogenous protein. Therefore, eluted 1 ml fractions (lanes 2-23, Figure 20) were pooled and concentrated to carry out further purification by SEC (Size Exclusion Chromatography) on Superdex200 in order to separate the oligomeric forms (Figure 21).

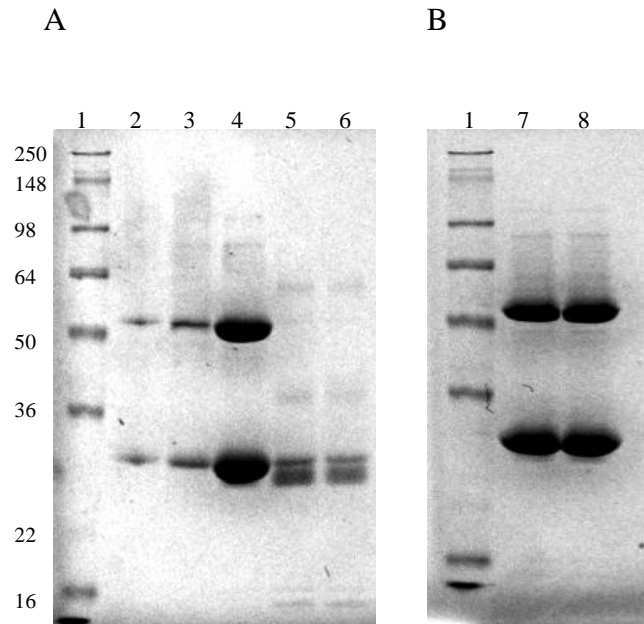


**Figure 21. SEC-based purification of CzrB in the presence of 0.1% UDM.**

*Elution of UV profile for SEC purification of CzrB carried out at 0.4 ml/min, the arrows indicate the fractions pooled for analysis on the SDS-PAGE (Figure 22) at 8ml, 12 ml and 20 ml.*

The purification by size exclusion on Superdex200 resulted in a largely symmetrical peak with an initial considerable shoulder and a significant tail. The former suggested the possibility of aggregation or poor resolution between higher oligomeric states of CzrB, whereas the tail can indicate protein degradation but may also result from overloading of one of the retention mechanism of the column. The fractions throughout the peak were analysed by SDS-PAGE (see Figure 22A), which showed that even samples corresponding to the apex of the SEC peak (lane 4, Figure 22A) were not homogenous but consisted of similar amounts of monomeric and dimeric proteins. Failure to separate different oligomeric forms of CzrB might be caused by a detergent “belt” around protein molecules that interferes with the support material of the column and causes a delay in elution. Figure 22B illustrates that the monomer-dimer ratio did not change upon heating of the sample. The inability to separate monomeric and dimeric populations of CzrB might be expected to have an

impact on crystallisation trials, because non-homogeneous populations may inhibit the formation of protein crystals.



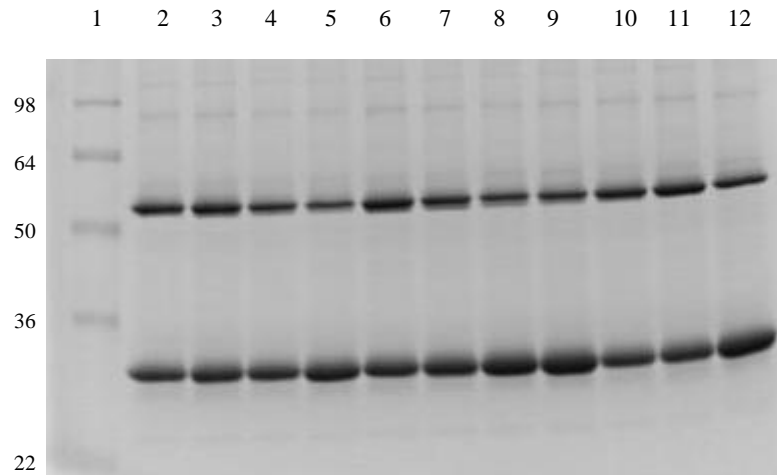
**Figure 22. SDS-PAGE analysis of fractions from SEC purification of CzrB in 50 mM Bis-Tris pH 6.8, 0.1% UDM (A) and analysis of an eluted sample before and after thermal denaturation (B).**

*Lanes; 1: molecular weight markers (values in kDa are presented beside each protein band), 2-3: fractions pooled from the “shoulder” before the main peak, 4: fraction from the top of the peak, 5-6: fractions pooled from the “tail” following the main peak, 7: SEC purified CzrB after incubation at 80°C for 30 sec, 8: SEC purified CzrB without heating.*

### 3.5.5. Effect of different buffers on dimerisation

The tendency of proteins to aggregate into oligomeric forms is influenced by factors such as the protein’s native structure or the chemical environment in which the protein was prepared or stored. In order to determine if the dimerisation of CzrB is influenced by buffer composition, several different buffer conditions were tested. The identification by Thermofluor assay of conditions that indicated a higher stability of the protein (50 mM Bis-Tris pH 6.0, 100 mM NaCl and 0.1% UDM; (Figure 19)) could indicate that dimer formation is favoured, thus increasing the stability of the CzrB protein. CzrB that had previously been purified and concentrated to 10 mg/ml in the above conditions was incubated for 1 h at a 1:20

ratio (v/v) of protein to buffer, followed by analysis of the oligomeric state of the protein by SDS-PAGE (Figure 23).



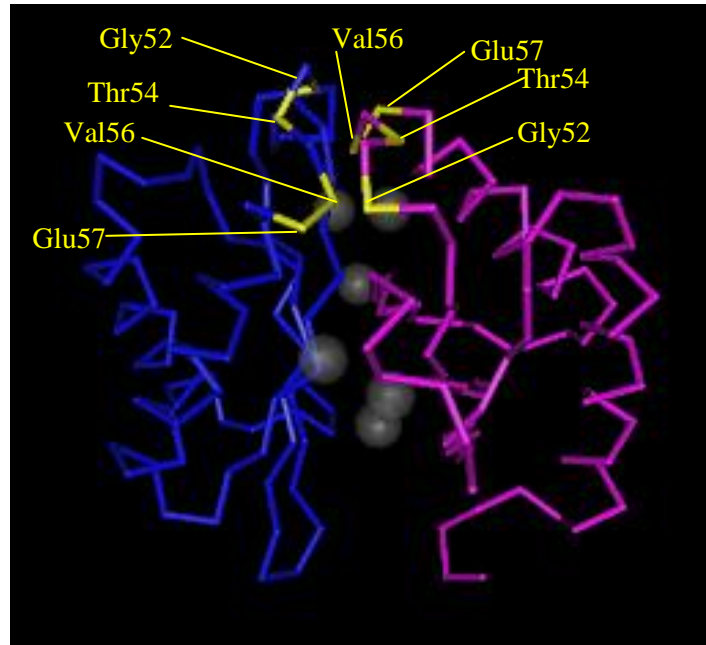
**Figure 23. SDS-PAGE analysis of the influence of buffer composition on CzrB dimerisation.**

Lanes; 1: molecular weight marker (values in kDa are presented beside each protein band), subsequent lane numbers represent buffers used in the dimerisation test; 2: 50 mM Tris-HCl pH 8.0, 3: 50 mM Tris-HCl pH 7.5, 4: 50 mM Bis-Tris pH 6.8, 5: 50 mM Bis-Tris pH 6.8, 6: 50mM Tris-HCl pH 8.0, 50 mM KCl, 7: 50 mM Bis-Tris pH 6.8, 100 mM NaCl, 8: 50 mM Bis-Tris pH 6.8, 100 mM NaCl, 9: 50 mM Bis-Tris pH 6.8, 50 mM KCl, 10: 50 mM Tris-HCl pH 8.0, 50 mM NaCl, 11: 50 mM Tris-HCl pH 7.5, 50 mM NaCl, 12: 50 mM Bis-Tris pH 6.0, 100 mM NaCl.

No significant differences were observed in the proportion of dimer formation, suggesting that pH changes and the addition of salts do not have a major influence on the dimerisation of CzrB.

The basis for dimer formation by CzrB was determined upon solving the structure of its cytoplasmic domain (Cherezov *et al.*, 2008): Glu57(N)-Gly52(O) and Val56(N)-Thr54(O) hydrogen bonds, and the highly hydrophobic region involving Val50, Pro55, Val56, Ala59, His60, Val83 and Pro85 results in a distinctive, inverted V-shape to the dimeric molecule. A structure of the homodimer assembly of the CzrB cytoplasmic domain upon zinc binding is shown in figure 24. The zinc binding leads to the two protomers, which are splayed apart in the apo form, snap together along the flat surfaces of the opposing ellipsoids (Cherezov *et al.*, 2008). The fact that the dimerisation does not involve disulfide bridges eliminates the possibility of

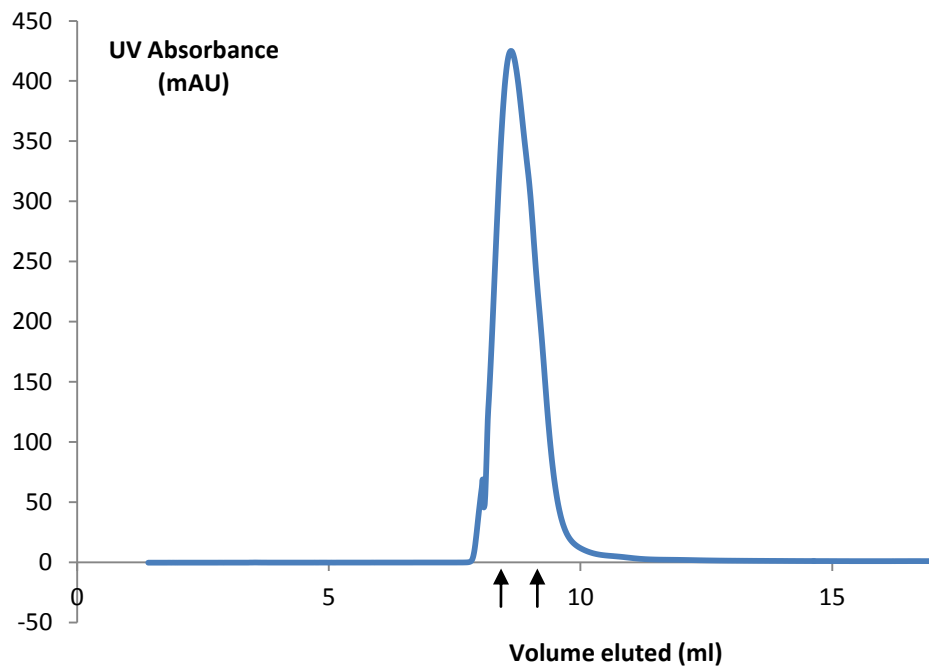
using a reducing agent like TCEP or DTT (Burns *et al.*, 1991, Chao and Fu, 2004) to obtain a monodisperse population.



**Figure 24. Structure of cytoplasmic domain of CzrB with bound zinc.**

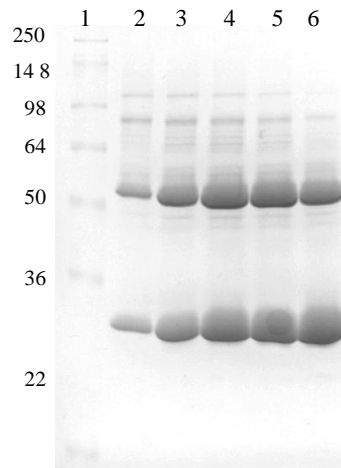
*Monomers are presented in blue and purple, the pairs of aminoacids involved in hydrogen bonds are marked in yellow and zinc ions are shown as grey spheres.*

Another approach investigated to separate monomeric and dimeric CzrB forms was the use of a Superdex75 (SD75) size exclusion column (Figure 25). SD75 is designed to separate proteins of molecular size from 3 to 70 kDa, while the molecular size of monomeric CzrB in UDM is thought to oscillate around 70 kDa.



**Figure 25.** UV chromatogram representing SEC analysis of CzrB purification on SD75. Arrows indicate the region of the peak from which samples were analysed by SDS-PAGE (see Figure 14).

While the size exclusion chromatogram revealed a symmetrical peak for CzrB, with no noticeable shoulder indicating the occurrence of aggregation or the presence in the sample of a higher molecular weight population (Figure 25). SDS-PAGE analysis nevertheless revealed the presence of both monomeric and dimeric protein in samples collected from SEC (Figure 26).



**Figure 26. SDS-PAGE analysis of the Superdex75 SEC purification.**

*Lanes; 1: molecular marker (values in kDa are presented beside each protein band), 2-5: 0.2 ml eluted fractions sampled from the top of the peak (within arrowed region) in Figure 24.*

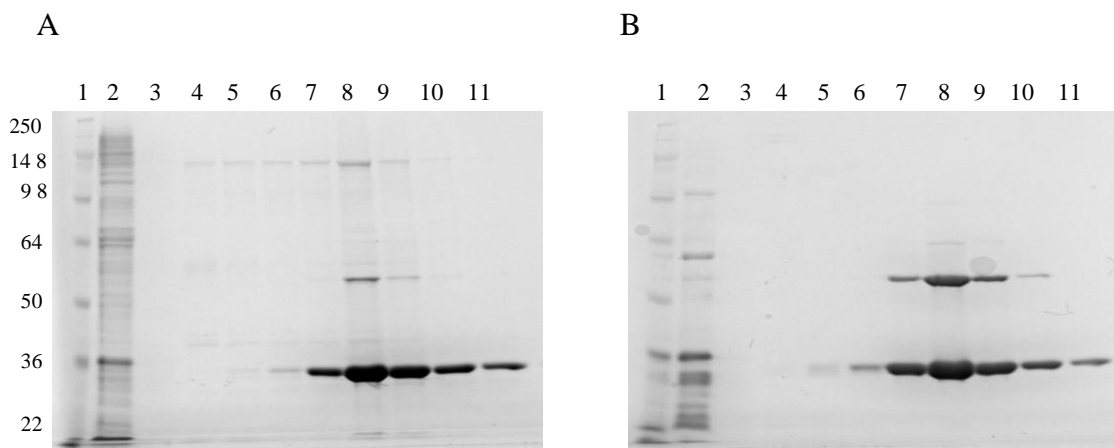
Figure 26 indicates that size exclusion chromatography appeared to increase the dimer:monomer ratio in the sample, as if the higher purity of the protein favours dimer formation by CzrB. SDS-PAGE analysis of membrane proteins is also known to be unable to break detergent micelles, however, so the possibility remains that CzrB exists in solution as a dimer and the equilibrium observed in Figure 26 is a consequence of inefficient reduction by SDS. The previous observation that the majority of the protein occurred in monomeric form (Figure 20) might also have resulted from denaturation of dimeric proteins into monomers during the thermal treatment used for solubilisation, whereas further purification steps allowed the protein to reassume its natural, dimeric state.

### 3.5.6. Extraction of CzrB from membranes after fractionation

The inability to separate the oligomeric forms of CzrB and successfully crystallise the protein raised concerns about the methods used for its extraction from cell membranes, and particularly the possible effect on the stability of CzrB of the harsh thermal treatment conditions used (see 2.2.6.2), notwithstanding its thermophilic origin (Spada *et al.*, 2002).

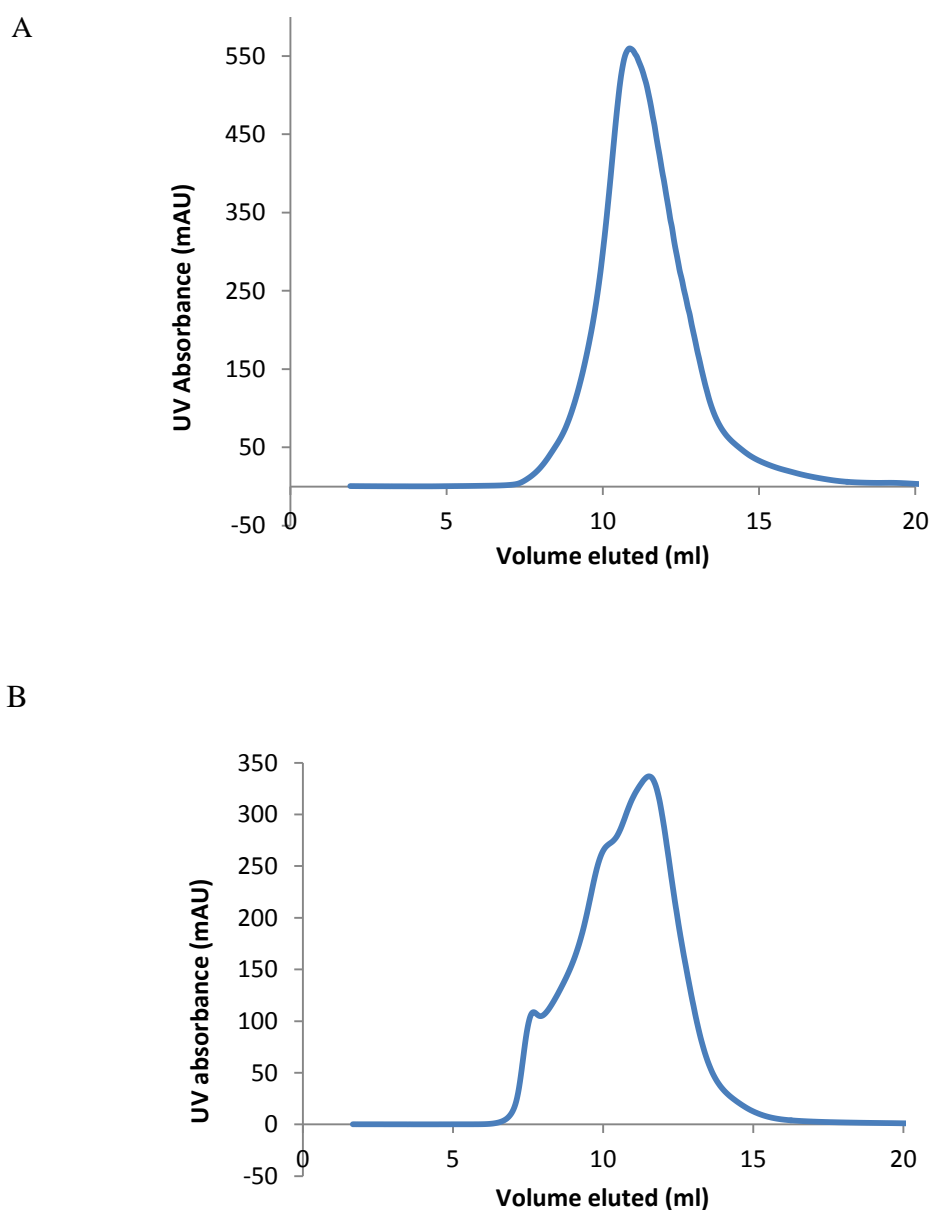
A method of cell fractionation by ultracentrifugation and extraction of CzrB from the membrane fraction (see 2.2.6.3, 2.2.6.4; (Drew *et al.*, 2008b)) was

investigated to separate soluble proteins from the membrane proteins. Fractionation also ensures the isolation of fully functional, membrane-embedded proteins from proteins successfully translated but not resident in the cell membrane. Furthermore, the use of a cell disruptor allows better recovery of the recombinant protein and tighter regulation of sample temperature than in the case of sonication, with a low and constant temperature throughout the extraction process expected to benefit protein stability. IMAC purification of protein samples extracted using the two methods (Figure 27) led to the conclusion that fractionation and extraction in 1% DDM at 4°C slightly favoured formation of the dimer (Figure 27B). This in turn may indicate that increased stability of CzrB correlates with dimer formation.



**Figure 27. SDS-PAGE analysis of comparison of purification of CzrB extracted by thermal extraction from cell extracts (A) and by cell fractionation followed by extraction at low temperature (B).**

*Lanes; 1: molecular weight markers (values in kDa are presented beside each protein band), 2: flow-through, 3-5: washes with 20 mM, 50 mM, 75 mM and 100 mM imidazole, respectively, 7-11: elution with buffer containing 250 mM imidazole*

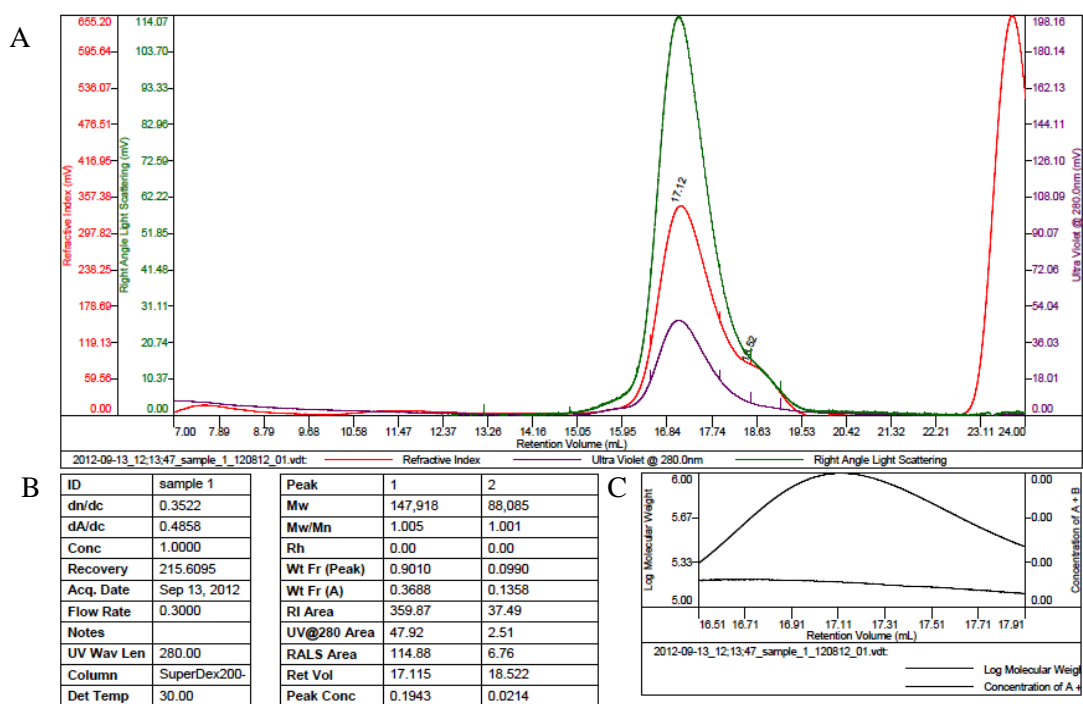


**Figure 28. SEC purification analysis of CzrB extracted using a thermal treatment (A) and at 4°C (B).**

*Comparison of elution profiles for SEC purification of CzrB extracted by different methods carried out at 0.4 ml/min. The eluted peaks are at their maximum at 11 ml (A) and 12 ml (B).*

Further purification with SEC showed that protein extracted using a thermal treatment produced a more symmetrical peak, suggesting the presence of a more homogenous sample (Figure 28). Based on these results, protein extracted using the thermal method (Figures 27A and 28A) should be more suitable for

crystallisation, yet no crystallisation hits were obtained in during extensive crystallisation screen. In order to complete the assessment of the new CzrB extraction approach, the protein extracted at low temperature was also analysed by TETRA detector, which revealed that the majority of the CzrB protein existed in dimeric form (Figure 29).



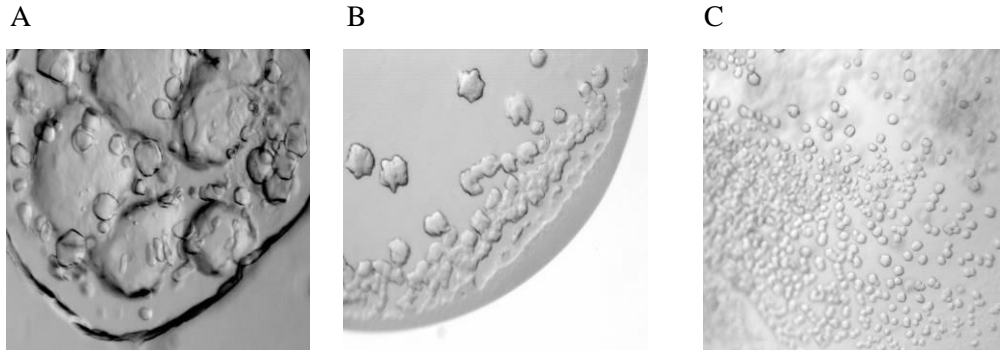
**Figure 29. TETRA detector analysis of CzrB, prepared without thermal treatment, in the presence of 0.1% UDM.**

Chromatogram peaks represent detection with UV (purple), RALS (green), RI (red) (A), table on the left provide information about the run including sample concentration, flow rate and sample, table on the right provides values to calculate molecular weight; [it is calculated by multiplying MW and Wt Fr(A), which indicates the percentage of the MW that is fraction A (protein of interest)] (B), log of MW against NWF (normalised weight fraction) (C).

### 3.5.7. Crystal analysis by X-ray

After setting up crystallisation trials with protein prepared as described in sections 2.2.6.3, 2.2.6.4, preliminary hits were observed in vapour diffusion and LCP trials (Figure 30). In the vapour diffusion plates, crystals were characterised by a high sensitivity to change in temperature and low stability, resulting in “melting” of

the crystals Figure 30B and C), whereas crystals obtained in LCP exhibited poorly defined edges but were more stable than those obtained using the vapour diffusion approach (Figure 30A).



**Figure 30. Preliminary crystallisation hits observed within 2 days in drops set up by LCP (A) and vapour-diffusion (B and C).**

*Crystals were observed in:*

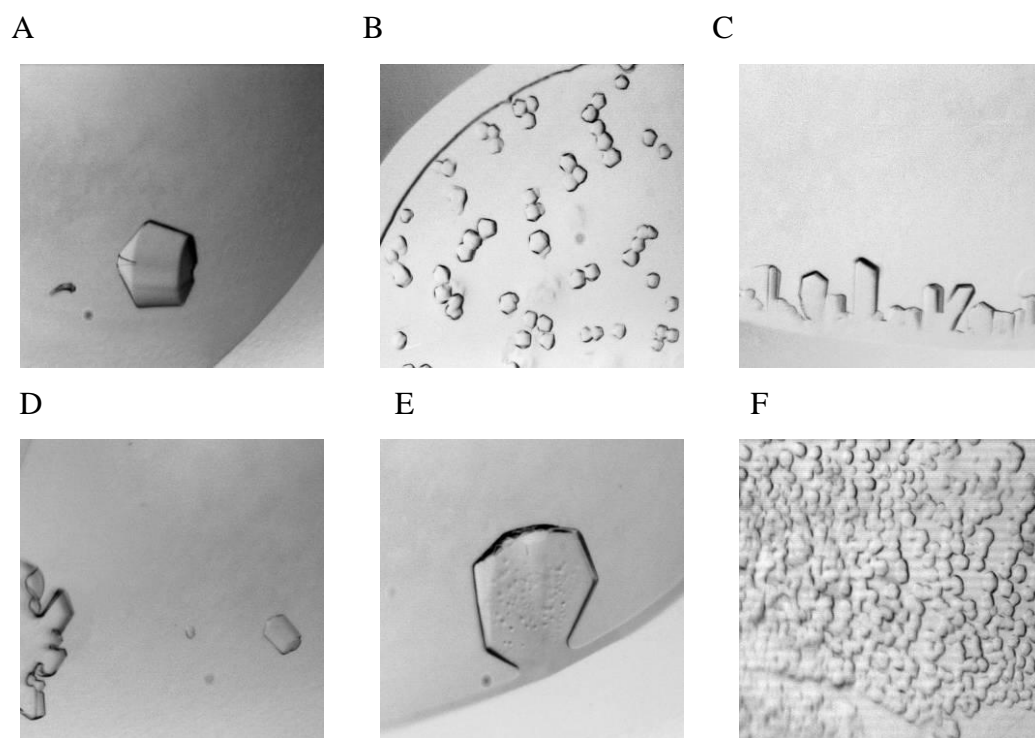
*A: 0.1 M magnesium chloride, 0.1 M HEPES pH 7.5, 30% PEG 400;*

*B: 2.75 M ammonium chloride, 0.025 M Bis-Tris pH 7.0;*

*C: 24% PEG 1500, 0.1 M sodium cacodylate pH 6.5.*

The instability of the former crystals suggested that further optimisation of crystallisation or even the protein preparation process was necessary. Only well-ordered crystals can produce a conclusive diffraction pattern (Chiu *et al.*, 2000, Palm and Colman, 1974) and the unstable, non-uniform shaped crystals observed are most likely the result of poor packaging of protein molecules within the crystals.

Upon incubation of the crystallisation plates for a number of weeks, however, bigger, more stable and better defined crystals were observed in a number of screening conditions (Figure 31).



**Figure 31. Examples of crystals grown from UDM-purified CzrB at 20°C during the HTC crystallisation screening.**

*Crystals were obtained in several conditions:*

*A: 31% pentaerythritol ethoxylate, 0.2 M ammonium formate, 0.1 M Tris pH 7.0;*

*B: 2.75 M ammonium chloride, 0.025 M Bis-Tris pH 7.0;*

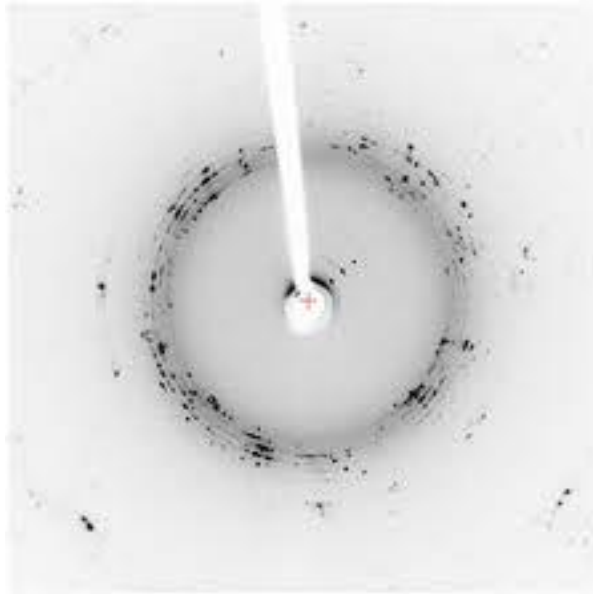
*C: 0.1 M sodium chloride, 0.1 M MES pH 6.5, 30% PEG 400;*

*D: 18% PEG 4000, 0.1 M potassium chloride, 0.1 M Bis-Tris pH 6.0;*

*E: 0.05 M magnesium acetate, 0.05 M sodium acetate pH 5.0, 28% PEG 400;*

*F: 22% PEG 3000, 0.25 M magnesium formate, 1 M sodium cacodylate pH 6.5.*

Harvesting of crystals was carried out by Dr Isabel Moraes of MPL, following which 16 crystals were “shot” using the I024 X-ray diffraction beamline at the Diamond Light Source. Analysis of the diffraction pattern revealed them to be of very poor resolution, which was assumed to be caused by protein-detergent or detergent crystals as the recorded reflection spots also appeared to be smudged instead of the more regular and sharp spots typically associated with protein crystals (Figure 32).



**Figure 32. An example of a diffraction pattern from UDM-purified CzrB crystals obtained.**

*Data was collected at I24, Diamond Light Source (DLS), Oxford, UK. Data was acquired using Pilatus3 6M detector.*

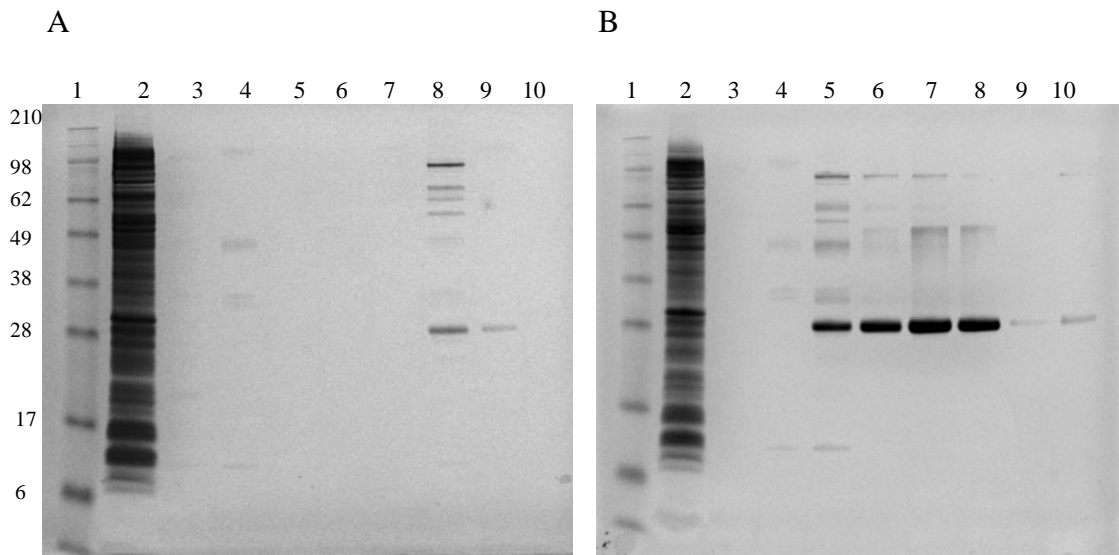
### **3.5.8. Purification of CzrB in DDM and DM**

Due to the lack of success in X-ray analysis, which indicated the presence of either protein-detergent or detergent crystals, the UDM used for purification up to this point in the work was now changed for other detergents. The presence of detergent crystals is consistent with results obtained with TETRA detector (see Figure 18) and reports that maltoside-based detergents (DDM, DM, UDM) tend to concentrate in 30 kDa and 50 kDa cut-off concentrators (Urbani and Warne, 2005).

A new approach was undertaken to replace UDM with DDM and DM. DDM and DM belong to the same group of detergents, non-ionic maltosides that differ from UDM in the length of the alkyl chain. UDM has a chain of 10 carbons while DDM's chain is longer and DM's shorter by one carbon each. Choosing quite similar detergents to UDM allows the introduction of less drastic changes, which can result in the improvement of the protein stability and crystal formation (Wang *et al.*, 2003, Wei *et al.*, 2004).

Preparation of CzrB in the two new detergents was carried out according to the same extraction and purification protocols discussed previously (see 3.5.4,

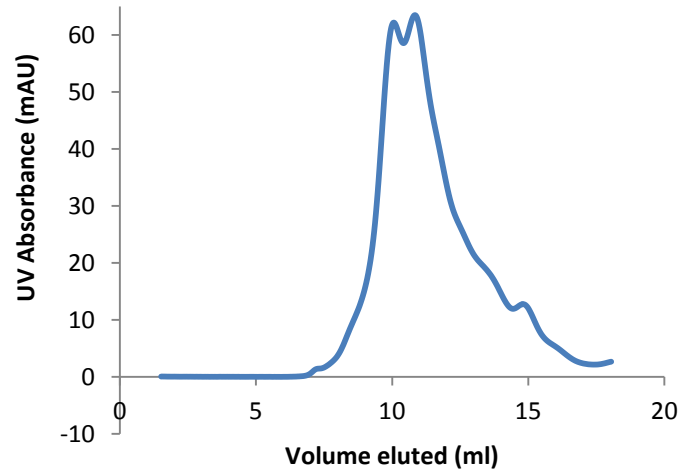
Figure 33). Therefore, it was surprising to see that the use of DDM resulted in a very low yield of purified protein after IMAC (Figure 33A).



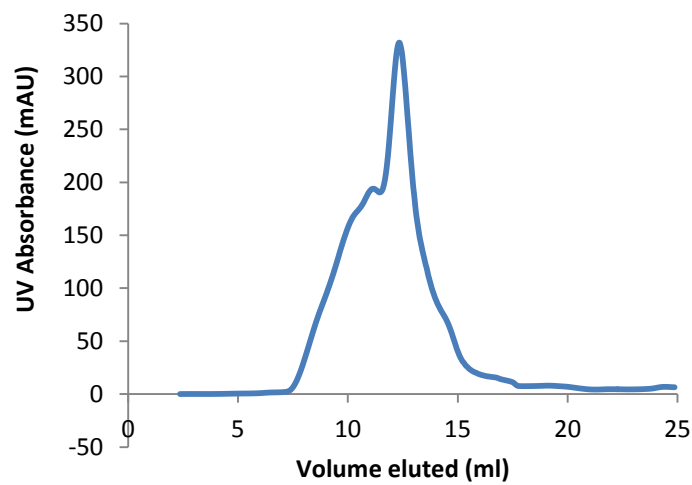
**Figure 33. CzrB purification in the presence of 0.03 % DDM (A) and 0.26 % DM (B).** Lanes; 1: molecular weight markers (values are presented in kDa beside each protein band), 2: flow-through from Ni-NTA resin, 3-6: washes with buffer containing 20 mM, 50 mM, 75 mM and 100 mM imidazole, 7-11: elution with buffer containing 250 mM imidazole.

Further purification of the protein by size exclusion (Figure 34) in the presence of 0.03% DDM or 0.26% DM revealed a highly asymmetric peak in the former case, but with some indication of possible resolution of dimeric and monomeric forms of the protein (Figure 34A). SEC of DM-purified CzrB resulted in a sharp peak with an extensive front tail, most likely indicating considerable aggregation of the protein. Detergents with shorter alkyl chains, such as DM, tend to be of harsher nature and may lead more often to protein instability (Privé, 2007, Tate, 2010).

A



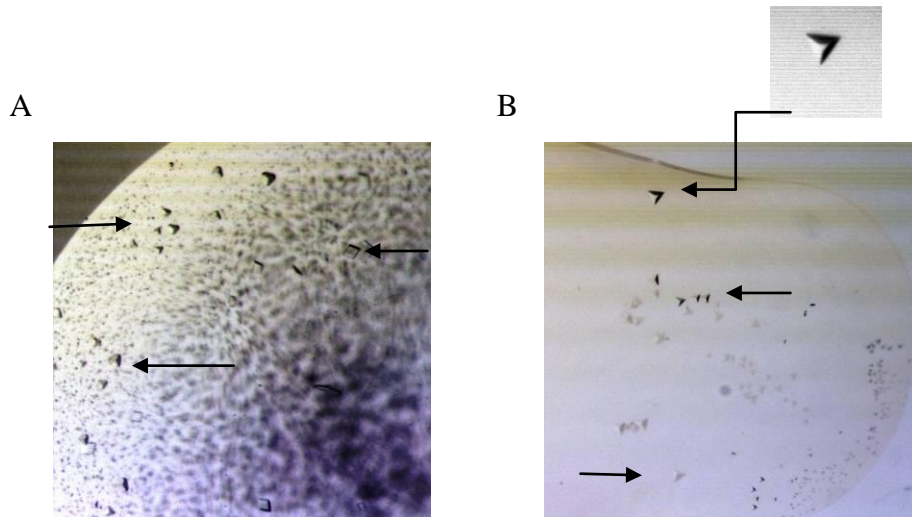
B



**Figure 34.** SEC analysis of CzrB purified in the presence of DDM (A) and DM (B).

Nevertheless, the best resolved and most central elution fractions were collected from the SEC analysis and concentrated for crystallisation screens. Trials were set up by vapour diffusion in 96-well sitting drop wells using MemGold and MemSys (see 3.5.2).

Within two days, two MemGold conditions produced mostly triangular and some rhomboid crystals in plates set up with CzrB purified in 50 mM Bis-Tris pH 6.8, 150 mM KCl, 0.03% DDM buffer (Figure 35).



**Figure 35. Triangular crystals observed in crystallisation trials set up with CzrB purified in DDM.**

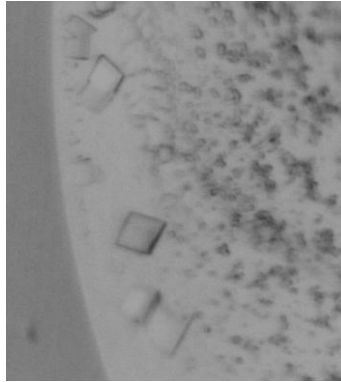
*Crystals were obtained in two conditions:*

*A: 0.1 M sodium chloride, 0.12 M Tris pH 9.4, 20% PEG 400;*

*B: 0.36 M sodium chloride, 0.1% w/v sodium azide, 0.015 M sodium phosphate pH 7, 9% PEG 400.*

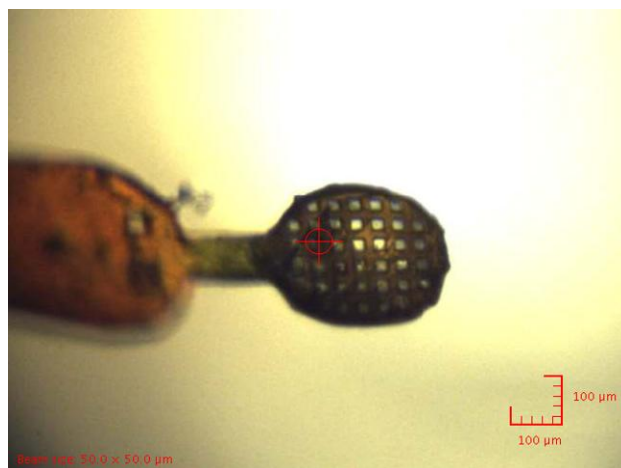
Efforts to reproduce the crystals involved purification of CzrB in the presence of DDM, which resulted in slightly improved yields but with the protein preparations characterised by extensive aggregation and precipitation during both purification and concentration. Eventually, approximately 0.5 mg of protein was purified in the presence of DDM from 10 L of bacterial culture. In order to regenerate the triangular crystals observed previously, an optimisation screen was set up in which conditions oscillated around the original successful conditions, but without success. However, rhomboid-shaped crystals appeared within two days of setting up the trials in control plates containing the commercial screen used to obtain the initial hit (Figure 36). This is not uncommon as optimisation screens prepared *de novo* often lead to failure in reproducing crystals due to changes that can occur in the purchased screens during storage or due to differences in making up solutions (Luft

*et al.*, 2014). The appearance of rhomboid-shaped crystals, however, appeared to confirm the tendency of CzrB to form crystals.



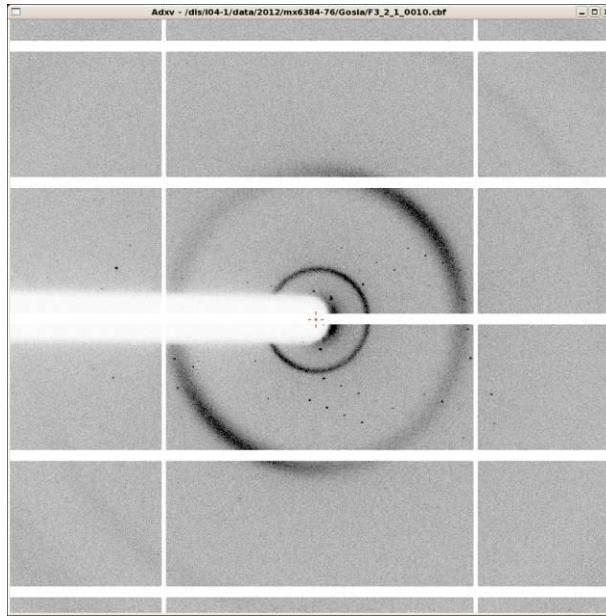
**Figure 36.** Crystals obtained in the control well during the optimisation screen.

Initially, crystals were harvested for analysis at the I04 beamline with a beam size of 50 x 50  $\mu\text{m}$  (Figure 37). The diffraction pattern was typical of that of a protein crystal, with sharp, distinctive spots arranged in a regular manner (Figure 38). Eventually, several crystals were tested and found to diffract from 30  $\text{\AA}$  to as high as 7  $\text{\AA}$ , which is a good starting point for optimising crystals of membrane proteins.



**Figure 37.** MicroMesh loop from MiTeGen with harvested crystal ready for the X-ray analysis inside the beam I04-1.

*The red crosshair indicates the position of the crystal.*



**Figure 38. The diffraction pattern obtained from X-ray analysis of crystals grown in vapour diffusion from DDM-purified CzcB.**

*Data was collected at I04, Diamond Light Source (DLS), Oxford, UK.*

*Data was acquired using Pilatus 6M-F detector.*

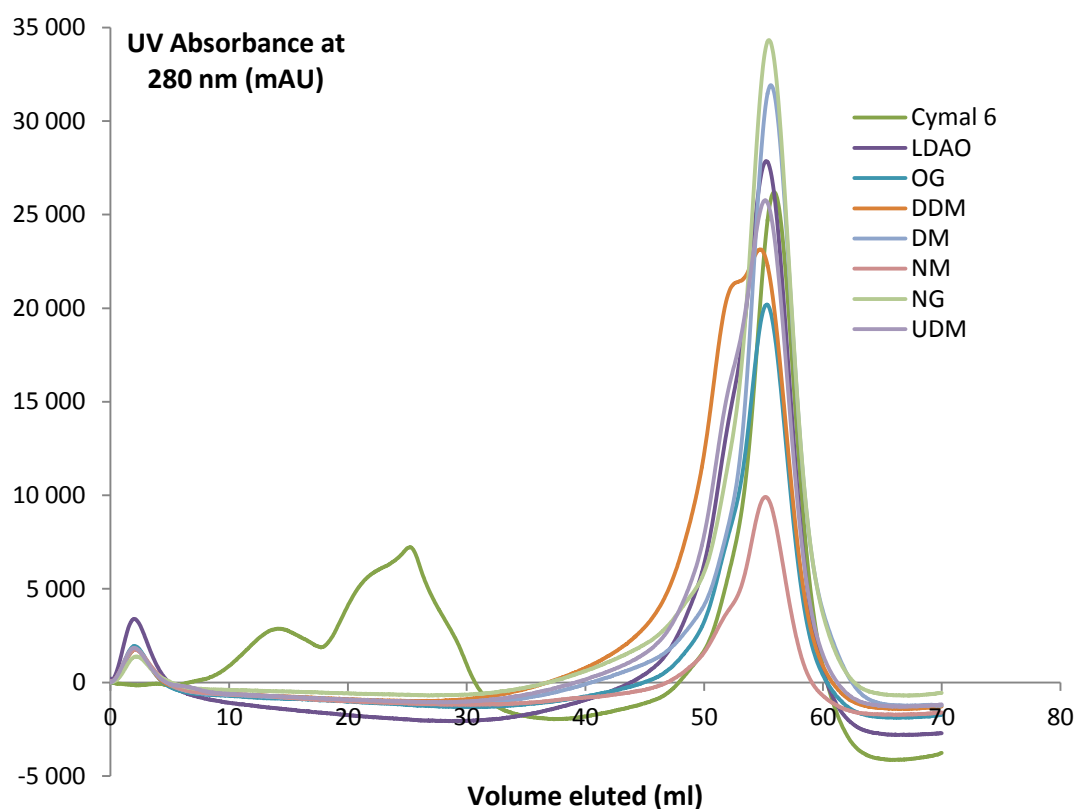
Difficulties in handling CzcB in the presence of DDM, manifested by aggregation and precipitation, suggested that protein is unstable in this particular detergent. It was surprising, therefore, to obtain crystals as apparent instability makes it unlikely for proteins to form crystals (Bolen, 2004, Ericsson *et al.*, 2006, Fan *et al.*, 2011).

The images collected on the I04-1 beamline were processed using MOSFLM software. However, the improvement in the diffraction and resolution allowed the space group and unit-cell parameters to be established (space group: R32 and unit cell  $a = b = 146.30$ ,  $c = 514.30$ ,  $90, 90, 120$ ), which led to the conclusion that the protein crystals were of the common *E. coli* contaminant AcrB (Moraes, I (2013) e-mail to Malgorzata Wronska, 08 April). AcrB is a multidrug efflux pump, which has a natural polyhistidine cluster that acts as an NTA-binding site and can lead to its co-purification with recombinant membrane proteins during their IMAC purification. The first report of AcrB spontaneously binding to a nickel resin came in 1999 by Zgurskaya (Zgurskaya and Nikaido, 1999); a critical assessment of the problem posed by AcrB in membrane protein crystallography surfaced nearly a

decade later from Veessler (2009). AcrB is particularly problematic for membrane crystallographers as even amounts undetectable by common analysis techniques are sufficient for it to crystallise and it crystallises successfully in a variety of conditions. The problems caused by ArcB contamination can be circumvented through the use of a protease cleavable his-tag on proteins of interest, expression in *acrB*<sup>-</sup> strains or by using a step purification system (Veessler *et al.*, 2008). The molecular size of AcrB is around 114 kDa; on reflection, upon looking at the analysis of CzcB purification in the presence of DDM (Figure 33), quite a striking band was detected around 98 kDa marker, which could be the AcrB contaminant. However, the band is nearly as strong as the band indicating a monomer of CzcB, which would rather suggest that it is an oligomer of CzcB, especially since the band is also present during purification in the presence of DM although not as prominent.

### **3.5.9. Screening of the effect of detergents on CzcB by analytical HPLC**

UDM-purified CzcB was incubated with Cymal-6, LDAO, OG, DDM, DM, NM, NG and UDM detergents, followed by analysis using a Shimadzu HPLC to establish the most suitable detergents to ensure pure and homogenous protein.



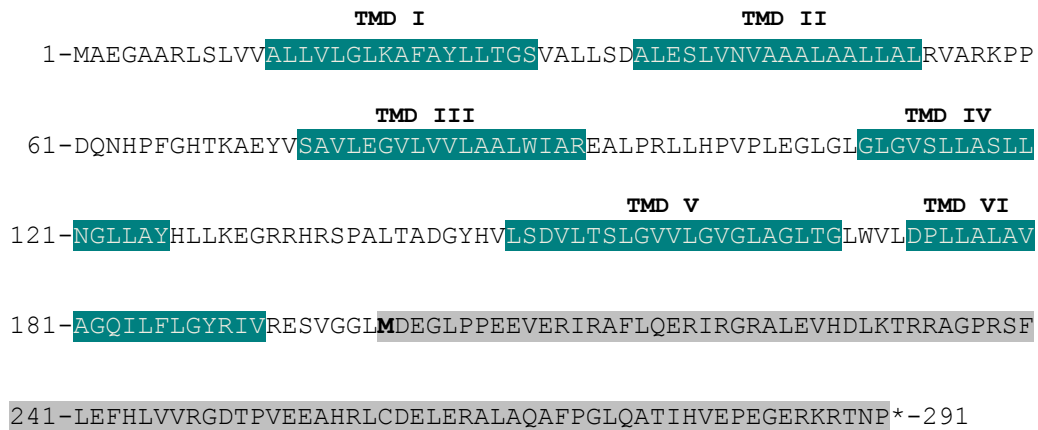
**Figure 39. Determination of monodispersity of CzrB in different detergents using HPLC.**

*Chromatograms representing CzrB in different detergents are shown in different colours and summarised on the right corner of the figure. Data were collected using a Prominence HPLC (Shimadzu).*

UV trace analysis of the effect of each detergent on the monodispersity of CzrB revealed that the most symmetrical peaks were obtained in NG (**light green**), DM (**light blue**) and OG (**ocean blue**), which contain short alkyl chains and small micelle sizes, but they also exhibit a tendency to denature the protein (Figure 39). DDM (**orange**) and Cymal-6 (**dark green**) exhibited extensive aggregation, whereas DDM and LDAO showed weak resolution between proteins assumed to be dimeric and monomeric (see Figure 39).

### 3.6. Construction of CzcB variants and preliminary expression screening at OPPF

In order to avoid further problem of co-purification of AcrB, a new construct of CzcB with a protease-cleavable hexahistidine tag was constructed. This was carried out, together with generation of a number of additional CzcB constructs, during a research visit at OPPF, Harwell, UK. The constructs are described in Figure 41 and included the full-length protein, the cytosolic domain (Met<sub>200</sub>-Pro<sub>291</sub>) and constructs containing only 4 or 5 transmembrane domains, as opposed to the 6 found in the native protein. Figure 40 shows the sequence of CzcB with 6 distinctive transmembranes and soluble domain.



**Figure 40. Amino acid sequence of CzcB.**

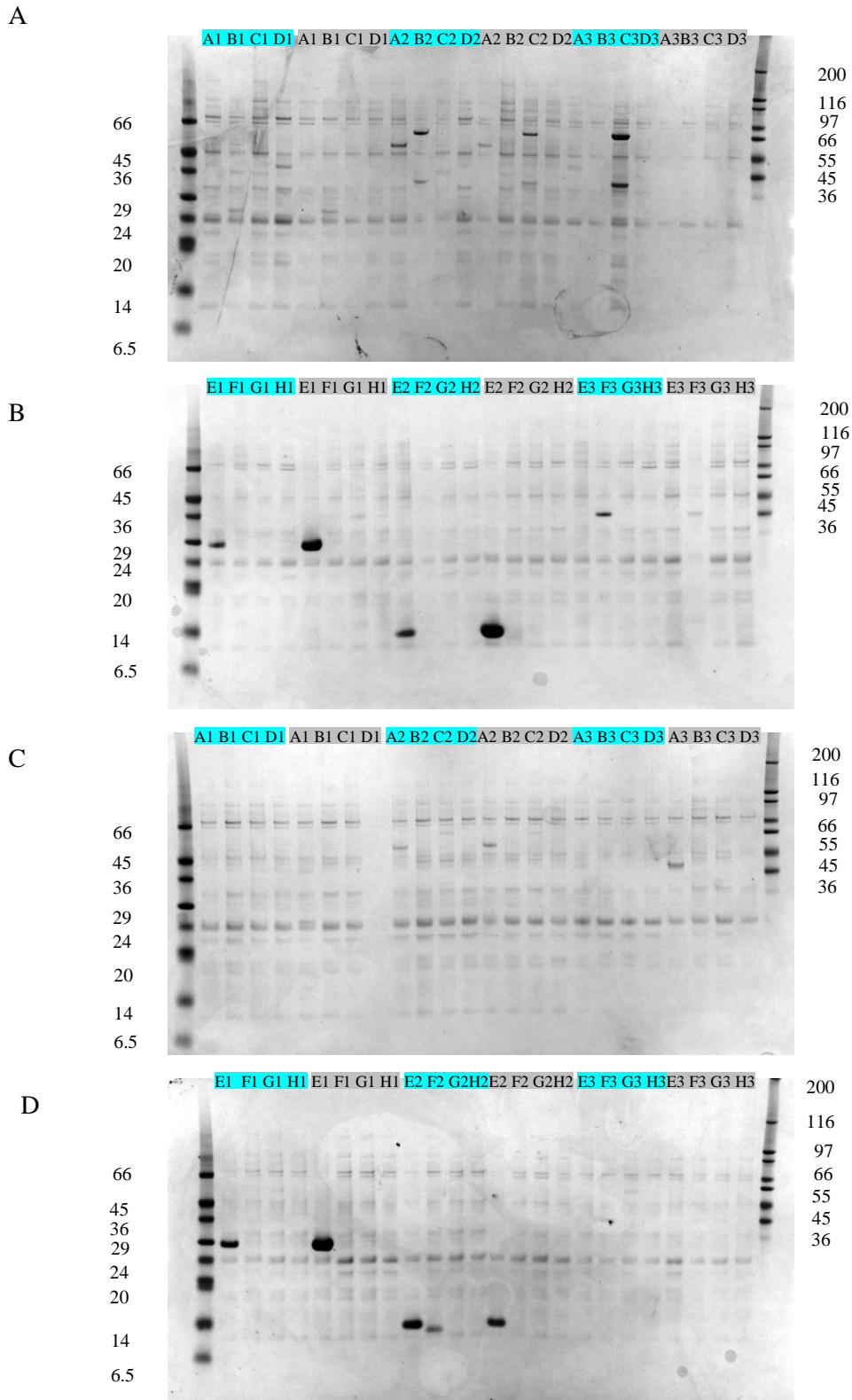
*The predicted transmembrane domains (TMD) are indicated by numbers and highlighted in teal, the cytoplasmic domain is highlighted in grey.*

The full-length *czcB* gene was then cloned into a variety of vectors in an attempt to improve expression and solubility to generate suitable protein products for crystallisation.

| Well | Gene name            | aa_N-aa_C | Vector         | MW-POI | MW-TAG | Total MW | Expression |
|------|----------------------|-----------|----------------|--------|--------|----------|------------|
| A1   | <i>czrB</i>          | 2-291     | pOPINE         | 31.109 | 1      | 32.109   | Green      |
| B1   | <i>czrB</i>          | 2-291     | pOPINF         | 31.109 | 1      | 32.109   | Green      |
| C1   | <i>czrB</i>          | 2-291     | pOPINM         | 31.109 | 42.5   | 73.609   | Green      |
| D1   | <i>czrB</i>          | 2-291     | pOPINS3C       | 31.109 | 12     | 43.109   | Green      |
| E1   | <i>czrB</i>          | 2-291     | pOPINS         | 31.109 | 12     | 43.109   | Pink       |
| F1   | <i>czrB</i>          | 2-291     | pOPINTRX       | 31.109 | 14     | 45.109   | Red        |
| G1   | <i>czrB</i>          | 2-291     | pOPINE-3C-eGFP | 31.109 | 29.6   | 60.709   | Red        |
| H1   | <i>czrB</i>          | 2-291     | pOPINE-3C-HALO | 31.109 | 33     | 64.109   | Red        |
| A2   | <i>czrB</i>          | 2-291     | pOPINJ         | 31.109 | 26     | 57.109   | Green      |
| B2   | <i>czrB</i>          | 2-291     | pOPINneo Flag  | 31.109 | 1      | 32.109   | Green      |
| C2   | <i>czrB</i>          | 2-291     | pOPINHALO7     | 31.109 | 33     | 64.109   | Green      |
| D2   | <i>czrB</i> $\Delta$ | 2-199     | pOPINE         | 20.528 | 1      | 21.528   | Green      |
| E2   | <i>czrB</i> $\Delta$ | 2-199     | pOPINF         | 20.528 | 1      | 21.528   | Pink       |
| F2   | <i>czrB</i> $\Delta$ | 2-199     | pOPINE-3C-eGFP | 20.528 | 29.6   | 50.128   | Pink       |
| G2   | $\Delta$ <i>czrB</i> | 200-291   | pOPINE         | 10.468 | 1      | 11.468   | Red        |
| H2   | $\Delta$ <i>czrB</i> | 200-291   | pOPINF         | 10.468 | 1      | 11.468   | Red        |
| A3   | $\Delta$ <i>czrB</i> | 200-291   | pOPINE-3C-eGFP | 10.468 | 29.6   | 40.068   | Green      |
| B3   | <i>czrB4tm</i>       | 2-131     | pOPINE         | 13.552 | 1      | 14.552   | Red        |
| C3   | <i>czrB4tm</i>       | 2-131     | pOPINF         | 13.552 | 1      | 14.552   | Pink       |
| D3   | <i>czrB4tm</i>       | 2-131     | pOPINE-3C-eGFP | 13.552 | 29.6   | 43.152   | Red        |
| E3   | <i>czrB5tm</i>       | 2-171     | pOPINE         | 18.867 | 1      | 19.867   | Red        |
| F3   | <i>czrB5tm</i>       | 2-171     | pOPINF         | 18.867 | 1      | 19.867   | Pink       |
| G3   | <i>czrB5tm</i>       | 2-171     | pOPINE-3C-eGFP | 18.867 | 29.6   | 48.467   | Green      |
| H3   | eGFP                 |           |                |        |        |          |            |

**Figure 41. Summary of the CzrB expression screen carried out using constructs generated at the OPPF.**

Column headings; Well: position of the plasmid in the 96-well plate, Gene name: gene given name, aa\_N-aa\_C: CzrB amino acids encoded within the gene, Vector: plasmid into which the gene was cloned, MW-POI, MW-TAG and Total MW: molecular weight of the “protein of interest”, recombinant tag and final product, respectively, Expression: result of expression screen in *E. coli* (green - protein of predicted MW was observed on SDS-PAGE, pink - MW of observed protein on SDS-PAGE differed from that predicted for the construct, red – no protein was detectable on SDS-PAGE. A list of pOPIN vectors and fusion tags with explanations is available in Appendix (page 171).



**Figure 42. SDS-PAGE analysis of preliminary expression screen of CzrB constructs in *E. coli*.**

Various versions of CzrB listed in Figure 41 were produced by induction with IPTG in the following *E. coli* strains: C41 (marked in turquoise), C43 (marked in grey)(A,B), Lemo21 (marked in turquoise), Rosetta RpLI (marked in grey) (C,D).

The results of initial expression experiments are shown in Figure 41 and 42. The pOPIN vector platform for HTP cloning and expression screening was originally developed by OPPF for use with soluble proteins (Berrow *et al.*, 2007). Screening of expression of the CzrB variants resulted in fewer than 50% of constructs being successfully expressed, which may be due to the fact that the platform was not designed for membrane proteins and some protocols are still being developed. Proteins of the expected size were observed in the case of 7 of the 11 full-length constructs. However, a number of the constructs were successfully expressed in only one or two of the four *E. coli* strains screened. Most constructs lacking the cytoplasmic domain failed either to express or to solubilise, with two of these (C3 and F3) yielding proteins different from the expected size. In the lane representing C3 we can observe two strong bands around 30 and 60 kDa instead of the expected 14.5 kDa, whereas lane F3 shows a band around 36 kDa mark instead of the predicted size of 19.9 kDa. The MW in both cases is approximately two times higher and so may be the result of multimerisation of the protein products. Only one of the constructs lacking a cytoplasmic domain (G3) was detected as a faint protein band when expressed by *E. coli* Lemo21.

Only construct B2 was detected on SDS-PAGE as two protein bands that corresponded to the molecular weights of the monomer and dimer products, which is surprising considering CzrB exists as a dimer. The intensity of the E2 protein product and its molecular size (approx. 14 kDa) might suggest it to be the cytoplasmic domain, which could result from an error during the HTP cloning and expression, though this would have to be checked by sequencing the vector before further analysis of this clone. It could also be an effect of the protein travelling further on the SDS-PAGE due to different amount of SDS binding different proteins or the fact that lack of soluble domain inhibits the molecule from unfolding, leading to greater electrophoretic mobility.

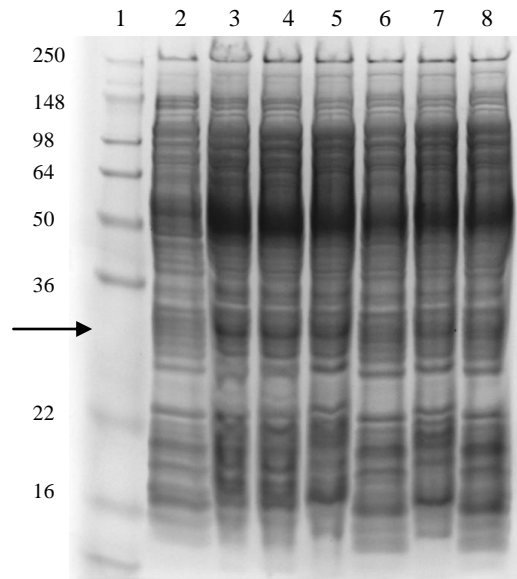
The main objective of generating a series of new CzrB constructs was to obtain a construct with a cleavable hexahistidine tag to avoid co-purification of AcrB (see 3.5.8) and it was also hoped to produce a more suitable crystallisation target in terms of stability and homogeneity. Overall, 6 full-length constructs of CzrB with cleavable tags were successfully expressed, including hexahistidine, MBP, SUMO,

HALO7, TRX and FLAG tags. Unfortunately, it was not possible to use the corresponding construct with a GFP tag as no expression was obtained with this vector, which was a limitation for using HTP technology developed for membrane proteins (Drew *et al.*, 2008b). However, the GFP-tagged construct containing 5 TM's gave a positive expression result, which could be used as an alternative to the full-length protein in the future.

Initially, the pOPINFCzrB vector was chosen for large scale protein overexpression in *E. coli* C43 to supply crystallisation trials.

### **3.7. Crystallisation of CzrB expressed from pOPINFCzrB**

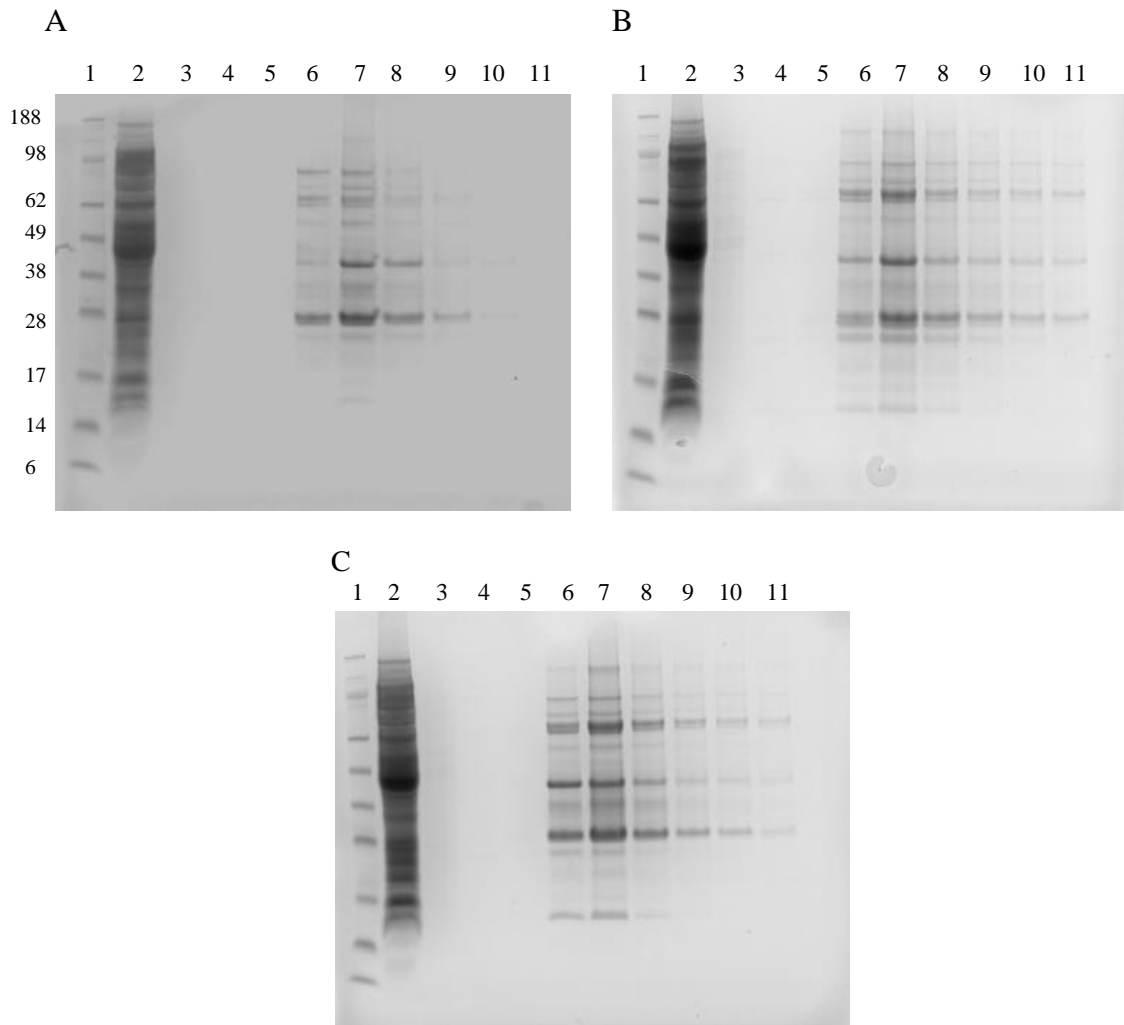
CzrB expressed in pOPINF was one of the targets selected for scale-up from the pool of developed constructs. After expression of CzrB in *E. coli*, cell fractionation was carried out using the usual protocol (section 2.2.6.3). Subsequently, several detergents including DDM, LDAO, UDM, FC-12, Cymal-6, DM and Triton-X-100 were investigated to determine which would most successfully solubilise CzrB from cell membranes (Figure 43). Figure 43 reveals that DDM, FC-12 and UDM exhibited similar solubilisation efficiencies, whereas the efficiency of DM in extracting CzrB was slightly lower and LDAO, Cymal and Triton-X-100 performed more poorly. Since DDM was the most efficient in extracting the protein from the membrane, it was chosen for use in the subsequent solubilisation protocol (see 2.2.6.4).



**Figure 43. SDS-PAGE analysis of the efficiency of various detergents in solubilising CzrB.**

*Lanes; 1: molecular weight markers (values are presented in kDa beside each protein band), 2: LDAO, 3: DDM, 4: UDM, 5: FC-12, 6: Cymal-6, 7: DM, 8: Triton-X-100. The arrow indicates the expected MW of CzrB.*

Based on the results of analytical HPLC performed using Shimadzu it was decided to use smaller detergents in CzrB purification since they showed the best monodisperse and symmetrical UV trace with no sign of protein aggregation. Therefore, purification by IMAC was investigated in parallel in the presence of 1% OG, 0.26% DM and 0.6% NG (Figure 44).



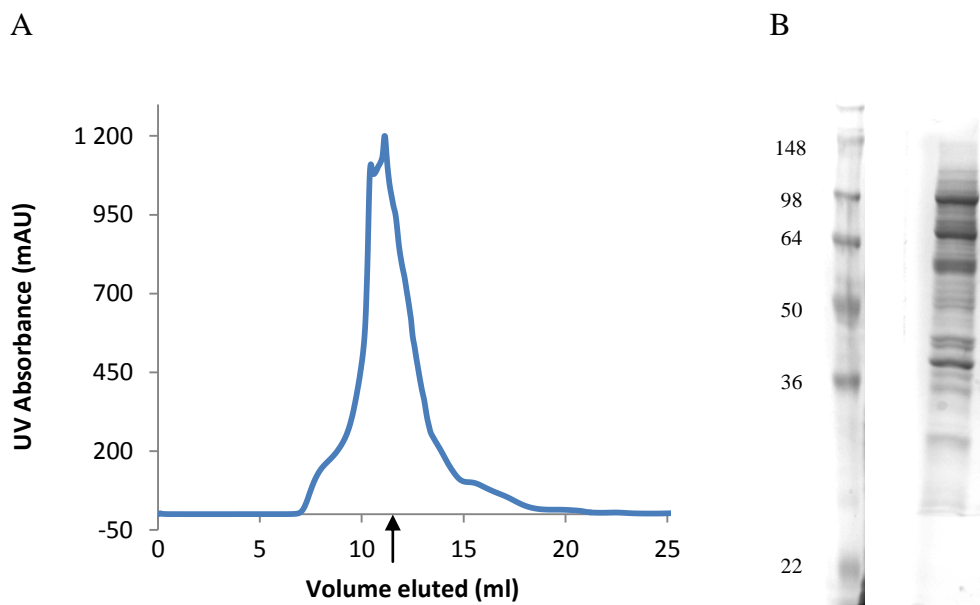
**Figure 44. SDS-PAGE analysis of IMAC purification in the presence of OG (A), DM (B) or NG (C).**

*Lanes; 1: molecular marker (values are presented in kDa beside each protein band), 2: resin flow-through, 3: binding buffer wash, 4: 35 mM imidazole wash, 5: 50 mM imidazole wash, 6-11: 500 mM imidazole elution fractions.*

Comparison of the purified fractions did not reveal any significant differences in purity of the samples (Figure 44) but the SDS-PAGE analysis demonstrated the presence of higher oligomeric forms of the protein present in each of the detergent samples. Moreover, since the three purifications were carried out with equal amounts of protein extract, it appeared that purification in the presence of OG resulted in a lower protein yield. Upon overnight dialysis to remove the histidine tag, extensive precipitation was noted, and complete loss of protein in OG and NG

samples. It was concluded, therefore, that the detergents were too harsh for prolonged incubation with CzcB. The third protein sample containing DM was further purified using SEC, resulting in poor separation and intensive heterogeneity, as indicated by numerous peaks in the UV chromatogram (data not shown). It was not possible therefore to use this sample in crystallisation trials.

In order to reduce the exposure of the protein to the denaturing environment created by OG and NG detergents, CzcB purification, including removal of the N-terminal hexahistidine tag, was carried out in the presence of DDM. This was followed by exchange of CzcB into an NG-containing buffer during SEC (Figure 45).



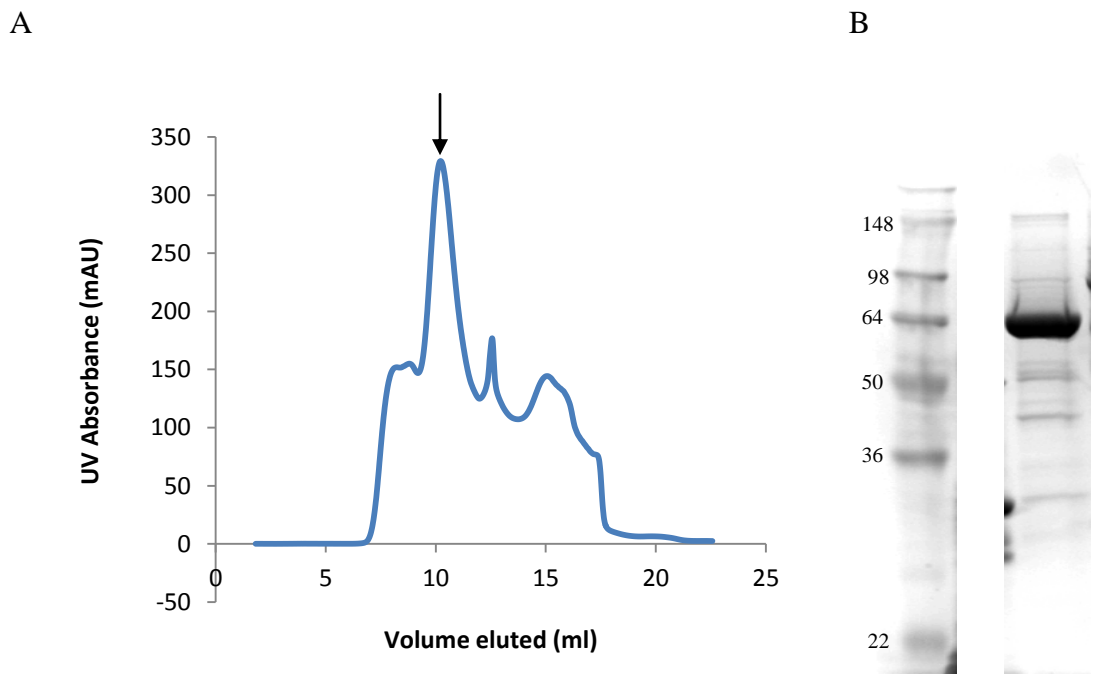
**Figure 45. SEC analysis of DDM-purified CzcB, with exchange into buffer containing 0.2% NG.**

*UV analysis of CzcB elution (A) and SDS-PAGE analysis of a fraction sampled from the arrowed point in the middle of the peak next to molecular marker (values are presented in kDa beside each protein band)(B).*

Size-exclusion chromatography (Figure 45A) revealed an asymmetrical peak and a fraction collected from the middle of the peak showed multiple protein bands of lower and higher MW than the full-length molecule on SDS-PAGE (Figure 45B), indicating possible oligomerisation but also the occurrence of degradation of

the target protein. It was concluded that CzrB obtained from the purification in the presence of NG was not suitable for crystallisation trials due to its obvious heterogeneity. The lack of monodispersity of the sample was confirmed by TETRA detector analysis (not shown).

Subsequently, CzrB was purified by IMAC, followed by the histidine-tag removal using 3C protease; those both steps were carried out in the presence of UDM. UDM appeared the most preferred detergent with the previously used construct but its use was discontinued after initial crystals were discovered to have high detergent content (Figure 32).



**Figure 46. SEC analysis of UDM-purified CzrB with exchange into buffer containing 0.2% NG.**

*UV analysis of CzrB elution (A) and SDS-PAGE analysis of a fraction pooled from the middle of the peak arrowed in A, next to molecular marker (values are presented in kDa beside each protein band)(B).*

The chromatogram (Figure 46A) demonstrates that a dominant, symmetrical peak preceded by a shoulder indicating possible aggregation and

followed by lower molecular weight peaks, potentially representing poor resolution between monomer and dimer forms of the protein, or degradation.

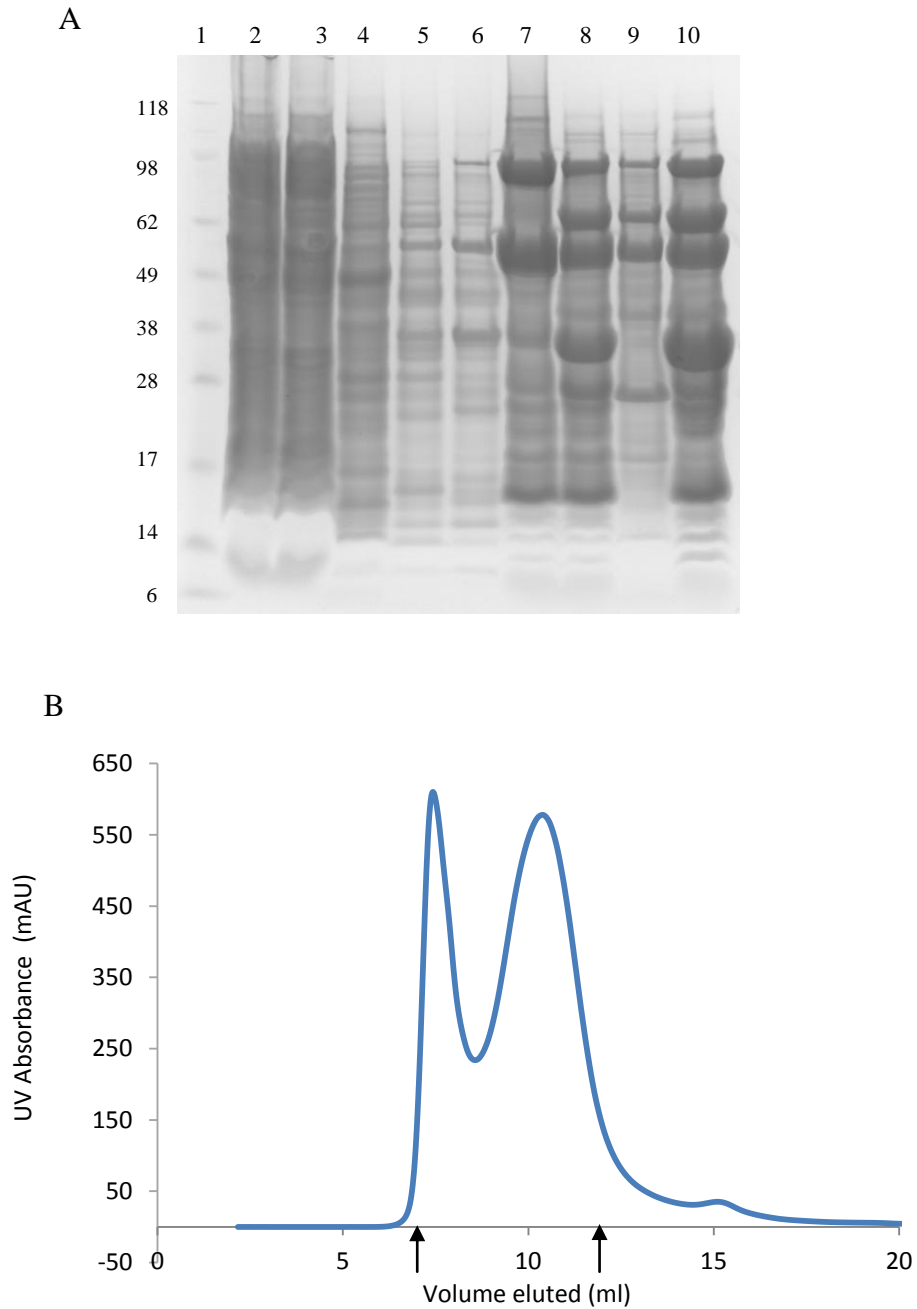
Figure 46B represents SDS-PAGE analysis of the middle fraction from the main peak (arrowed in Figure 46A). The analysed sample is characterised by a high degree of purity and appears to be exclusively dimeric. However, given the multiple peaks in the chromatogram, crystallisation trials could be set up using only fractions of CzrB from the top of the main peak to avoid heterogeneity in the protein preparation as much as possible. Crystallisation trials were therefore set up in vapour diffusion plates using MemGold and MemSys (Molecular Dimensions), and containing 20 mg/ml CzrB, but no crystals resulted.

### **3.8. Crystallisation of CzrB expressed from pOPINEHALO7CzrB**

There were concerns about the N-terminal position of His-tag in CzrB expressed from pOPINFCzrB and its effect on the behaviour of the protein through interfering with the signal peptide. According to an analysis of the effect of hexahistidine tag location on the expression of recombinant proteins (Gordon *et al.*, 2008), the position of the His-tag had no effect on 3 of 20 analysed proteins, whereas 4 expressed better with a C-terminal tag and 12 with an N-terminal tag. Since it is not possible to predict the effects of tag location *in silico*, and to avoid possible complications arising from the N-terminal tag and leader peptide, expression of CzrB with a C-terminal, removable hexahistidine-tag was also investigated.

The protein was produced in *E. coli* C43 containing pOPINHALO7CzrB, grown in overnight autoinducing medium Power Prime Broth at 25°C, according to the protocol developed at OPPF. It was also produced from the same vector in *E. coli* Lemo21 cells grown in autoinducing medium PASM-5052 containing additional L-rhamnose for increased yields. Expression in *E. coli* L21 in the presence of different concentrations of L-rhamnose resulted in no detectable difference in protein yields (data not shown) and so *E. coli* C43 was used for large-scale expression.

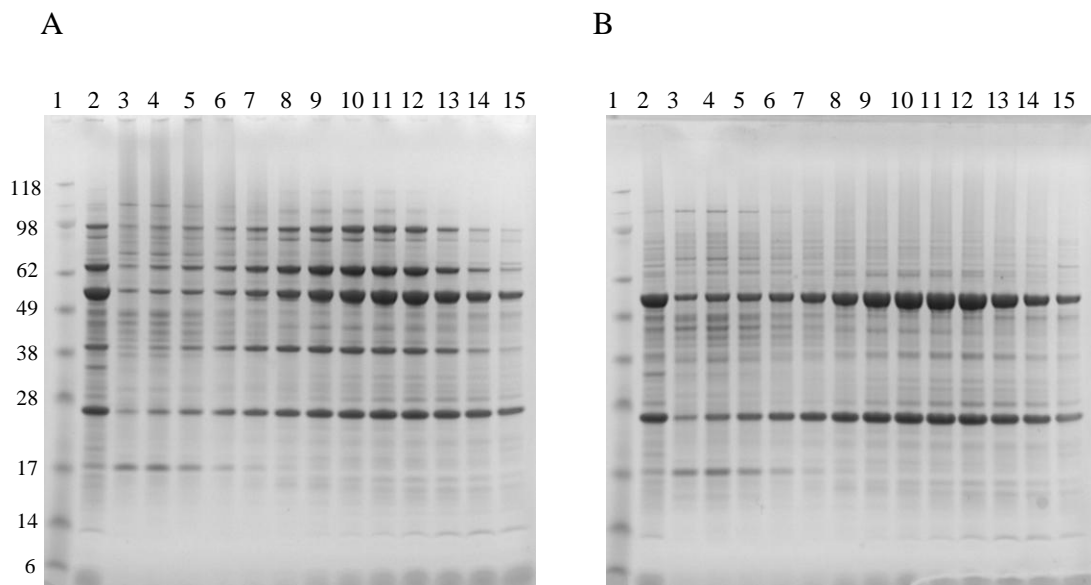
Extraction was carried out in the presence of DDM and purification (Figure 47A), including SEC (Figure 47B), in the presence of UDM.



**Figure 47. SDS-PAGE analysis of IMAC purification and His-tag removal from CzrB (A) and SEC purification of CzrB in the presence of 0.09% UDM (B).**

*Lanes in A; 1: molecular marker (values are presented in kDa beside each protein band), 2: solubilised membranes, 3: IMAC resin flow-through, 4: binding buffer wash, 5: 35 mM imidazole wash, 6: 50 mM imidazole wash, 7: eluted fraction with 250 mM imidazole buffer, 8: cleavage with 3C protease, 9: flow-through of reverse IMAC, 10: elution of uncleaved protein and 3C protease.*

SEC purification yielded a broad peak with distinctive separation, which could indicate separation of two populations: monomer and dimer (Figure 47B), but it is more likely a mix of cleaved and uncleaved protein, which would be consistent with SDS-PAGE. As Figure 47 A, lane 8 shows the presence of uncleaved protein, it seems like there was a poor separation of tagged and untagged molecules, in which case the first peak on SEC chromatogram would likely be the uncleaved protein.



**Figure 48. SDS-PAGE analysis of size exclusion purification of CzrB expressed in pOPINHALO7.**

*Fractions were analysed with (B) and without denaturation with 2-ME (A).*

*Lanes; 1: molecular marker, 2: IMAC-purified CzrB, 3-15: alternate SEC fractions (0.2 ml) collected throughout the peaks in Figure 47B between 7<sup>th</sup> and 12<sup>th</sup> ml of eluted volume (marked by arrows).*

CzrB purified in the presence of UDM appears to exist in a combination of monomeric and dimeric forms, with numerous higher oligomeric forms also visible on SDS-PAGE (Figure 48). Addition of 2-beta-mercaptoethanol and heat treatment led to the disappearance of higher oligomers, leaving the protein in predominantly dimeric form but with monomers also present. The fact that higher CzrB oligomers are destroyed by heat treatment suggests that they are less stable than dimer, which is still present in the PAGE after samples were denatured.

Moreover, crystallisation trials performed with this protein again yielded crystals identified as ArcB, confirming the observation from Figure 47A that removal of the His-tag was unsuccessful due to incomplete cleavage by the 3C protease and poor performance of the reverse IMAC; however, this issue can most likely be resolved with further optimisation of protease cleavage. Others have reported that 3C exhibits high sensitivity to detergents (Vergis and Wiener, 2011), whereas the enterokinase protease was affected by very few of 94 tested detergents, so this protease appears to be a better choice for use with membrane protein constructs.

Since it was identified during this project that oligomerisation of CzrB poses a serious obstacle in obtaining a homogenous protein preparation suitable for crystallisation trials, it was decided to generate a CzrB variant that is unable to dimerise, as described in section 2.2.2.2.1. By replacing amino acids that are vital in creation of dimer with alanine residues, it is hoped to obtain a CzrB molecule that is exclusively monomeric and thus to increase the chance of obtaining a homogenous protein population and so the chance of successful crystallisation of CzrB. This part of the project work is ongoing.

## 4. RESULTS FROM STRUCTURAL STUDIES OF TRIGGER FACTOR FROM *PSYCHROBACTER FRIGIDICOLA*

### 4.1. Taking crystallographic approach to investigate TFPf

Trigger factor from *Psychrobacter frigidicola* (TFPf) spiked our interest as a target for crystallographic studies as a result of a previous comparison with the well-characterised trigger factor from *E. coli* in our group (see 1.2.2). The earlier study was the first characterisation of TFPf and showed it to exist as a monomer in contrast with its homologue from *E. coli*, which tends to dimerise. Moreover, TFPf did not exhibit reduced refolding efficiency or impede export to the periplasm of co-expressed recombinant protein at high TF concentrations, which leads to increased potential of the psychrophile protein as a folding aid in cell-free recombinant protein expression systems (Robin *et al.*, 2009). The conclusion drawn from this study was that the two homologues may vary in their quaternary structure, leading to differences in their chaperone activity. The close connection between protein structure, function and mechanism of action was the principal motivation for this project, with the goal being to gain an understanding of the basis for the differences between the chaperones from *E. coli* and *P. frigidicola* through X-ray crystallography.

Trigger factor plays an important role as a molecular chaperone in protein biosynthesis. The *E. coli* protein has been crystallised and its structure solved (Ferbitz *et al.*, 2004). It consists of three domains and adopts an unique dragon-like structure with a central body from which two arms, a head and a tail region extend, which allows its binding to the ribosome protein L23 (Kramer *et al.*, 2002) (see 1.2.2). Nascent polypeptides interact with hydrophobic binding sites present within trigger factor upon exiting the ribosomal tunnel, which prevents their premature folding (Gamerdinger and Deuerling, 2014).

The fundamental approaches in crystallisation of soluble proteins are the same as for membrane proteins. The success of crystallisation relies greatly on its purity and stability.

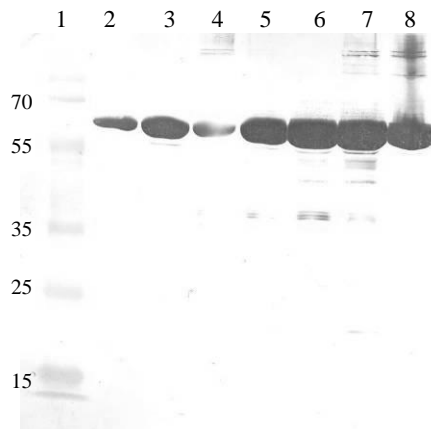
#### 4.2. Expression of full-length TFPf for crystallisation studies

Previously in this group (Robin *et al.*, 2009) the full-length *tig* gene from *P. frigidicola* was cloned into the p15ara expression vector consisting of a pACYC184 plasmid containing the complete arabinose promoter cassette from pBADHisB to ensure a tight control of background expression. A low basal level of expression and a rapid increase in the rate of synthesis of the target gene after induction with arabinose are characteristic for pBADara expression systems and can result in high yields of recombinant protein during expression (Guzman *et al.*, 1995).

Generally, the choice of the most appropriate method to express a recombinant protein can be a time-consuming process. Efforts in optimising the expression conditions include reducing the stress imposed on a bacterial cell, therefore avoiding the accumulation of misfolded proteins, inappropriate modifications and aggregation, leading in turn to higher yields of functional protein (Sørensen *et al.*, 2003, Sørensen and Mortensen, 2005a, Bonander and Bill, 2012, Hunt, 2005). The use of appropriately engineered *E. coli* strains for recombinant protein expression not only ensures a lack of pathogenic features but it can also provide sustained overexpression. Trigger factor was expressed in this work in *E. coli* BL21, which is commonly used *E. coli* host strain for expression, with an outstanding record in the expression of standard recombinant proteins (Studier, 2014). *E. coli* BL21 is also quite robust and grows rapidly in a variety of media.

Expression of recombinant proteins at high rates causes stresses to the host bacterial cells, resulting in impairment of the protein folding machinery and processes, which leads in turn to association of protein intermediates in the cell into amorphous aggregates called inclusion bodies (Hoffmann and Rinas, 2004). The temperature at which expression is carried out is one of the main factors that contribute to the formation of inclusion bodies. Lowering the temperature to 25°C, from the preferred *E. coli* growth temperature of 37°C, during TFPf expression facilitates the production of soluble, properly folded protein molecules (Gasser *et al.*, 2008). In this study, use of this lower temperature did not significantly slow the growth of *E. coli*, resulting in a good cell density upon harvest. In order to establish the optimal time for TFPf production (Figure 49), expression was carried out for 16 h

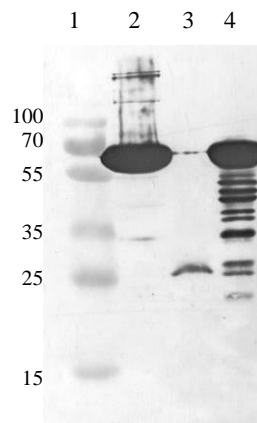
and protein yields were analysed at hourly intervals. Figure 49 reveals a gradual increase in protein yield from 0-5 h, after which the yield remained the same until harvest at 16 h. The recombinant trigger factor produced in this work has an expected molecular size of 50.6 kDa and contains a C-terminal histidine tag. It was noted during SDS-PAGE analysis that the recombinant TFPf migrated consistently and reproducibly at 60-65 kDa. This apparent discrepancy might be caused by low SDS binding to the recombinant protein, leading to smaller negative charge and a decrease in mobility of the molecule (Hjemel and Chrambach, 1981). The deviation between the predicted MW (50.6 kDa) and that observed upon SDS-PAGE might also be due to the fact that the mobility of molecules in the gel is more a function of molecular size rather than of molecular mass, which means that it is affected not only by weight but also by charge of the molecule (Sallantin *et al.*, 1990). One of the factors supporting successful production of recombinant protein is cell growth, which was monitored by observing the cell's optical density at 600 nm. This provides information about the density of the bacterial culture during the expression. Prolonged expression can cause the death of the host cell, which may lead to the protein loss due to cell disruption.



**Figure 49. Western blot analysis of yield of TFPf production after up to 16 h induction.** Lanes 1: molecular weight marker; 2-8: protein yields after 1, 2, 3, 4, 5, 6 and 16 h, respectively, of induction.

### 4.3. Extraction and purification of trigger factor for crystallisation

Successful expression needs to be followed by effective extraction of recombinant proteins from the host cell. Optimal cell lysis conditions maximise the yield of functional product, whereas poor extraction reduces protein yields and often affects protein quality. Trigger factor is a soluble, cytosolic protein, which is released with other cytosolic proteins upon bacterial cell lysis. Enzymatic lysis using r-lysozyme and CellLytic preceded by a freeze-thaw procedure efficiently extracted the majority of the bacterially-produced recombinant protein (Figure 50). r-Lysozyme (Novagen) is a stable and highly efficient recombinant enzyme, which hydrolyses the *N*-acetylmuramide linkages in the bacterial cell wall. CellLytic (Sigma-Aldrich) also supports degradation of bacterial wall; moreover, the presence of non-denaturing zwitterionic detergents in the reagent enhances the stability of extracted proteins.



**Figure 50. Western blot analysis of TFPf present in expressed and extracted samples.** Lanes; 1: molecular weight marker (values are presented in kDa beside each protein band), 2: whole cells producing TFPf, 3: ~30 kDa protein used as a control, 4: soluble cell extract.

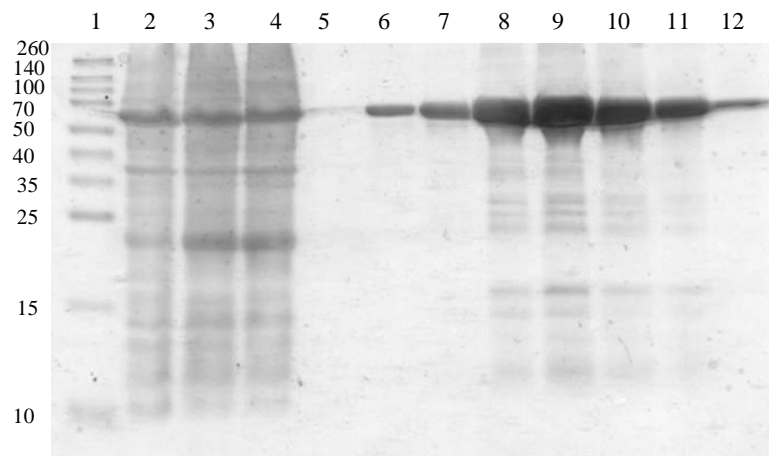
The efficiency of extraction appears high, from the comparable levels of protein in the two samples, but the extracted protein was clearly proteolysed during the extraction process. Detection of the apparent degradation products by western blotting using antibody against His-tag indicated that the degradation occurred at the

protein's non-tagged N-terminus, leading to the accumulation of products of various molecular sizes smaller than the full-length protein.

#### **4.4. Purification of full-length TFPf**

##### **4.4.1. Immobilised Metal Affinity Chromatography**

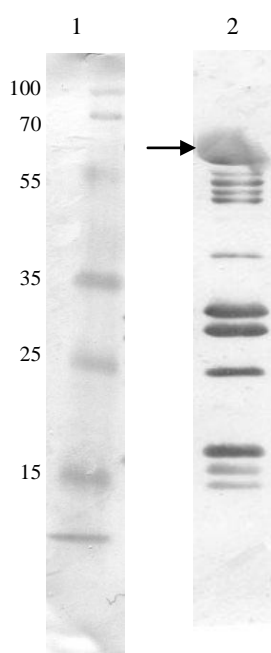
The purification of trigger factor by IMAC posed significant difficulties due to the occurrence of the extensive degradation observed in Figure 50 (lane 4). IMAC purification is based on the interaction between an affinity tag, in this case a hexahistidine tag, and transition metal cations immobilised on a matrix present in the purification column. Consecutive histidine residues are efficiently bound to the IMAC residue, followed by removal of non-tagged proteins by washing and elution of the His-tagged protein in a high imidazole concentration buffer due to the high affinity of imidazole for metal ions (Gaberc-Porekar and Menart, 2001). Purification using polyhistidine tags and metal affinity can result in high yields of recombinant products of ~ 95% purity and often one-step IMAC purification is sufficient to obtain high quality protein (Liu *et al.*, 2006, Ardao *et al.*, 2006). However, IMAC purification of TFPf met with significant obstacles due to co-purification of degraded products that retained a histidine tag and so continued to bind to the nickel matrix efficiently (Figure 51).



**Figure 51. SDS-PAGE Coomassie Blue analysis of the summary of TFPf purification by IMAC.**

*Lanes; 1: molecular weight marker (values are presented in kDa beside each protein band), 2: whole cell extract 3: soluble protein fraction, 4: insoluble protein fraction from E. coli cells producing TFPf, 5: IMAC column flow-through, 6-12: eluted fractions collected from HPLC after IMAC purification and buffer exchange.*

Figure 51 shows result of TFPf purification by IMAC followed by buffer exchange into 50 mM Tris-HCl, 150 mM NaCl, pH 8.0 in order to remove imidazole, where lanes 6 to 12 represent 1 ml fractions collected based on the chromatogram (not shown). As degraded products appeared to be present in the purified protein samples, this was confirmed by further analysis by Western Blot (Figure 52, lane 2). Several lower molecular weight protein bands of high-intensity were detected by immunoblotting in a purified sample of TFPf indicating that degraded fragments retained histidine tag and co-purified with the full-length protein (Figure 52).

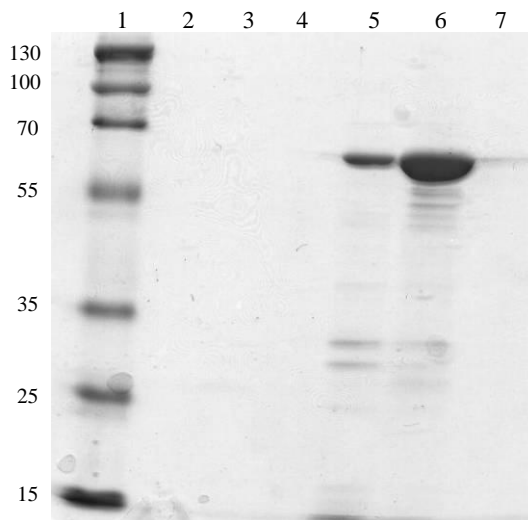


**Figure 52. Western Blot analysis of IMAC-purified TFPf.**

*Lanes; 1: molecular marker (values are presented in kDa beside each protein band), 2: IMAC purified TFPf. The C-terminally His-tagged TFPf was detected using a monoclonal anti-polyHistidine antibody. The arrow indicates the expected full-length molecule.*

#### **4.4.2. Ion-exchange chromatography (IEC)**

In order to achieve a more homogenous sample that would be more suited to crystallisation trials, IEC was used as a second purification step following IMAC (Figure 53). Ion-exchange is commonly used as a first step purification due to its simplicity, high capacity and relatively fast procedure; however it is also used for “polishing” of partly purified samples (Selkirk, 2004). It is based on the interaction of charged molecules coupled to the column matrix with oppositely charged sample molecules in the mobile phase. As a protein’s charge is determined by its amino acid sequence, it was anticipated that the lower molecular weight molecules visible in Figure 52 might possess a different charge from the full-length target protein and so be separated using IEC.



**Figure 53. Coomassie Blue-stained SDS-PAGE analysis of purification of TFPf by Ion Exchange Chromatography (IEC).**

*Lanes; 1: 1: molecular marker (values are presented in kDa beside each protein band), 2: Resource Q column flow-through, 3: wash with 25 mM NaCl buffer, 4: wash with 50 mM NaCl buffer, 5: wash with 100 mM NaCl buffer, 6: wash with 250 mM NaCl buffer, 7: wash with 500 mM NaCl buffer.*

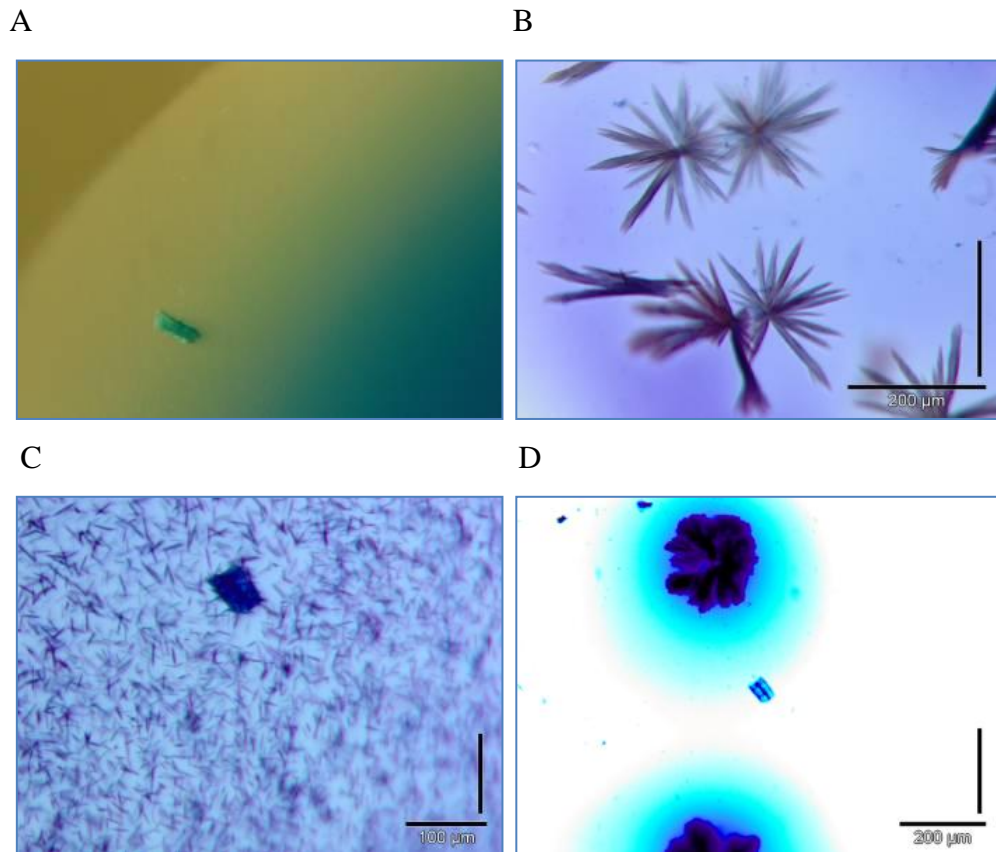
Results of IEC purification are presented in Figure 53; lanes 5 and 6 represent eluted TFPf. SDS-PAGE analysis reveals that even though a majority of the eluted protein fractions consisted of the full-length TFPf, degradation products were still detected. Therefore, the second purification step did not completely overcome the degradation problem, although it significantly improved the purity of the sample. Even though initial crystallisation screening is often conducted with a protein sample of 90-95 % purity as judged by SDS-PAGE (Benvenuti and Mangani, 2007), the presence of impurities in protein samples can inhibit crystal growth. On the other hand, heterogenous protein samples (due to degradation, modification or binding of ligands) can sometimes act as nucleation centres, allowing the growth of crystals around them (McPherson *et al.*, 1996). Nucleation and growth are two inseparable steps in the crystallisation of macromolecules. Nucleation is a critical transition for the molecules, during which they cross from a completely disordered state to an orderly one, which is followed by spontaneous growth. Often, the crystallisation process is a test of protein purity in itself and if it leads to crystal growth it allows

identification of crystallisation conditions. In cases where crystals are not suitable for X-ray analysis or produce low resolution diffraction, optimisation efforts then typically focuses on increasing sample purity. Therefore, after obtaining purified TFPf, crystallisation trials were commenced despite the impurities in the protein preparation.

#### **4.5. Preliminary crystallisation trials of trigger factor from *P. frigidicola***

Initial crystallisation trials were set up with IMAC- and IEC-purified TFPf at a concentration of 10 mg/ml using the vapour-diffusion sitting-drop method. Vapour-diffusion is a popular method for early screening of crystallisation conditions and one that can be set up manually quite easily when there is no access to high-throughput (HTP) approach. Trials were set up using a wide range of conditions from commercial screens JCSG Suite Core (I, II, III, IV), designed by Hampton Research, at 4°C, 15°C and room temperature.

Initial hits were obtained at 15°C in several conditions: 20% MDP, 0.1 M Tris-HCl pH 7, 0.1 M ADP (Figure 54A), which produced a single rod-shaped crystal; 0.1 M Na cacodylate, 0.2 M Ammonium sulphate, 0.2 M NaCl (Figure 54B), resulting in star-like clusters of crystals; and 0.1 M CHES pH 9.5, 1M K/Na tartrate, 0.2 M Li sulphate (Figure 54C and Figure 54D), where the formation of small needles was observed.



**Figure 54. Result of TFpf initial crystallisation screening.**

*Crystals appeared in several conditions:*

*A: 20% MDP, 0.1 M Tris-HCl pH 7, 0.1 M ADP;*

*B: 0.1 M Na cacodylate, 0.2 M Ammonium sulphate, 0.2 M NaCl;*

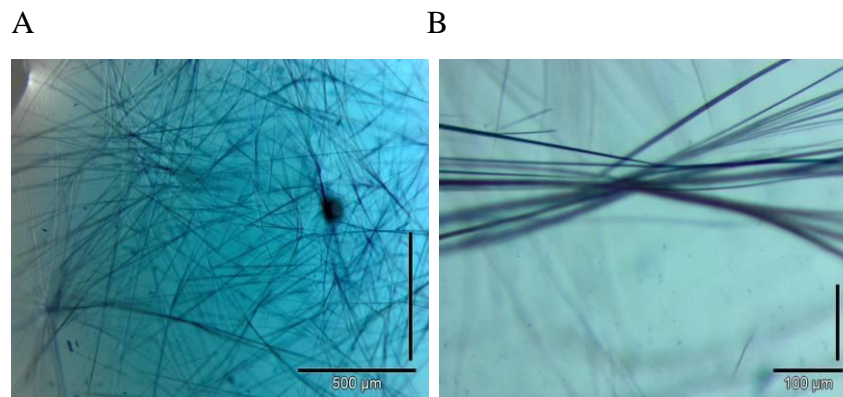
*C and D: 0.1 M CHES pH 9.5, 1 M K/Na tartrate, 0.2 M Li sulphate.*

*Size bars represent 100 µm in Figure 54 (C) and 200 µm (B, D).*

Crystals were stained using IZIT Crystal Dye (Hampton Research), which enters protein crystals and colours them blue, whereas salt crystals, which do not possess large solvent channels, remain clear. The rod-shaped crystal (Figure 54A) and needle crystals (Figure 54C) were clearly stained by the dye; within a few days, the latter gathered into urchin-like formations (Figure 54D).

#### 4.6. Optimisation of preliminary hits

Well-packed 3D single crystals are the prerequisite for generating high-resolution diffraction data. Thus, the conditions of the single crystal initial hit (20% MDP, 0.1 M Tris-HCl pH 7, 0.1 M ADP; Figure 54A) were used to reproduce the rod-shaped crystal by designing an optimisation screen that varied the concentration of the precipitant (MDP) and the additive (ADP). This led to the appearance of long, thin needles (Figure 55) within a week of setting up trials, which formed in several conditions containing MDP as precipitant and often reached the length of the crystallisation drop. Staining with Izit Dye appeared to confirm that the needles were protein crystals.



**Figure 55. Needle crystals obtained through optimisation of the initial hits.**

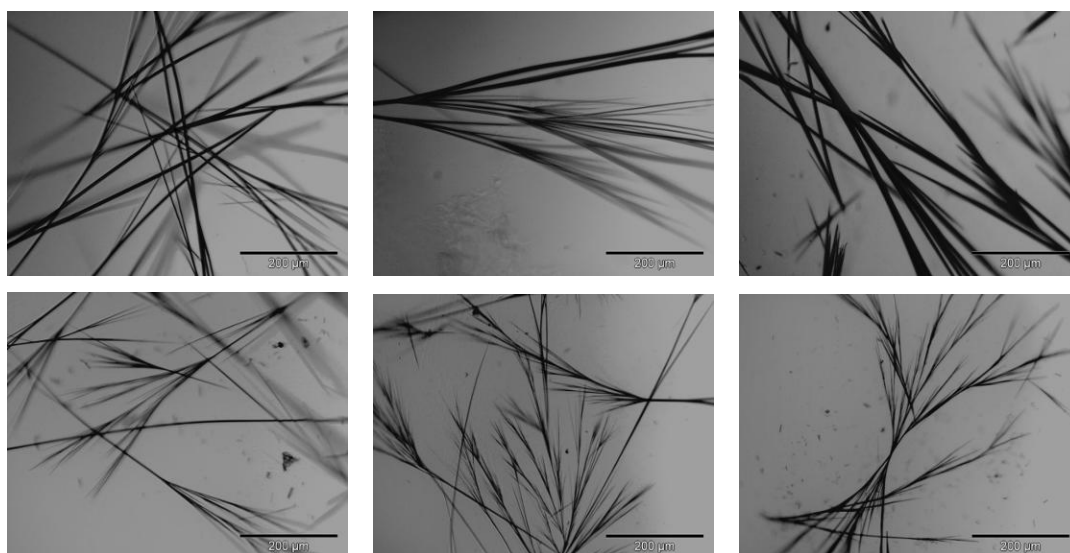
*Crystals were obtained under 30% MDP, 0.1 M Tris-HCl pH 7, 0.1 M ADP in A and B, respectively. Size bars represent 500 µm in (A) and 100 µm in (B).*

The observed needles were very thin and fragile; however, since they were obtained in the early stages of crystallisation, they were thought to indicate the protein's ability to crystallise and raised hopes for possible improvement of their quality through further optimisation. Optimisation is a process of informed design based on utilising the initial conditions of screening experiments to produce crystals of sufficient size and quality for X-ray diffraction. It is based on changing environmental variables (such as temperature, size of drop, crystallisation method) and chemical variables (type and concentration of precipitant, use of additives or ligands and the solution pH) to refine the crystallisation growth dynamic (McPherson

and Kuznetsov, 2014). Needle crystals rarely result in successful data collection, although they are a good starting point to improve the morphology and packing of a crystal. The needle crystals obtained (Figure 55) were difficult to harvest, unstable and easily broken, which suggested their unsuitability for X-ray analysis because it was anticipated they would be highly prone to radiation damage and not sufficient for data collection. Nevertheless, further optimisation was carried out in the hope that it would lead to crystals of different appearances being obtained, thus allowing high-resolution diffraction data to be collected.

#### 4.6.1. Optimisation of needle crystals

The screen in which precipitant (MDP) concentration (20, 25, 30, and 35%) and the concentration of additive (ADP) (0.1, 0.2 M) were varied led to the appearance of needle-crystals. Additionally, the effect of changing the buffer (Tris-HCl) pH (7.0, 7.5, 8.0, 8.5, 9.0, and 9.5) was evaluated, as well as the effect of reducing the target protein concentration, in order to improve the crystallisation drop dynamic and quality of grown crystals. The screening yielded more needle crystals (Figure 56) in a variety of pH values, precipitant concentrations and at 4°C and 15°C.

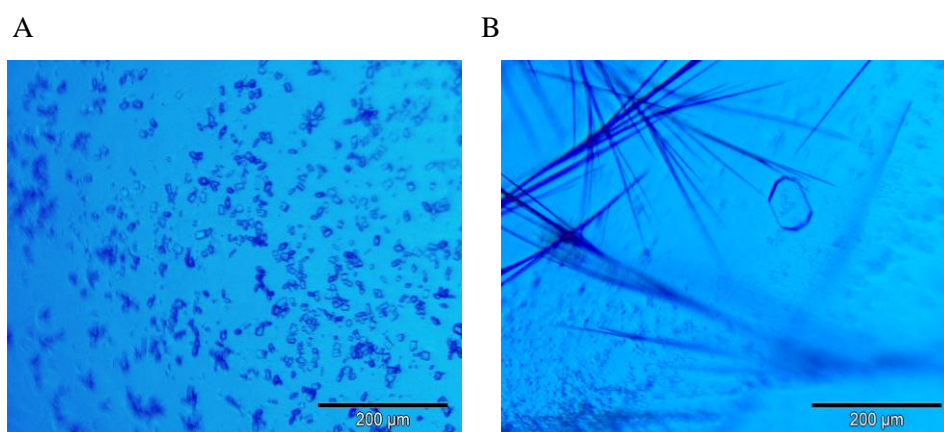


**Figure 56. Examples of needle crystals obtained at various pH values during optimisation screening of TFpf.**

*All crystals were obtained using MDP as precipitant. Size bars represent 200 µm.*

The rate of crystallisation is especially dependent on nucleation kinetics but also on protein concentration, pH, temperature and precipitant and, despite improvements in the understanding of the kinetics of crystal growth, it is still a trial-and-error approach. Optimal conditions leading to crystal nucleation often do not support subsequent growth of crystals (Bergfors, 2009). This as a result of nucleation and appearance of crystals is more likely to occur at high supersaturation level, while slow and orderly crystal growth is favoured by supersaturation at low levels. This often means that crystal growth is very rapid, causing poor packing and low quality of crystals (Chayen, 1998). Overall, the attempts for optimisation revealed no significant changes in the features of the previously obtained needle crystals. As the needle-shaped crystals appeared to be too small for diffraction, there was no opportunity to verify if they were protein crystals as even the use of a microfocus beamline was thought likely to damage the crystals too extensively to allow structure determination.

The next step in the optimisation process was screening the initial condition (20% MDP, 0.1 M Tris-HCl pH 7, 0.1 M ADP) with an additive screen (Hampton Research), which consists of a range of small molecules that can enhance protein stability and crystallisation by affecting protein-protein and also protein-solvent interactions. Examples of crystals that arose from the additive screen are presented in the Figure 57; both examples were obtained with  $\text{BaCl}_2$  as an additive.



**Figure 57. Crystals obtained with MDP as precipitant during the additive screen.** In both photos the additive is  $\text{BaCl}_2$ . Size bars represent 200  $\mu\text{m}$ .

Crystals in Figure 57A were small, cubic-shaped and numerous, whereas in Figure 57B a single octagonal crystal (~100  $\mu\text{m}$ ) with well-defined edges can be seen. As the crystals remained clear after the use of the IZIT Crystal Dye, they were thought likely to be salt crystals. However, in Figure 57B the appearance of needle crystals can again be observed.

#### **4.6.2. Seeding of needle crystals to improve their quality**

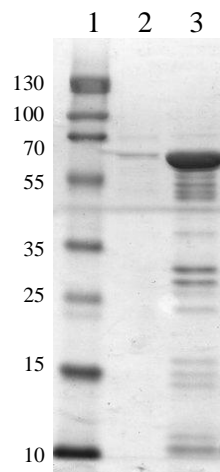
Seeding is a powerful tool for the separation of nucleation and growth using previously nucleated crystals, often of low quality (Bergfors, 2003). It involves seeding new crystallisation drops equilibrated at lower levels of supersaturation with existing crystals in the hope that these will trigger the growth of better-formed, bigger and more stable crystals. The thin crystal needles were difficult to visualise under the microscope but staining with the IZIT dye led to the dye entering crystal cavities and turning the needles blue. Needle crystals dyed by IZIT dye were used for seeding, therefore, as it was anticipated that the dilution necessary for seeding would diminish any possible effect of the dye on the subsequent nucleation process. The needles dissolved upon dilution into the mother liquor, however, while the use of this seeding stock in fresh crystallisation drops resulted in immediate formation of new needle crystals, which formed into extensive flower-like branches. The behaviour of the crystals in the drops during seeding triggered the suspicion that the previously observed crystals were, in fact, crystals of the IZIT dye. The suspicion was also supported by the wide variety of conditions in which the needle crystals appeared, the fact that the crystals were only visible after addition of dye and previous reports by other researchers of the appearance of needle- or flake-shaped crystals of IZIT (Bukrinsky and Poulsen, 2001).

Consequently, a repeated effort to identify new conditions for crystal formation was undertaken with a wide screen of buffers and precipitants, including Hampton JCSG Suite Core (I, II, III, IV) screen and a range of buffers (NaAC pH 3.3, NaAC pH 4.6, MES pH 6.0, HEPES pH 7.0, Tris-HCl pH 8.0 and Tris-HCl pH 8.5) and precipitants (KCl, LiCl and NaCl) produced *de novo*. Trials were set up at 4°C, 15°C and room temperature, and in the presence of a wide variety of additives

(Hampton Research). However, despite the extent of the conditions screened, no crystallisation hits were observed.

#### 4.7. Size Exclusion Chromatography (SEC)

As discussed earlier (see 3.1.1 and 4.4.2), in order to identify hits or to improve low quality crystals, researchers must often step back a number of steps (e.g., protein purification) in order to prepare higher quality or more stable protein and ultimately advance towards stable crystals. The extensive degradation observed upon concentration of TFPf using Vivaspin 30 kDa MWCO, as revealed during SDS-PAGE analysis (Figure 58), was thought likely to be a contributory factor in the failure to identify crystallisation hits. Coomassie-stained SDS-PAGE samples revealed a significant proportion of eluted, His-tagged protein to be of lower molecular weight than that expected for the TFPf sample, which might be a source of difficulties with subsequent crystallisation as the presence of impurities can inhibit crystal formation. Therefore, efforts were now focused on reducing or preventing degradation of the TFPf protein before proceeding with further crystallisation trials.



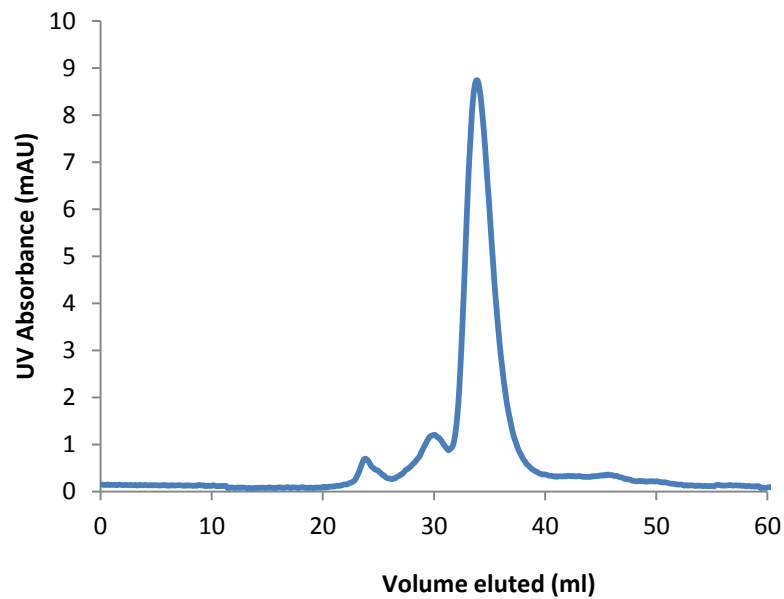
**Figure 58. SDS-PAGE analysis of purified and concentrated TFPf showing extensive degradation of the protein.**

*Lanes; 1: molecular marker (values are presented in kDa beside each protein band), 2: purified TFPf, 3: concentrated TFPf.*

Heterogeneously expressed proteins produced in *E. coli* can be subject to degradation by native cellular proteases. Proteolysis of the protein of interest is one of the major bottlenecks in recombinant protein production, resulting in decreased yields and contamination by degradation products in downstream applications (Sørensen and Mortensen, 2005b). Most efforts to control proteolysis are applied either at the cellular level during cultivation, such as the choice of host strain, reduced cultivation temperature or improved aeration, or at the protein level by altering the protein sequence to eliminate protease-susceptible sites or by incorporation of fused partners to increase solubility and/or stability of the target protein (Esposito and Chatterjee, 2006).

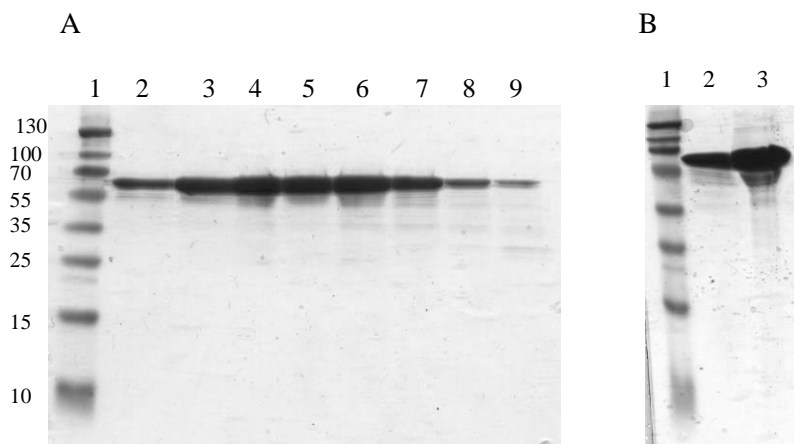
As is evident from Figure 50, TFPf degradation did not occur during expression and so it became necessary to prevent its degradation during extraction from the expressing bacterial cells. This can be achieved by carrying out extraction at low temperatures and using a fast protocol, as well as by using protease inhibitors. When TFPf was extracted in the presence of an EDTA-free protease inhibitor cocktail (Roche), and all extraction steps were carried out on ice except for cell lysis using lyzosome at room temperature, the extent of degradation of the target protein was not reduced (data not shown).

The final step in TFPf purification, which employed an AKTExpress system, was carried by SEC as described in section 2.2.19.3. However, since trigger factor from *P. frigidicola* has been proved to exist as monomer (Robin *et al.*, 2009), purification was expected to separate the full-length, monomeric TF from its degradation products. The gel filtration revealed numerous peaks that were not very well resolved (data not shown), which was consistent with the extensive degradation observed upon SDS-PAGE (Figure 58). Fractions corresponding to the peak representing TFPf were collected and subjected to a second SEC run, which resulted in a single, well-resolved and symmetrical peak that suggested successful separation from the degradation products (Figure 59). Further SDS-PAGE analysis showed a cleaner protein band without visible degradation products, which confirmed that a protein sample of high purity (Figure 60) had been obtained.



**Figure 59. Further purification of trigger factor to increase purity.**

Elution UV profile for SEC purification representing a re-run of TF purification carried out at 0.5 ml/min. The peak reached its maximum at 35 ml.



**Figure 60. SDS-PAGE analysis of (A) TFPf purified by SEC, (B) followed by vivaspin concentration.**

*Lanes (A); 1: molecular weight marker (values are presented in kDa beside each protein band), 2-9: fractions collected from the SEC peak (Figure 59). Lanes (B); 1: molecular weight marker, 2: SEC purified TFPf, 3: concentrated TFPf.*

In order to further ensure a high purity and homogeneity of the sample, only fractions from the top of the peak (*e.g.* lanes 3-6, Figure 60A) were pooled and used in crystallisation trials. The purified TFPf was concentrated (Figure 60B, lane 3)

to 10-30 mg/ml and crystallisation trials were set up with the protein immediately afterwards.

Crystallisation trials were performed using Hampton JCSG Suite Core (I, II, III, IV) screen and screens designed *de novo* using two of the most commonly used precipitants: PEG and Ammonium sulphate (Table 4).

| Precipitant                                     | Concentration | 100 mM Tris-HCl pH |      |     |     |
|---|---------------|--------------------|------|-----|-----|
| PEG 1.900 - 2.000                               | 15%           | 8                  | 7.07 | 6.1 | 5.2 |
|   | 20%           |                    |      |     |     |
|   | 25%           |                    |      |     |     |
|   | 30%           |                    |      |     |     |
| PEG 3350  | 15%           | 8                  | 7.07 | 6.1 | 5.2 |
|   | 20%           |                    |      |     |     |
|   | 25%           |                    |      |     |     |
|   | 30%           |                    |      |     |     |
| PEG 8000  | 15%           | 8                  | 7.07 | 6.1 | 5.2 |
|   | 20%           |                    |      |     |     |
|   | 25%           |                    |      |     |     |
|   | 30%           |                    |      |     |     |
| (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> | 3.5M          | 8                  | 7    | 6   |     |
|   | 3.0M          |                    |      |     |     |
|   | 2.5M          |                    |      |     |     |
|   | 2.0M          |                    |      |     |     |

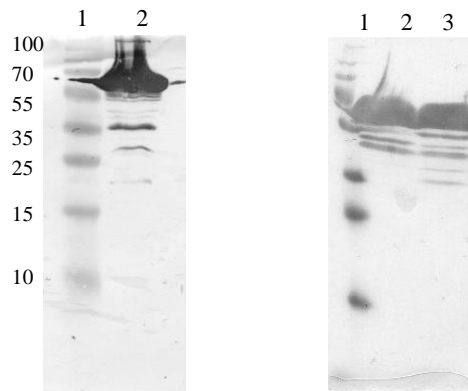
**Table 4. Summary of conditions used to set up trials of TFPf.**

Conditions presented in Table 3 were used to set up trials with TF at 20°C. It was observed that precipitation occurred in most wells containing PEG after trials were set up even though the appropriate 25 mg/ml concentration of protein was determined by PCT Pre-crystallisation test (Hampton). Very little precipitation occurred in wells in which (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> was precipitant, even up to 3.0 M (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> concentration. Trials were also set up using (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> at 3.25 M and an additional precipitant (0.3 M Malonate or 0.1 M and 0.2 M KCl or NaCl) in Tris buffer at pH 6, 7, 8. No visible crystallisation hits were obtained in any trials using this TFPf protein preparation.

#### 4.8. Expression, extraction and purification of trigger factor with a removable His<sub>6</sub>-tag (HIS-TEV-TFPf)

The hexahistidine tag is a valuable tool in IMAC, which is typically the first step in purification of recombinant proteins. However, the presence of His-tag at a protein terminus can not only impede the formation of crystals (Bergfors, 2009, DeLucas, 2009) but can also affect the native structure of the target protein (Chant *et al.*, 2005, Perron-Savard *et al.*, 2005).

In order to eliminate the possibility of the His-tag interfering with crystal growth, trigger factor was expressed with a TEV-cleavable His-tag as described in section 2.2.5.2. Expression of the HIS-TEV-TFPf construct from the pACYC184HisTEVtigPF vector in *E. coli* BL21 (DE3) was carried out by autoinduction at 25°C. Autoinduction usually involves longer production times than traditional IPTG-based approaches and the optimum expression length for HIS-TEV-TFPf was determined to be 16 h (data not shown).



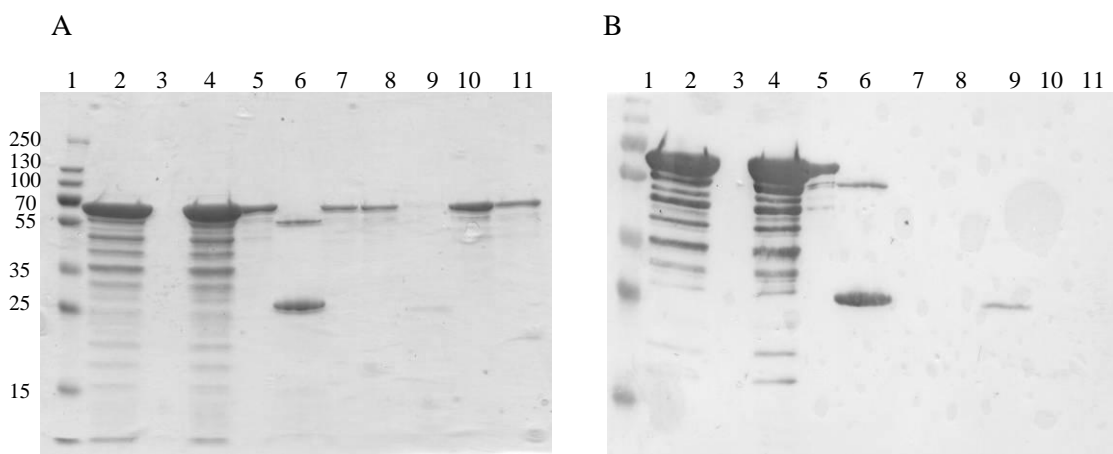
**Figure 61.** Western blot analysis of expression of HIS-TEV-TFPf expressed by autoinduction in ZYP-5052 at 25°C for 16 h (A); expressed and extracted HIS-TEV-TFPf (B).

Lanes (A); 1: molecular weight marker (values presented in kDa beside each band), 2: expressed TFPf. Lanes (B); 1: MW marker, 2: expressed TFPf, 3: extracted TFPf.

Figure 61A shows the occurrence of degradation during expression, whereas in Figure 61B, it can be observed that there was no significant increase noted in the level of degradation after lysis of *E. coli* cells. Cells were opened enzymatically using the same protocol as previously (see 2.2.6.1) but the change in

TF construct appeared to result in a significant improvement in stability and quality of the overall sample.

Purification was carried on an AKTExpress by IMAC (see 2.2.8.1.1), followed by SEC on HiLoad (see 2.2.8.1.4). The His-tag was removed by overnight incubation with AcTEV protease and the cleaved protein was obtained upon reverse IMAC on a 1-ml column (see 2.2.8.1.2). The results of purification presented in Figure 62 indicate, however, that the His-free form of trigger factor was also prone to degradation similar to the His-tagged form of the protein.



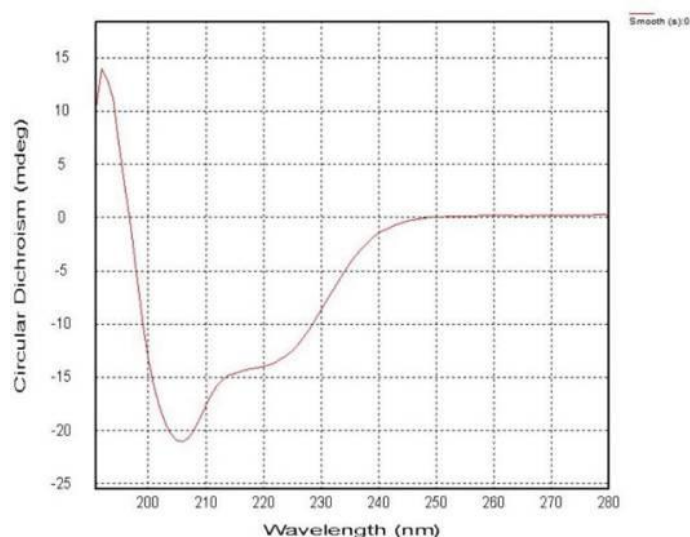
**Figure 62. Purification of HIS-TEV-TFPf. SDS-PAGE analysis of Coomassie stained (A) and immunoblotting (B).**

*Lanes; 1: molecular weight marker (values are presented in kDa beside each protein band), 2 and 4: TF purified by IMAC, 5: TF purified by SEC, 6: control protein, 7-8: TF after His-tag cleavage, 9: elution of His-tagged TEV protease, 10-11: TF concentrated on Vivaspin.*

Figure 62 shows successful removal of the histidine tag from TFPf using TEV protease. The protein band present in lanes 7 and 8 in the Coomassie-stained SDS-PAGE (Figure 62A) was not apparent in the immunoblot (Figure 62B), indicating that the His-free TFPf is undetectable by the His-tag-binding antibody. Lane 9 of the immunoblot (Figure 62B) reveals the presence of TEV protease that is itself His-tagged, a faint band is also visible in Coomassie stained PAGE. The purification and tag removal was characterised by significant loss of TFPf protein.

#### 4.9. Analysis of trigger factor by circular dichroism

CD spectroscopy is a valuable tool to examine the secondary structure and folding motifs of purified proteins and it plays an important role in structural investigation of macromolecules (Whitmore and Wallace, 2008). Purified HIS-TEV-TFPf sample was analysed by CD at the University of Limerick (see 2.2.9.2), and exhibited negative bands at ~222 and ~206 nm, whereas the positive band was seen at ~194 nm which is characteristic of proteins in which alpha helices dominate the structure (Figure 63). Figure 63 also reveals that the spectrum is limited and disturbed due to the extensive spectral noise, which most likely arises from a high absorption by the buffer. However, the detection of the positive peak confirmed the high degree of order of the helical protein.



**Figure 63. CD spectrum of HIS-TEV-TFPf in 10 mM sodium phosphate pH 7.0.**

CD spectrum was recorded at the wavelength between 190 and 280 nm and exhibited negative bands at ~206 and ~222 nm.

The theoretical content of  $\alpha$ -helices in HIS-TEV-TFPf was derived from its amino acid sequence and predicted, with the use of GOR4, HNN, Temp-prot and SOPMA algorithms, as 57.05%, 49.66%, 48.7% and 59.5%, respectively. Calculations of the actual secondary structure  $\alpha$ -helical content at the different spectrum range was carried out using CDNN software (CD Spectra Deconvolution v

2.1; (Böhm *et al.*, 1992), see 2.2.9.2; Figure 64). The  $\alpha$ -helical content of HIS-TEV-TFPf was found to vary between 43.5% and 62.4%, which was in agreement with the predicted values.

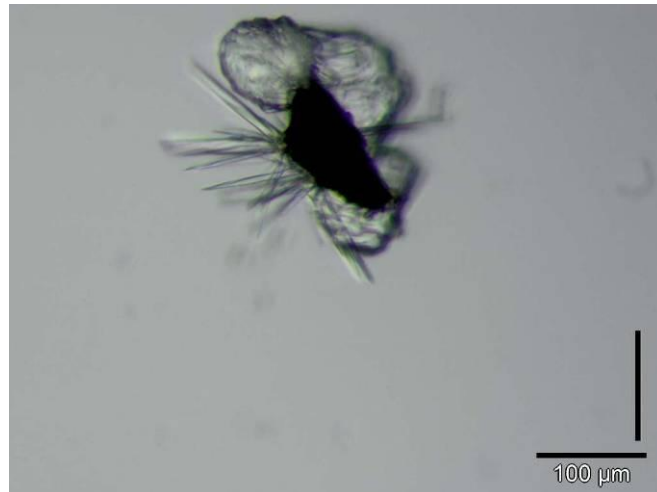
| Secondary structure element | 190-260 nm | 195-260 nm | 200-260 nm | 205-260 nm | 210-260 nm |
|-----------------------------|------------|------------|------------|------------|------------|
| $\alpha$ -helix             | n.d        | 43.5%      | 49.8%      | 62.4%      | 55.6%      |
| Antiparallel                | n.d        | 8.4 %      | 5.6%       | 3.7%       | 4.2%       |
| Parallel                    | n.d        | 5.0%       | 4.6%       | 3.7%       | 5.0%       |
| $\beta$ -turn               | n.d        | 16.1%      | 15.1%      | 12.9%      | 13.5%      |
| Random coil                 | n.d        | 15.3%      | 16.9%      | 17.0%      | 21.0%      |
| Total sum                   | n.d        | 88.3%      | 91.9%      | 99.8%      | 99.2%      |

**Figure 64. Secondary structure elements (% of amino acids) of TF in different CD spectrum ranges.**

The CD analysis shows that TF exhibits a predominantly helical structure, as predicted for its *E. coli* and *V. cholera* homologues. Moreover, this analysis served as a confirmation of the correct folding of the protein prior to initiating crystallisation experiments.

#### **4.10. Crystallisation of HIS-TEV-TFPf**

Initial crystallisation trials with purified HIS-TEV-TFPf were set up at a protein concentration of 10 mg/ml using a vapour-diffusion sitting-drop method and as per protocol established previously with TFPf (see 2.2.10.1). Trials were set up using a wide range of conditions from commercial screens JCSG Suite Core (I, II, III, IV; Hampton Research) at 4°C, 15°C and room temperature. Approximately 300 trials were set up which proved very difficult due to the extensive degradation of the construct, leading to reduced yields of protein.



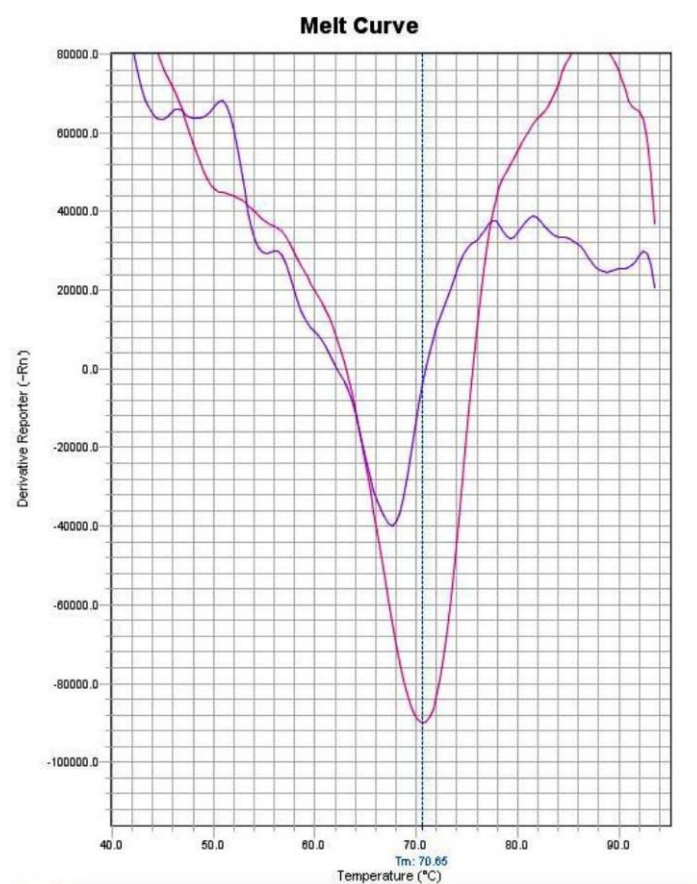
**Figure 65. Crystals that appeared in the JCSG screen (Hampton Research).**  
*Conditions: 0.1 M Na/K Phosphate, 25% 1, 2- Propanediol, 10% Glycerol.*

The appearance of crystals (Figure 65) was followed by X-ray diffraction analysis at the Swiss Light Synchrotron (SLS), carried out by Professor Tewik Soulimane, which revealed the crystals to be salt. Efforts to obtain crystals of TFPf were severely limited by the extensive degradation of the protein and its instability, which can negatively influence nucleation.

#### **4.10.1. Stability check on Thermofluor using SybrOrange**

The research stay at MPL was also used to perform additional analysis of TF (Thermofluor assay and TETRA detector assay-not shown) and to set up crystallisation trials with TF using HTP technology.

The stability of proteins in solution is important for crystallogenes and it can be very specific for each individual protein. Protein stability is often directly correlated with specific detergents but also with temperature, pH, buffer composition, and particular salts or additives. The Thermofluor assay allows the screening of many different conditions to identify those favourable for protein stability. Protein stability is indicated by  $T_m$  (melting temperature) and can be estimated using Sypro Orange, a fluorescent dye which binds to hydrophobic regions in soluble proteins (Figure 66).



**Figure 66. Summary of the melting curve demonstrating a shift in  $T_m$  for HIS-TEV-TFPf upon changing buffer, indicating greater protein stability.**

*Curves: purple - 50 mM BisTris pH 6.0, 150 mM NaCl; pink - 50 mM BisTris pH 6.0, 300 mM NaCl buffer in which the highest  $T_m$  was noted for the protein.*

The stability screen was inconclusive for TFPf while two suitable conditions were identified for HIS-TEV-TFPf, with increased stabilities ( $T_m$  of 68°C and 70°C, respectively) observed in 50 mM BisTris pH 6.0, 150 mM NaCl and 50 mM BisTris pH 6.0, 300 mM NaCl. No shift was observed as the protein did not display a fluorescent signal in the preliminary conditions. The conditions chosen to purify HIS-TEV-TFPf were in the former buffer with lower salt concentration to avoid possible formation of salt crystals in the crystallisation drops.

#### **4.10.2. Crystallisation of HIS-TEV-TFPf at MPL**

HIS-TEV-TFPf crystallisation trials were set up as described in section 2.2.10.1 with 20 mg/ml, 15 mg/ml and 7.5 mg/ml protein in 96 well plates at 100 nL

: 100 nL and 100 nL : 50 nL ratios of protein to buffer and using commercial screens Wizard 1&2, Wizard 3&4, JCSG+ at 4°C and 20°C. Most wells with 20 mg/ml TF displayed precipitation, which indicates very high supersaturation. Following reduction of the protein concentration, a significant drop in the number of wells with precipitation was noted but crystallisation hits were still not observed. Crystallisation wells must display a level of moderate supersaturation for spontaneous nucleation to occur (Figure 6), after which the protein concentration will decrease, thereby resulting in supersaturation reaching a lower level (the “metastable zone”) allowing the growth of crystals. The phase diagram offers a theoretical basis for crystallogenesis. It provides in a relatively simple way the knowledge necessary to grow protein crystals but it is not always possible to establish an accurate phase diagram for individual proteins as reaching equilibrium can take several months and requires large amounts of protein.

Failure to obtain crystallisation hits with HIS-TEV-TFPf during the research visit at MPL was despite screening of nearly 500 conditions. Therefore, no further optimisation, crystallisation or X-ray diffraction of this particular protein construct was attempted.

#### **4.11. Construction of TF variants and preliminary expression screening at OPPF**

Following the establishment of a collaborative venture with OPPF, a high-throughput structural biology facility for vector construction and expression screening, new truncated versions of TF and constructs of individual TF domains were designed for expression and analysis. This was to investigate whether cropping TF would result in better ordered component parts and increase the chances of obtaining stable, crystallisation-suitable product(s).

Eight constructs were designed, amplified and cloned into vectors encoding a variety of fusion tags known to improve protein production or increase the likelihood of successful purification. After cloning, proteins were expressed in *E. coli* strains Lemo21 and Rosetta pLacI by IPTG induction in Power Broth and by autoinduction in Overnight Express Instant TB. Figure 67 shows a summary of the

designed constructs, including the tags used with each protein construct and the results of the preliminary screening of expression of each construct.

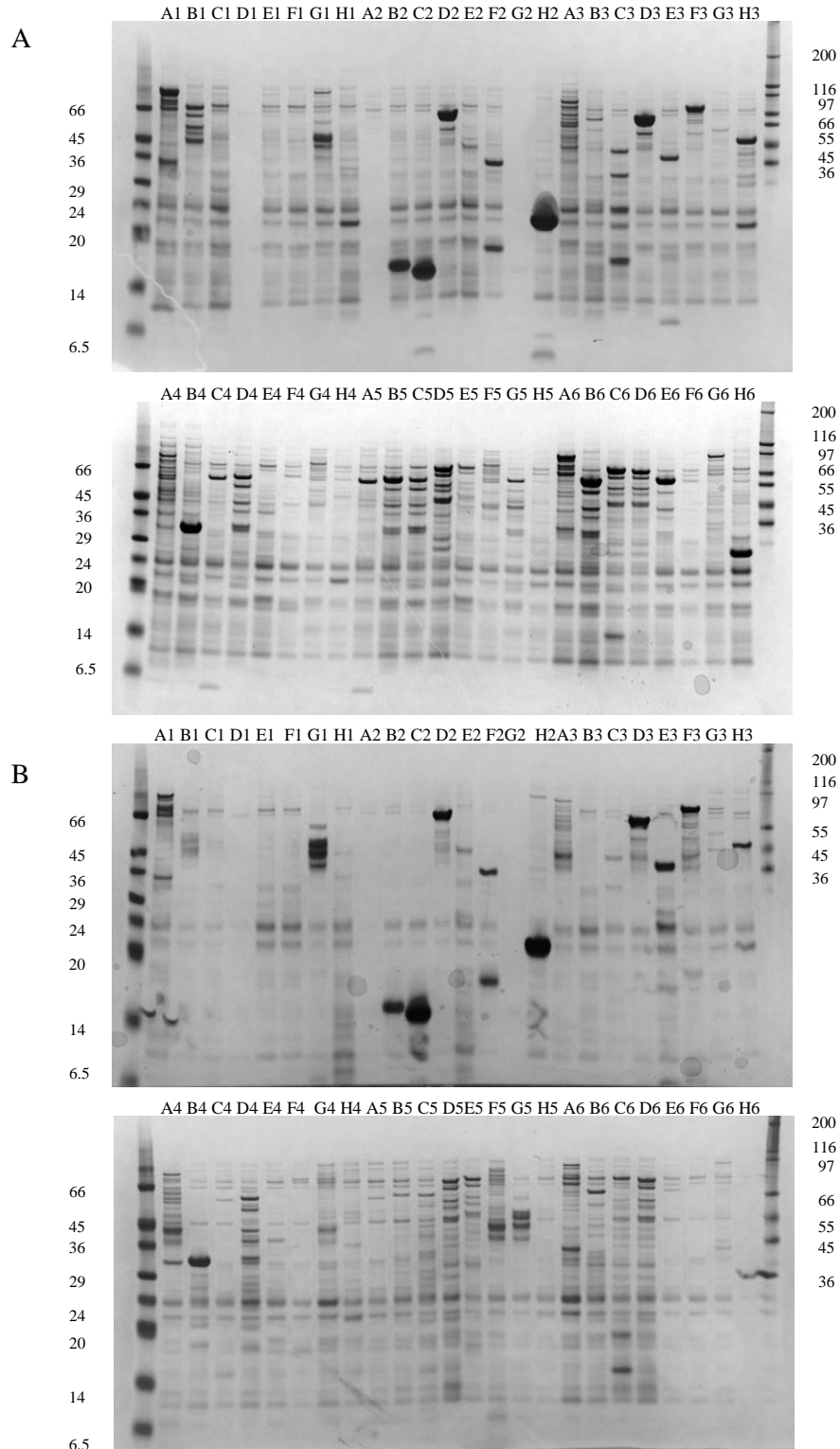
| Well | Gene name | aa_N-aa_C | Vector    | MW-POI | MW-tag | Total MW | Expression |
|------|-----------|-----------|-----------|--------|--------|----------|------------|
| A1   | eGFP      |           |           |        |        |          |            |
| B1   | tigPF-C   | 250-447   | pOPINM    | 22.694 | 43.7   | 66.394   |            |
| C1   | tigPF-C   | 250-447   | pOPINMSYB | 22.694 | 16     | 38.694   |            |
| D1   | tigPF-C   | 250-447   | pOPINS3C  | 22.694 | 13.7   | 36.394   |            |
| E1   | tigPF-C   | 250-447   | pOPINF    | 22.694 | 1.7    | 24.394   |            |
| F1   | tigPF-C   | 250-447   | pOPINE    | 22.694 | 0.8    | 23.494   |            |
| G1   | tigPF-P   | 150-250   | pOPINM    | 10.808 | 43.7   | 54.508   |            |
| H1   | tigPF-P   | 150-250   | pOPINMSYB | 10.808 | 16     | 26.808   |            |
| A2   | tigPF-P   | 150-250   | pOPINS3C  | 10.808 | 13.7   | 24.508   |            |
| B2   | tigPF-P   | 150-250   | pOPINF    | 10.808 | 1.7    | 12.508   |            |
| C2   | tigPF-P   | 150-250   | pOPINE    | 10.808 | 0.8    | 11.608   |            |
| D2   | tigPF-N   | 2-150     | pOPINM    | 16.901 | 43.7   | 60.601   |            |
| E2   | tigPF-N   | 2-150     | pOPINMSYB | 16.901 | 16     | 32.901   |            |
| F2   | tigPF-N   | 2-150     | pOPINS3C  | 16.901 | 13.7   | 30.601   |            |
| G2   | tigPF-N   | 2-150     | pOPINF    | 16.901 | 1.7    | 18.601   |            |
| H2   | tigPF-N   | 2-150     | pOPINE    | 16.901 | 0.8    | 17.701   |            |
| A3   | tigPF-PC  | 150-447   | pOPINM    | 33.77  | 43.7   | 77.47    |            |
| B3   | tigPF-PC  | 150-447   | pOPINMSYB | 33.77  | 16     | 49.77    |            |
| C3   | tigPF-PC  | 150-447   | pOPINS3C  | 33.77  | 13.7   | 47.47    |            |
| D3   | tigPF-PC  | 150-450   | pOPINF    | 33.77  | 1.7    | 35.47    |            |
| E3   | tigPF-PC  | 150-447   | pOPINE    | 33.77  | 0.8    | 34.57    |            |
| F3   | tigPF-NP  | 2-250     | pOPINM    | 27.82  | 43.7   | 71.52    |            |
| G3   | tigPF-NP  | 2-250     | pOPINMSYB | 27.82  | 16     | 43.82    |            |
| H3   | tigPF-NP  | 2-250     | pOPINS3C  | 27.82  | 13.7   | 41.52    |            |
| A4   | tigPF-NP  | 2-250     | pOPINF    | 27.82  | 1.7    | 29.52    |            |
| B4   | tigPF-NP  | 2-250     | pOPINE    | 27.82  | 0.8    | 28.62    |            |
| C4   | tigPF     | 5-438     | pOPINE    | 48.927 | 0.8    | 49.727   |            |
| D4   | tigPF     | 5-438     | pOPINF    | 48.927 | 1.7    | 50.627   |            |
| E4   | tigPF     | 5-438     | pOPINMSYB | 48.927 | 16     | 64.927   |            |
| F4   | tigPF     | 5-438     | pOPINS3C  | 48.927 | 13.7   | 62.627   |            |
| G4   | tigPF     | 5-438     | pOPINM    | 48.927 | 43.7   | 92.627   |            |
| H4   | tigPF     | 5-438     | pOPINW    | 48.927 | 2.7    | 51.627   |            |
| A5   | tigPF     | 5-438     | pOPINFS   | 48.927 | 2.7    | 51.627   |            |
| B5   | tig PF    | 2-438     | pOPINE    | 49.455 | 0.8    | 50.255   |            |
| C5   | tig PF    | 2-438     | pOPINF    | 49.455 | 1.7    | 51.155   |            |
| D5   | tig PF    | 2-438     | pOPINTRX  | 49.455 | 14.7   | 64.155   |            |

|    |        |       |            |        |      |        |  |
|----|--------|-------|------------|--------|------|--------|--|
| E5 | tig PF | 2-438 | pOPINS3C   | 49.455 | 13.7 | 63.155 |  |
| F5 | tig PF | 2-438 | pOPINM     | 49.455 | 43.7 | 93.155 |  |
| G5 | tig PF | 2-438 | pOPINFS    | 49.455 | 2.7  | 52.155 |  |
| H5 | tig PF | 2-438 | pOPINW     | 49.455 | 2.7  | 52.155 |  |
| A6 | tig PF | 2-438 | pOPINHALO7 | 49.455 | 35.7 | 85.155 |  |
| B6 | tigPF  | 2-447 | pOPINF     | 50.497 | 1.7  | 52.197 |  |
| C6 | tigPF  | 2-447 | pOPINTRX   | 50.497 | 14.7 | 65.197 |  |
| D6 | tigPF  | 2-447 | pOPINS3C   | 50.497 | 13.7 | 64.197 |  |
| E6 | tigPF  | 2-447 | pOPINFS    | 50.497 | 2.7  | 53.197 |  |
| F6 | tigPF  | 2-447 | pOPINW     | 50.497 | 2.7  | 53.197 |  |
| G6 | tigPF  | 2-447 | pOPINHALO7 | 50.497 | 35.7 | 86.197 |  |
| H6 | tigPF  | 2-447 | pOPINE     | 50.497 | 0.8  | 51.297 |  |

**Figure 67. Summary of the TF $Pf$  expression screen carried out using constructs generated at OPPF.**

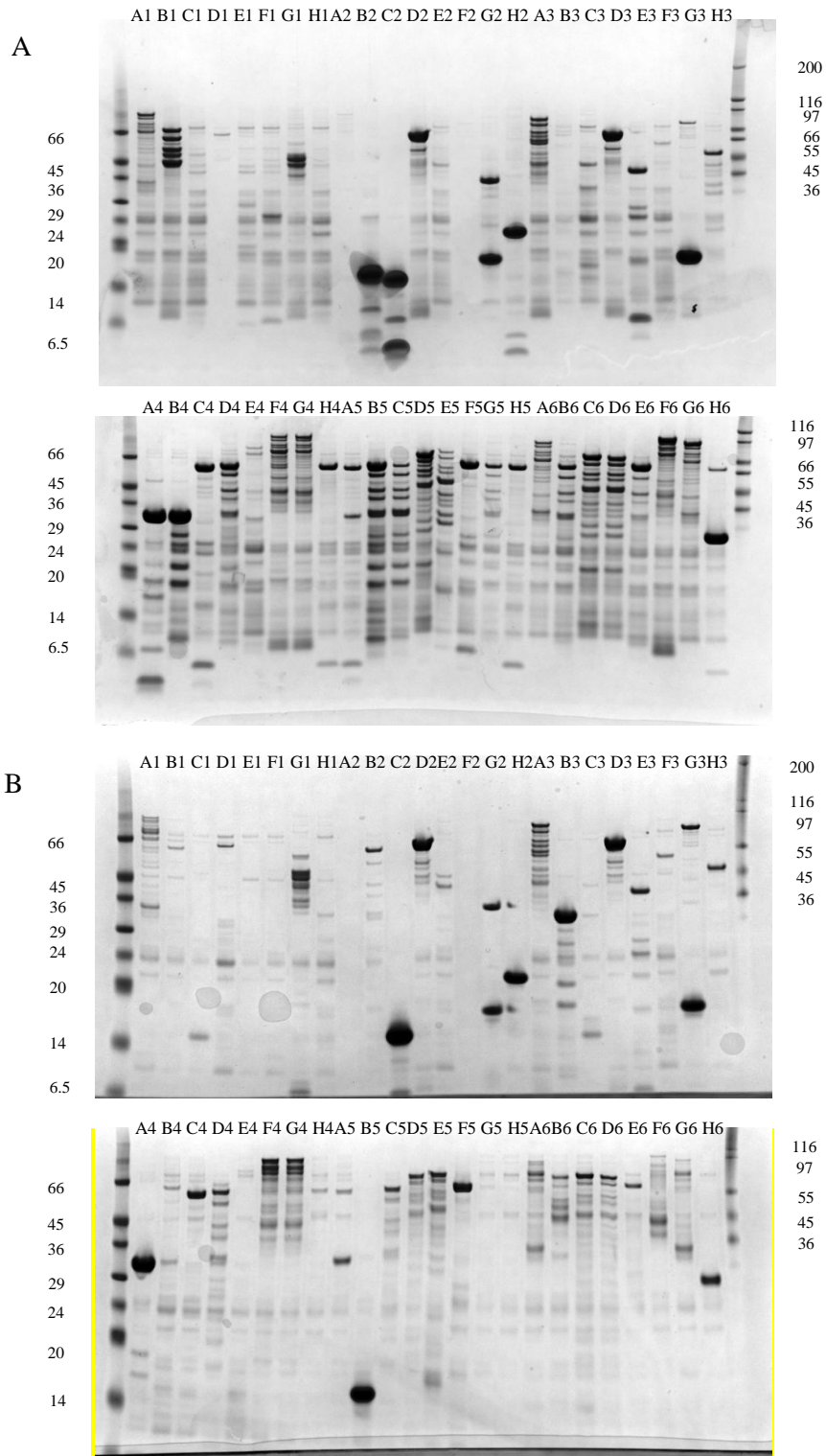
Column headings; Well: position of the plasmid in the 96-well plate, Gene name: given name of gene, aa\_N-aa\_C: TF amino acids encoded in the relevant gene, Vector: plasmid into which the gene was cloned, MW-POI, MW-TAG and Total MW: molecular weight of the protein of interest, recombinant tag and final product in kDa, respectively, Expression: results of initial expression screen in *E. coli* (green - protein of predicted MW was observed on SDS-PAGE; pink - MW of protein different to that predicted was observed on SDS-PAGE, red - no protein was observed on SDS-PAGE. A list of pOPIN vectors and fusion tags with explanations is available in Appendix (page 171).

The incorporation of fusion tags is also used to improve the production of recombinant proteins in *E. coli*, particularly those prone to insolubility or aggregation. A variety of fusion proteins which differ in size, structure and function have been employed to enhance protein expression, solubility and purification (Bird, 2011). Our experience with the previous TF $Pf$  construct showed that it did not encounter difficulties in expression and high TF yields were obtained; therefore the fusion tags were utilised in this approach in an attempt to reduce degradation of the expressed product (see Figure 58 and 62). Proteolysis is part of maintaining the homeostasis within a cell and often leads to the removal of incorrectly folded and unrecognisable proteins, which can include highly expressed recombinant products. Generating protective fusions has been shown to protect numerous target proteins from degradation (Martinez *et al.*, 1995, Butt *et al.*, 2005).



**Figure 68. SDS-PAGE analysis of preliminary expression screen followed by IMAC purification of TFPf constructs expressed in *E. coli* Lemo21.**

Expression was carried out in PB by IPTG induction for 16 h (A) and by autoinduction in Overnight Express<sup>TM</sup> Instant TB Medium for 24 h (B). Labels represent constructs listed in Figure 67, values in kDa are presented beside each band of protein marker next to the pictures.



**Figure 69. SDS-PAGE analysis of preliminary expression screen followed by purification by IMAC of TFp constructs expressed in *E. coli* Rosetta pLacI.** Expression was carried out in PB by IPTG induction for 16 h (A) and by autoinduction in Overnight Express<sup>TM</sup> Instant TB Medium for 24 h (B). Labels represent constructs listed in Figure 67, values in kDa are presented beside each band of protein marker next to the pictures.

Figure 67 reveals that 32 of the 47 constructs were successfully expressed with detectable proteins of molecular weights predicted using Vector-NTI. Of the remaining 15 constructs, 7 were expressed but yielded products of molecular weight different to those predicted and the remaining 8 constructs did not yield detectable product. This latter group of 8 proteins consisted mostly of constructs of single domains (tigPF-C, tigPF-N, tigPF-P), of which half were fused to the MsyB tag, which could indicate a problem with this particular tag. The differences in predicted and obtained molecular weight are most likely a consequence of a human error during cloning or expression. However, there is also a possibility that those detected proteins are *E. coli* proteins that co-purified, which could be simply determined in the future by immunoblotting. Co-purification of *E. coli* proteins with a histidine-tagged recombinant protein is quite common (Bolanos-Garcia and Davies, 2006), and the co-purifying protein can be misinterpreted as the protein of interest, especially when the expression level of the latter is poor.

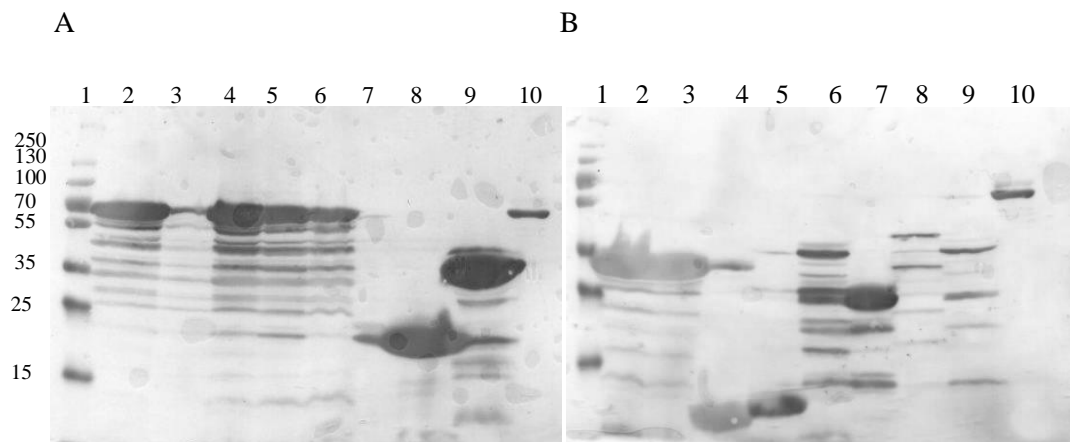
Figures 68 and 69 show the SDS-PAGE analysis of the expression experiments summarised in Figure 67. The majority of constructs expressed well and generated high yields of protein. Initially, truncated constructs of full-length TF, expressed from the pOPINF vector with an N-terminal cleavable His-tag, were thought most likely to be suitable for crystallisation studies as cropping the protein at the C-terminus was thought likely to reduce its flexibility, thereby making it more stable and prone to crystallise. However, the tigPF 2-438 and tig 5-438 constructs (Figures 68 and 69; wells C5 and D4, respectively) exhibited degradation during their initial expression, indicating likely unsuitability for crystallisation trials.

Proteins representative of each group of constructs (Figure 70) were selected for further analysis to identify likely candidates for crystallisation and to observe the possible changes in degradation resulting from truncation of the protein. Therefore, constructs with histidine tags on C- and N-terminus were selected if they were expressed in the preliminary screen to be able to compare the effect of truncations on protein behaviour. In two cases, those of the N-domain and PC-domain, N-terminus constructs had to be substituted with one tagged with SUMO (Figure 70 and 71).

|    | Gene name | aa_N-aa_C | Vector   | MW   |
|----|-----------|-----------|----------|------|
| 1  | tigPF     | 2-447     | pOPINF5  | 53.2 |
| 2  | tigPF     | 2-447     | pOPINF   | 52.2 |
| 3  | tigPF     | 2-438     | pOPINF5  | 52.2 |
| 4  | tigPF     | 2-438     | pOPINF   | 51.2 |
| 5  | tigPF     | 5-438     | pOPINF   | 50.6 |
| 6  | tigPF     | 5-438     | pOPINE   | 49.7 |
| 7  | tigPF-N   | 2-150     | pOPINF   | 18.6 |
| 8  | tigPF-N   | 2-150     | pOPINS3C | 30.6 |
| 9  | tigPF-NP  | 2-250     | pOPINF   | 29.5 |
| 10 | tigPF-NP  | 2-250     | pOPINE   | 28.6 |
| 11 | tigPF-P   | 150-250   | pOPINE   | 11.6 |
| 12 | tigPF-P   | 150-250   | pOPINF   | 12.5 |
| 13 | tigPF-C   | 250-447   | pOPINE   | 23.5 |
| 14 | tigPF-C   | 250-447   | pOPINF   | 24.4 |
| 15 | tigPF-PC  | 150-447   | pOPINS3C | 47.5 |
| 16 | tigPF-PC  | 150-447   | pOPINE   | 34.6 |

**Figure 70. Summary of representatives of each construct identified for crystallisation trials.**

Column headings; Gene name: gene given name, aa\_N-aa\_C: TF amino acids encoded in the relevant gene, Vector: plasmid into which the gene was cloned, Total MW: molecular weight of the final product in kDa. A list of pOPIN vectors and fusion tags with explanations is available in Appendix (page 171).

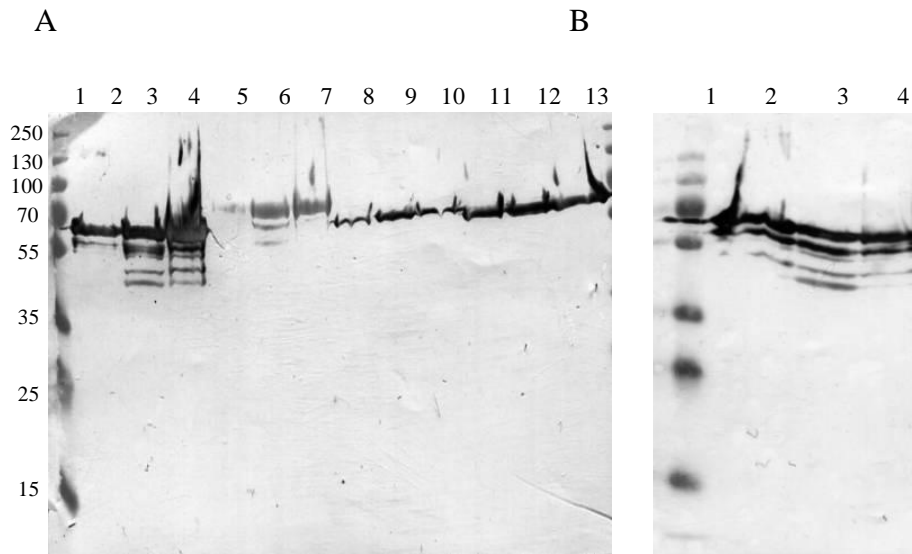


**Figure 71. Immunoblot analysis of expression of TFpf constructs.**

Lanes (A); 1: molecular weight marker (values presented in kDa beside each band), lanes represent expressed construct specified in Figure 70; 2: 1, 3: 2, 4: 3, 5: 4, 6: 5, 7: 6, 8: 7, 9: 8, 10: WB control. Lanes (B): 2: 9, 3:10, 4: 11, 5: 12, 6: 13, 7: 14, 8: 15, 9: 16, 10: WB control.

As discussed previously (see 4.7), a key to successful crystallisation is the availability of pure, stable and homogeneous protein. The preliminary production of representative TF constructs was carried out under standard conditions involving IPTG induction and yielded large amounts of protein for most constructs (Figure 70). However, the majority of constructs exhibited considerable degradation of the protein of interest during expression. Samples in lanes 7 and 8 representing constructs of N-domain showed a lesser degree of degradation, likewise for samples in lanes 11 and 12 representing constructs of domain-P, whereas the constructs of PC-domains present in lanes 15 and 16 were degraded to the extent that the full-length product was no more intense than the other bands. This may indicate that much of the degradation seen in the protein is linked to the C-domain, but this would require further analysis to confirm. On the other hand, in case of N-domain, the construct in lane 8 was expressed with the SUMO tag, which might have had a positive influence on the degradation problem, although this did not appear to lead to any improvement for the PC-domain (Figure 71B, lane 8).

Likewise, many full-length constructs of TF, and truncated ones expressed with other fusion tags (SUMO, MBP, TRX, HALO7), exhibited degradation during expression, poor solubility or molecular weights that differed from those predicted. Figure 72 summarises the analysis of 5 constructs based on a full-length TF product that appeared the most promising of the initially screened constructs (data not shown). It appears from this investigation that the presence of fusion tags SUMO and TRX (lanes 8-10 and 11-13, respectively) improved product quality since no evidence of degradation was noted. The analysis also revealed a significant proportion of protein to be insoluble in most cases, which could possibly be improved by further optimisation.



**Figure 72. Immunoblot analysis of TF full-length constructs expressed in several different vector formats.**

*Lanes (A); 1: molecular weight marker (values presented in kDa beside each band), 2-4: tig 2-447 pOPINE, 5-7: tig 2-447 pOPINW, 8-10: tig 2-447 pOPINS3C, 11-13: tig 2-447 pOPINTRX. Lanes (B); 2-4: tig 5-438 pOPINE. The order of samples in each construct is: whole cells, soluble and insoluble fractions.*

From the search for a suitable candidate protein for the crystallisation trials carried out to date, it appears that a TRX tag might be beneficial for up-scaling and crystallisation of TF (Figure 72, lanes 11-13). Immunoblot analysis indicates a good yield of protein in this case and no detectable degradation during the extraction process.

## 5. OVERVIEW AND FINAL CONCLUSIONS

Protein crystallisation is a significant bottleneck in structure determination; crystallisation of membrane proteins poses a particular set of challenges, which is evident in their poor representation in the form of solved structures in the Protein Data Bank (PDB). Despite recent dramatic advances in the areas of automation and miniaturisation that impact on crystallisation and enable hundreds of crystallisation conditions to be screened using only nanolitres of product (Luft *et al.*, 2011, Moraes *et al.*, 2014), the proportion of known membrane proteins with solved structures remains very small (Baker, 2010). One of the main limitations in many HTP platforms is the challenge of obtaining high quality recombinant protein that maintains the native protein form. The research presented in this thesis provides a basis for successful recombinant expression of the two proteins of interest and demonstrates systematic progress towards obtaining protein suitable for growing diffraction-quality crystals for their analysis.

The first protein, CzrB from *Thermus thermophilus*, belongs to the CDF family of metal transporters, which is highly underrepresented in PDB as only one structure of a full-length protein representative – the *E. coli* zinc transporter encoded by *yiiP* – has been determined to date by X-ray crystallography (Lu and Fu, 2007). Additionally, two three-dimensional structures of cytoplasmic domains of CDF family members, from *Thermus thermophilus*, as investigated in the present work (Cherezov *et al.*, 2008) and *Thermatoga maritima* (Hattori *et al.*, 2007), have been submitted to PDB to date. The monomeric structure of cytoplasmic domain from *T. maritima* TM0876<sub>206-306</sub> consists of two  $\alpha$ -helices and three mixed  $\beta$ -strands, similarly as determined in this group Zn<sup>2+</sup>-bound and apo-form structures of the CzrB cytosolic domain (Cherezov *et al.*, 2008) (Cherezov), but in TM0876<sub>206-306</sub>  $\beta$ 1 is distorted by the presence of Pro233 (Higuchi *et al.*, 2009).

The present study illustrates that the full-length membrane-bound CzrB, while previously demonstrated to exist as a homodimer (Cherezov *et al.*, 2008), occurs in a mixture of monomeric and dimeric forms in detergent-solubilised solutions. This observation, which has previously been noted within the CDF membrane protein family (Russell and Soulimane, 2012), is believed to result from a

reduction in the stability of the protein upon its removal from its native membrane environment. The occurrence of the CzrB monomer-dimer mix was confirmed in this work using SDS-PAGE, SEC and TETRA detector analyses and it was established that changes in detergents, buffers, pH or the expression construct did not significantly increase the stability of CzrB or completely eliminate heterogeneity from the protein samples. A more monodisperse and mostly monomeric sample was observed on SDS-PAGE, however, upon solubilisation of the protein at high temperature, which suggested better suitability for crystallisation trials, however did not generate any hits. Possible conclusion was drawn that the higher stability of CzrB is related to the dimeric state and CzrB had tendency to eventually, despite the solubilisation method, exhibit both monomeric and dimeric forms. Nonetheless, the consistently observed heterogeneity of protein states in purified preparations is thought to be a major contributory factor in the subsequent inability to obtain high quality protein crystals of CzrB for analysis.

Current knowledge and hypotheses about the structures of other molecules in this CDF transporter family are based on the single structure of *E. coli* family member; therefore solving the structure of full-length CzrB from *T. thermophilus* would likely provide a significant additional insight into this group of proteins. The *E. coli* YiiP transporter shares the molecular architecture of 6 transmembrane regions and a cytosolic domain (CTD) predicted for most CDF family transporters, including the present CzrB. While it has been hypothesised that the cytoplasmic domain plays the role of a metallochaperone-like molecule in CDF transporters (Lu *et al.*, 2009), the mechanism of action of the protein is not well understood. The cytoplasmic domain from *T. thermophilus* has been demonstrated to bind zinc, leading to a considerable conformational change in the protein structure, but the concept of cytoplasmic domain as metallochaperone is not consistent throughout the entire CDF family of transporters. Russell *et al.* (2012) have reported that a CDF family member from the marine bacterium *Maricaulis maris* (MmCDF3) can bind divalent cations despite lacking a cytoplasmic domain. They could also measure the affinity of the membrane protein for specific substrates using isothermal titration calorimetry (ITC), which had previously been used in thermodynamic studies of YiiP binding of Zn, Cd and Hg (Chao and Fu, 2004).

In clear structural differences within the family, YiiP exists as a homodimer and its cytosolic domain is thought to play a major role in stabilising the dimer, whereas CzrB appears to lose stability upon its removal from the cell membrane and to exist as a heterogeneous monomer-dimer mixture. Similarly, while monodisperse YiiP can be obtained by HPLC, which might indicate that YiiP is more stable than other family members or that its stability is enhanced by the presence of glycerol in the purification buffer, similar attempts at purification of CzrB, including with the inclusion of glycerol, failed to yield a monodisperse solution or to improve crystallisation.

The heterogeneity of CzrB preparations observed throughout the project was believed to be related to the lack of stability of the protein. The occurrence of oligomeric forms of membrane proteins is often detergent-dependent and, to avoid this oligomerisation, it can be necessary to replace the detergent used in protein solubilisation for the purification process. While several detergents were tested in this work to improve the stability of the CzrB dimer, none resulted in a detectable increase in sample monodispersity. It has been reported that the use of mixed micelles improved monodispersity in the MmCDF3 CDF family transporter (Russell *et al.*, 2012) but this was not attempted in this project. In the published study, Russell reported that mixed micelles created using FC12:DDM ratios of 3:1, 1:1 and 1:3 M were used during purification and that samples with a majority or equimolar amounts of FC12 continued to exhibit large aggregation peaks on SEC, whereas a clear separation of aggregated protein and non-aggregated protein was achieved in the micelles formed in 3:1 DDM: FC12 (Russell *et al.*, 2012). Given the sequence homology and similar behaviour of the MmCDF3 and CzrB, the creation of mixed micelles seems as a promising approach to obtain more monodisperse population of CDF family protein and could be profitable when applied in this project even though there was no report to date of this method resulting in solving the structure of MmCDF3.

Another approach used to improving the stability of metal transporters during crystallisation has been to carry out trials in the presence of zinc. It has been previously reported by Coudray and co-workers (Coudray *et al.*, 2013) that low zinc concentrations ( $> 10 \mu\text{M}$ ) can enhance the production of well-ordered crystals,

whereas higher concentrations (from 10  $\mu\text{M}$  to greater than 100  $\mu\text{M}$ ) can reduce crystal quality or inhibit crystal growth. In addition, Kolaj (2008) demonstrated that the addition of 10 mM  $\text{ZnCl}_2$  to purified CzrB led to a decrease in crystal size and changes in morphology. These findings, however, suggest that the presence of zinc is more likely to influence the state of existing crystals rather than to lead to their initial formation and so is unlikely to overcome the major bottleneck observed in the present work of initiating the formation of original protein crystals.

Moreover, the ability to measure the stability of integral membrane proteins such as transporters, which are characterised by challenging activity assays in their detergent-solubilised forms, is therefore very valuable. The fluorescence-based thermal stability assay used in this study allows the examination of the influence of factors external to the protein molecule itself, such as buffers, detergents, pH or ligands on its stability. By monitoring the shift in the melting temperature ( $T_m$ ) the general integrity of a membrane protein can be established in a wide variety of conditions. This method of screening the shift in  $T_m$  was successfully adopted by Tate and Schertler to increase the stability of the turkey  $\beta 1$ -adrenergic receptor (Tate and Schertler, 2009). Using an alanine scanning approach to generate a library of protein variants for screening, they obtained an increase in  $T_m$  of 21°C in a mutated protein, which was successfully crystallised. In this work, the thermofluor assay carried out using CMP dye to detect buffer conditions under which an increase in the  $T_m$  of CzrB was noted, indicating that a higher stability could be achieved in this manner. However, as CMP binds to cysteines in the protein sequence and the only cysteine (Cys261) in the CzrB sequence resides within the cytoplasmic domain, the  $T_m$  may not be a good indicator of the integrity of the folded state of the whole protein molecule in the case of CzrB.

The difficulties of crystallising membrane proteins *in surfo* are also mostly related to their loss of stability when removed from the cell membrane and to the limited protein-protein contacts in PDCs (Privé, 2007). As discussed previously (section 1.6.2), crystallisation *in surfo* generates type II 3D crystals that are often characterised by poor diffraction. This is due to the reduction in the essential protein-protein contacts due to the presence of the detergent micelle “belt” which covers the hydrophobic surface of the protein. As a consequence, *in surfo*

crystallisation is critically dependent on the choice of detergent as a determinant of the shape and size of the detergent micelle, which is crucial for crystal formation. CzcB was previously purified in a variety of detergents to identify the optimal detergent in this regard (Kolaj, 2008), which revealed maltosides, and UDM in particular, to be best suited to mimicking the protein's natural environment. Therefore, the majority of crystallisation trials in this project were carried out in UDM, which yielded crystals with a high detergent content. These temperature-sensitive crystals diffracted poorly, which was likely due to their high detergent content, a characteristic difficulty encountered when analysing relatively small-sized molecules resident in large detergent micelles. Large detergent micelles, often present in excess after concentration of protein samples, restrict protein-protein contacts and may inhibit the formation of crystals (Seddon *et al.*, 2004). In the context of CzcB, therefore, the results indicate the need to carry out an additional, more extensive screen of a wider range of detergents suitable for its crystallisation and that greater effort should be directed towards the removal of excess detergent from samples in future analyses.

Membrane proteins with large cytoplasmic domains have a significant advantage over those lacking such a domain in having a larger area available to form the protein-protein contacts that are essential for crystal growth (Carpenter *et al.*, 2008). One successful approach to enhance protein-protein contacts within PDCs is co-crystallisation with an appropriate protein-binding chaperone, which involves engineering the scaffolds of a membrane protein to increase its exposed and accessible surface area. The formation of the chaperone:membrane protein complex not only enlarges the hydrophilic surface area available to form crystal contacts but also typically increases the solubility of the target protein. This is also an alternative approach for proteins without a suitable cytoplasmic or extracellular domain may be the insertion of stable soluble protein into the transmembrane loop of a membrane protein. A successful example of this technique lies in the insertion of T4 lysozyme into the third intracellular loop of G protein-coupled receptors (GPCRs), which has contributed to the determination of the structures of several members of this important class of receptors (Cherezov *et al.*, 2007, Kruse *et al.*, 2012, Wu *et al.*, 2012). Insertion of the additional protein has, unsurprisingly however, been found to

alter membrane protein function in some cases, such as the  $\beta$ 2AR construct ultimately used for structure determination and in which case the added T4 lysozyme led to binding of the GPCR to G-proteins (Rasmussen *et al.*, 2011b). A thermostabilised mutant of apocytochrome b562(RIL) has also been used in a similar manner and it may be better suited to this role as an insertion partner judging by its associated success in determination of the structure of human A(2A) adenosine receptor to very high resolution, 1.8 Å (Liu *et al.*, 2012). It may be concluded, therefore, that the use of this soluble-protein-insertion approach requires an extensive effort both in molecular engineering and in screening for the most suitable candidate that balances successful crystallisation of the protein of interest with retaining its biological activity. Insertion of highly crystallisable T4 lysozyme could potentially increase a chance of CzcB to form crystals through improving the stability of the protein once removed from the cellular membranes. It is believed that increased stability of CzcB would lead to more homogenous sample, most likely in dimeric form.

Increasing the hydrophilic surface of a membrane protein to enhance its crystallisation can also be achieved using antibody fragments, especially Fv and Fab fragments to date. The first success in solving a protein co-crystal structure with an antibody fragment was that of bacterial cytochrome *c* oxidase with an Fv molecule, as reported by Iwata in 1995 (Iwata *et al.*, 1995). The structure revealed, intriguingly, that there were no direct interactions between the membrane protein molecules and that Fv fragments were instead involved in all crystal contacts. Subsequently, co-crystallisation with Fab or Fv fragments derived from monoclonal antibodies led to successful high resolution crystal structures of many membrane proteins, including respiratory complexes (Hunte *et al.*, 2000), ion channels (Zhou *et al.*, 2001, Dutzler *et al.*, 2003, Hibbs and Gouaux, 2011), membrane transporters (Shaffer *et al.*, 2009, Fang *et al.*, 2009) and GPCRs (Rasmussen *et al.*, 2007). It is now recognised that antibody fragments not only play a role in improving protein packing during crystallisation but they can also influence the conformational flexibility of the membrane protein, which consequently will increase its stability (Hunte and Michel, 2002). More recently, Kobilka *et al.* succeeded in isolating camelid single-chain antibody fragments (nanobodies) that were characterised by

exceptionally high affinity and specificity against cavities in human  $\beta(2)$  adrenergic receptor that were inaccessible to larger antibodies and fragments, leading to the first active state crystal structure of the receptor being achieved in complex with the nanobody (Rasmussen *et al.*, 2011a). This approach proved fundamental in stabilising active conformations and obtaining well-diffracting crystals of  $\beta 2$ -AR (Rasmussen *et al.*, 2011b, Steyaert and Kobilka, 2011)

Approximately 200 of solved structures of membrane proteins deposited in PDB to date have been obtained using the method of lipid cubic phase (LCP) (Caffrey, 2015). LCP is a continuously developing alternative to traditional crystallisation methods that is based on mimicking the environment of the lipid bilayer to attempt to ensure the stability and native conformation of a membrane protein of interest. The use of LCP in this project was limited to work periods spent at MPL and yielded a single crystal, albeit too small to analyse or to photograph. The appearance of such a crystal remains promising, however, and continued exploration of this crystallisation method with CzcB is further justified by the increasing number of structures that have solved using LCP. A structure of the *ba3*-type cytochrome *c* oxidase from *T. thermophilus* was solved by Tiefenbrunn to high resolution, by using a lipid cubic phase approach the diffraction was increased from 2.3 to 1.8 Å and enabled an improved understanding of the mode of action of the *ba3*-oxidase system (Tiefenbrunn *et al.*, 2011). As few as 24 transporter structures have been solved using this approach, however, while a structure of the vitamin B12 transporter BtuB solved to 1.95 Å using *in meso* grown crystals was found to differ in several details from its counterpart grown using a detergent-based (*in surfo*) method (Cherezov *et al.*, 2006). The improved stability assured by the LCP approach, and increased potential to retain their native conformation, could prove particularly advantageous for molecules such as membrane transporters, which undergo conformational changes upon substrate binding.

Meanwhile in this project, crystals obtained *in surfo* from DDM-purified CzcB diffracted to 7 Å and were later identified as AcrB, a co-purificant commonly encountered with His-tagged membrane proteins expressed in *E. coli*. AcrB is a recurring problem in membrane protein crystallography for a number of reasons, foremost amongst these the fact that its transcription is up-regulated by the stress

associated with heterologous protein production in *E. coli* and the significant affinity its exhibits to Ni<sup>2+</sup> matrices typically used to purify histidine-tagged recombinant proteins (Psakis *et al.*, 2009). Consequently, a cleavable His-tag CzrB construct was generated in this work to greatly reduce the possibility of co-purification of AcrB with the target protein. Initial trials with the new construct indicated incomplete removal of the *E. coli* contaminant from purified CzrB preparations, however, leading to the conclusion that only *E. coli* *acrB*- strains should be used in future expression studies to completely avoid co-purification of AcrB.

A more radical redesign of *czrB* expression constructs was also undertaken in this project in a further attempt to obtain diffraction quality crystals. This involved the expression of a variety of protein molecules that differed in their fusion tags or number of TMDs. Two full-length CzrB constructs, with N-terminus His-tag and C-terminus Halo7 tag, were examined in order to improve homogeneity of the sample resulting once again in a mixture of monomer-dimer population. It would be interesting to examine homogeneity in CzrB constructs lacking the cytoplasmic domain. One such construct, a truncated version of CzrB containing 5 TMDs and a C-terminal GFP tag, may prove fruitful in future analyses by facilitating close monitoring of the monodispersity of the protein in a HTP set-up thanks to GFP tag. As the absence of the soluble domain could decrease the solubility of the protein, the ability to scan solubilising conditions efficiently will be important. The lack of the cytoplasmic domain could also result in a more monodisperse sample but as crystallisation in surfactant can pose significant difficulties due to limited protein-protein contacts, the use of LCP will be considered for crystallisation trials involving the 5-TMD-CzrB construct in order to provide a more stable and native-like environment for crystal formation.

In another potentially promising approach, the non-dimerising CzrB designed towards the end of this study might be expected to exhibit reduced sample heterogeneity and to lead to a breakthrough in obtaining a monodisperse sample. Even though CzrB exists naturally as a homodimer, its highest stability is most likely achieved in its dimeric state, while the membrane-bound CzrB undergoes significant conformational changes during zinc binding and release. Therefore, by disabling the

native dimeric form, it may be possible to obtain a more stable product, still structurally and functionally relevant, which is more suitable for crystallisation.

Heterogeneity similarly caused problems in crystallisation of the second target, the soluble trigger factor protein from *Psychrobacter frigidicola*, though in this case due to degradation of the protein during its extraction. While the molecular basis for the degradation remains unresolved, significant improvements in sample purity were achieved using size exclusion chromatography. However, it also failed to lead to production of protein crystals.

The most extensively studied ribosome-bound chaperone is *E. coli* trigger factor (TF), which adopts a characteristic elongated dragon-shaped structure (Ferbitz *et al.*, 2004). However to date, a number of trigger factor structures from other species have also been determined, including the proteins from *Vibrio cholera* (Ludlam *et al.*, 2004), *Deinococcus radiodurans* (Schlünzen *et al.*, 2005, Baram *et al.*, 2005) and *Thermotoga maritima* (Martinez-Hackert and Hendrickson, 2007).

Trigger factor binds to bacterial ribosomes in order to assist folding of nascent polypeptides once they emerge from the ribosome tunnel, but it is also present in the cytosol where it is understood to be involved in the assembly of protein complexes and promoting protein refolding. TF interacts with polypeptides through four hydrophobic binding sites, which show high flexibility in their architecture in order to enable attachment to diverse substrates (Deuerling *et al.*, 1999). This conformational flexibility (Hoffmann *et al.*, 2010) may have contributed to difficulties in crystallising the *P. frigidicola* trigger factor in the present work. Significant differences have been noted between the crystallographic conformations of *Thermotoga maritima* TF in its apo state and in dimeric complex form with the ribosomal protein S7, while structural data also reveal conformational changes upon *T. maritima* TF ribosomal binding (Thomas *et al.*, 2013). It is possible that the conformation of TF when bound to the ribosome or its nascent protein substrate favours stability and therefore crystallisation. In 2004, a 2.7 Å crystal structure of *E. coli* trigger factor was solved in which its ribosome-binding domain was in complex with the large ribosomal subunit from *Haloarcula marismortui* (Ferbitz *et al.*, 2004).

Conversely, TF shows remarkable three-dimensional organisation between different species, with the N-terminal and PPIase domains, which are

adjacent in the primary sequence, separated in the structure by the C-terminal domain, which is critically stabilised by the N-terminal-PPIase inter-domain linker (Hoffmann *et al.*, 2012). *TFPf* is likely to exhibit a similar macromolecular structure, which may lead to increased flexibility and impair crystallisation. In the case of *Vibrio cholera* trigger factor (VCTF), the structure was solved using a protein variant lacking 44 C-terminal amino acids (Ludlam *et al.*, 2004), which may have reduced the flexibility of the molecule and increased its propensity to crystallise. Similar truncated constructs are commonly used in X-ray crystallographic studies to reduce flexibility and increase stability of a target, such as in the case of *Thermatoga maritima* TF, where high-resolution crystal structures were achieved from N- (TF<sub>N</sub> aa<sub>1-116</sub>) and C- (TF<sub>C</sub> aa<sub>243-404</sub>) encompassing domains (Martinez-Hackert and Hendrickson, 2007). Another approach at reducing flexibility used in crystallising full-length trigger factor from *T. maritima* involved its complexation with ribosomal protein S7, which allowed to a diffraction of 3.5/4.0Å to be achieved. At this resolution, secondary structure was readily interpretable and high-resolution domains of *T. maritima* and a structure of *Bacillus stearothermophilus* S7 were used to assist in model building (Martinez-Hackert and Hendrickson, 2009).

Therefore, further efforts to advance *P. frigidicola* TF crystallisation in this project included adoption of a HTP platform to produce a series of full-length and component domain constructs. Affinity tags including SUMO, TRX, MBP and FLAG were investigated to increase the stability of the target protein or to reduce its degradation upon expression and/or extraction. Recombinant proteins are often subjected to stress response-related proteolytic degradation in their host cells, affecting the yield of recombinant protein expression (Demain and Vaishnav, 2009). Fusion of recombinant tags at the N-terminus has been found to protect some proteins from degradation (Malakhov *et al.*, 2004), while fusion with, e.g., MBP can also support translocation of a target protein to a different cellular compartment in order to reduce its exposure to the protease-rich cytosol (Nikaido, 1994).

Preliminary expression and purification analyses of various domains of trigger factor in this study indicated a particular susceptibility of the C-domain to degradation. Interestingly, the structure of the homologous *Vibrio cholera* TF was solved in a truncated version lacking 44 amino acids from its C-domain, suggesting

also that a *TFPf* construct lacking all or some of its C-terminus might exhibit increased stability. In an additional lead to achieving full-length, stable protein in this work, a construct encoding a thioredoxin tag yielded a highly pure, apparently stable product that warrants further investigation as a candidate to supply crystallisation studies.

In a parallel approach, it is proposed to produce the full-length *TFPf* by secretion to the periplasm and to the extracellular environment. In the work to date, its high-level expression in the *E. coli* cytoplasm was associated with extensive degradation of the translated product. Expression in the periplasm is likely to be associated with lower proteolytic activity and simplified purification due to the less crowded periplasmic environment (Mergulhão and Monteiro, 2007). Secretion into the culture medium, meanwhile, might further simplify purification and, importantly, reduce proteolysis by removing the need for a cell lysis. Extracellular secretion can be achieved in several ways, including fusion to a native extracellular or outer membrane carrier protein such as OpmA or YebF (Zhang *et al.*, 2006), co-expression of lysis-promoting proteins such as bacteriocin release protein (BRP), or the use of so-called “leaky”, mutated *E. coli* strains (Sommer *et al.*, 2010). In the case of trigger factor, the former approach will likely be preferred, with a cleavable fusion partner, in order to minimise exposure of the translated product to cellular proteases and maximise the chances of obtaining stable, full-length product.

Overall, the work presented in this thesis provides a solid base for the expression in *E. coli* of stable, homogenous CzrB and *TFPf* proteins and their crystallisation, leading ultimately to solving their three-dimensional structures. Elucidation of the structure of CzrB has the potential to shed light on its previously characterised beneficial effects on host *E. coli* cells during recombinant protein production, as well as advancing fundamental knowledge of CDF family transporters, particularly in thermophilic species. Gaining structural information on *P. frigidicola* TF, meanwhile, may enable elucidation of the basis of its functional differences from its *E. coli* homologue and improve our understanding of the role of this chaperone in polypeptide folding in cold-adapted species.

**Bibliography:**

- Agathos, S. N. 2010. Insect cell culture. *Manual of Industrial Microbiology and Biotechnology*, 212-222.
- Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. A. & Stevens, R. C. 2008. Microscale fluorescent thermal stability assay for membrane proteins. *Structure*, 16, 351-359.
- Arachea, B. T., Sun, Z., Potente, N., Malik, R., Isailovic, D. & Viola, R. E. 2012. Detergent selection for enhanced extraction of membrane proteins. *Protein Expression and Purification*, 86, 12-20.
- Ardao, I., Benaiges, M. D., Caminal, G. & Álvaro, G. 2006. One step purification-immobilization of fucose-1-phosphate aldolase, a class II DHAP dependent aldolase, by using metal-chelate supports. *Enzyme and Microbial Technology*, 39, 22-27.
- Arinaminpathy, Y., Khurana, E., Engelman, D. M. & Gerstein, M. B. 2009. Computational analysis of membrane proteins: the largest class of drug targets. *Drug Discovery Today*, 14, 1130-1135.
- Arnold, T. & Linke, D. 2008. The use of detergents to purify membrane proteins. In: Coligan, J. E. (ed.) *Current Protocols in Protein Science*.
- Auld, D. S. 2001. Zinc coordination sphere in biochemical zinc sites. *Biometals*, 14, 271-313.
- Baker, M. 2010. Making membrane proteins for structures: a trillion tiny tweaks. *Nature Methods*, 7, 429-434.
- Baneyx, F. 1999. Recombinant protein expression in *Escherichia coli*. *Current Opinion in Biotechnology*, 10, 411-421.
- Baram, D., Pyetan, E., Sittner, A., Auerbach-Nevo, T., Bashan, A. & Yonath, A. 2005. Structure of trigger factor binding domain in biologically homologous complex with eubacterial ribosome reveals its chaperone action. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 12017-12022.
- Bentley, W. E., Mirjalili, N., Andersen, D. C., Davis, R. H. & Kompala, D. S. 1990. Plasmid-encoded protein: the principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnology and Bioengineering*, 35, 668-681.
- Benvenuti, M. & Mangani, S. 2007. Crystallization of soluble proteins in vapor diffusion for x-ray crystallography. *Nature Protocols*, 2, 1633-1651.
- Berg, J. M. & Godwin, H. A. 1997. Lessons from zinc-binding peptides. *Annual Review of Biophysics and Biomolecular Structure*, 26, 357-371.
- Bergfors, T. 2003. Seeds to crystals. *Journal of Structural Biology*, 142, 66-76.
- Bergfors, T. M. 2009. *Protein Crystallization*, La Jolla, CA, IUL Biotechnology Series.
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. 2000. The Protein Data Bank. *Nucleic Acids Research*, 28, 235-242.
- Bernaodat, F., Frelet-Barrand, A., Pochon, N., Dementin, S., Hivin, P., Boutigny, S., Rioux, J.-B., Salvi, D., Seigneurin-Berny, D., Richaud, P., Joyard, J., Pignol,

- D., Sabaty, M., Desnos, T., Pebay-Peyroula, E., Darrouzet, E., Vernet, T. & Rolland, N. 2011. Heterologous expression of membrane proteins: choosing the appropriate host. *Public Library Of Science ONE*, 6, e29191.
- Berndt, U., Oellerer, S., Zhang, Y., Johnson, A. E. & Rospert, S. 2009. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1398-1403.
- Berrow, N. S., Alderton, D., Sainsbury, S., Nettleship, J., Assenberg, R., Rahman, N., Stuart, D. I. & Owens, R. J. 2007. A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Research*, 35, e45.
- Bilderback, D. H., Elleaume, P. & Weckert, E. 2005. Review of third and next generation synchrotron light sources. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 38, 773-797.
- Bill, R. M., Henderson, P. J. F., Iwata, S., Kunji, E. R. S., Michel, H., Neutze, R., Newstead, S., Poolman, B., Tate, C. G. & Vogel, H. 2011. Overcoming barriers to membrane protein structure determination. *Nature Biotechnology*, 29, 335-340.
- Bird, L. E. 2011. High throughput construction and small scale expression screening of multi-tag vectors in *Escherichia coli*. *Methods*, 55, 29-37.
- Böhm, G., Muhr, R. & Jaenicke, R. 1992. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein Engineering*, 5, 191-195.
- Bolanos-Garcia, V. M. & Davies, O. R. 2006. Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1760, 1304-1313.
- Bolen, D. W. 2004. Effects of naturally occurring osmolytes on protein stability and solubility: issues important in protein crystallization. *Methods*, 34, 312-322.
- Bonander, N. & Bill, R. M. 2012. Optimising yeast as a host for recombinant protein production. In: Bill, R. M. (ed.) *Recombinant Protein Production in Yeast*. New York: Humana Press.
- Bowie, J. U. 2001. Stabilizing membrane proteins. *Current Opinion in Structural Biology*, 11, 397-402.
- Bowman, J. 2006. The Genus Psychrobacter. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E. (eds.) *The Prokaryotes*. Springer, New York.
- Bowman, J. P., Cavanagh, J., Austin, J. J. & Sanderson, K. 1996. Novel psychrobacter species from antarctic ornithogenic soils. *International Journal of Systematic Bacteriology*, 46, 841-848.
- Bruinsma, J. J., Jirakulaporn, T., Muslin, A. J. & Kornfeld, K. 2002. Zinc ions and cation diffusion facilitator proteins regulate ras-mediated signaling. *Developmental Cell*, 2, 567-578.
- Brumshtein, B., Greenblatt, H. M., Futerman, A. H., Silman, I. & Sussman, J. L. 2008. Control of the rate of evaporation in protein crystallization by the 'microbatch under oil' method. *Journal of Applied Crystallography*, 41, 969-971.

- Bukrinsky, J. T. & Poulsen, J.-C. N. 2001. pH, conductivity and long-term stability in the crystal screen solutions. *Journal of Applied Crystallography*, 34, 533-534.
- Burnette, W. N. 1981. "Western Blotting": Electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. *Analytical Biochemistry*, 112, 195-203.
- Burns, J. A., Butler, J. C., Moran, J. & Whitesides, G. M. 1991. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *The Journal of Organic Chemistry*, 56, 2648-2650.
- Butt, T. R., Edavettal, S. C., Hall, J. P. & Mattern, M. R. 2005. SUMO fusion technology for difficult-to-express proteins. *Protein Expression and Purification*, 43, 1-9.
- Caffrey, M. 2015. A comprehensive review of the lipid cubic phase or in meso method for crystallizing membrane and soluble proteins and complexes. *Acta Crystallogr F Struct Biol Commun*, 71, 3-18.
- Caffrey, M. & Cherezov, V. 2009. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols*, 4, 706-731.
- Campbell, E. A., Tupy, J. L., Gruber, T. M., Wang, S., Sharp, M. M., Gross, C. A. & Darst, S. A. 2003. Crystal structure of *Escherichia coli*  $\sigma$ E with the cytoplasmic domain of its anti- $\sigma$  RseA. *Molecular Cell*, 11, 1067-1078.
- Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. 2008. Overcoming the challenges of membrane protein crystallography. *Current Opinion in Structural Biology*, 18, 581-586.
- Carrion-Vazquez, M., Oberhauser, A. F., Fisher, T. E., Marszalek, P. E., Li, H. & Fernandez, J. M. 2000. Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Progress in Biophysics and Molecular Biology*, 74, 63-91.
- Çelik, E. & Çalık, P. 2012. Production of recombinant proteins by yeast cells. *Biotechnology Advances*, 30, 1108-1118.
- Chant, A., Kraemer-Pecore, C. M., Watkin, R. & Kneale, G. G. 2005. Attachment of a histidine tag to the minimal zinc finger protein of the *Aspergillus nidulans* gene regulatory protein AreA causes a conformational change at the DNA-binding site. *Protein Expression and Purification*, 39, 152-159.
- Chao, Y. & Fu, D. 2004. Thermodynamic studies of the mechanism of metal binding to the *Escherichia coli* zinc transporter YjiP. *Journal of Biological Chemistry*, 279, 17173-17180.
- Chayen, N. E. 1998. Comparative studies of protein crystallization by vapour-diffusion and microbatch techniques. *Acta Crystallographica Section D: Biological Crystallography*, 54, 8-15.
- Chayen, N. E. 2004. Turning protein crystallisation from an art into a science. *Current Opinion in Structural Biology*, 14, 577-583.
- Chayen, N. E. 2005. Methods for separating nucleation and growth in protein crystallisation. *Progress in Biophysics and Molecular Biology*, 88, 329-337.
- Chayen, N. E. & Saridakis, E. 2008. Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods*, 5, 147-153.

- Chen, R. 2012. Bacterial expression systems for recombinant protein production: *E. coli* and beyond. *Biotechnology Advances*, 30, 1102-1107.
- Cherezov, V. 2011. Lipidic cubic phase technologies for membrane protein structural studies. *Current Opinion in Structural Biology*, 21, 559-566.
- Cherezov, V., Höfer, N., Szebenyi, D. M. E., Kolaj, O., Wall, J. G., Gillilan, R., Srinivasan, V., Jaroniec, C. P. & Caffrey, M. 2008. Insights into the mode of action of a putative zinc transporter CzcB in *Thermus thermophilus*. *Structure*, 16, 1378-1388.
- Cherezov, V., Rosenbaum, D. M., Hanson, M. A., Rasmussen, S. G., Thian, F. S., Kobilka, T. S., Choi, H.-J., Kuhn, P., Weis, W. I. & Kobilka, B. K. 2007. High-resolution crystal structure of an engineered human  $\beta$ 2-adrenergic G protein-coupled receptor. *Science*, 318, 1258-1265.
- Cherezov, V., Yamashita, E., Liu, W., Zhahnina, M., Cramer, W. A. & Caffrey, M. 2006. In Meso Structure of the Cobalamin Transporter, BtuB, at 1.95 Å Resolution. *Journal of Molecular Biology*, 364, 716-734.
- Chernov, A. A. 1997. Crystals built of biological macromolecules. *Physics Report*, 288, 61-75.
- Chiu, M. L., Nollert, P., Loewen, M. e. C., Belrhali, H., Pebay-Peyroula, E., Rosenbusch, J. P. & Landau, E. M. 2000. Crystallization in cubo: general applicability to membrane proteins. *Acta Crystallographica Section D: Biological Crystallography*, 56, 781-784.
- Conklin, D. S., Culbertson, M. R. & Kung, C. 1994. Interactions between gene products involved in divalent cation transport in *Saccharomyces cerevisiae*. *Molecular and General Genetics*, 244, 303-311.
- Coudray, N., Valvo, S., Hu, M., Lasala, R., Kim, C., Vink, M., Zhou, M., Provasi, D., Filizola, M., Tao, J., Fang, J., Penczek, P. A., Ubarretxena-Belandia, I. & Stokes, D. L. 2013. Inward-facing conformation of the zinc transporter YiiP revealed by cryoelectron microscopy. *Proceedings of the National Academy of Sciences*, 110, 2140-2145.
- Cregg, J. M., Cereghino, J. L., Shi, J. & Higgins, D. R. 2000. Recombinant protein expression in *Pichia pastoris*. *Molecular biotechnology*, 16, 23-52.
- Cronan, J. E. 2006. A family of arabinose-inducible *Escherichia coli* expression vectors having pBR322 copy control. *Plasmid*, 55, 152-157.
- D'Arcy, A., Elmore, C., Stihle, M. & Johnston, J. E. 1996. A novel approach to crystallising proteins under oil. *Journal of Crystal Growth*, 168, 175-180.
- Dalbey, R. E., Wang, P. & Kuhn, A. 2011. Assembly of bacterial inner membrane proteins. *Annual Review of Biochemistry*, 80, 161-187.
- Dąbrowski, S., Olszewski, M., Piątek, R. & Kur, J. 2002. Novel thermostable ssDNA-binding proteins from *Thermus thermophilus* and *T. aquaticus*—expression and purification. *Protein Expression and Purification*, 26, 131-138.
- De Las Peñas, A., Connolly, L. & Gross, C. A. 1997. The  $\sigma$ E-mediated response to extracytoplasmic stress in *Escherichia coli* is transduced by RseA and RseB, two negative regulators of  $\sigma$ E. *Molecular Microbiology*, 24, 373-385.
- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. 1985. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3Å resolution. *Nature*, 318, 618-624.

- DeLucas, L. 2009. *Membrane Protein Crystallization*, Burlington, MA, Academic Press.
- DeLucas, L. J., Bray, T. L., Nagy, L., McCombs, D., Chernov, N., Hamrick, D., Cosenza, L., Belgovskiy, A., Stoops, B. & Chait, A. 2003. Efficient protein crystallization. *Journal of Structural Biology*, 142, 188-206.
- Demain, A. L. & Vaishnav, P. 2009. Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, 27, 297-306.
- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A. & Bukau, B. 1999. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400, 693-696.
- Diehn, M., Bhattacharya, R., Botstein, D. & Brown, P. O. 2006. Genome-scale identification of membrane-associated human mRNAs. *Public Library of Science Genetics*, 2, e11.
- Drew, D., Klepsch, M. M., Newstead, S., Flaig, R., De Gier, J. W., Iwata, S. & Beis, K. 2008a. The structure of the efflux pump AcrB in complex with bile acid. *Molecular Membrane Biology*, 25, 677-682.
- Drew, D., Newstead, S., Sonoda, Y., Kim, H., von Heijne, G. & Iwata, S. 2008b. GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in *Saccharomyces cerevisiae*. *Nature Protocols*, 3, 784-798.
- Dumon-Seignovert, L., Cariot, G. & Vuillard, L. 2004. The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purification*, 37, 203-206.
- Dutzler, R., Campbell, E. B. & MacKinnon, R. 2003. Gating the Selectivity Filter in ClC Chloride Channels. *Science*, 300, 108-112.
- Ellis, C. D., Wang, F., MacDiarmid, C. W., Clark, S., Lyons, T. & Eide, D. J. 2004. Zinc and the Msc2 zinc transporter protein are required for endoplasmic reticulum function. *Journal of Cell Biology*, 166, 325-335.
- Ellis, R. J. & Minton, A. P. 2006. Protein aggregation in crowded environments. *Biological Chemistry*, 387, 485-497.
- Ericsson, U. B., Hallberg, B. M., DeTitta, G. T., Dekker, N. & Nordlund, P. 2006. Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Analytical Biochemistry*, 357, 289-298.
- Eshaghi, S., Hedrén, M., Nasser, M. I. A., Hammarberg, T., Thornell, A. & Nordlund, P. 2005. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Science*, 14, 676-683.
- Esposito, D. & Chatterjee, D. K. 2006. Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology*, 17, 353-358.
- Faham, S., Boulting, G. L., Massey, E. A., Yohannan, S., Yang, D. & Bowie, J. U. 2005. Crystallization of bacteriorhodopsin from bicelle formulations at room temperature. *Protein Science*, 14, 836-840.
- Faham, S. & Bowie, J. U. 2002. Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *Journal of Molecular Biology*, 316, 1-6.

- Fan, J., Heng, J., Dai, S., Shaw, N., Zhou, B., Huang, B., He, Z., Wang, Y., Jiang, T., Li, X., Liu, Z., Wang, X. & Zhang, X. C. 2011. An efficient strategy for high throughput screening of recombinant integral membrane protein expression and stability. *Protein Expression and Purification*, 78, 6-13.
- Fang, Y., Jayaram, H., Shane, T., Kolmakova-Partensky, L., Wu, F., Williams, C., Xiong, Y. & Miller, C. 2009. Structure of a prokaryotic virtual proton pump at 3.2 Å resolution. *Nature*, 460, 1040-1043.
- Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E. & Ban, N. 2004. Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. *Nature*, 431, 590-596.
- Gaberc-Porekar, V. & Menart, V. 2001. Perspectives of immobilized-metal affinity chromatography. *Journal of Biochemical and Biophysical Methods*, 49, 335-360.
- Gamerding, M. & Deuerling, E. 2014. Trigger factor flexibility. *Science*, 344, 590-591.
- Garavito, R. M. & Ferguson-Miller, S. 2001. Detergents as tools in membrane biochemistry. *Journal of Biological Chemistry*, 276, 32403-32406.
- Garman, E. F. & Owen, R. L. 2005. Cryocooling and radiation damage in macromolecular crystallography. *Acta Crystallographica Section D: Biological Crystallography*, 62, 32-47.
- Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P. & Lang, C. 2008. Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microbial Cell Factories*, 7, 11.
- Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L.-Y. 2004. A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society*, 126, 8933-8939.
- Gordeliy, V. I., Labahn, J., Moukhametzanov, R., Efremov, R., Granzin, J., Schlesinger, R., Buldt, G., Savopol, T., Scheidig, A. J., Klare, J. P. & Engelhard, M. 2002. Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. *Nature*, 419, 484-487.
- Gordon, E., Horsefield, R., Swarts, H. G. P., de Pont, J. J. H. H. M., Neutze, R. & Snijder, A. 2008. Effective high-throughput overproduction of membrane proteins in *Escherichia coli*. *Protein Expression and Purification*, 62, 1-8.
- Gräslund, S., Sagemark, J., Berglund, H., Dahlgren, L.-G., Flores, A., Hammarström, M., Johansson, I., Kotenyova, T., Nilsson, M., Nordlund, P. & Weigelt, J. 2008. The use of systematic N- and C-terminal deletions to promote production and structural studies of recombinant proteins. *Protein Expression and Purification*, 58, 210-221.
- Grisshammer, R. 2006. Understanding recombinant expression of membrane proteins. *Current Opinion in Biotechnology*, 17, 337-340.
- Guerrero, S., Hecht, H.-J., Hofmann, B., Biebl, H. & Singh, M. 2001. Production of selenomethionine-labelled proteins using simplified culture conditions and generally applicable host/vector systems. *Applied Microbiology and Biotechnology*, 56, 718-723.

- Gutmann, D. A., Mizohata, E., Newstead, S., Ferrandon, S., Henderson, P. J., Van Veen, H. W. & Byrne, B. 2007. A high-throughput method for membrane protein solubility screening: The ultracentrifugation dispersity sedimentation assay. *Protein Science*, 16, 1422-1428.
- Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. 1995. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *Journal of Bacteriology*, 177, 4121-4130.
- Hartl, F. U.-H. M. 2002. Molecular Chaperones in the Cytosol: from Nascent Chain to Folded Protein. *Science*, 295, 1852-1858.
- Hartl, F. U., Bracher, A. & Hayer-Hartl, M. 2011. Molecular chaperones in protein folding and proteostasis. *Nature*, 475, 324-332.
- Hattori, M., Tanaka, Y., Ishitani, R. & Nureki, O. 2007. Crystallization and preliminary X-ray diffraction analysis of the cytosolic domain of a cation diffusion facilitator family protein. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63, 771-773.
- Helenius, A., McCaslin, D. R., Fries, E. & Tanford, C. 1979. Properties of detergents. *Methods in enzymology*, 56, 734-749.
- Henderson, R., Baldwin, J. M., Ceska, T. A., Zemlin, F., Beckmann, E. & Downing, K. H. 1990. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of Molecular Biology*, 213, 899-929.
- Henne, A., Bruggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., Johann, A., Lienard, T., Gohl, O., Martinez-Arias, R., Jacobi, C., Starkuviene, V., Schlenczeck, S., Dencker, S., Huber, R., Klenk, H.-P., Kramer, W., Merkl, R., Gottschalk, G. & Fritz, H.-J. 2004. The genome sequence of the extreme thermophile *Thermus thermophilus*. *Nature Biotechnology*, 22, 547-553.
- Henningsen, R., Gale, B. L., Straub, K. M. & DeNagel, D. C. 2002. Application of zwitterionic detergents to the solubilization of integral membrane proteins for two-dimensional gel electrophoresis and mass spectrometry. *Proteomics*, 2, 1479-1488.
- Hibbs, R. E. & Gouaux, E. 2011. Principles of activation and permeation in an anion-selective Cys-loop receptor. *Nature*, 474, 54-60.
- Higuchi, T., Hattori, M., Tanaka, Y., Ishitani, R. & Nureki, O. 2009. Crystal structure of the cytosolic domain of the cation diffusion facilitator family protein. *Proteins: Structure, Function, and Bioinformatics*, 76, 768-771.
- Hjmel, L. M. & Chrambach, A. 1981. Electrophoresis and electrofocusing in detergent containing media: A discussion of basic concepts. *Electrophoresis*, 2, 1-11.
- Hofer, N., Kolaj, O., Li, H., Cherezov, V., Gillilan, R., Wall, J. G. & Caffrey, M. 2007. Crystallization and preliminary X-ray diffraction analysis of a soluble domain of the putative zinc transporter CzcB from *Thermus thermophilus*. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 63, 673-677.
- Hoffmann, A., Becker, A.-m., Zachmann-Brand, B., Deuerling, E., Bukau, B. & Kramer, G. 2012. Concerted action of the ribosome and the associated

- chaperone trigger factor confines nascent polypeptide folding. *Molecular Cell*, 48, 63-74.
- Hoffmann, A., Bukau, B. & Kramer, G. 2010. Structure and function of the molecular chaperone Trigger Factor. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1803, 650-661.
- Hoffmann, F. & Rinas, U. 2004. Stress induced by recombinant protein production in *Escherichia coli*. *Advances in Biochemical Engineering / Biotechnology*, 89, 73-92.
- Hsu, E., Osslund, T., Nybo, R., Chen, B.-L., Kenney, W. C., Morris, C. F., Arakawa, T. & Narhi, L. O. 2006. Enhanced stability of recombinant keratinocyte growth factor by mutagenesis. *Protein Engineering Design and Selection*, 19, 147-153.
- Hu, N.-J., Iwata, S., Cameron, A. D. & Drew, D. 2011. Crystal structure of a bacterial homologue of the bile acid sodium symporter ASBT. *Nature*, 478, 408-411.
- Huang, C. J., Lin, H. & Yang, X. 2012. Industrial production of recombinant therapeutics in *Escherichia coli* and its recent advancements. *Journal of Industrial Microbiology and Biotechnology*, 39, 383-399.
- Huang, K.-x., Badger, M., Haney, K. & Evans, S. L. 2007. Large scale production of *Bacillus thuringiensis* PS149B1 insecticidal proteins Cry34Ab1 and Cry35Ab1 from *Pseudomonas fluorescens*. *Protein Expression and Purification*, 53, 325-330.
- Huber, D., Boyd, D., Xia, Y., Olma, M. H., Gerstein, M. & Beckwith, J. 2005. Use of thioredoxin as a reporter to identify a subset of *Escherichia coli* signal sequences that promote signal recognition particle-dependent translocation. *Journal of Bacteriology*, 187, 2983-2991.
- Hunt, I. 2005. From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression and Purification*, 40, 1-22.
- Hunte, C., Koepke, J., Lange, C., Roßmanith, T. & Michel, H. 2000. Structure at 2.3 Å resolution of the cytochrome bc<sub>1</sub> complex from the yeast *Saccharomyces cerevisiae* co-crystallized with an antibody Fv fragment. *Structure*, 8, 669-684.
- Hunte, C. & Michel, H. 2002. Crystallisation of membrane proteins mediated by antibody fragments. *Current Opinion in Structural Biology*, 12, 503-508.
- Hunte, C., von Jagow, G. & Schagger, H. 2003. *Membrane protein purification and crystallization: a practical guide*, London, Academic Press.
- Ikonomou, L., Schneider, Y.-J. & Agathos, S. 2003. Insect cell culture for industrial production of recombinant proteins. *Applied Microbiology and Biotechnology*, 62, 1-20.
- Iwata, S. 2003. *Methods and results in crystallization of membrane proteins*, La Jolla, CA, IUL Biotechnology Series.
- Iwata, S., Ostermeier, C., Ludwig, B. & Michel, H. 1995. Structure at 2.8 Å resolution of cytochrome c oxidase from *Paracoccus denitrificans*. *Nature*, 376, 660-669.

- Jimbo, K., Inoue, J., Masuda, T., Shibata, T. & Mikawa, T. 2007. Purification and characterization of the *Thermus thermophilus* HB8 RecX protein. *Protein Expression and Purification*, 51, 320-323.
- Kambe, T., Narita, H., Yamaguchi-Iwai, Y., Hirose, J., Amano, T., Sugiura, N., Sasaki, R., Mori, K., Iwanaga, T. & Nagao, M. 2002. Cloning and characterization of a novel mammalian zinc transporter, zinc transporter 5, abundantly expressed in pancreatic  $\beta$  cells. *Journal of Biological Chemistry*, 277, 19049-19055.
- Kang, H. J., Lee, C. & Drew, D. 2013. Breaking the barriers in membrane protein crystallography. *The International Journal of Biochemistry & Cell Biology*, 45, 636-644.
- Kawamoto, J., Kurihara, T., Kitagawa, M., Kato, I. & Esaki, N. 2007. Proteomic studies of an Antarctic cold-adapted bacterium, *Shewanella livingstonensis* Ac10, for global identification of cold-inducible proteins. *Extremophiles*, 11, 819-826.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R., Wyckoff, H. & Phillips, D. 1958. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181, 662-666.
- Kim, J., Kim, Y.-G. & Lee, G. 2012. CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Applied Microbiology & Biotechnology*, 93, 917-930.
- Kim, Y., Babnigg, G., Jedrzejczak, R., Eschenfeldt, W. H., Li, H., Maltseva, N., Hatzos-Skintges, C., Gu, M., Makowska-Grzyska, M., Wu, R., An, H., Chhor, G. & Joachimiak, A. 2011. High-throughput protein purification and quality assessment for crystallization. *Methods*, 55, 12-28.
- Kobayashi, S., Miyabe, S., Izawa, S., Inoue, Y. & Kimura, A. 1996. Correlation of the OSR/ZRCI gene product and the intracellular glutathione levels in *Saccharomyces cerevisiae*. *Biotechnology and Applied Biochemistry*, 23, 3-6.
- Kolaj, O. 2008. *Investigation of the structure, function and biotechnological potential of CzrB from Thermus thermophilus*. PhD Thesis, University of Limerick, Ireland.
- Korczynska, J., Hu, T. C., Smith, D. K., Jenkins, J., Lewis, R., Edwards, T. & Brzozowski, A. M. 2007. Microscale vapour diffusion for protein crystallization. *Acta Crystallographica Section D: Biological Crystallography*, 63, 1009-1015.
- Kost, T. A., Condreay, J. P. & Jarvis, D. L. 2005. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nature Biotechnology*, 23, 567-575.
- Kramer, G., Rauch, T., Rist, W., Vorderwulbecke, S., Patzelt, H., Schulze-Specking, A., Ban, N., Deuerling, E. & Bukau, B. 2002. L23 protein functions as a chaperone docking site on the ribosome. *Nature*, 419, 171-174.
- Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., Green, H. F., Liu, T., Chae, P. S., Dror, R. O., Shaw, D. E., Weis, W. I., Wess, J. & Kobilka, B. K. 2012. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, 482, 552-556.

- Krzeslak, J., Braun, P., Voulhoux, R., Cool, R. H. & Quax, W. J. 2009. Heterologous production of *Escherichia coli* penicillin G acylase in *Pseudomonas aeruginosa*. *Journal of Biotechnology*, 142, 250-258.
- Kumar, S., Tsai, C. J. & Nussinov, R. 2000. Factors enhancing protein thermostability. *Protein Engineering*, 13, 179-191.
- Kunji, E. R., Slotboom, D.-J. & Poolman, B. 2003. *Lactococcus lactis* as host for overproduction of functional membrane proteins. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1610, 97-108.
- Kuroda, M., Hayashi, H. & Ohta, T. 1999. Chromosome-determined zinc-responsible operon *czr* in *Staphylococcus aureus* strain 912. *Microbiology and Immunology*, 43, 115-125.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, 227, 680-685.
- Lakshminpathy, S. K., Tomic, S., Kaiser, C. M., Chang, H.-C., Genevaux, P., Georgopoulos, C., Barral, J. M., Johnson, A. E., Hartl, F. U. & Etchells, S. A. 2007. Identification of nascent chain interaction sites on trigger factor. *Journal of Biological Chemistry*, 282, 12186-12193.
- Lama, L., Nicolaus, B., Calandrelli, V., Romano, I., Basile, R. & Gambacorta, A. 2001. Purification and characterization of thermostable xylose(glucose) isomerase from *Bacillus thermoantarcticus*. *Journal of Industrial Microbiology and Biotechnology*, 27, 234-240.
- Landau, E. M. & Rosenbusch, J. P. 1996. Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 14532-14535.
- Lau, F. W. & Bowie, J. U. 1997. A method for assessing the stability of a membrane protein. *Biochemistry*, 36, 5884-5892.
- le Maire, M., Champeil, P. & Møller, J. V. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1508, 86-111.
- Lewis, L. & Caplan, P. E. 1950. The shoe-fitting fluoroscope as a radiation hazard. *California medicine*, 72, 26.
- Li, D., Lee, J. & Caffrey, M. 2011. Crystallizing membrane proteins in lipidic mesophases. A host lipid screen. *Crystal Growth & Design*, 11, 530-537.
- Li, F., Shen, A. & Amanullah, A. 2010. Cell culture processes in monoclonal antibody production. *mAbs*, 2, 466-479.
- Li, L. & Kaplan, J. 2001. The yeast gene *MSC2*, a member of the cation diffusion facilitator family, affects the cellular distribution of zinc. *Journal of Biological Chemistry*, 276, 5036-5043.
- Lindemann, L. & Hoener, M. C. 2005. A renaissance in trace amines inspired by a novel GPCR family. *Trends in Pharmacological Sciences*, 26, 274-281.
- Liu, M., Wang, X., Yin, C., Zhang, Z., Lin, Q., Zhen, Y. & Huang, H. 2006. One-step on-column purification and refolding of a single-chain variable fragment (scFv) antibody against tumour necrosis factor  $\alpha$ . *Biotechnology and Applied Biochemistry*, 43, 137-145.
- Liu, W., Chun, E., Thompson, A. A., Chubukov, P., Xu, F., Katritch, V., Han, G. W., Roth, C. B., Heitman, L. H., Ijzerman, A. P., Cherezov, V. & Stevens, R. C.

2012. Structural Basis for Allosteric Regulation of GPCRs by Sodium Ions(). *Science (New York, N.y.)*, 337, 232-236.
- Lizak, C., Fan, Y. Y., Weber, T. C. & Aebi, M. 2011. N-linked glycosylation of antibody fragments in *Escherichia coli*. *Bioconjugate Chemistry*, 22, 488-496.
- Lluis, M. W., Godfroy, J. I. & Yin, H. 2012. Protein engineering methods applied to membrane protein targets. *Protein Engineering Design and Selection*.
- Lu, M., Chai, J. & Fu, D. 2009. Structural basis for autoregulation of the zinc transporter YiiP. *Nat Struct Mol Biol*, 16, 1063-1067.
- Lu, M. & Fu, D. 2007. Structure of the zinc transporter YiiP. *Science*, 317, 1746-1748.
- Luche, S., Santoni, V. & Rabilloud, T. 2003. Evaluation of nonionic and zwitterionic detergents as membrane protein solubilizers in two-dimensional electrophoresis. *Proteomics*, 3, 249-253.
- Luckow, V. A., Lee, S., Barry, G. & Olins, P. 1993. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *Journal of Virology*, 67, 4566-4579.
- Ludlam, A. V., Moore, B. A. & Xu, Z. 2004. The crystal structure of ribosomal chaperone trigger factor from *Vibrio cholerae*. *Proceedings of the National Academy of Sciences of the United States of America*, 101, 13436-13441.
- Luecke, H., Schobert, B., Richter, H.-T., Cartailier, J.-P. & Lanyi, J. K. 1999. Structure of bacteriorhodopsin at 1.55 Å resolution. *Journal of Molecular Biology*, 291, 899-911.
- Luft, J. R., Newman, J. & Snell, E. H. 2014. Crystallization screening: the influence of history on current practice. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 70, 835-853.
- Luft, J. R., Snell, E. H. & DeTitta, G. T. 2011. Lessons from high-throughput protein crystallization screening: 10 years of practical experience. *Expert Opinion on Drug Discovery*, 6, 465-480.
- Luirink, J. & Sinning, I. 2004. SRP-mediated protein targeting: structure and function revisited. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1694, 17-35.
- Maier, T., Ferbitz, L., Deuerling, E. & Ban, N. 2005. A cradle for new proteins: trigger factor at the ribosome. *Current Opinion in Structural Biology*, 15, 204-212.
- Malakhov, M., Mattern, M., Malakhova, O., Drinker, M., Weeks, S. & Butt, T. 2004. SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *Journal of Structural and Functional Genomics*, 5, 75-86.
- Martinez-Hackert, E. & Hendrickson, W. A. 2007. Structures of and interactions between domains of trigger factor from *Thermotoga maritima*. *Acta Crystallographica Section D: Biological Crystallography*, 63, 536-547.
- Martinez-Hackert, E. & Hendrickson, W. A. 2009. Promiscuous Substrate Recognition in Folding and Assembly Activities of the Trigger Factor Chaperone. *Cell*, 138, 923-934.

- Martinez, A., Knappskog, P. M., Olafsdottir, S., Døskeland, A. P., Eiken, H. G., Svebak, R. M., Bozzini, M., Apold, J. & Flatmark, T. 1995. Expression of recombinant human phenylalanine hydroxylase as fusion protein in *Escherichia coli* circumvents proteolytic degradation by host cell proteases. Isolation and characterization of the wild-type enzyme. *Biochemical Journal*, 306, 589-597.
- Mashaghi, A., Kramer, G., Bechtluft, P., Zachmann-Brand, B., Driessen, A. J. M., Bukau, B. & Tans, S. J. 2013. Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature*, 500, 98-101.
- McPherson, A. 2004. Introduction to protein crystallization. *Methods*, 34, 254-265.
- McPherson, A. & Kuznetsov, Y. G. 2014. Mechanisms, kinetics, impurities and defects: consequences in macromolecular crystallization. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 70, 384-403.
- McPherson, A., Malkin, A. J., Kuznetsov, Y. G. & Koszelak, S. 1996. Incorporation of impurities into macromolecular crystals. *Journal of Crystal Growth*, 168, 74-92.
- Médigue, C., Krin, E., Pascal, G., Barbe, V., Bernsel, A., Bertin, P. N., Cheung, F., Cruveiller, S., D'Amico, S., Duilio, A., Fang, G., Feller, G., Ho, C., Mangenot, S., Marino, G., Nilsson, J., Parrilli, E., Rocha, E. P. C., Rouy, Z., Sekowska, A., Tutino, M. L., Vallenet, D., von Heijne, G. & Danchin, A. 2005. Coping with cold: The genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. *Genome Research*, 15, 1325-1335.
- Mergulhão, F. & Monteiro, G. 2007. Periplasmic Targeting of Recombinant Proteins in *Escherichia coli*. In: van der Giezen, M. (ed.) *Protein Targeting Protocols*. Humana Press.
- Merz, F., Boehringer, D., Schaffitzel, C., Preissler, S., Hoffmann, A., Maier, T., Rutkowska, A., Lozza, J., Ban, N., Bukau, B. & Deuerling, E. 2008. Molecular mechanism and structure of Trigger Factor bound to the translating ribosome. *The EMBO Journal*, 27, 1622-1632.
- Mierau, I., Leij, P., van Swam, I., Blommestein, B., Floris, E., Mond, J. & Smid, E. J. 2005. Industrial-scale production and purification of a heterologous protein in *Lactococcus lactis* using the nisin-controlled gene expression system NICE: the case of lysostaphin. *Microbial Cell Factories*, 4, 15.
- Miroux, B. & Walker, J. E. 1996. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *Journal of Molecular Biology*, 260, 289-298.
- Mohan, C., Kim, Y. G., Koo, J. & Lee, G. M. 2008. Assessment of cell engineering strategies for improved therapeutic protein production in CHO cells. *Biotechnology Journal*, 3, 624-630.
- Monné, M., Chan, K. W., Slotboom, D. J. & Kunji, E. R. 2005. Functional expression of eukaryotic membrane proteins in *Lactococcus lactis*. *Protein Science*, 14, 3048-3056.
- Montanini, B., Blaudez, D., Jeandroz, S., Sanders, D. & Chalot, M. 2007. Phylogenetic and functional analysis of the cation diffusion facilitator (CDF)

- family: improved signature and prediction of substrate specificity. *BMC Genomics*, 8, 107.
- Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. 2014. Membrane protein structure determination — the next generation. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1838, 78-87.
- Mould, R. F. 1993. *A century of X-rays and radioactivity in medicine: with emphasis on photographic records of the early years*, Boca Raton, FL, CRC Press.
- Narayanan, A., Ridilla, M. & Yernool, D. A. 2011. Restrained expression, a method to overproduce toxic membrane proteins by exploiting operator–repressor interactions. *Protein Science*, 20, 51-61.
- Newby, Z. E. R., O'Connell, J. D., Gruswitz, F., Hays, F. A., Harries, W. E. C., Harwood, I. M., Ho, J. D., Lee, J. K., Savage, D. F., Miercke, L. J. W. & Stroud, R. M. 2009. A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nature Protocols*, 4, 619-637.
- Newstead, S., Drew, D., Cameron, A. D., Postis, V. L. G., Xia, X., Fowler, P. W., Ingram, J. C., Carpenter, E. P., Sansom, M. S. P., McPherson, M. J., Baldwin, S. A. & Iwata, S. 2011. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide–proton symporters, PepT1 and PepT2. *The EMBO Journal*, 30, 417-426.
- Nies, D. H., Nies, A., Chu, L. & Silver, S. 1989. Expression and nucleotide sequence of a plasmid-determined divalent cation efflux system from *Alcaligenes eutrophus*. *Proceedings of the National Academy of Sciences of the United States of America*, 86, 7351-7355.
- Nikaido, H. 1994. Maltose transport system of *Escherichia coli*: An ABC-type transporter. *FEBS Letters*, 346, 55-58.
- Nollert, P., Qiu, H., Caffrey, M., Rosenbusch, J. P. & Landau, E. M. 2001. Molecular mechanism for the crystallization of bacteriorhodopsin in lipidic cubic phases. *FEBS Letters*, 504, 179-186.
- O'Brien, E. P., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. 2012. Trigger factor slows co-translational folding through kinetic trapping while sterically protecting the nascent chain from aberrant cytosolic interactions. *Journal of the American Chemical Society*, 134, 10920-10932.
- Oberai, A., Joh, N. H., Pettit, F. K. & Bowie, J. U. 2009. Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 17747-17750.
- Oliver, D. C. & Paetzel, M. 2008. Crystal structure of the major periplasmic domain of the bacterial membrane protein assembly facilitator YidC. *Journal of Biological Chemistry*, 283, 5208-5216.
- Oshima, T. & Imahori, K. 1974. Description of *Thermus thermophilus* (Yoshida and Oshima) comb. nov., a nonsporulating thermophilic bacterium from a Japanese thermal spa. *International Journal of Systematic Bacteriology*, 24, 102-112.
- Outten, C. E. O. H. T. V. 2001. Femtomolar sensitivity of metalloregulatory proteins controlling zinc homeostasis. *Science*, 292, 2488-2492.

- Palm, W. & Colman, P. 1974. Preliminary X-ray data from well-ordered crystals of a human immunoglobulin G molecule. *Journal of Molecular Biology*, 82, 587-594.
- Palmiter, R. D. & Findley, S. D. 1995. Cloning and functional characterization of a mammalian zinc transporter that confers resistance to zinc. *The EMBO Journal*, 14, 639.
- Pan, S. H. & Malcolm, B. A. 2000. Reduced background expression and improved plasmid stability with pET vectors in BL21 (DE3). *BioTechniques*, 29, 1234-1238.
- Park, E. & Rapoport, T. A. 2012. Mechanisms of Sec61/SecY-mediated protein translocation across membranes. *Annual Review of Biophysics*, 41, 21-40.
- Paulsen, I. T. & Saier, M. H., Jr. 1997. A novel family of ubiquitous heavy metal ion transport proteins. *The Journal of Membrane Biology*, 156, 99-103.
- Peng, Q. & Li, H. 2009. Domain insertion effectively regulates the mechanical unfolding hierarchy of elastomeric proteins: toward engineering multifunctional elastomeric proteins. *Journal of the American Chemical Society*, 131, 14050-14056.
- Perron-Savard, P., De Crescenzo, G. & Moual, H. L. 2005. Dimerization and DNA binding of the *Salmonella enterica* PhoP response regulator are phosphorylation independent. *Microbiology*, 151, 3979-3987.
- Peternel, S. & Komel, R. 2010. Isolation of biologically active nanomaterial (inclusion bodies) from bacterial cells. *Microbial Cell Factories*, 9, 66.
- Petranovic, D. & Nielsen, J. 2008. Can yeast systems biology contribute to the understanding of human disease? *Trends in Biotechnology*, 26, 584-590.
- Piette, F., D'Amico, S., Struvay, C., Mazzucchelli, G., Renaut, J., Tutino, M. L., Danchin, A., Leprince, P. & Feller, G. 2010. Proteomics of life at low temperatures: trigger factor is the primary chaperone in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. *Molecular Microbiology*, 76, 120-132.
- Piette, F., Struvay, C. & Feller, G. 2011. The protein folding challenge in psychrophiles: facts and current issues. *Environmental Microbiology*, 13, 1924-1933.
- Porath, J. 1992. Immobilized metal ion affinity chromatography. *Protein Expression and Purification*, 3, 263-281.
- Porath, J., Carlsson, J., Olsson, I. & Belfrage, G. 1975. Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature*, 258, 598-599.
- Privé, G. G. 2007. Detergents for the stabilization and crystallization of membrane proteins. *Methods*, 41, 388-397.
- Psakis, G., Polaczek, J. & Essen, L.-O. 2009. AcrB et al.: Obstinate contaminants in a picogram scale. One more bottleneck in the membrane protein structure pipeline. *Journal of Structural Biology*, 166, 107-111.
- Raschle, T., Hiller, S., Yu, T.-Y., Rice, A. J., Walz, T. & Wagner, G. 2009. Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *Journal of the American Chemical Society*, 131, 17777-17779.

- Rasmussen, S. G. F., Choi, H.-J., Fung, J. J., Pardon, E., Casarosa, P., Chae, P. S., DeVree, B. T., Rosenbaum, D. M., Thian, F. S., Kobilka, T. S., Schnapp, A., Konetzki, I., Sunahara, R. K., Gellman, S. H., Pautsch, A., Steyaert, J., Weis, W. I. & Kobilka, B. K. 2011a. Structure of a nanobody-stabilized active state of the  $\beta(2)$  adrenoceptor. *Nature*, 469, 175-180.
- Rasmussen, S. G. F., Choi, H. J., Rosenbaum, D. M., Kobilka, T. S., Thian, F. S., Edwards, P. C., Burghammer, M., Ratnala, V. R. P., Sanishvili, R., Fischetti, R. F., Schertler, G. F. X., Weis, W. I. & Kobilka, B. K. 2007. Crystal structure of the human  $\beta_2$  adrenergic G-protein-coupled receptor. *Nature*, 450, 383-387.
- Rasmussen, S. G. F., DeVree, B. T., Zou, Y., Kruse, A. C., Chung, K. Y., Kobilka, T. S., Thian, F. S., Chae, P. S., Pardon, E., Calinski, D., Mathiesen, J. M., Shah, S. T. A., Lyons, J. A., Caffrey, M., Gellman, S. H., Steyaert, J., Skiniotis, G., Weis, W. I., Sunahara, R. K. & Kobilka, B. K. 2011b. Crystal Structure of the  $\beta(2)$ Adrenergic Receptor-Gs protein complex. *Nature*, 477, 549-555.
- Ravaud, S., Stjepanovic, G., Wild, K. & Sinning, I. 2008. The crystal structure of the periplasmic domain of the *Escherichia coli* membrane protein insertase YidC contains a substrate binding cleft. *Journal of Biological Chemistry*, 283, 9350-9358.
- Rengachari, S., Bezerra, G. A., Riegler-Berket, L., Gruber, C. C., Sturm, C., Taschler, U., Boeszoermenyi, A., Dreveny, I., Zimmermann, R., Gruber, K. & Oberer, M. 2012. The structure of monoacylglycerol lipase from *Bacillus sp.* H257 reveals unexpected conservation of the cap architecture between bacterial and human enzymes. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1821, 1012-1021.
- Rhodes, G. 2010. *Crystallography made crystal clear: a guide for users of macromolecular models*, London, Academic Press.
- Robin, S., Togashi, D. M., Ryder, A. G. & Wall, J. G. 2009. Trigger factor from the psychrophilic bacterium *Psychrobacter frigidicola* is a monomeric chaperone. *Journal of Bacteriology*, 191, 1162-1168.
- Rosano, G. L. & Ceccarelli, E. A. 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. *Frontiers in Microbiology*, 5, 172-206.
- Russell, D., Kolaj-Robin, O. & Soulimane, T. 2012. Maricaulis maris cation diffusion facilitator: Achieving homogeneity through a mixed-micelle approach. *Protein Expression and Purification*, 85, 173-180.
- Russell, D. & Soulimane, T. 2012. Evidence for zinc and cadmium binding in a CDF transporter lacking the cytoplasmic domain. *FEBS Letters*, 586, 4332-4338.
- Sahdev, S., Khattar, S. & Saini, K. 2008. Production of active eukaryotic proteins through bacterial expression systems: a review of the existing biotechnology strategies. *Molecular and Cellular Biochemistry*, 307, 249-264.
- Saio, T., Guan, X., Rossi, P., Economou, A. & Kalodimos, C. G. 2014. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science*, 344, 597-611.
- Sallantin, M., Huet, J.-C., Demartean, C. & Pernollet, J.-C. 1990. Reassessment of commercially available molecular weight standards for peptide sodium

- dodecyl sulfatepolyacrylamide gel electrophoresis using electroblotting and microsequencing. *Electrophoresis*, 11, 34-36.
- Sambrook, J., Russell, D. W. & Russell, D. W. 2001. *Molecular cloning: a laboratory manual*, Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.
- Samuelson, J. C., Chen, M., Jiang, F., Moller, I., Wiedmann, M., Kuhn, A., Phillips, G. J. & Dalbey, R. E. 2000. YidC mediates membrane protein insertion in bacteria. *Nature*, 406, 637-641.
- Savitzky, A. & Golay, M. J. E. 1964. Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627-1639.
- Schlünzen, F., Wilson, D. N., Tian, P., Harms, J. M., McInnes, S. J., Hansen, H. A., Albrecht, R., Buerger, J., Wilbanks, S. M. & Fucini, P. 2005. The binding mode of the trigger factor on the ribosome: implications for protein folding and SRP interaction. *Structure*, 13, 1685-1694.
- Seddon, A. M., Curnow, P. & Booth, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1666, 105-117.
- Selkirk, C. 2004. Ion-Exchange Chromatography. In: Cutler, P. (ed.) *Protein Purification Protocols*. New York: Humana Press
- Seneque, O., Bonnet, E., Joumas, F. L. & Latour, J. M. 2009. Cooperative metal binding and helical folding in model peptides of treble-clef zinc fingers. *Chemistry*, 15, 4798-4810.
- Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. 2008. Conformational thermostabilization of the  $\beta$ 1-adrenergic receptor in a detergent-resistant form. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 877-882.
- Shaffer, P. L., Goehring, A., Shankaranarayanan, A. & Gouaux, E. 2009. Structure and Mechanism of a Na<sup>+</sup>-Independent Amino Acid Transporter. *Science*, 325, 1010-1014.
- Shimamura, T., Weyand, S., Beckstein, O., Rutherford, N. G., Hadden, J. M., Sharples, D., Sansom, M. S. P., Iwata, S., Henderson, P. J. F. & Cameron, A. D. 2010. Molecular basis of alternating access membrane transport by the sodium-hydantoin transporter Mhp1. *Science*, 328, 470-473.
- Snell, G. 2012. X-Ray Sources and High-Throughput Data Collection Methods. *Methods in Molecular Biology*, 93-141.
- Sommer, B., Friehs, K. & Flaschel, E. 2010. Efficient production of extracellular proteins with *Escherichia coli* by means of optimized coexpression of bacteriocin release proteins. *Journal of Biotechnology*, 145, 350-358.
- Sørensen, H. P. & Mortensen, K. K. 2005a. Advanced genetic strategies for recombinant protein expression in *Escherichia coli*. *Journal of Biotechnology*, 115, 113-128.
- Sørensen, H. P. & Mortensen, K. K. 2005b. Soluble expression of recombinant proteins in the cytoplasm of *Escherichia coli*. *Microbial Cell Factories*, 4, 1.
- Sørensen, H. P., Sperling-Petersen, H. U. & Mortensen, K. K. 2003. Production of recombinant thermostable proteins expressed in *Escherichia coli*: completion of protein synthesis is the bottleneck. *Journal of Chromatography B*, 786, 207-214.

- Spada, S., Pembroke, T. & Wall, G. 2002. Isolation of a novel *Thermus thermophilus* metal efflux protein that improves *Escherichia coli* growth under stress conditions. *Extremophiles*, 6, 301-308.
- Stahlberg, H., Fotiadis, D., Scheuring, S., Rémiqy, H., Braun, T., Mitsuoka, K., Fujiyoshi, Y. & Engel, A. 2001. Two-dimensional crystals: a powerful approach to assess structure, function and dynamics of membrane proteins. *FEBS Letters*, 504, 166-172.
- Stathopoulos, P. B., Scholz, G. A., Hwang, Y.-M., Rumfeldt, J. A. O., Lepock, J. R. & Meiering, E. M. 2004. Sonication of proteins causes formation of aggregates that resemble amyloid. *Protein Science*, 13, 3017-3027.
- Steyaert, J. & Kobilka, B. K. 2011. Nanobody stabilization of G protein-coupled receptor conformational states. *Current Opinion in Structural Biology*, 21, 567-572.
- Stock, D., Perisic, O. & Löwe, J. 2005. Robotic nanolitre protein crystallisation at the MRC Laboratory of Molecular Biology. *Progress in Biophysics and Molecular Biology*, 88, 311-327.
- Stroud, R. M. 2011. New tools in membrane protein determination. *F1000 Biology Reports*, 3, 8.
- Studier, F. W. 1991. Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. *Journal of Molecular Biology*, 219, 37-44.
- Studier, F. W. 2005. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification*, 41, 207-234.
- Studier, F. W. 2014. Stable expression clones and auto-induction for protein production in *E. coli*. In: Chen, Y. W. (ed.) *Structural Genomics: General Applications, Methods in Molecular Biology*. New York: Humana Press.
- Subramaniam, S. & Henderson, R. 2000. Molecular mechanism of vectorial proton translocation by bacteriorhodopsin. *Nature*, 406, 653-657.
- Swartz, J. R. 2001. Advances in *Escherichia coli* production of therapeutic proteins. *Current Opinion in Biotechnology*, 12, 195-201.
- Tate, C. 2010. Practical Considerations of Membrane Protein Instability during Purification and Crystallisation. In: Mus-Veteau, I. (ed.) *Heterologous Expression of Membrane Proteins*. New York: Humana Press.
- Tate, C. G. & Schertler, G. F. X. 2009. Engineering G protein-coupled receptors to facilitate their structure determination. *Current Opinion in Structural Biology*, 19, 386-395.
- Taticek, R. & Shuler, M. 1997. Effect of elevated oxygen and glutamine levels on foreign protein production at high cell densities using the insect cell-baculovirus expression system. *Biotechnology and Bioengineering*, 54, 142-152.
- Terpe, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 60, 523-533.
- Terpe, K. 2006. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*, 72, 211-222.

- Thomas, Andrew S., Mao, S. & Elcock, Adrian H. 2013. Flexibility of the Bacterial Chaperone Trigger Factor in Microsecond-Timescale Molecular Dynamics Simulations. *Biophysical Journal*, 105, 732-744.
- Thomas, D. H., Rob, A. & Rice, D. W. 1989. A novel dialysis procedure for the crystallization of proteins. *Protein Engineering*, 2, 489-491.
- Tiefenbrunn, T., Liu, W., Chen, Y., Katritch, V., Stout, C. D., Fee, J. A. & Cherezov, V. 2011. High Resolution Structure of the ba3 Cytochrome c Oxidase from *Thermus thermophilus* in a Lipidic Environment. *PLOS ONE*, 6, e22348.
- Trivedi, M. V., Laurence, J. S. & Siahhan, T. J. 2009. The role of thiols and disulfides in protein chemical and physical stability. *Current Protein & Peptide Science*, 10, 614.
- Tulumello, D. V. & Deber, C. M. 2012. Efficiency of detergents at maintaining membrane protein structures in their biologically relevant forms. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818, 1351-1358.
- Ujwal, R. & Bowie, J. U. 2011. Crystallizing membrane proteins using lipidic bicelles. *Methods*, 55, 337-341.
- Ulbrandt, N. D., Newitt, J. A. & Bernstein, H. D. 1997. The *E. coli* signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell*, 88, 187-196.
- Unger, T. & Peleg, Y. 2012. Recombinant protein expression in the baculovirus-infected insect cell system. *Methods in Molecular Biology*, 800, 187-199.
- Urbani, A. & Warne, T. 2005. A colorimetric determination for glycosidic and bile salt-based detergents: applications in membrane protein research. *Analytical Biochemistry*, 336, 117-124.
- van der Lelie, D., Schwuchow, T., Schwidetzky, U., Wuertz, S., Baeyens, W., Mergeay, M. & Nies, D. H. 1997. Two-component regulatory system involved in transcriptional control of heavy-metal homeostasis in *Alcaligenes eutrophus*. *Molecular Microbiology*, 23, 493-503.
- Vedadi, M., Niesen, F. H., Allali-Hassani, A., Fedorov, O. Y., Finerty, P. J., Wasney, G. A., Yeung, R., Arrowsmith, C., Ball, L. J., Berglund, H., Hui, R., Marsden, B. D., Nordlund, P., Sundstrom, M., Weigelt, J. & Edwards, A. M. 2006. Chemical screening methods to identify ligands that promote protein stability, protein crystallization, and structure determination. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 15835-15840.
- Veesler, D., Blangy, S., Cambillau, C. & Sciara, G. 2008. There is a baby in the bath water: AcrB contamination is a major problem in membrane-protein crystallization. *Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, 64, 880-885.
- Vergis, J. M. & Wiener, M. C. 2011. The variable detergent sensitivity of proteases that are utilized for recombinant protein affinity tag removal. *Protein Expression and Purification*, 78, 139-142.
- Vinothkumar, K., Edwards, P. & Standfuss, J. 2013. Practical Aspects in Expression and Purification of Membrane Proteins for Structural Analysis. In: Schmidt-Krey, I. & Cheng, Y. (eds.) *Electron Crystallography of Soluble and Membrane Proteins*. New York: Humana Press.

- Vrancken, K. & Anné, J. 2009. Secretory production of recombinant proteins by *Streptomyces*. *Future Microbiology*, 4, 181-188.
- Wacker, M., Linton Dennis, Hitchen Paul C., Nita-Lazar, M., Haslam Stuart M., North Simon J., Panico Maria, Morris Howard R., Dell Anne, W., W. B. & Markus, A. 2002. N-Linked glycosylation in *Campylobacter jejuni* and its functional transfer into *E. coli*. *Science*, 298, 1790-1793.
- Wagner, S., Klepsch, M. M., Schlegel, S., Appel, A., Draheim, R., Tarry, M., Högbom, M., van Wijk, K. J., Slotboom, D. J., Persson, J. O. & de Gier, J.-W. 2008. Tuning *Escherichia coli* for membrane protein overexpression. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 14371-14376.
- Walden, H. 2010. Selenium incorporation using recombinant techniques. *Acta Crystallographica Section D: Biological Crystallography*, 66, 352-357.
- Wang, D.-N., Safferling, M., Lemieux, M. J., Griffith, H., Chen, Y. & Li, X.-D. 2003. Practical aspects of overexpressing bacterial secondary membrane transporters for structural studies. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1610, 23-36.
- Wei, Y., Li, H. & Fu, D. 2004. Oligomeric State of the *Escherichia coli* Metal Transporter YiiP. *Journal of Biological Chemistry*, 279, 39251-39259.
- White, S. 1998-2014. *Membrane Proteins of Known Structure* [Online]. Available: <http://blanco.biomol.uci.edu/mpstruc> [Accessed 21 September 2014].
- Whitmore, L. & Wallace, B. A. 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers*, 89, 392-400.
- Wiener, M. C. 2004. A pedestrian guide to membrane protein crystallization. *Methods*, 34, 364-372.
- Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G. W., Vardy, E., Liu, W., Thompson, A. A., Huang, X.-P., Carroll, F. I., Mascarella, S. W., Westkaemper, R. B., Mosier, P. D., Roth, B. L., Cherezov, V. & Stevens, R. C. 2012. Structure of the human [kgr]-opioid receptor in complex with JDTC. *Nature*, 485, 327-332.
- Wurm, F. M. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology*, 22, 1393-1398.
- Xiong, A. & Jayaswal, R. K. 1998. Molecular characterization of a chromosomal determinant conferring resistance to zinc and cobalt ions in *Staphylococcus aureus*. *Journal of Bacteriology*, 180, 4024-4029.
- Yang, X., Huang, J., Jiang, Y. & Zhang, H.-S. 2009. Cloning and functional identification of two members of the ZIP (Zrt, Irt-like protein) gene family in rice (*Oryza sativa* L.). *Molecular Biology Reports*, 36, 281-287.
- Yeh, C.-M., Kao, B.-Y. & Peng, H.-J. 2009. Production of a recombinant type 1 antifreeze protein analogue by *L. lactis* and its applications on frozen meat and frozen dough. *Journal of Agricultural and Food Chemistry*, 57, 6216-6223.
- Zgurskaya, H. I. & Nikaido, H. 1999. Bypassing the periplasm: reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 7190-7195.

- Zhang, G., Brokx, S. & Weiner, J. H. 2006. Extracellular accumulation of recombinant proteins fused to the carrier protein YebF in *Escherichia coli*. *Nature Biotechnology*, 24, 100-104.
- Zhao, K. Q., Hurst, R., Slater, M. R. & Bulleit, R. F. 2007. Functional protein expression from a DNA based wheat germ cell-free system. *Journal of Structural and Functional Genomics*, 8, 199-208.
- Zheng, S., Ponder, M. A., Shih, J. Y. J., Tiedje, J. M., Thomashow, M. F. & Lubman, D. M. 2007. A proteomic analysis of *Psychrobacter articus* 273-4 adaptation to low temperature and salinity using a 2-D liquid mapping approach. *Electrophoresis*, 28, 467-488.
- Zhou, Y., Morais-Cabral, J. H., Kaufman, A. & MacKinnon, R. 2001. Chemistry of ion coordination and hydration revealed by a K<sup>+</sup> channel-Fab complex at 2.0 Å resolution. *Nature*, 414, 43-48.

## APPENDIX

### pOPIN Suite Vectors

| Vector         | Tag                                |
|----------------|------------------------------------|
| pOPINE         | <b>POI-C-HIS<sub>6</sub></b>       |
| pOPINF         | N-HIS-3C- <b>POI</b>               |
| pOPINM         | N-HIS-MBP-3C- <b>POI</b>           |
| pOPINS3C       | N-HIS-SUMO- <b>POI</b>             |
| pOPINS         | N-HIS-SUMO-3C- <b>POI</b>          |
| pOPINTRX       | N-HIS-TRX-3C- <b>POI</b>           |
| pOPIN-3C-eGFP  | <b>POI-3C-eGFP-C-HIS</b>           |
| pOPIN-3C-HALO7 | <b>POI-3C-HALO7</b>                |
| pOPINEHALO7    | N-HIS-HALO7-3C- <b>POI</b>         |
| pOPINJ         | N-HIS-GST-3C- <b>POI</b>           |
| pOPINneoFLAG   | <b>POI-3C-FLAG-HIS<sub>8</sub></b> |
| pOPINMSYB      | N-HIS-MSYB-3C- <b>POI</b>          |
| pOPINW         | <b>POI-3C-FLAG-HIS<sub>6</sub></b> |
| pOPINFS        | N-HIS-3C- <b>POI-C-STREPII</b>     |

**POI** = protein of interest, **N-HIS** = N-terminal His6 tag, **C-His** = C-terminal His6 tag, **C-His8** = C-terminal His8 tag, **3C** = Rhinovirus 3C protease site, **GST** = Glutathione-S-transferase, **MBP** = Maltose binding Protein, **STREPII** = strepavidin binding peptide II, **SUMO** = small ubiquitin like modifier, **MSYB** = *E. coli* MsyB, **TRX** = *E. coli* Thioredoxin reductase, **HALO** = Halo7 tag, **eGFP** = enhanced Green Fluorescence Protein, **FLAG** = FLAG octapeptide.

### Primers for construction of pET28b+ vector expressing non-dimerising CzrB

Restriction sites are shown in italics, stop codons are shown in bold, the hexahistidine tag is underlined and point mutations are highlighted in grey.

---

**czrBNcoIF** 5'-ATAATAATACCATGGCTATGGCCGAAGGCGCCGCCC-3'

**czrBGluThrToAlaR1** 5'-CTCCTCCACGGGCGCGTCCGCCCCGCACCACGAG-3'

**czrBGluThrToAlaF2** 5'-CTCGTGGTGCGGGCGGACGCGCCCCGTGGAGGAG-3'

**czrBHisHindIIIR** 5'-TATAAGCTTTCATCAATGATGATGATGATGATGGGG  
TTTGTCCGCTTCC-3'