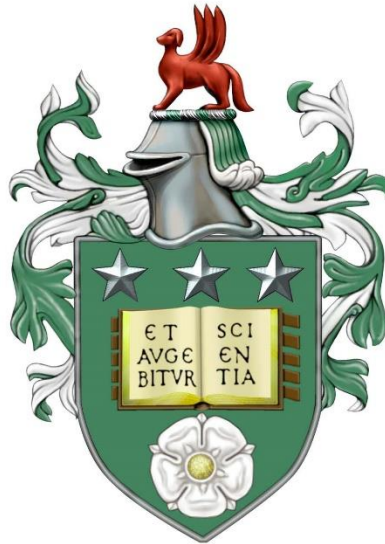


Production and Characterization of ArAE Family Members including Putative Efflux Transporters (PETs) from Bacteria and Aluminium Activated Malate Transporters (ALMTs) from Plants



Antony James Palmer

Submitted in accordance with the requirements for the degree of Doctor of

Philosophy

The University of Leeds

School of Biomedical Sciences

August 2016

The candidate confirms that the work submitted is his/her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

© 2016 The University of Leeds and Antony James Palmer

The right of Antony James Palmer to be identified as Author of this work has been asserted by him in accordance with the Copyright, Designs, and Patents Act 1988

Acknowledgements

This thesis is dedicated to the memory of Steve Baldwin. Steve was the reason I came to Leeds and my inspiration to complete this thesis. I am sure that if he had still been able to participate in the preparation of this thesis, it would be much improved for his help. I hope he would have found it worthy. I will always remember his kindness and support, and if I can have even half of his relentless enthusiasm and positivity in life I will be doing well.

I wish to thank those whose words and support meant so much, and those who have shared stories and memories of Steve and all those who attended the conference organised in Leeds. It was heartening to hear that this is not the end for his legacy. In particular I would like to thank Sheena Radford for encouraging words and Peter Henderson for being a calming influence as well as a repository of knowledge.

And now to the more conventional acknowledgements. I wish to thank my supervisors for their guidance, clarity of thought, and constant patient support. Their constructive criticism, comments, and feedback, provided me with direction when I found myself lacking. The BBSRC should be thanked for their funding, organising events and trips, and facilitating PIPS. Thanks go to YCR for allowing me to come for an internship placement, and for Katherine Scott for accepting me. I learnt a lot.

Members of Level 6 and Level 9 are thanked for creating an encouraging working environment Vincent Agboh, Dave Carrier, Peter Oatley, Catherine Qi, Amelia Lesiuk, Clair Phillips, Cheng Ma, Zhenyu Hao, Andrea Rawlings, Emily

III

Caseley, Tim Munsey, Ceasar Stanislaus, Lin Rong, Nita Shah, Maren Thompson, Craig Wilkinson, Jack Wright, Sarah Sabir. Continue eating cakes, playing squash, and playing games, all. You joined in with all our pranks and enhanced my competitive spirit, and remember that there's nothing wrong with complicated. Special thanks go to David Sharples for superlative technical support, training on the cell disruptor, and performing the fermentations within this thesis. Even more special thanks go to Vincent Postis without whose knowledge and support I could not have learned and performed even half of the experiments herein.

I could not have made it through without my friends. Vince A, Grace H, Jess H, Jess M, Jess B, Sarah S, Aaron M, Claudia M, Nita S, Dave C, Aimee L, Jamie K, Hannah S, Tom N, Jack G, James R, Ruth N, Laura M, Eleanor E, Fran R, Liz I, Chloe B, Rachel O, Debs R, Liz K. You've all made it special. In addition, much of my sanity has remained due to members of the Leeds ultimate frisbee team, and the wider frisbee community I've been able to be part of.

As with all theses, many people provided important materials. Thanks go to Prof. George Lomonosoff (John Innes Centre) for providing the pEAQ-HT and pEAQ-GFP vectors for plant expression, Dr Carine De Marcos Lousa for providing *Nicotiana benthamina* seeds, Vince Agboh for preparing the pL51-ALMT plasmid, and Mike Harrison for fulfilling the thankless role of examiner of my transfer and providing useful comments and questions.

A number of people here at Leeds should be thanked for their contributions to the experimental work presented here. James Ault performed protein identification by Mass Spectrometry, Nasir Khan assisted in CD measurements,

IV

Iain Manfield for use of the Optim in the Centre for Biomolecular Interactions, Chi Trinh supported the Crystallography, Martin Fuller provided technical assistance in EM, while firstly Robert Kolodziejczyk and more recently Maren Thompsen helped greatly with the SEC-MALLS operation and analysis and particularly Maren was a well of knowledge and advice.

Samantha Aspinall, Martha Smith, and Lucy Parker in the Graduate School were invaluable on occasions too numerous to recall here, and also provided regular opportunities for free food and drink.

I would like to thank my family, who have always had and unwavering belief in me and offered unconditional support.

Vince, we have one last toast to make.

Abstract

The focus of this study is proteins from the ArAE family, specifically two subfamilies: firstly, the ALMT family of plant membrane channels, which have numerous vital roles in plants, and secondly, the bacterial family first described as PET inner membrane exporters. Constructs were created for expression of firstly, the C-terminal domain (CTD) of wheat ALMT1 in *E. coli* and secondly, three full-length ALMTs (ALMT from wheat and ALMT5, ALMT9 from *Arabidopsis*) in *Nicotiana benthamiana* for structural and biochemical studies. Unfortunately, it appears that the CTD is not an independent soluble domain as originally thought, and this domain is now predicted to contain transmembrane helices, making it unsuitable for study in the manner planned. Similarly, production of full length ALMTs was unsuccessful as when extraction and purification was attempted, the protein degraded. Thirdly, a bacterial member of the ArAE family was expressed and characterised: AaeB from *E. coli*, along with its putative binding partner, AaeA. This was shown to form a complex with and be vital for the stability of AaeB. A strategy was devised for expression, solubilisation, and purification of these proteins and once they were obtained in pure form they were subjected to a range of biochemical and structural experiments. Microcrystals of AaeA were produced, towards a strategy for structural determination by X-ray crystallography. The first Electron Microscopy examination was performed on complexes of AaeB, and negative stain classes were produced. The first *in silico* homology model of AaeA has been produced and validated by CD spectroscopy, providing a range of insights. The complex formed by AaeA and AaeB has also been probed by crosslinking and SEC-MALLS analysis, suggesting a 6:2 or 6:3 stoichiometry. Together this has furthered our understanding of this poorly characterised membrane protein family and provided a set of clones and protocols for future studies.

Table of Contents

ACKNOWLEDGEMENTS	II
ABSTRACT	V
TABLE OF CONTENTS.....	VI
LIST OF FIGURES.....	XIII
LIST OF TABLES.....	XX
ABBREVIATIONS	XXII
CHAPTER 1 INTRODUCTION	1
1.1 Compartmentalisation and Membranes	2
1.1.1 Biological membranes and membrane proteins.....	2
1.1.2 Transport across membranes	4
1.1.3 Channels and transporters.....	5
1.1.3.1 Channels.....	5
1.1.3.2 Transporters.....	8
1.2 ArAE Family of Acid Exporters (including plant ALMTs)	12
1.3 ALMT Family Members in Plants, Function, Topology, Mechanism, and Structure	14
1.3.1 Food security.....	15
1.3.2 Aluminium Resistance.....	16
1.3.2.1 Overcoming aluminium toxicity on acid soils.....	16
1.3.2.2 TaALMT1 in Aluminium Resistance	16
1.3.2.3 Other ALMTs involved in aluminium-resistance	17
1.3.3 Structural and topological studies of TaALMT1 and AtALMT1	18

VII

1.3.3.1	C-terminal domain	19
1.3.3.2	Al-binding.....	20
1.3.3.3	Phosphorylation.....	21
1.3.4	Other roles of ALMTs.....	22
1.3.4.1	Other root related functions	22
1.3.5	Guard Cell Movements.....	25
1.3.6	Other Roles.....	29
1.3.6.1	Malate storage and homeostasis.....	29
1.3.6.2	Cell elongation and nutrient storage	29
1.3.6.3	Fruit Flavour	30
1.3.6.4	Inorganic Anion Homeostasis	30
1.3.6.5	GABA Signalling.....	31
1.4	Bacterial members of the ArAE family (PETs)	33
1.4.1	AaeA and AaeB – the aromatic acid exporter	33
1.5	MdtO/SdsQ - the Sulfa-drug Exporter	36
1.6	Challenges in membrane protein expression and structural studies	38
1.1.1	Expression	39
1.6.1	Structural Studies of Bacterial Proteins to Understand Eukaryotic Targets.....	44
1.6.2	Solubilisation	46
1.6.3	Purification	48
1.6.4	Crystallisation	49
1.6.5	Nuclear Magnetic Resonance (NMR) Spectroscopy	51
1.6.6	Electron Microscopy.....	52
1.7	Aims of this project	56
CHAPTER 2 MATERIALS AND METHODS		57
2.1	Chemicals and Reagents	58
2.2	Bacterial Growth and Expression Media.....	58
2.2.1	Media for growth	58

VIII

2.3	Bacterial Strains.....	62
2.3.1	Preparation of Chemically-Competent <i>Escherichia coli</i> Cells.....	62
2.3.2	Heat Shock Transformation of <i>Escherichia coli</i> Cells	63
2.3.3	Plasmid Maintenance in Bacteria	63
2.3.4	ALMT cDNA	64
2.4	Molecular Biology.....	65
2.4.1	PCR.....	65
2.4.2	Quikchange™ Site-Directed Mutagenesis	68
2.4.3	Cleaning of PCR Products.....	68
2.4.4	Colony PCR.....	68
2.4.5	Restriction Endonuclease Digestion of DNA	69
2.4.6	Agarose Gel Separation of DNA	69
2.4.7	Agarose Gel Extraction of DNA	70
2.4.8	Dephosphorylation of DNA Fragments	70
2.4.9	Ligation of DNA	70
2.4.10	Plasmid Purification and Preparation: Mini-preparation	71
2.4.11	DNA Sequencing.....	71
2.5	Protein Expression.....	72
2.5.1	Expression Screening in Small Scale Cultures	72
2.5.2	Large Scale Expression	72
2.5.3	Membrane Preparation	73
2.5.4	Detergent Screening for Membrane Protein Solubilisation.....	73
2.5.5	Purification of Proteins by Immobilised Metal Affinity Chromatography	74
2.5.6	Protein Estimation by the BCA Method	74
2.5.7	Protein Estimation by Absorbance at A280	74
2.5.8	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis.....	75
2.5.9	SDS-Polyacrylamide Gel Coomassie Staining	76
2.5.10	SDS-Polyacrylamide Gel Silver Staining	76
2.5.11	Dot Blotting and Western Blotting of Proteins	76

IX

2.5.12	Densitometry Protein Estimation of Coomassie-stained SDS-PAGE gels, or Western blots	78
2.6	Blue Native Polyacrylamide Gel Electrophoresis	78
2.7	Glutaraldehyde Cross Linking	78
2.8	Expression of ALMTs in Plants	79
2.8.1	Plant Growth Conditions	79
2.8.2	Transformation of agrobacteria	79
2.8.3	Agrobacteria tumefaciens cultures	79
2.8.4	Agrobacteria tumefaciens infiltration	80
2.8.5	Protein extraction from tobacco leaves	80
2.8.6	Imaging of GFP in leaf	81
2.9	Measurement of Protein Stability in Varying Buffer Conditions	81
2.10	Protein Structural Homology Modelling with Phyre2	81
2.11	Circular Dichroism Spectroscopy	83
2.12	Size Exclusion Chromatography-Multi Angle Laser Light Scattering (SEC-MALLS) analysis of the oligomeric states of membrane proteins	83
2.13	Electron Microscopy and Single Particle Analysis	83
2.14	Screening for Crystallography	84
CHAPTER 3 PRODUCTION OF THE C-TERMINAL DOMAIN OF TALMT1		85
3.1	Introduction	86
3.2	Production of the CTD of TaALMT1	87
3.2.1	Vector construction - pTP1 and pTP2	87
3.2.2	Expression trials from pTP1 and pTP2	91
3.2.2.1	Initial expression trial	91
3.2.2.2	Solubility Screen	92
3.2.2.3	Lysis screen	96
3.3	Production of a Truncated Construct for the CTD	97
3.3.1	Mutagenesis to Truncate pTP1 and pTP2	97

3.3.2	Expression of Truncated Constructs	99
3.3.3	Testing additives to aid solubility.....	103
3.3.3.1	Attempted purification of the solubilised truncated CTD	104
3.3.4	Denaturation and Refolding.....	106
3.4	Discussion and Conclusions	108
CHAPTER 4 EXPRESSION OF ALMTS IN TOBACCO		111
4.1	Introduction.....	112
4.1.1	Choice of plant expression and the pEAQ-HT vector	113
4.1.2	Proteins chosen from ALMT family for expression	115
4.2	Cloning.....	117
4.2.1	Primer design	117
4.2.2	Vector construction	118
4.3	GFP expression in tobacco as a proof of principle.....	123
4.4	Transient expression of His-tagged ALMTs via agrobacteria-mediated agroinfiltration in tobacco	
	127	
4.4.1	An initial expression trial of AtALMT5C	128
4.4.2	Improving fractionation with a high-density sucrose buffer	132
4.4.3	Improving expression conditions.....	134
4.4.4	Improving ALMT extraction	139
4.4.5	Test of rapid purification of ALMT5N.....	143
4.5	Discussion and Conclusion	147
CHAPTER 5 EXPRESSION AND PURIFICATION OF BACTERIAL ARAE EXPORTERS.....		151
5.1	Molecular biology.....	152
5.1.1	Identification of candidate genes	152
5.1.2	Design and Preparation of a Suitable Expression Vector.....	153
5.1.2.1	Design of pTTQ18-TP	153
5.1.2.2	Construction of pTTQ18-TP	156

XI

5.1.3	Production of expression vectors.....	157
5.1.3.1	Amplification of target genes	157
5.1.3.2	Insertion of ORFs into the expression vector pTTQ18-TP.....	158
5.1.4	Expression of bacterial genes in <i>E. coli</i>	160
5.1.4.1	Solubilisation and purification trials for So-TTQ.....	161
5.1.4.2	Ec-TTQ expression and solubilisation	162
5.2	Construction of vectors for expression of AaeAB, AaeA, and AaeA Δ TM.....	164
5.2.1	Production of AaeB co-expressed with AaeA.....	167
5.2.1.1	Optimisation of expression conditions for AaeA-AaeB.	167
5.2.1.2	Assessment of IPTG- vs Auto-induction for optimised yield.....	168
5.2.1.3	Solubilisation of AaeB co-expressed with AaeA	169
5.2.1.4	Purification of AaeB	170
5.2.2	Expression of AaeA and AaeA Δ TM.....	175
5.2.2.1	Initial expression conditions screening for AaeA and AaeA Δ TM.....	175
5.2.2.2	Assessing the membrane localisation and solubility of AaeA.....	176
5.2.2.3	Purification of AaeA	178
5.2.2.4	Production of sufficient AaeA for structural and biochemical studies	179
5.2.2.5	Assessing the membrane localisation and solubility of AaeA Δ TM	182
5.3	Summary and Conclusions	185
CHAPTER 6 BIOCHEMICAL AND STRUCTURAL INVESTIGATIONS OF AAEA AND AAEB.....		187
5.1	<i>In silico</i> homology modelling of AaeA and bioinformatics analysis	188
6.1.1	The RLS Motif	192
6.1.2	Conservation and Surface Charge	194
6.2	Examination of buffer conditions for optimal stability of AaeA.	196
6.3	Circular Dichroism measurement of AaeA.....	199
6.3.1	Thermal Melt	201
6.3.2	Optimised buffer	202
6.4	X-ray Crystallography of AaeA	205

XII

6.5	Electron Microscopy of AaeB	211
6.5.1	EM classes.....	212
6.5.2	Gold labelling	216
6.6	Studies on the oligomeric state of AaeA and AaeB	219
6.6.1	Crosslinking of AaeB.....	220
6.6.2	Native PAGE Gel of AaeB	224
6.6.3	Crosslinking of AaeA	225
6.7	Assessment of the Oligomeric State of AaeB by SEC-MALLS.....	227
6.7.1	SEC-MALLS analysis of AaeB	230
6.7.2	SEC-MALLS analysis of AaeA	235
6.8	Summary and conclusions.....	239
CHAPTER 7 GENERAL DISCUSSION AND CONCLUSIONS		243
7.1	Strategies for overexpression of membrane proteins for structural studies	244
7.1.1	The CTD of ALMT1	244
7.1.2	Full length ALMTs.....	245
7.1.3	Bacterial ArAE Proteins	246
7.2	Future work.....	251
7.2.1	Continuation of work on AaeA-AaeB	251
7.2.1.1	An alternative to detergent.....	253
7.2.2	Study of ALMTs	254
7.2.3	Study of Bacterial Homologues.....	256
7.3	Summary	258
CHAPTER 8 REFERENCES		261

List of Figures

FIGURE 1.1 - DIFFUSION OF LIPID-SOLUBLE COMPOUNDS ACROSS MEMBRANES TAKES PLACE WITH A NET FLUX FROM REGIONS OF HIGH CONCENTRATION TO REGIONS OF LOW CONCENTRATION.....	5
FIGURE 1.2 - DIAGRAMMATIC REPRESENTATION OF THE ACTION OF CHANNELS AND TRANSPORTERS.....	7
FIGURE 1.3 - PREDICTED OVERALL ARCHITECTURE OF ARAE FAMILY MEMBERS FROM PLANTS AND BACTERIA.	13
FIGURE 1.4 - STRUCTURE OF MALATE AND ALUMINIUM COMPLEXED BY MALATE. STRUCTURE OF THE COMPLEX FROM TASHIRO <i>ET AL.</i> ⁶⁶	17
FIGURE 1.5 - PHYLOGENETIC TREE OF ALMTS WITH KNOWN FUNCTIONS PLUS FAMILY MEMBERS FROM <i>ARABIDOPSIS</i> WITH UNKNOWN FUNCTION.	24
FIGURE 1.6 - SCHEMATIC OF ION CURRENTS IN STOMATAL OPENING AND CLOSING.....	26
FIGURE 1.7 THE PRIMARY SUBSTRATE OF AAEB, P-HYDROXYBENZOIC ACID.	33
FIGURE 1.8 – HYDROPHOBICITY PLOTS FOR AAEA AND AAEB.	34
FIGURE 1.9 - SCHEMATIC REPRESENTATION OF A RANGE OF MFP CRYSTAL STRUCTURES.	35
FIGURE 1.10 – CUMULATIVE NUMBER OF MEMBRANE PROTEIN STRUCTURES SOLVED, SHOWING AN EXPONENTIAL INCREASE OVER TIME.....	39
FIGURE 1.11 DETERGENT SOLUBILISATION OF MEMBRANE PROTEINS.	48
FIGURE 1.12 - EFFECT OF DIFFERING MICELLE SIZES ON MEMBRANE PROTEIN VISIBILITY AND CRYSTAL CONTACTS.	51
FIGURE 1.13 - CROSS SECTION OF V-ATPASE SHOWING IMPROVEMENTS IN RESOLUTION BETWEEN 2009-2016.....	54
FIGURE 3.1 - OVERALL MODEL OF THE DOMAIN ORGANISATION OF ALMT PROTEINS. PREPARED BY STEVE BALDWIN.	86
FIGURE 3.2 – TP1 AND TP2 VECTORS	88

XIV

FIGURE 3.3 – DIGESTS TO SHOW PTP1 CONSTRUCTION.....	89
FIGURE 3.4 - RESTRICTION DIGESTS OF PTP1 AND PTP2 TO CONFORM CORRECT CLONING.	90
FIGURE 3.5 – DOT BLOT AND WESTERN BLOT OF HIS-TAGGED PROTEIN EXPRESSED FROM PTP1.....	91
FIGURE 3.6 - SELECTION OF WESTERN BLOTS COMPARING TOTAL AND SOLUBLE FRACTIONS FROM TP1 EXPRESSED OVER A RANGE OF TIME-POINTS IN EITHER BL21 GOLD OR BL21 STAR IN LB MEDIA.....	93
FIGURE 3.7 - DOT BLOT ASSESSING THE EFFECTIVENESS OF CHEMICAL LYSIS WITH TRITON AGAINST LYSIS IN A CELL DISRUPTOR.....	96
FIGURE 3.8 OVERVIEW OF MUTAGENESIS STRATEGY WITH PTP1 AS AN EXAMPLE.....	97
FIGURE 3.9 -TRUNCATED VECTORS OVERVIEW.....	98
FIGURE 3.10 - CHARACTERISATION OF PRODUCTS PRODUCED BY QUIKCHANGE MUTAGENESIS.	99
FIGURE 3.11 –BLOTS TO ASSESS TRUNCATED TP1_T.....	100
FIGURE 3.12 - WESTERN BLOT TO ASSESS EXPRESSION OF TP1_T IN BL21 STAR USING AUTOINDUCTION FOR THREE TIME POINTS.....	101
FIGURE 3.13 – ASSESSMENT OF THE EFFECTS OF ADDITIVES TO THE SOLUBILITY OF TP1_T.	103
FIGURE 3.14 - A PURIFICATION OF A SAMPLE OF TP1_T CONTAINING TRITON, EDTA, DTT.	105
FIGURE 3.15 – ASSESSMENT OF REFOLDING TP1.	107
FIGURE 3.16 - PHOTOGRAPH OF RE-FOLDED TP1 AFTER DIALYSIS. ALL MATERIAL PRECIPITATES.....	107
FIGURE 3.17 – PUTATIVE TM HELICES IN THE CTD OF TAALMT1.....	109
FIGURE 3.18 - SEQUENCES OF TWO PUTATIVE TM HELICES IDENTIFIED BY DREYER <i>ET AL.</i> ⁴⁵	110

FIGURE 4.1 - SCHEMATIC VIEW OF PEAQ-HT SEQUENCE, HIGHLIGHTING KEY ELEMENTS.	115
FIGURE 4.2 PHYLOGENETIC TREE OF ALMTS WITH KNOWN FUNCTIONS PLUS FAMILY MEMBERS FROM ARABIDOPSIS WITH UNKNOWN FUNCTION. PROTEINS STUDIED HEREIN ARE BOXED.....	116
FIGURE 4.3 - GRADIENT PCR WITH A RANGE OF ANNEALING TEMPERATURES TO AMPLIFY ATALMT9N.....	119
FIGURE 4.4 PCRS OF TAALMT1C (1C), ATALMT5N (5C), ATALMT9C (9C), ATALMT5N (5N), ATALMT9N (9N).	119
FIGURE 4.5 -DIGESTS OF BLUNT VECTORS.....	120
FIGURE 4.6 – 1% AGAROSE GEL OF DIGESTED FINAL CONSTRUCTS FOR PLANT EXPRESSION.	121
FIGURE 4.7 - 1% AGAROSE GEL SHOWING DOUBLY DIGESTED ALMT1C AND ALMT1N. .	121
FIGURE 4.8 - 1% AGAROSE GEL OF 5 ML SAMPLES OF COLONY PCR TO CONFIRM THE PRESENCE OF INSERTS IN QUADRUPPLICATE.	122
FIGURE 4.9 - FLUORESCENCE IMAGE OF A SET OF LEAVES INFILTRATED WITH AGROBACTERIUM TUMEFACIENS LBA4404 TRANSFORMED WITH PEAQ-GFP.	124
FIGURE 4.10 - FLUORESCENCE IMAGE OF TWO TUBES CONTAINING PROTEIN EXTRACTED FROM A SET OF LEAVES INFILTRATED WITH INFILTRATION MEDIUM (BLANK), OR AGROBACTERIUM TUMEFACIENS LBA4404 TRANSFORMED WITH PEAQ-GFP (GFP).	125
FIGURE 4.11 - FLUORESCENCE IMAGE OF SDS-PAGE GEL OF EXTRACTED SAMPLES FROM LEAVES.....	126
FIGURE 4.12 – FLUORESCENCE IMAGE OF A TOBACCO LEAF INFILTRATED WITH TWO PEAQ CONSTRUCTS.....	128
FIGURE 4.13 - DOT BLOT OF TWO PAIRS OF PELLET/SUPERNATANT FRACTIONS EXTRACTED FROM EACH OF SIX SAMPLES OF LEAVES INFILTRATED WITH ALMT5C AFTER 5 DAYS OF EXPRESSION.	129
FIGURE 4.14 – WESTERN BLOT OF MEMBRANES PREPARED FROM SIX LEAF SAMPLES INOCULATED WITH VARYING O.D. OF AGROBACTERIA AND EXPRESSING ATALMT5C. TWO BIOLOGICAL REPLICATES WERE TAKEN FOR EACH O.D. ₆₀₀ TESTED.	131

FIGURE 4.15 - SAMPLES OF PROTEIN EXTRACTED FROM LEAVES INFILTRATED WITH THREE O.D. ₆₀₀ OF AGROBACTERIA HARBOURING PEAQ-TAALMT1N HARVESTED AFTER 8 DAYS OF EXPRESSION.....	133
FIGURE 4.16 - WESTERN BLOT OF TOTAL PROTEIN EXTRACTED FROM TOBACCO LEAVES EXPRESSING TAALMT1C, TAALMT1N, ATALMT5C, ATALMT9C, ATALMT9N.....	138
FIGURE 4.17 - WESTERN BLOT OF 20 ML OF SAMPLES EXTRACTED FROM TOBACCO LEAVES EXPRESSING TAALMT1C (1C), TAALMT1N (1N), ATALMT5C (5C), ATALMT9C (9C), ATALMT9N (9N) EXTRACTION FIRST WITH BUFFER AND SECONDLY WITH BUFFER CONTAINING 1% TRITON X-100.....	139
FIGURE 4.18 - WESTERN BLOT OF 20 ML OF THREE SAMPLES OF LEAF CONTAINING ALMT5N, INOCULATED WITH AGROBACTERIA AT O.D. ₆₀₀ = 0.1, 0.5, OR 1.0, EACH EXTRACTED TWICE WITH BUFFER CONTAINING TRITON X-100.....	140
FIGURE 4.19 - WESTERN BLOT OF ALMT5N WITH HIS-STANDARDS TO ASSESS EXPRESSION LEVELS.	142
FIGURE 4.20 - DOT BLOT OF PROTEIN EXTRACTED FROM TOBACCO LEAVES EXPRESSING ATALMT5N.....	143
FIGURE 4.21 - ALMT5N PURIFICATION.	144
FIGURE 4.22 - WESTERN BLOT OF 20 ML OF FRACTIONS OF A ATALMT5N PURIFICATION, WITHOUT STANDARDS.....	145
FIGURE 4.23 - SILVER STAINED SDS-PAGE GEL OF ALMT5N PURIFICATION.	145
FIGURE 5.1 - SCHEMATIC OF THE PTTQ18-TP EXPRESSION VECTOR.....	155
FIGURE 5.2 - 0.5% AGAROSE GELS OF DIGESTED SAMPLES OF THE TWO VECTORS USED TO CREATE THE EXPRESSION PLASMID USED IN THIS STUDY.	156
FIGURE 5.3 - TOUCHDOWN PCR TO AMPLIFY THREE ORFS FROM <i>S. ONEIDENSIS</i> (SO), <i>D. VULGARIS</i> (DV), AND <i>E. COLI</i> (EC) GENOMIC DNA.	157
FIGURE 5.4 - DOUBLE DIGESTS TO PREPARE FRAGMENTS FOR CONSTRUCTION OF EXPRESSION VECTORS.....	159
FIGURE 5.5 - DOT BLOT OF 3 ML SAMPLES FROM SMALL SCALE EXPRESSION TESTS OF TWO BACTERIAL CONSTRUCTS.....	160
FIGURE 5.6 - SOLUBILISATION AND PURIFICATION OF SO-TTQ (20 ML).	161

FIGURE 5.7 - SOLUBILISATION OF EC-TTQ (20 ML).	163
FIGURE 5.8 - SCHEMATIC TO SHOW CONSTRUCTS.....	165
FIGURE 5.9 – VECTOR CONSTRUCTION.	166
FIGURE 5.10 - DIGEST TO TEST THE EXPRESSION VECTOR AAEA IN PET28(B) ON A 1% AGAROSE-TAE GEL.	166
FIGURE 5.11 - DOT BLOT ASSESSING CULTURE CONDITIONS FOR AAEA-AAEB.	167
FIGURE 5.12 - GRAPH MONITORING OPTICAL DENSITY OF BACTERIAL CULTURES EXPRESSING AAEB, CO EXPRESSED WITH AAEA, AFTER IPTG-INDUCTION.	168
FIGURE 5.13 - WESTERN BLOT OF 20 ML OF PROTEIN EXTRACTED FROM SIX CULTURES OF BL21 GOLD EXPRESSING AAEB.	169
FIGURE 5.14 - WESTERN BLOTS SHOWING THAT AAEB CAN ONLY BE SOLUBILISED IN DETERGENT WHEN CO-EXPRESSED WITH AAEA.....	170
FIGURE 5.15 - 20 ML SAMPLES FROM EACH FRACTION OF A PURIFICATION OF AAEB. ...	171
FIGURE 5.16 - PEPTIDES IDENTIFIED IN MASS SPECTROMETRY ANALYSIS OF BANDS IN PURIFIED SAMPLES OF AAEB CO-EXPRESSED WITH AAEA.	172
FIGURE 5.17 - COOMASSIE STAINED SAMPLES OF THREE PURIFICATIONS OF AAEB, RUN ON SDS-PAGE GELS.	174
FIGURE 5.18 - DOT BLOT TO ASSESS INITIAL EXPRESSION TRIAL FOR AAEA AND AAEA Δ TM.	175
FIGURE 5.19 - WESTERN BLOT OF A FRACTIONATION OF CELLS TO LOCATE AAEA.....	176
FIGURE 5.20 - WESTERN BLOT OF ATTEMPTED SOLUBILISATION OF AAEA.	177
FIGURE 5.21 - PURIFICATION OF AAEA.....	179
FIGURE 5.22 - WESTERN BLOT OF SEVERAL SAMPLES OF AAEA EXPRESSED IN BL21 GOLD USING LB MEDIUM.....	180
FIGURE 5.23 - DOT BLOT OF SUPERNATANT AND PELLETS FROM CENTRIFUGATION OF AAEA Δ TM.....	182

XVIII

FIGURE 5.24 - DOT BLOT TO MONITOR SOLUBILISATION OF INCLUSION BODIES OF AAEAΔTM.....	183
FIGURE 6.1 - SECONDARY STRUCTURE AND DISORDER PREDICTION BY PHYRE2 FOR AAEA.	189
FIGURE 6.2 - MODEL OF AAEA PRODUCED BY PHYRE2.....	191
FIGURE 6.3 - RLS MOTIF ALIGNMENT FOR SEVERAL MFP PROTEINS FROM SEVERAL FAMILIES.....	193
FIGURE 6.4 – EXAMINATION OF THE AAEA MODEL.....	195
FIGURE 6.5 - EXAMPLE GRAPHS PRODUCED FROM OPTIM DATA.....	197
FIGURE 6.6 - SUMMARY GRAPHS FROM OPTIM REPORTING T_M (A-C) AND T_{AGG} (D-F) OVER A RANGE OF PH (A, D), NA CL CONCENTRATION (B, E), AND GLYCEROL % (C, F) IN THE BUFFER.	198
FIGURE 6.7 – CD ANALYSIS OF AAEA.	200
FIGURE 6.8 - SECONDARY STRUCTURE PREDICTION BY CD AS COMPARED TO PHYRE2 MODEL.....	204
FIGURE 6.9 - A RANGE OF IMAGES OF REPRESENTATIVE RESULTS FROM SCREENING CONDITIONS FOR GROWTH OF CRYSTALS.....	206
FIGURE 6.10 – REPRESENTATIVE EXAMPLES OF POTENTIAL HITS FROM CRYSTAL SCREENING.	208
FIGURE 6.11 - A PAIR OF REPRESENTATIVE MICROGRAPHS AT 30,000X MAGNIFICATION FROM AAEB NEGATIVE STAIN EM GRIDS.....	211
FIGURE 6.12 - EM CLASS AVERAGES OF AAEB IMAGED AT 30,000X MAGNIFICATION.	212
FIGURE 6.13 – A) EM CLASS AVERAGES FROM AAEB TAKEN AT 50,000X MAGNIFICATION. B) SELECTED CLASSES THAT SHOW A MORE DISTINCT MORPHOLOGY.....	213
FIGURE 6.14 - EXAMPLE MICROGRAPHS OF AAEB NEGATIVE STAIN EM GRIDS IMAGED AT 30,000X MAGNIFICATION.	214
FIGURE 6.15 - SECOND SET OF EM CLASS AVERAGES FROM AN INDEPENDENTLY PREPARED SAMPLE.	215

XIX

FIGURE 6.16 – EM NEGATIVE STAIN IMAGES OF SAMPLES OF AAEB LABELLED WITH GOLD.	217
FIGURE 6.17 - COOMASSIE STAINED SDS-PAGE GEL OF 20 ML OF IMAC-PURIFIED AAEB.	219
FIGURE 6.18 – CROSSLINKING OF AAEB.....	221
FIGURE 6.19 - CROSSLINKING OF AAEB.	223
FIGURE 6.20 – WESTERN BLOT OF A NATIVE PAGE GEL.....	224
FIGURE 6.21 - WESTERN BLOT OF 20 ML SAMPLES OF PURIFIED AAEA CROSSLINKED WITH GLUTARALDEHYDE FOR 15 MINUTES.	225
FIGURE 6.22 - CROSSLINKING OF PURIFIED DDM-SOLUBILISED AAEA	226
FIGURE 6.23 - SEC-MALLS ANALYSIS OF CARBONIC ANHYDRASE STANDARD TO ASSESS CALIBRATION.	229
FIGURE 6.24 - SEC-MALLS ANALYSIS OF PURIFIED AAEB.....	231
FIGURE 6.25 - SEC-MALLS ANALYSIS OF AAEA MIXED WITH AAEB.....	235
FIGURE 6.26 - SEC-MALLS ANALYSIS OF PURIFIED AAEA.....	238
FIGURE 7.1 - OVERALL SCHEMATIC PROPOSED FOR THE COMPLEX AAEAB AS COMPARED TO ACRA-B-TOLC	250

List of Tables

TABLE 1 - SUMMARY OF KNOWN FUNCTIONS AND LOCALISATION OF ALMTS FROM SEVERAL SPECIES	23
TABLE 2.1 - LYSOGENY BROTH (LB) MEDIUM	58
TABLE 2.2 - SUPER BROTH (SB) MEDIUM	58
TABLE 2.3 - LB MEDIUM WITHOUT SALT FOR USE IN AUTOINDUCTION.....	59
TABLE 2.4 - SB MEDIUM WITHOUT SALT FOR USE IN AUTOINDUCTION.....	59
TABLE 2.5 - 5X M9 SALTS FOR USE IN AUTOINDUCTION	59
TABLE 2.6 - 20X NSPC FOR USE IN AUTOINDUCTION.....	59
TABLE 2.7 - 50X 5052 FOR USE IN AUTOINDUCTION	60
TABLE 2.8 - LB FOR AUTOINDUCTION.....	60
TABLE 2.9 - SB FOR AUTOINDUCTION	60
TABLE 2.10 - M9 FOR AUTOINDUCTION	61
TABLE 2.11 - LIST OF PRIMERS USED FOR THE TRUNCATION OF PTP1 AND PTP2.....	66
TABLE 2.12 - LIST OF PRIMERS USED FOR CONSTRUCTION OF PLANT EXPRESSION CONSTRUCTS.	67
TABLE 2.13 - LIST OF PRIMERS USED FOR THE AMPLIFICATION OF BACTERIAL GENES.....	67
TABLE 2.14 - PRIMERS FOR SEQUENCING	71
TABLE 2.15 - COMPOSITION OF 12% SDS-PAGE GELS.....	75
TABLE 3.1 – TABLE LISTING EXPRESSION CONDITIONS ATTEMPTED FOR TP1 AND TP2....	95
TABLE 3.2 - TABLE LISTING EXPRESSION CONDITIONS ATTEMPTED FOR TP1_T AND TP2_T	102

TABLE 4.1 - SUMMARY OF CONDITIONS USED FOR EXPRESSING HIGH YIELDS OF PROTEIN FROM CPMV- AND PEAQ-BASED EXPRESSION VECTORS FROM RECENT PAPERS ..	134
TABLE 4.2 - REVISED CONDITIONS FOR SCREENING ALMT EXPRESSION IN TOBACCO	135
TABLE 4.3 –FULL LIST OF CONDITIONS TESTED FOR ALMT EXPRESSION TRIALS.	136
TABLE 5.1 – INFORMATION ON GENES IDENTIFIED AS HOMOLOGOUS TO ALMT FAMILY MEMBERS	153
TABLE 6.1 - TABLE OF CONDITIONS WITHIN THE WELLS OF PUTATIVE HITS IN CRYSTAL SCREEN.	209
TABLE 6.2 - SUMMARY OF SEC-MALLS RESULTS FOR AAEB ON S200 COLUMN	232
TABLE 6.3 - SUMMARY OF SEC-MALLS RESULTS FOR AAEB ON WTC-MP030S5 COLUMN	233
TABLE 6.4 - SUMMARY OF SEC-MALLS RESULTS FOR AAEA PLUS AAEB ON WTC-MP030S5 COLUMN	235
TABLE 6.5 - SUMMARY OF SEC-MALLS RESULTS FOR AAEA ON S200 COLUMN	236
TABLE 6.6 - SUMMARY OF SEC-MALLS RESULTS FOR AAEA ON WTC-MP030S5 COLUMN	237

Abbreviations

A ₂₈₀	Absorbance at 280 nm
ALMT	Aluminium Activated Malate Transporter
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BLAST	Basic Local Alignment Search Tool
bp	base pairs
C12E8	Octaethylene Glycol Monododecyl Ether
CD	Circular Dichroism
CMC	Critical Micelle Concentration
DDM	n-Dodecyl β -D-maltoside
DM	n-Decyl β -D-maltoside
DMNG	Decyl Maltose Neopentyl Glycol
DMSO	Dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EM	Electron Microscopy
GABA	γ -Aminobutyric acid
GFP	Green Fluorescent Protein
His	Histidine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

XXIII

HRP	Horseradish peroxidase
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
kbp	kilobasepairs
kDa	kiloDaltons
LB	Lysogeny broth
LDAO	N,N-Dimethyldodecylamine N-oxide
LMNG	Lauryl Maltose Neopentyl Glycol
Mr	Relative Molecular Mass
mRNA	messenger RNA
MW	Molecular Weight
NTA	Nitrilotriacetic acid
OD ₆₀₀	Optical Density measured at 600 nm
β -OG	octyl glucoside
ORF	Open Reading Frame
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PDB	Protein Databank
PDBTM	Protein Databank of Transmembrane Proteins
PEG	Polyethylene glycol
RNA	Ribonucleic acid
SB	Super Broth
SDS	Sodium dodecyl sulfate

XXIV

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC-MALLS	Size exclusion chromatography - multi angle laser light scattering
TAE	Tris-acetic acid-EDTA
TBS	Tris buffered saline
TBST	Tris buffered saline-Tween 20
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TM	Transmembrane
TMHMM	Transmembrane Hidden Markov Model
Tris	Tris(hydroxymethyl)aminomethane
v/v	volume/volume
V_{\max}	Maximum velocity of transport
w/v	weight/volume
WT	Wildtype

Chapter 1

Introduction

1.1 Compartmentalisation and Membranes

1.1.1 Biological membranes and membrane proteins

All cells are surrounded by a membrane, and in many ways it is the membrane that defines the boundary of life: whatever is inside the plasma membrane is part of a living organism and whatever is outside is not. This phospholipid bilayer has a hydrophobic core that only permits small, uncharged molecules such as CO₂ to cross, while molecules such as water, charged inorganic ions, and large polar molecules are impermeable. Thus, membranes perform a vital role for any organism or cell, allowing cells to contain and control their DNA, proteins, and metabolic activity, while permitting vital chemicals to be contained at useful concentrations and keeping out undesirable compounds. A plasma membrane surrounds every cell of every organism and eukaryotes also contain additional intracellular compartments separated by membranes. This compartmentalisation enables a single cell to have several different environments within it. For example, lysosomes are maintained at an acidic pH to enable efficient degradation¹, and separate organelles are available for photosynthesis, in the case of chloroplasts², or respiration in the case of mitochondria³, enabling each to have controlled conditions, prevent futile cycles of opposing pathways, and to keep reactive oxygen species contained. However, while separation and compartmentalisation brings many benefits, cells and organelles also require contact with their immediate environment in a controlled way that does not compromise the integrity of the membrane and for this task they have a specialised group of transport proteins embedded in their membranes.

Membrane proteins are involved with many vital roles within cells including bioenergetics, signal transduction, ion and solute transmission, and catalysis. There are many such proteins: the human genome has over 1000 transporters alone (<http://www.membranetransport.org>). Consequently, and unsurprisingly, they are key targets for medical research, chiefly because dysfunction in membrane proteins is the basis of many diseases⁴⁻⁶. For example, cystic fibrosis is known to be caused by deficiency in control of chloride transport⁷, and membrane transporters are the target for many pharmacological therapies such as proton pump inhibitors to treat gastro-intestinal problems⁸, and current research is attempting to target bacterial transporters involved with removal of antibiotics in order to reduce multidrug resistance⁹. A commonly used drug class is “beta blockers”, which bind to the β -Adrenergic G-protein-coupled receptor (GPCR) to protect the heart against heart attacks and cardiac arrhythmias, and the first human GPCR crystal structure was of the β 2 Adrenergic Receptor bound to the inverse agonist carazolol¹⁰. Another drug class that targets a GPCR is serotonin-specific reuptake inhibitors (SSRIs), which bind to the serotonin transporter SERT¹¹ and are used to treat depression. Of course, drugs are not the sole driver of membrane protein research, with the notable area of plant membrane biology being important to the understanding and generation of improved crops, whether by conventional breeding or bioengineering¹².

Membrane proteins are an important target for study: they comprise up to one third of genes in the genomes of sequenced organisms,¹³ over two thirds of current drugs target membrane proteins¹⁴, and even outside of the context of drug development they are involved in numerous vital processes. However,

despite this importance there have been relatively few membrane protein structures solved¹⁵. This is largely due to the fact that membrane proteins are inherently hard to produce, stabilise, and characterise structurally⁴. The membrane volume is only a small fraction of a cell, and so they are usually expressed at lower levels than their soluble counterparts. They are hard to purify, and in the native state require a lipid environment to remain in their stable three-dimensional structure but these lipids can interfere with biochemical experiments as well as structural studies. We still lack a detailed understanding of the structure and function of most membrane protein families, how they work, how they move, and how they are regulated. This understanding will require the determination of many more high-resolution structures to develop a more complete understanding of the structure-function relationships encoded within membrane protein sequences.

1.1.2 Transport across membranes

Membranes are essential to contain and define a cell and its organelles and, conversely, transport across these membranes is vital to enable metabolites to enter, waste products to be removed, and signals to be transmitted. Some molecules, such as urea and steroid hormones such as testosterone, are sufficiently hydrophobic to permit them to cross the membrane un-aided, by simple diffusion¹⁶. This takes place down a concentration gradient by the process of random movement of molecules, towards equilibrium, as seen in **Figure 1.1**. This process requires no external energy input, and will continue to happen until equilibrium is reached, but cannot move solutes up a concentration

gradient, and does not permit the transport of hydrophilic substrates. For these tasks, specialised membrane proteins are required.

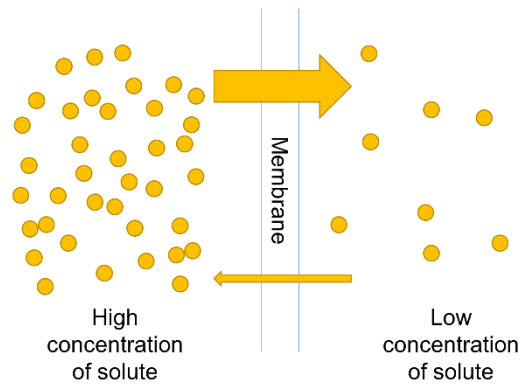


Figure 1.1 - Diffusion of lipid-soluble compounds across membranes takes place with a net flux from regions of high concentration to regions of low concentration.

1.1.3 Channels and transporters

Channels and transporters are required for cells to mediate the movement of hydrophilic compounds across membranes and to move compounds up a concentration gradient. There are many such molecules that are important to cross into or out of the cell, for example signalling molecules¹⁷, metabolites such as glucose¹⁸ and nucleosides¹⁹, and equilibration of water content in cells²⁰. Thus, these channels and transporters are important for growth, metabolic processes, and sensory perception.

1.1.3.1 Channels

Channels, seen in **Figure 1.2A**, act as gates within the membrane; they have a hollow, hydrophilic core that forms the channel pore and permits ions, water, and other small molecules to pass through them into the cell by a process of facilitated diffusion, which takes place down the electrochemical gradient in the same manner as simple diffusion. They generally have a relatively simple

mechanism with a central gated pore to allow transport and often have very high rates of transfer (up to 10^8 molecules per second²¹) as when they are in the open state they are limited only by the rate of diffusion. Aquaporins, for example, enable an almost instantaneous equilibration of water across membranes. Flux is determined by two factors, the chemical potential, resulting from a difference in concentration across the membrane, and, for charged particles, the membrane potential, or electrical potential, resulting from the distribution of charge across a membrane. These factors combined are called the electrochemical potential.

While the overall mechanism of action of channels is conceptually simple, the details of individual channels are more complex. Channel opening and closing can be stimulated by three stimuli, firstly, ligand binding²² (which may be a non-covalent ligand such as acetylcholine or a covalent modification such as phosphorylation); secondly, changes in membrane potential in the case of voltage-gated channels²³, which contain a charged domain that moves in response to voltage changes and causes a conformational change in the pore; and thirdly, mechanical tension may activate channels, for example those found in smooth muscle or that control turgor in plants²⁴.

Whatever the method of activation and de-activation, the process of gating itself can take place in one of three general ways. Firstly, gating can be via a small discrete area, for example at the selectivity filter, secondly, it can come about from a larger, general conformational change, or thirdly a channel may contain a blocking domain that moves into the pore to seal it in a closed state²⁵. This gating control is vital for the cell, as without it channels would permanently dissipate electrochemical gradients established across the membrane.

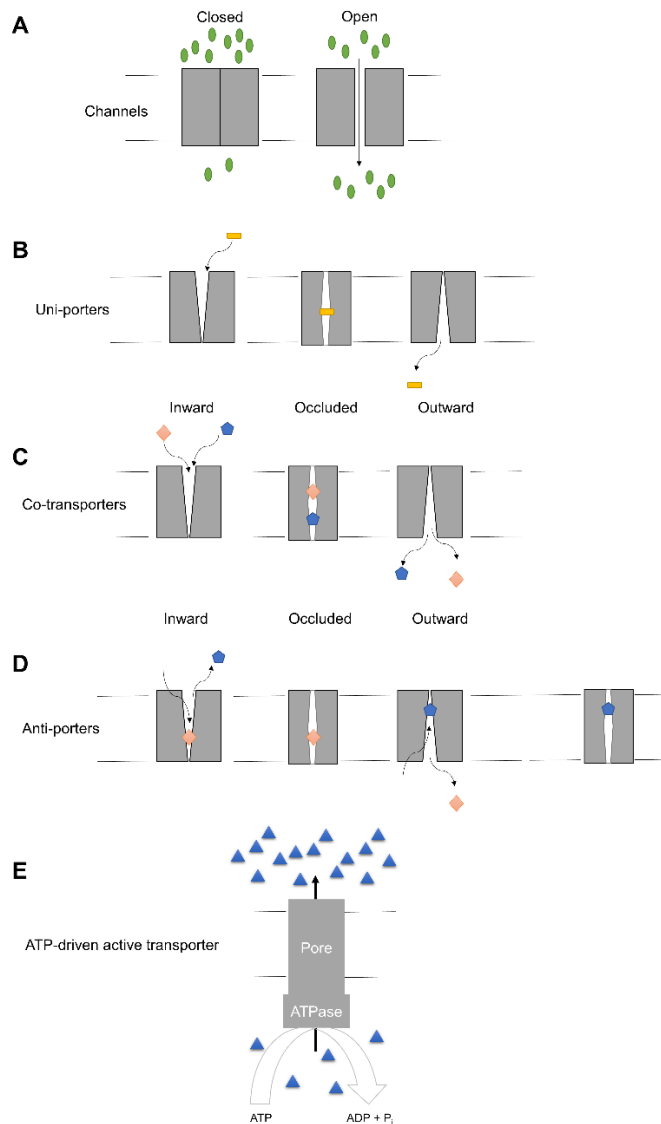


Figure 1.2 - Diagrammatic representation of the action of channels and transporters.

A) Channels switch from a closed state to an open state to allow transport down a concentration gradient. B) uniporters permit facilitated diffusion in a similar manner to channels, but transport is via several conformational changes. C) Co-transport and D) antiport is shown with proteins cycling through different states in order to transport their substrate. This can be a case of both substrates moving down a concentration gradient but also enables secondary active transport against an electrochemical gradient. E) ATP-driven Transporter. Hydrolysis of ATP drives transport of substrate up a concentration gradient. ATPase domain hydrolyses ATP and this energy is used to drive transport, leading to primary active transport, for example in the V-ATPase or ABC transporters.

An example of such channels is aquaporins. These proteins serve as channels in the transfer of water across the membrane and conserved in bacteria, plants, and animals²⁰. Channel gating has been investigated in the spinach aquaporin PIP2;1 protein, with crystallography revealing both open and closed conformations, with large loops moving to occlude the pore in the closed state and phosphorylation promoting closing of the pore during drought stress²⁶ and this tight control of water loss plays a large role in survival during drought stress.

1.1.3.2 Transporters

Transporters are also used by cells to move solutes from one side of a membrane to the other, but often have a more complicated mechanism of action than channels. These proteins often have two “gates” within their structure, one at either side of the membrane, which are never both open at the same time, but can be both closed resulting in an occluded state whereby the substrate is located in an internal cavity with no access to either side of the membrane. These gates sequentially open and close as the protein shifts through several intermediate states. These states allow substrate binding, movement through the membrane, and substrate release.

1.1.3.2.1 Facilitated diffusion by uniporters

Facilitated diffusion through uniporters is conceptually similar to transport by channels. It does not require energy input to move ions or molecules across the membrane and the membrane proteins merely provide a path across the membrane through which the substrate can travel down the electrochemical gradient. Uniporters, seen in **Figure 1.2B**, can provide this type of transport to their specific substrates. For example, GLUT1 transports glucose across

mammalian plasma membranes²⁷ as a Michaelis-Menten enzyme relying on the specific binding of the substrate to the protein. Transport is driven by conformational changes between three states, one state open to the outside, allowing it to pick up glucose, an occluded state, which is closed to both sides of the membrane, and a third state open to the inside allowing it to release the glucose into the cell²⁷. Like a channel, transport still depends on the concentration gradient, but unlike a channel transport is saturable due to the mechanism which requires multiple conformational state changes, and thus there is a limited maximum flux per protein.

1.1.3.2.2 Active transport

While channels and facilitative transporters only permit substrates to move across membranes down their electrochemical gradients, active transporters can move a substrate against its concentration gradient, using an external source of energy. This can be achieved by primary active transport – that is, coupling transport to the hydrolysis of ATP. This is the case, for example, in ABC transporters²⁸: where ATP hydrolysis drives conformational changes in the molecule to allow the transport of substrate against its own electrochemical gradient, or in the V-ATPase, which is involved in pumping protons to acidify compartments²⁹, depicted in **Figure 1.2E**. This enables transport to be concentrative, rather than equilibrative, so that a chemical species can be accumulated to much higher concentrations than found in the environment in order to, for example, scavenge precious nucleosides from the environment³⁰, or to maintain the vital metabolite glutamine concentrations in the cytoplasm at ~2-20 mM against an external concentration of 0.2-0.8 mM³¹. A unique transport system in bacteria is the Phosphotransferase System (PTS) which is

used for uptake of carbohydrates and uses a set of interconnected enzymes to couple translocation and phosphorylation of carbohydrates and the energy for this is provided by hydrolysis of phosphoenolpyruvate³².

A second method is secondary active transport: coupling the transport of the substrate up a concentration gradient to transport of another molecule down its concentration gradient, which is used in the case of co-transporters and antiporters, seen in **Figure 1.2C-D**. This relies on the initial generation of an electrochemical gradient for the second species, which is generally established by a transporter coupled directly to ATP hydrolysis, and so the transport is indirectly coupled to ATP hydrolysis. Generally, the energy source for secondary active transport is the electrochemical gradient of either protons or sodium ions³³.

Transporters generally have several states through which they move during the transport cycle, and because of this rates of translocation are lower than the rate for channels, typically around 1000 molecules per second but sometimes as low as one molecule per second³⁴. These changes in conformation involve alternately opening gates at either side of the membrane and it is vital that both gates are never both open at the same time, as this would, in effect, produce an open channel and thus a current in the opposite direction with several orders of magnitude higher than the active transport. The transport kinetics of transporters are again Michaelis-Menten meaning that increased substrate concentration leads to increased rate of transport only up to the point that the transporter is saturated and then the transport rate reaches a plateau. However, a cell can add more capacity by upregulating expression, if a higher rate of transfer is required.

An example of a transporter is the NRT1.1 from *Arabidopsis thaliana*, belonging to the Major Facilitator Superfamily (MFS) of proton-driven antiporters. This protein provides nitrate uptake from the soil and has two modes, a low-affinity mode that is used in times when nitrate is plentiful and a high-affinity mode that is used in low-nitrate conditions, allowing for rapid response to changing nitrate levels³⁵. A combination of crystallographic and binding studies suggested that the switch is driven by phosphorylation, which drives the protein into a more-flexible conformation, allowing for faster transfer rates³⁵. Additionally, binding of the herbicide chlorate was seen, an important commercial implication in addition to the fundamental structural biology.

1.2 ArAE Family of Acid Exporters (including plant ALMTs)

The Aromatic Acid Exporter (ArAE) family of membrane proteins is found throughout plants, and bacteria, and also in yeast and protozoa and were first identified in 2000³⁶. The most important members of this family so far characterised include the ALMT family of anion channels found ubiquitously in plants³⁷ as well as bacterial members, which were initially described as Putative Efflux Transporters (PETs)³⁸. Since their identification, the Transporter Classification Database³⁹ has classified these proteins as “2.A.85 The Aromatic Acid Exporter (ArAE) Family”. The plant ALMTs are a family of ion channels, predicted to contain six TM helices followed by a large soluble domain, while the bacterial members are predicted to contain an internal repeat, and thus be comprised of two sequential sections of six TM helices each followed by soluble domains (shown in **Figure 1.3**). As such, plant ALMTs can perhaps be considered ‘half transporter’ versions of bacterial PETs and it has been postulated that if bacterial PETs act as monomers then ALMTs are likely to form dimers³⁶. A similar situation can be found in the SWEET transporters. Plant members of this family contain a pair of 3-TM helices in their structure, joined by a linker, while bacterial members (so-called semiSWEETs) contain a single 3-TM bundle and require homo-dimerization^{40,41}.

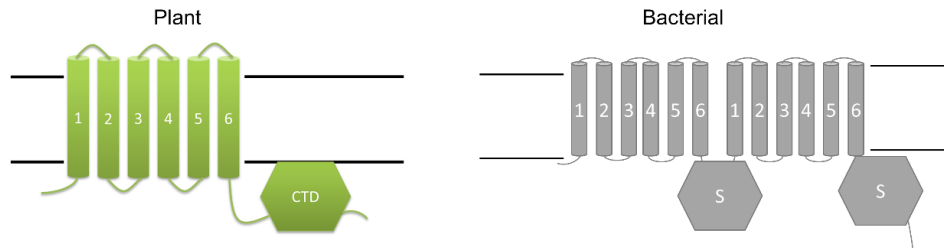


Figure 1.3 - Predicted overall architecture of ArAE family members from plants and bacteria.

The plant ArAE family members (ALMTs) were predicted to contain six TM helices and a soluble C-terminal domain (CTD), while the bacterial genes were predicted to be formed of a duplication with two sets of six TM helices each followed by a soluble domain (S).

Specifically, the *Arabidopsis* ALMTs that were identified in the initial paper concerning PETs were: ALMT6 and 12, whose functions are described in the following section, and ALMT5, 7, 10, and 11, whose functions have not yet been studied. The ALMTs have been more widely characterised, but remain poorly understood at a structural level. The bacterial PETs, similarly to ALMTs, have not been well studied and no structural studies have been performed. One bacterial family members has been characterised as an aromatic carboxylic acid efflux pump³⁸, and a second is identified as a sulfa-drug efflux pump⁴² but otherwise the bacterial family members remain poorly studied.

1.3 ALMT Family Members in Plants, Function, Topology, Mechanism, and Structure

The first protein family studied here is the **Aluminium activated Malate Transporter (ALMT)** family of ion channels, found ubiquitously within plants. Family members have been shown to have a wide variety of vital roles in the plant (reviewed in detail below) and there is a growing appreciation for their central role in many processes, but structural and mechanistic data are scarce and often conflicting. Thus, the ALMT family represents a high value but as-yet under-studied family of plant membrane proteins that are in particular need of high-quality structural data. The following section will review the physiological importance of ALMTs, and give an overview of the current state of biochemical, topological, and structural data for ALMTs.

The ALMT family is found ubiquitously in genomes throughout the plant kingdom and was named after the first member of the family to be characterised (*TaALMT1*) was found to be involved in aluminium resistance in wheat, *Triticum aestivum*⁴³. The first characterised homologue in *Arabidopsis thaliana*, *AtALMT1*, was similarly found to be involved in Al-resistance⁴⁴. However, the gene family in this species contains 13 other members, some of which have already shown to have a wide range of other roles³⁷. Indeed, the fact that species have so many homologues was the first indication that the ALMT family was of high importance – for example poplar contains 22 members⁴⁵, rice has nine members⁴⁵, and soybean has 31 members⁴⁶. To date, the ALMT family has been shown be central to processes such as control of stomatal aperture^{47,48} and anion homeostasis⁴⁹. Furthermore, increasing attention is

being paid due to their potential role in economically valuable traits such as fruit flavour⁵⁰ and grain filling⁵¹, and more recently they have been shown to be key mediators of GABA signalling⁵² and in supplying carbon to symbiotic soil bacteria⁴⁶. Identification of the molecular actors in these processes is already helping guide marker-assisted breeding⁵³.

1.3.1 Food security

The issue of food security is of current importance as the world faces what the UN describes as “an unprecedented convergence of pressures”⁵⁴. Namely, a growing world population, with demand for food expected to double by 2050; pressures for alternative land use including production of biofuels and urbanisation; a changing climate with varying water availability; and declining stocks of available resources such as oil and phosphate rock⁵⁵. A further constraint on yields is the presence of acid soils and the aluminium toxicity provoked by these. As ALMTs have emerged as key molecular actors in vital processes linked to this and other aspects of plant physiology, their importance to potential food security issues and GM crops has grown, including current research that is examining the potential for transgenic crops to combat acid soils⁵⁶. A brief survey of their importance to the trait of Aluminium Resistance will now be given as an example.

1.3.2 Aluminium Resistance

1.3.2.1 Overcoming aluminium toxicity on acid soils

Acid soils are prevalent worldwide, comprising around half of all potentially arable land⁵⁷. In these soils, aluminium ions become solubilised and damage crops via root growth inhibition⁵⁸ and, to compound the problem, nutrients such as phosphate become less available⁵⁹. Several plant species have been identified as aluminium-resistant and they use a variety of mechanisms including thickening of cell walls⁶⁰, active transport of aluminium away from sensitive organs⁶¹, or, more prominently, organic acid exudation, chiefly either by release of malate, citrate, or oxalate^{62,63}, which will now be discussed in the following section.

1.3.2.2 TaALMT1 in Aluminium Resistance

Studies on crosses of near isogenic Al-resistant and Al-sensitive wheat (*Triticum aestivum*) cultivars identified a gene responsible for malate exudation from root tips, providing the primary mechanism of aluminium resistance. Subsequent electrophysiological studies using *Xenopus* oocytes^{43,64,65} first characterised TaALMT1 as a malate channel activated by Al. TaALMT1 is constitutively expressed in the root apices of Al-resistant wheat and malate chelates Al³⁺ forming a 2:2 complex with the trivalent aluminium ions (seen in **Figure 1.4**), thus encasing the ions and rendering them non-toxic⁶⁶. This enables longer root growth and greater yields compared to a sensitive cultivar grown on acid soils. In addition to protection from Al³⁺ toxicity, malate extrusion has the benefit of increasing phosphate availability in the soil – since Al³⁺ binds and complexes phosphate⁶⁷. This is part of a host of processes activated in plants for improved phosphorus usage^{59,68}. Heterologous expression of TaALMT1 in cultured tobacco cells, *Xenopus* oocytes, and transgenic rice

plants has shown efflux of malate activated by the presence of Al^{3+} and expression confers Al-resistance to tobacco cells⁴³. Importantly, transgenic expression of *TaALMT1* in Al-sensitive barley plants rendered them resistant to aluminium toxicity⁶⁹. The transgenic plants grown on acid soil displayed root growth similar to that seen in neutral soils and a doubling in yield when expressing just the single gene⁶⁸, making it a powerful tool for transgenic crop development.

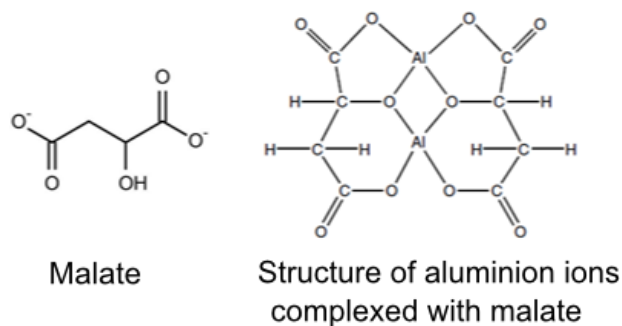


Figure 1.4 - Structure of malate and aluminium complexed by malate. Structure of the complex from Tashiro *et al.*⁶⁶.

1.3.2.3 Other ALMTs involved in aluminium-resistance

Since the characterisation of *TaALMT1*, several ALMTs from other species including oilseed rape, *Arabidopsis*, rye, and soybean have been shown to be vital for Al-resistance and characterised (see **Table 1** and **Figure 1.5**). These channels are activated by Al^{3+} and in some cases their expression is also upregulated upon sensing Al^{3+} ⁷⁰.

AtALMT1 is a malate channel critical for Al-resistance in *Arabidopsis*, expressed at the plasma membrane of the epidermal cells of the root tip⁴⁴. This protein shows 41% sequence identity and 63% similarity to *TaALMT1*. However, in contrast, *AtALMT1* is not constitutively expressed; instead, expression is

upregulated by aluminium^{70,71}, mediated by the transcription factors STOP1, STOP2 and WRKY46, which also regulate other genes critical for tolerance of acid soils⁷²⁻⁷⁴. A recent study showed that a chimeric protein made from the N-terminal half of TaALMT1 and the C-terminal half of AtALMT1 was functional, and actually showed higher malate export than either of the native channels⁷⁵. Their study also noted that there were significant differences between expression of ALMTs in *Xenopus oocytes*⁷⁶ as compared to in tobacco BY-2 cells and cautioned that heterologous expression in animal cells may not be a faithful representation of reality⁷⁵.

In addition, two genes with 95% sequence identity to one another (*BnALMT1*, *BnALMT2*) were identified in oilseed rape (*Brassica napus*)^{77,78}, as well as *GmALMT1* in soybean (*Glycine max*), and *ScALMT1* in rye (*Secale cereale*)^{79,80}. These genes are expressed at root tip plasma membranes, and the corresponding proteins are permeable to malate and are activated by aluminium, showing a widespread mechanism for Al-resistance among plant species. In addition, *GmALMT1* channel activity has also been shown to be regulated by pH changes and phosphorous concentration⁸¹. It is expected that as more species are studied more Al-responsive genes will be found.

1.3.3 Structural and topological studies of TaALMT1 and AtALMT1

Although a sizable and growing number of ALMTs have been physiologically and functionally characterised (as discussed throughout this section), much less is understood about their structure and mechanism, with even aspects of their basic topology remaining controversial. AtALMT1 and TaALMT1 remain the most comprehensively characterised of ALMTs, with studies covering mutagenesis of individual residues, deletion of various sections, swapping of

domains, fusions with YFP, and pharmacological approaches taken. A discussion of these studies will now be given.

1.3.3.1 C-terminal domain

The C-terminal domain (CTD) is the most studied region of ALMT1 and is known to be involved in regulation of ALMTs and truncation of it in TaALMT1 and AtALMT1 abolishes both basal and Al³⁺-activated anion currents, indicating this domain is vital for channel function⁸². It has been speculated that the CTD may be involved in oligomerisation vital for channel function; and it is likely be involved in gating the channel, although this requires further study to confirm. The CTD contains conserved motifs including a sequence “**F-(X)₄-W-(X)₂-E-X-L**” at the very start (residues 213-223 in TaALMT1), which has been identified to be crucial for GABA-binding (discussed in more detail below); and the “WEP” motif (residues 282-284 in TaALMT1) whose function has not been established, but is likely to be important in the formation of tertiary structure. In addition, the T323A mutant has significantly higher basal and aluminium-activated activity⁸². Threonine, uniquely, has a secondary βC making it conformationally restricted and it was speculated that perhaps the increased flexibility caused by mutation leads to this enhancement in malate currents. Discovering the structural basis for this will be important in enabling future genetic engineering to optimise the development of transgenic species.

1.3.3.2 Al-binding

Three specific acidic residues (E274, D275 and E284) in the CTD of TaALMT1 were shown by mutagenesis to be essential for channel activation by Al³⁺⁸³. The authors speculated that Al³⁺ binds directly to this domain via these three residues, causing a conformational change that leads to channel activation. However, later studies have cast significant doubt on this hypothesis. The final glutamate of the proposed binding site (E284) is present in another homologue, AtALMT12, as E276. Mutagenesis of this residue in AtALMT12 led to a loss of channel function, and as AtALMT12 is not Al-activated this suggests that this residue is unlikely to have a specific role in Al-sensing in TaALMT1 either. Indeed, a later study involving a comprehensive mutagenesis of 43 acidic residues within TaALMT1 showed that it was possible to generate an identical phenotype by mutagenesis at many different locations including residues predicted to be within the transmembrane region and located at either side of the membrane, and even mutagenesis of neutral residues can produce a protein that lacks Al-responsiveness⁸⁴. This suggests that the change in activity that is brought about by mutagenesis, rather than being a specific diminishment in the binding of Al, is likely to cause a more fundamental structural change. Indeed, the final Glu identified in the original study is part of the highly conserved WEP motif that is present throughout all ALMTs, including those that show no activation by Al it is therefore implausible for this residue to be vital to the specific activation by Al.

A second potential binding motif was identified near the N-terminus. It was discovered that Al-responsive homologues contain a motif with the sequence [D,E]-[H,K,R]-X-[K,R]-[D,E]-X-X-X-[D,E] (residues 4-12 in TaALMT1) that is not present in those homologues that do not interact with Al and mutation of

individual residues led to a loss of AI-responsiveness⁸⁴. However, when the entire region was deleted in another study AI activation was only reduced rather than removed, showing this region is not essential⁷⁵. The binding site for AI remains uncharacterised, and it is likely that AI-activation requires several regions of the protein in both the N- and C-terminal halves.

1.3.3.3 Phosphorylation

Kinase inhibitors have been shown to inhibit the activation of *TaALMT1* by Al^{3+} and mutation of putative phosphorylation site S384 within the CTD has been shown to abolish transport by the channel when expressed in *Xenopus* oocytes for both basal and activated currents suggesting that phosphorylation is a prerequisite for any transport⁸². Depending on what topology (discussed below) is chosen, this residue is alternately placed either inside or outside the cell, prompting speculation about how this could be phosphorylated if it were indeed extracellular. Were the topology of ALMT1 shown to place this residue outside of the cell, phosphorylation could possibly be mediated by cell wall associated kinase WAK1, which has been shown to be upregulated by Al^{3+} treatment in *Arabidopsis*⁸⁵, although this would not account for activity of *TaALMT1* in *Xenopus*, which necessarily do not contain such a kinase. Conversely, mutation of the three serine residues that constitute putative phosphorylation sites in a homologue *AtALMT12* does not affect transport, suggesting that phosphorylation is not universally required for ALMT activity⁸⁶.

1.3.4 Other roles of ALMTs

As mentioned above, ALMTs are a multigenic family and have many diverse roles in addition to acid soil tolerance. A review of the current understanding of these roles beyond Al-tolerance, and the structural and mechanistic basis of channel activity will now be given.

1.3.4.1 Other root related functions

A large number of ALMTs have been shown to be root-localised and Al-responsive, but other homologues have been linked with wider roles, even within the root. *ZmALMT2* from maize (*Zea mays*) has been shown to be root localised and to release malate into the soil. Unlike *AtALMT1* and *TaALMT1*, however, this is not correlated with Al-resistance, but instead is likely to provide solubilisation for soil nutrients, such as phosphate, as discussed above for *TaALMT1*. As it is also found in vascular tissue, *ZmALMT2* could also play a role in the transport of organic acids or mineral anions in the xylem.⁸⁷

Another recently-revealed role of ALMTs is in providing carbon to symbionts such as rhizobia that form root nodules. *LjALMT4* is expressed at the parenchyma cells at root vascular bundles and releases dicarboxylic acids such as malate, fumarate, and succinate plus other inorganic anions which are then available for the bacteria to take up in exchange of fixed nitrogen⁴⁶. This finding raises the possibility that ALMTs may be responsible for other nutrient exchange with other symbiotic organisms such as mycorrhiza, as the molecular actors in carbon-nitrogen exchange remain unidentified. Mycorrhizal associations are formed very widely with 80% of surveyed plant species having mycorrhizal associations⁸⁸ and thus ALMTs, which are found ubiquitously and have many homologues with as-yet unknown function, are plausible candidates.

Table 1 - Summary of known functions and localisation of ALMTs from several species

Gene	Organism	Localisation	Al-activated?	Function
AtALMT1	<i>Arabidopsis thaliana</i>	Root cell plasma membranes	Y	Aluminium resistance
AtALMT6	<i>Arabidopsis thaliana</i>	Tonoplast	N	Guard cell malate currents
AtALMT9	<i>Arabidopsis thaliana</i>	Tonoplast	N	Chloride currents in guard cells and malate homeostasis
AtALMT12	<i>Arabidopsis thaliana</i>	Plasma membrane of guard cells	N	Stomatal closing
TaALMT1	Wheat, <i>Triticum aestivum</i>	Root cell plasma membranes	Y	Aluminium resistance
VvALMT9	Grape, <i>Vitis vinifera</i>	Tonoplast of berry mesocarp	N	Fruit flavour/vacuolar malate uptake
MdMA1	Apple, <i>Malus domestica</i>	Tonoplast	N	Fruit flavour/vacuolar malate uptake
BnALMT1	Oilseed rape, <i>Brassica napus</i>	Root cell plasma membranes	Y	Aluminium resistance
BnALMT2	Oilseed rape, <i>Brassica napus</i>	Root cell plasma membranes	Y	Aluminium resistance
ScALMT1	Rye, <i>Secale cereale</i>	Root cell plasma membranes	Y	Aluminium resistance
GmALMT1	Soybean, <i>Glycine max</i>	Root cell plasma membranes	Y	Aluminium resistance
HvALMT1	Barley, <i>Hordeum vulgare</i>	Root cell plasma membranes and guard cells	N	Maintaining turgor in growing cells, guard cell movements
ZmALMT1	Maize, <i>Zea mays</i>	Plasma membranes throughout plant	N	Inorganic anion homeostasis
ZmALMT2	Maize, <i>Zea mays</i>	Root plasma membrane	N	Constitutive malate efflux, not Al-related
LJALMT4	Lotus, <i>Lotus japonicus</i>	Root nodules	N	Symbiosis with rhizobia

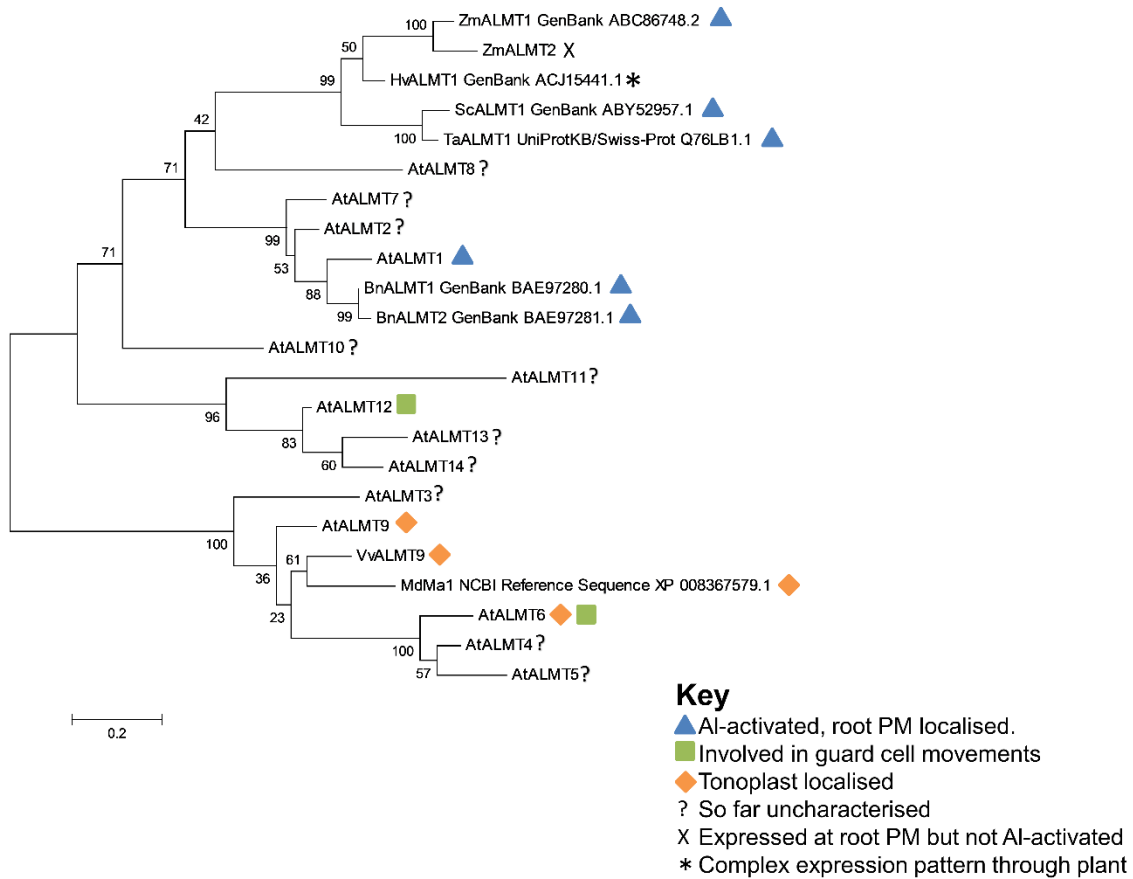


Figure 1.5 - Phylogenetic tree of ALMTs with known functions plus family members from *Arabidopsis* with unknown function.

Numbers on nodes denote likelihood of correctness. Tree made using MEGA6 software⁸⁹, sequences were aligned by the inbuilt MUSCLE functionality, and constructed with the Maximum Likelihood method and 500 iterations of bootstrapping. Taken from Palmer *et al.* (2016)

1.3.5 Guard Cell Movements

Another vital role played by some ALMT family members is found far away from the roots, within the leaves. Several homologues are molecular components of the guard cell movements that regulate gas exchange across leaf surfaces. Plants control CO₂ uptake and water loss by regulating the aperture of the stomatal pores. Three family members from *Arabidopsis*, *AtALMT6*, *AtALMT9*, and *AtALMT12*, are involved in opening and closing of stomata, with movement driven by osmotically active inorganic and organic ions⁹⁰. An overview of the process is seen in **Figure 1.6**. Upon stomatal opening, K⁺ enters the guard cell via voltage-gated inward rectifying potassium channels, driven by the electrochemical gradient maintained by ATP-driven proton pumping. Malate synthesis from stored starch provides a charge-balancing ion, and is taken up into the vacuole via *AtALMT6*, and similarly *AtALMT9* acts to permit entry of chloride counterions into the vacuole. The increase in solutes draws water into the cell down the water potential gradient, swelling the cell. The process is inverted during stomatal closing. The membrane is depolarised, prompted by the action of *AtALMT12/QUAC1* (Quick Anion Channel1) releasing malate rapidly, this allows K⁺ to flow out of the cell, accompanied by Cl⁻ and NO₃⁻ anions via *SLAC1* (Slow Anion Channel1). In addition, *AtALMT6* activity is regulated in part by the cytosolic malate concentration, and so malate can flow out of the vacuole to either be lost via *AtALMT12/QUAC1* or used in metabolism. This loss of osmotica drives a loss of water and closing of the stomata.

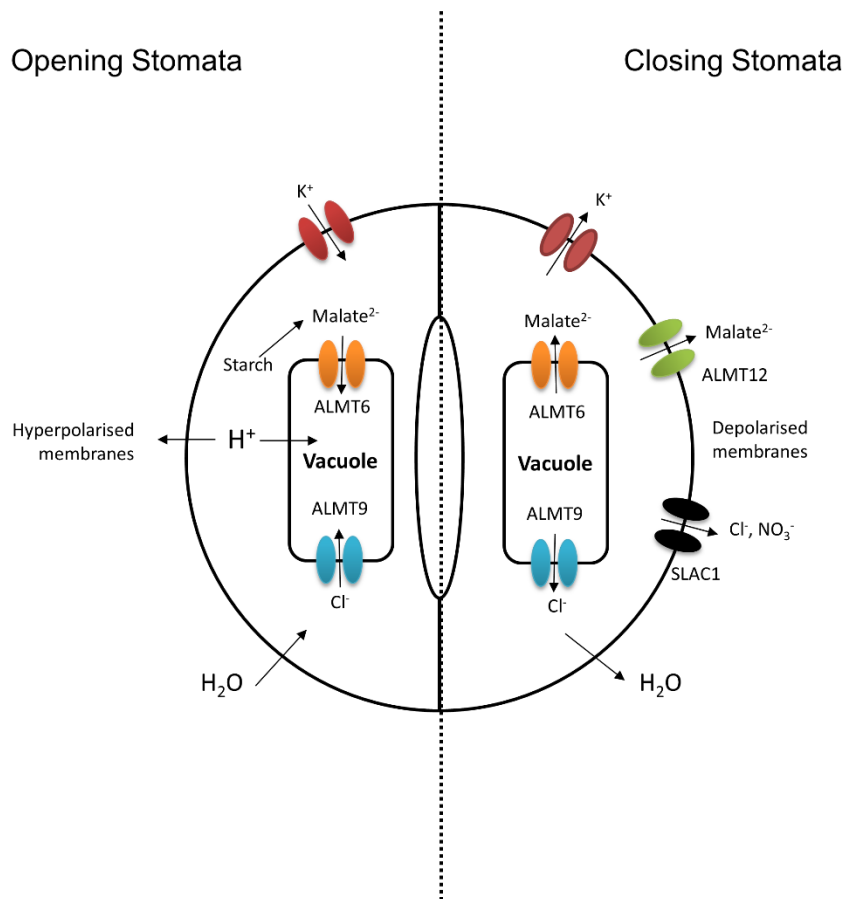


Figure 1.6 - Schematic of ion currents in stomatal opening and closing.

K^+ ions enter the cell during opening, and the charge is balanced by malate synthesis from starch, which is then taken up into the vacuole via ALMT6 and activates ALMT9 to permit Cl^- uptake into the vacuole. This increase in osmotic potential draws water into the cell, leading to swelling and stomatal opening. Stomatal closing, conversely, is driven by release of K^+ ions, which is permitted during membrane depolarisation and driven by ALMT12. Additionally, anions are released by SLAC1 over a longer period of time. Loss of osmolytes leads to concomitant loss of water and thus stomatal closing. Taken from Palmer *et al.* (2016)

AtALMT6 is expressed in guard cell vacuoles and is a malate channel, specific for divalent malate involved in stomatal movements. It is not aluminium activated but instead is controlled by light, ABA⁹¹, pH, and cytosolic malate concentration⁹². As transport is dependent on the concentration of malate in the cytosol and the tonoplast membrane potential, *AtALMT6* can mediate both malate uptake into and release from the vacuole in guard cells, with uptake

during stomatal opening, and release during stomatal closing as shown in **Figure 1.6**. Interestingly, in *Arabidopsis thaliana* expression is also seen in floral organs, suggesting another role yet to be elucidated. Knockout plants did not show phenotypic differences, indicating functional redundancy in vacuolar malate channels, perhaps from the action of *AtALMT9* and perhaps also *AtALMT5*, which has been shown to be expressed in guard cells, and to be closely related to *AtALMT6* as seen in the phylogenetic analysis in **Figure 1.5**.

AtALMT9 is a vacuolar chloride channel in guard cells and knockouts show impaired stomatal opening⁴⁸. It permits chloride to enter the vacuole, providing a charge-balance for K^+ in the same way as malate. Malate in the cytosol (synthesised from starch) activates the channel, meaning malate can act both as an osmolyte and as a signalling molecule in guard cells. Interestingly, kinetic data suggest *AtALMT9* is multimeric, with a number of subunits >2.5 with more recent work based on co-expressing WT and mutated *AtALMT9* suggesting that it forms a tetramer⁹³. Furthermore, cytosolic nucleotides such as ATP block the activity of *AtALMT9*, competing with malate for a binding site⁹⁴. This provides a level of regulation based on the activity of the H^+ -V-ATPases that maintain a hyperpolarised tonoplast. These consume ATP, provoking lowered cytosolic ATP concentrations, removing this block and thus permitting currents via *AtALMT9* and facilitating anion uptake into the vacuole. Mutagenesis studies identified the positively-charged Lys193, Arg200, and Arg215 in the fifth TM helix as being vital for channel function⁹³ but, similar to studies of *ALMT1*, the mechanistic relevance of these residues is unclear although they are speculated to be part of the permeation pathway, putatively formed by TM5. Indeed, *AtALMT9* is the only member whose N-terminal transmembrane domain

has been studied in detail by mutagenesis studies and a conserved pair of residues (Lys93 and Glu130), found in the first two helices were found to form a salt bridge vital for formation of a functional channel. The N-terminal domain of ALMTs are well conserved, with sometimes >50% identity, and many residues (such as the pair involved in a salt bridge) being strictly conserved and so findings such as this are likely to be widely applicable across all family members.

AtALMT12 (also known as *QUAC1*) is expressed in guard cell plasma membranes and operates as an R-type (Rapid-type) channel crucial for stomatal closing⁴⁷, allowing rapid malate release. One study also found localisation to endomembranes, although this is possibly an artefact of overexpression or evidence of regulation of the channel⁹⁵. The channel opens with fast kinetics upon membrane depolarisation, releasing malate into the apoplast in parallel with K^+ release through potassium channels to maintain the depolarisation. Moreover, both cytoplasmic malate and external malate causes increased activation of the channel and thus may represent a positive-feedback loop. Loss-of function mutants confer a wilted phenotype due to their impaired stomatal closing. Rather than being ligand-gated, as seen in *ALMT1*, *ALMT12* activity is voltage gated; however, the voltage sensor is yet to be identified, although it is likely in the CTD, which has been shown to be vital for regulation, but eleven deletion mutants missing sections of the CTD were all non-functional and so could not help locate the voltage sensing domain⁸⁶. A GFP fusion at the C-terminus also prevented voltage-gated deactivation of the channel, suggesting that the CTD is involved in closing the channel.

1.3.6 Other Roles

1.3.6.1 Malate storage and homeostasis

AtALMT9, in addition to its role in guard cells has been shown to be permeable to both chloride and malate and expressed strongly throughout leaf mesophyll tissue. It is proposed to have a role in homeostasis: ensuring that the concentration of malate – which plays an essential role in metabolism as part of the citric acid cycle – remains stable within the cytoplasm. Similar to *AtALMT6*, the channel is likely to work in both directions, by storing excess in the vacuole and releasing it when required to regulate osmotic potential and C-metabolism⁹⁶.

1.3.6.2 Cell elongation and nutrient storage

Barley (*Hordeum vulgare*) *HvALMT1* is a malate channel expressed in guard cells and in the root elongation zone, as well as in floral tissues and seeds⁹⁷. Although it has the greatest sequence similarity to *TaALMT1* in the barley genome and localises to the plasma membrane, it is not involved in Al-resistance but seems to have several distinct roles within the plant. *HvALMT1* over-expressing lines take longer to close their stomata⁹⁸ and RNAi knockouts show a similar wilted phenotype to *Atalmt12* knockouts⁵¹, so *HvALMT1* is proposed to be a functional homologue of *AtALMT12*, transporting malate from the guard cell. In expanding cells *HvALMT1* may help provide an osmotic balance and regulate turgor. Additionally, later studies have shown that this channel plays a role in seed development during acidification of the starchy endosperm, which is required for enzyme activity⁵¹. Rather than directly causing acidification itself (this is most plausibly caused by a H⁺-ATPase pumping protons from the surrounding aleurone layer) release of malate is suggested to act as a counterion for H⁺ and other positively charged nutrients such as K⁺,

helping to maintain electroneutrality and osmotic balance in a similar manner to the role of malate in guard cell movements. The significant difference in function between *HvALMT1* and *TaALMT1* despite strong sequence similarity again highlights that small differences in sequence can underlie large changes in function.

1.3.6.3 Fruit Flavour

Malic acid is an important component of apple taste, grape quality, and wine production: an economically-significant set of traits. A malic acid channel homologous to *AtALMT* genes has been shown to be responsible for the acidity of apples, and functions to accumulate malic acid in the vacuole: 90% of acid content of apples is malic acid.^{50,99} Similarly, grape berries contain an ALMT family member expressed in the tonoplast of berry mesocarp tissue responsible for malate and tartrate accumulation¹⁰⁰. Understanding of the action of these genes could be valuable for fruit and wine development⁹⁹. Interestingly, one allele has been found containing a premature stop codon within the CTD, causing a loss of function. This protein localises to the plasma membrane rather than tonoplast, suggesting that localisation is at least partly controlled by the CTD⁹⁹.

1.3.6.4 Inorganic Anion Homeostasis

ZmALMT1 from maize was one of the first family members to be described that did not have a role in aluminium resistance. It localises to plasma membranes throughout the plant, but is less permeable to organic anions and instead is likely involved in inorganic anion homeostasis and mineral nutrition⁴⁹.

1.3.6.5 GABA Signalling

A recently-discovered role of ALMTs is in mediating γ -aminobutyric acid (GABA) signalling in plants. This is the first evidence for GABA signalling in plants in addition to its established role as a metabolite within the TCA cycle⁵². GABA is a non-protein amino acid that accumulates in plant tissues in response to biotic and abiotic stresses, and has a central role in pollen tube growth and the regulation of root growth¹⁰¹. Recently, it has been shown that GABA's influence is exerted by interaction with ALMT proteins, and a putative GABA-binding motif has been identified in the CTD, falling just after the sixth transmembrane helix, on the extracellular face. GABA binding negatively regulates ion flux through the channel, i.e. decreasing carbon flux from roots in the case of *TaALMT1*.

Additionally, as plant anion equilibrium potentials are strongly positive and plant action potentials are generated by voltage gated ion channels, GABA inhibition of ALMTs will hyperpolarise membranes and decrease excitability. And as ALMTs can control membrane voltages, they indirectly control transport of other ions that rely on the membrane potential. A mechanism of long-distance electrical signalling in plants has been proposed to act via ALMTs. Malate and GABA regulate ALMT activity in opposite directions and can both be transported out of a cell to affect the activity of ALMTs in adjacent cells. This could mediate signalling in times of increased stress that produces ROS, as in these conditions 2-oxoglutarate is transported out of the mitochondria, and converted to GABA in the cytoplasm which is free to be released, or can otherwise re-enter the mitochondria to re-join the TCA cycle.

This new understanding of the role of GABA, like many other discoveries made relating to ALMTs raises even more questions. By what mechanism does GABA

regulate ALMTs? Where is the GABA binding site located in the ALMT structure, is it on the inside or outside of the cell? What implications does this have for ALMT topology? Are ALMTs functional multimers, and if so are these homomers-or heteromers? Mammalian GABA-receptors are heteromers with the GABA binding site found at the junction between proteins, is this reflected in ALMTs?

1.4 Bacterial members of the ArAE family (PETs)

The second sub-family of the AeAE family of interest within this study is the bacterial clade, seven members of which were initially identified as part of the PET family in 2000³⁶, alongside the plant ALMTs. The best studied of these bacterial members is AaeB, with the only other protein with an identified function being SdsQ. These two were first identified as part of a novel family of inner-membrane bacterial transporters, named Putative Efflux Transporters (PETs) in 2000³⁶, since classified as Aromatic Acid Exporters (ArAE). These two characterised proteins will be discussed below.

1.4.1 AaeA and AaeB – the aromatic acid exporter

AaeB is one of only two members of the PET family that has been characterised at all, albeit in a limited way, and its role has been described as a “metabolic relief valve”, involved in exporting aromatic acids such as p-hydroxybenzoic acid (pHBA) (shown in **Figure 1.7**), leading to the naming of it as Aromatic Acid Exporter B (AaeB)³⁸. In addition to this protective role, AaeB may also have a function in biofilm formation¹⁰². The gene is expressed when the cell is under some metabolic upset which makes it a target for biotechnological development as a sensor for build-up of toxic metabolites, for example in biofuel production^{103–105}.

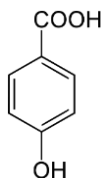


Figure 1.7 The primary substrate of AaeB, p-hydroxybenzoic acid.

AaeB is predicted to contain a duplicated repeat of six transmembrane helices, each followed by a soluble cytoplasmic domain, as previously mentioned, shown in **Figure 1.8B**. It is likely to form an inner-membrane efflux pump.

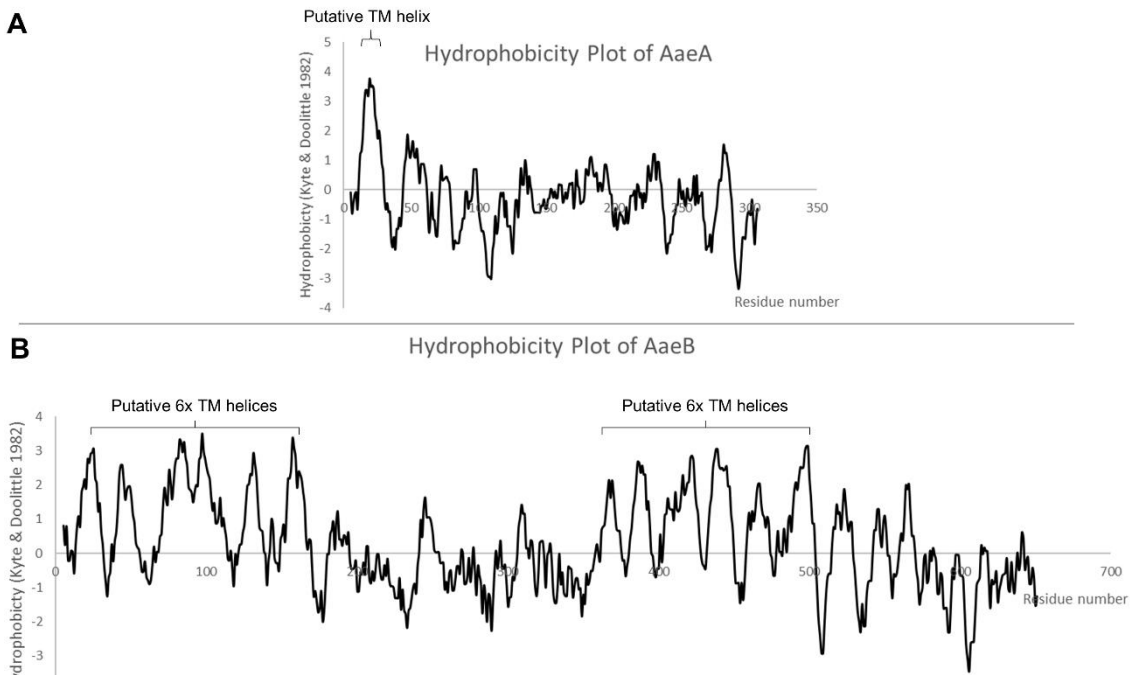


Figure 1.8 – Hydrophobicity plots for AaeA and AaeB.

A) hydrophobicity plot calculated by the Kyte and Doolittle method for AaeA. The predicted transmembrane helix is annotated. **B)** Hydrophobicity plot for AaeB. The two transmembrane domains are annotated.

Also expressed from the same operon as AaeB is a second protein, designated AaeA. From its sequence this is predicted to be part of the Membrane Fusion Protein (MFP) class, also called the Periplasmic Adaptor Proteins (PAPs)¹⁰⁶. Several of these (which associate with other IM transporters from families such as the RND or MFS families) have had crystal structures solved. These MFP/PAPs have a conserved overall architecture, shown in **Figure 1.9** with typically four domains, which extend out from the inner membrane: membrane

proximal domain (MPD), the β -barrel domain, the lipoyl domain, and the α -hairpin domain¹⁰⁷. They are anchored into the inner membrane either by a single hydrophobic transmembrane helix, indicated in **Figure 1.8A**, or by a lipid anchor.

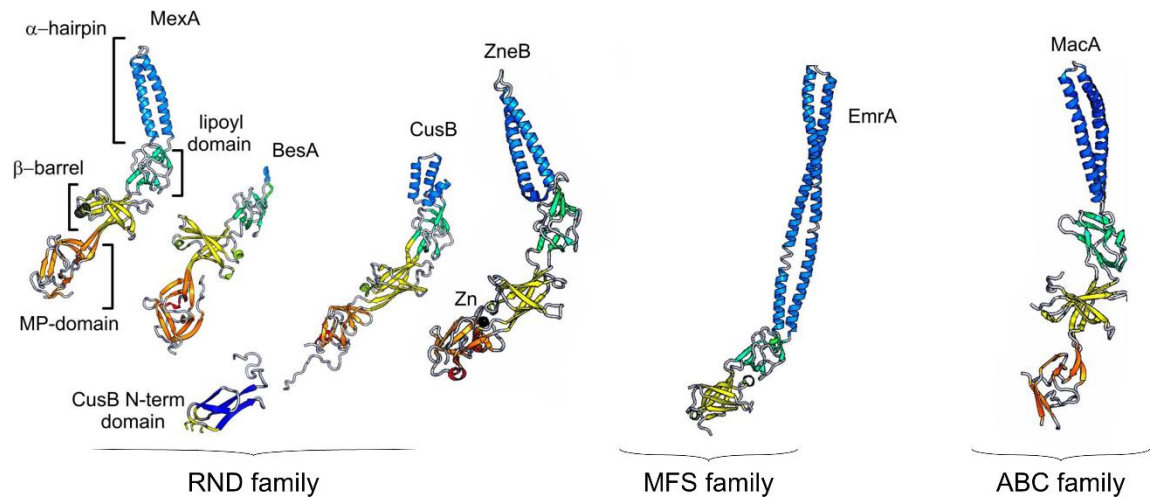


Figure 1.9 - Schematic representation of a range of MFP crystal structures.

Structures of MFPs that form complexes with three families of exporters, Resistance-Nodulation-Division (**RND**) Major Facilitator Superfamily (**MFS**), and ATP-Binding Cassette (**ABC**) are shown. The four canonical domains are: the membrane proximal domain (**MPD**), coloured orange; the **β -barrel** domain, coloured yellow; the **lipoyl** domain, coloured green; and the **α -hairpin** domain, coloured blue. BesA lacks the α -hairpin, while EmrA lacks a MPD, and CusB has an additional N-terminal domain. Figure adapted from Symmons *et al.* (2015)¹⁰⁷

These proteins form tripartite efflux pumps alongside inner membrane transporters (such as AaeB) and outer membrane channels (such as TolC) helping to bridge the gap between the membranes in Gram-negative bacteria such as *E. coli*¹⁰⁸, although MFP proteins have been identified in Gram-positive bacteria, which do not have an outer membrane, suggesting that these proteins have a wider role¹⁰⁹. Interestingly, although previous studies had shown that

both AaeA and AaeB were required to provide physiological protection against aromatic acid toxicity in *E. coli*, knockouts of TolC had only a minimal effect, suggesting that AaeA and AaeB do not form a complex with TolC in the manner of tripartite systems such as AcrA-AcrB-TolC¹¹⁰. This could also be because the substrate for transport in this case is a normal metabolic product that is a useful carbon source for the cell, and so export via the AaeB inner membrane transporter into the periplasm may leave the pHBA available for re-uptake and later use, rendering any export entirely out of the cell past the outer membrane unnecessary. Alternatively, a complex may be formed with another outer membrane protein such as the one found in the sdsQRP operon.

1.5 MdtO/SdsQ - the Sulfa-drug Exporter

The second bacterial ArAE family member identified in *E. coli* that has been studied to a degree is MdtO/SdsQ (sulfa drug sensitivity determinant), which was identified as a target of the LeuO transcription factor⁴². Expression of LeuO was shown to increase expression of the sdsRQP operon and as such increase resistance to sulfa drugs, while knockouts of the sdsRQP operon were more sensitive to sulfa drugs even in the presence of LeuO expression, suggesting that SdsRQP is a sulfa drug efflux pump⁴². Sulfa drugs have similar structural characteristics – such as a polar functional group attached to an aromatic ring – to pHBA, the substrate of AaeAB, which again shows that members of the bacterial PET family (or AeAE family) have similar aromatic substrates. Although the proteins of this operon have been shown to be involved in protection against sulfa drugs no further characterisation or isolation of the individual proteins has taken place, meaning that, like for AaeAB, there is room for much more exploration of these proteins and the wider family. Interestingly,

the operon contains proteins that are predicted to be an inner membrane transporter (SdsQ), a membrane fusion protein (SdsR), and an outer membrane channel (SdsP), analogous to bacterial tripartite efflux systems¹⁰⁸.

1.6 Challenges in membrane protein expression and structural studies

Membrane proteins represent around 30% of genes in organisms ¹¹¹, and over 50% of drug targets ¹¹². Thus, they form a class of high-priority targets for structural studies. Yet, despite this importance, less than 2% of published structures in the PDB are membrane proteins, although this number is growing as seen in **Figure 1.10**. This disparity is due to difficulties in expression and solubilisation of these proteins. Indeed, a minor but not inconsequential exemplar of the problems encountered in studies of membrane proteins is that they often run as a much more mobile band on SDS-PAGE than would be expected from their molecular weight, sometimes at a size as little as 65% of the expected^{113–115}. This complicates the analysis and identification of membrane proteins but is not a fatal flaw. However, there are many more significant challenges to overcome, which will be discussed in this section.

Membrane proteins are very challenging targets as a large proportion of their surface area is hydrophobic, and they are covered with membrane lipids in the native state; conversely, most biochemical and structural assays are carried out in aqueous solution, and consequently detergents, or other solubilisation methods, are required to solubilise and maintain the function of membrane proteins in solution. However, these detergents can present problems of their own. A brief survey of the major challenges in any studies of membrane proteins, expression, solubilisation, purification, and of the challenges specific to structural biology is described in the following section.

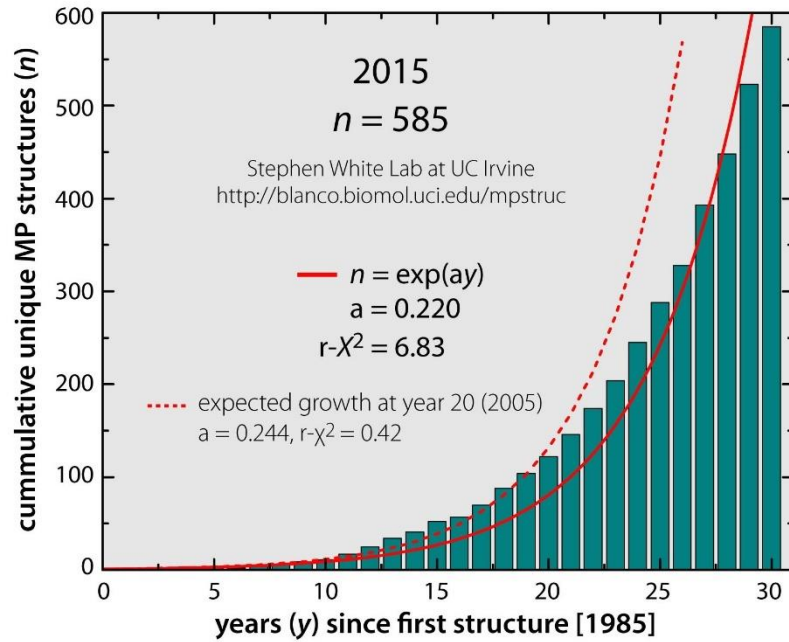


Figure 1.10 – Cumulative number of membrane protein structures solved, showing an exponential increase over time.

List maintained by the White lab at UC Irvine. Figure taken from <http://blanco.biomol.uci.edu/mpstruc/>. 620 unique proteins, from ~2000 coordinate files as of 05/06/2016

1.6.1 Expression

A primary problem when working with many membrane proteins is obtaining sufficient quantity of the protein of interest¹¹⁶. Membrane proteins are often found only at low levels in their native host, as they are localised within membranes, which comprise only a small total proportion of the cell, less than 0.1% (whether the cell membrane, or particular endomembranes)¹¹⁷. Although some systems can be obtained from the native host without overexpression, for example, the mitochondrial ATPase and *bc1* complex¹¹⁸ these systems are not the norm. As a result, heterologous overexpression can be used to generate a sufficient amount of material for biochemical and structural studies, which often require several milligrams of the protein of interest¹¹⁹. However, heterologous

expression can lead to aggregation of the protein and formation of inclusion bodies in the cytoplasm, and, depending on the target and expression system chosen, the expression host may lack important machinery for folding and post-translational modification required to produce fully-functional membrane protein^{120,121}. Thus, optimisation of expression is crucial in order to obtain sufficient material in an active state, and this is aided by the development high-throughput methodologies for heterologous expression which allow for screening of a large number of conditions on a small scale, rapidly¹²². However, at present, it is not possible to predict what conditions will prove successful at the outset of an investigation. In a study of three closely related proteins, OCTN1, 2, and 3, protein expression was optimised in order to produce the three proteins for structural studies but the best conditions were not the same for each of the three proteins despite their close evolutionary relationship and structure, including expression vector, fusion tag, codon optimisation, expression strain, growth medium, and duration of expression¹¹⁵. Similarly, when expressing several inositol polyphosphate sensing SPX-domains (which are soluble in isolation, but often part of membrane proteins) from a variety of eukaryotic proteins, the study authors screened 45 combinations of protein source organism, truncation, mutation, type and location of tag, and found that less than half of the conditions expressed well in *E. coli* and crystals were obtained from only three of the constructs¹²³. There are no straightforward criteria for improving the yield proteins when expressing them in bacteria, even for closely related proteins.

Choice of expression host is certainly a key factor for the success of protein expression¹²⁴. While bacterial expression is attractive due to its low cost, high

scalability, and ease of use, eukaryotic membrane protein targets are rarely possible to express in the many common *E. coli* expression strains¹²⁵. Bacterial expression hosts often have difficulty with several elements of expression. For example, codon bias, caused by factors such as *E. coli* rarely using certain codons, such as AUA for Ile, resulting in a shortage of the corresponding tRNA, stalling expression¹²⁶; or, problems with folding long complex sections of protein, which is difficult for bacterial machinery to process, followed by insertion into the membrane, which can overload the bacterial Sec translocon, with toxic effects¹²⁷. There are a variety of different tools available, such as modified strains of *E. coli*, to help overcome these problems, many of which are commercially available and several of which are surveyed by Indiveri *et al.*¹¹⁵.

If an investigator plans to use a eukaryotic expression system instead, several options exist. Established eukaryotic expression hosts available include, in order of increasing expense and similarity to human systems, the yeasts¹²⁸ *S. cerevisiae* and *P. pastoris*, the insect cell line Sf9¹²⁹, and various mammalian immortalised cell lines¹³⁰. There have even been recent developments in plant-based expression systems which have seen success in creating vaccines, enzymes and antibodies^{131,132}, and cell-free systems are available that can produce protein *in vitro*^{133,134}. Often, choosing a system for expression requires trade-offs to be made: although human-derived immortalised cells may be the closest option for producing a human membrane protein in a near-native environment, the cost may be prohibitively high, especially when considering the large cultures required to produce sufficient purified protein for X-ray crystallography as eukaryotic systems have, with few exceptions, much lower expression yields when compared to *E. coli*¹¹⁵. In fact, data from the PDB

shows that the second most common expression host (after *E. coli*) to successfully produce a protein structure is the insect cell line, *Spodoptera frugiperda*, which perhaps strikes a balance between cost, ease of use, and complexity of internal machinery¹³⁵.

A significant challenge for many proteins is the problem of toxicity.

Overexpression of a protein of interest may cause significant negative effects in the host cell, inhibiting protein synthesis of endogenous proteins through competition and potentially causing the cell to die as a result. This has the obvious side effect of reducing yields of heterologously expressed protein.

Alternatively, a protein could be toxic due to unregulated biological activity. For example, overexpression of an ion channel, which rests in the “open state”

could result in the host cell becoming ‘leaky’ as ions flow through the channels and destroy vital electrochemical gradients. As such, expression systems are often designed to prevent expression until it is desired. For example,

biotechnological bacterial expression systems are frequently based on the lac operon. The LacI repressor gene is expressed, which prevents any expression of the target protein until the addition of IPTG. This allows for careful control of induction of expression¹³⁶. A related system is the method of autoinduction¹³⁷.

This again takes advantage of the Lac repressor, by growing the cells in media containing three sugars. Glucose in the media provides a carbon source and allows the bacteria to grow to high density, while also ensuring that the lac-operon based repression is held on; then as glucose is used up (at the point that the cells have reach high density) the lactose within the media induces expression of the target gene; and during this expression period the glycerol within the media provides a carbon source as an alternative to lactose to ensure

it is not used up¹³⁷. By this method expression is prevented while the cells are grown to high density, then expression is induced and even if the target protein is toxic to the cells and they cannot express the protein to a high amount per cell, they are at such a high density that the yield per litre of culture is still significant¹¹⁶.

Sometimes when expression in *E. coli* produces a high yield the protein is expressed in inclusion bodies, which sequester the target protein in insoluble aggregates. One approach to solve this problem is refolding the inclusion bodies using a denaturant such as urea, before removing the denaturant and allowing the protein to re-fold into its correct conformation^{138,139}. Several helical membrane proteins have been refolded from inclusion bodies¹⁴⁰ including Light Harvesting Complex II¹⁴¹ and Diacylglycerol Kinase¹⁴². This can also be achieved with β -barrel proteins, the outer membrane protein precursor proOmpA, for example, can be produced from inclusion bodies solubilised in 6M urea before using in downstream experiments¹⁴³.

The use of protein chimeras has also been developed to enhance expression. Often the study of protein complexes can be enhanced by the production of chimeric constructs to increase complex stability. The Kv1.2 K⁺ channel was solved with a chimeric paddle domain transferred from Kv2.1¹⁴⁴. Similarly, when studying the structure of the tripartite AcrA-AcrB-TolC pump that spans from the inner to the outer membrane of *E. coli*, the proteins were expressed and fused with poly-glycine linkers in order to produce the first structure of the full complex and its quaternary organization¹¹⁰. Fusion of proteins that are stable but not directly related to the target protein such as a Maltose Binding Protein (MBP)^{145,146}, glutathione S-transferase (GST)¹⁴⁷, and thioredoxin (Trx)¹⁴⁸ can

also be used. MBP is particularly appropriate as it is from a membrane protein system, and forms part of the maltose/maltodextrin uptake system from *E. coli*¹⁴⁹. This will often generate a chimeric protein product in a greater yield that is more stable and soluble, and more amenable to study^{150,151}, and can often create a larger area of crystal contacts. Over 100 structures in the PDB have been produced using MBP-fusion proteins alone, with the number growing every year.

There are many other possible techniques, including expression of heat shock proteins and chaperones^{152,153}, use of engineered *E. coli* strains^{154,155}, and lowering the temperature of expression¹⁵⁶ among many more, and these techniques are used in combinations too numerous to review in totality, and indeed no approach is a guarantee of success¹⁵⁷.

1.6.2 Structural Studies of Bacterial Proteins to Understand Eukaryotic Targets

While prokaryotic proteins are important and deserving of study in their own right, they can also prove useful for the study of eukaryotic protein targets, especially given the intractability of many eukaryotic membrane proteins. Ideally, all studies of eukaryotic proteins would be done on the protein in question, but as discussed above, this is not always possible. As such, it is often easier to identify a homologue of the protein of interest from a bacterial species and then use that for structural experiments as they are usually more experimentally amenable and can be more frequently expressed to the high levels required for structural studies¹⁵⁸, taking advantage of the ability to grow *E. coli* on a large scale (for example in a 30-100 L fermentation vessel).

Although bacterial membrane proteins can differ significantly in primary sequence to their eukaryotic homologues, structural motifs are generally much more strongly conserved¹⁵⁹ and so there are often functional properties in common between the two proteins and the three dimensional structural information obtained from such experiments can inform modelling of the structure of the eukaryotic protein. This has been done, for example with human eukaryotic initiation factor 5A, based on two crystal structures of bacterial proteins 32-34% identical to the human protein¹⁶⁰, and understanding of bacteriorhodopsin has informed the study of rhodopsin-like proteins¹⁶¹. A recent study detailed a method for generating high-resolution structural models of alpha-helical membrane proteins from distant homologues¹⁶². Such models can help generate hypotheses, and guide the design and interpretation of experiments. For example, information about residues involved in transport in one protein can guide mutagenesis studies to investigate the same in another protein, and these experiments will drive an iterative cycle of testing and re-modelling.

An example of such an approach is seen the study of NupG from *E. coli*, a bacterial homologue of nucleotide transporters found in humans which are important for cancer chemotherapy, neurotransmission, cardiac activity, and viral infection^{163,164}. The study of the bacterial protein opened the way to elucidating the molecular mechanisms involved in transport for this family of membrane transporters³⁰. Molecular modelling has since been applied to nucleoside transporter proteins¹⁶⁵ based on a related crystal structure from *Vibrio cholerae*¹⁶⁶ alongside the previously acquired data for the NupG secondary structure. More recently, a protocol for protein production,

solubilisation, and purification was developed further to enable greater variety in affinity tagging and applied to a range of bacterial proteins with homology to human proteins¹⁶⁷.

Often a strategy used is to find a homologous protein within a thermophilic species such as *Thermus thermophilus*¹⁶⁸. Proteins from thermophiles are prime targets mainly because the enzymes and macromolecular complexes of thermophiles are more easily purified, manipulated, and more stable during the long time periods needed for crystallization and this has proved to be a successful route to the study of protein structures and 21% of the *T. thermophilus* proteome is now represented by a solved crystal structure¹⁶⁸. Of course, this approach is only appropriate if a closely-related protein can be found within the *T. thermophilus* genome.

1.6.3 Solubilisation

Once a sufficient level of expression has been achieved, then scaled up¹⁶⁹, and the cellular membranes have been isolated either by centrifugation¹¹³ or tangential flow filtration¹⁷⁰, a method must be found to extract the membrane protein from the lipidic environment, while maintaining the intact folded state and solubility¹⁷¹. This is made more complicated by the fact that in many instances the membrane lipids that naturally interact with the protein are vital for stability, and may play a role in protein activity such as is seen in the case of TRPV1, which has bound lipids that are vital for its mechanism¹⁷². There are several potential strategies and many possible reagents including the use of detergents¹⁷³, nanodiscs¹⁷⁴, SMALPs¹⁷⁵, and amphipols¹⁷⁶, so finding the optimal approach often can be an expensive process of trial and error – often

using significant amount of precious overexpressed protein. This process has been aided by the development of high-throughput screens to allow faster identification of conditions, using less material^{177–179}. One such method uses the dye 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM), which binds to cysteines that become exposed as the protein is heated to measure thermal unfolding alongside light scattering¹⁸⁰.

The most established method for solubilisation is the use of detergents. These amphipathic molecules can solubilise the lipid bilayer and coat the hydrophobic surfaces of membrane proteins, rendering them stable in aqueous solution¹⁷³. Each detergent has a Critical Micellar Concentration (CMC), dependant on the structure of the molecule, such as alkyl chain length and the nature of the polar head group and the CMC varies with pH, temperature, and other factors. Below the CMC the detergent exists as monomers in solution, but when the concentration of detergent is raised above the CMC, spherical micelles spontaneously form, driven by the hydrophobic effect. **Figure 1.11A** shows the process of adding detergent at above the CMC to a sample of protein held in a membrane. The protein is spontaneously extracted from the membrane and coated in a shell of detergent, which is of a size approximately equal in mass to the mass of the protein¹⁸¹. Also shown in **Figure 1.11B** are structures of several non-ionic or zwitterionic detergents routinely used for membrane protein solubilisation. Also shown for comparison is SDS, an anionic detergent used for much harsher solubilisation of proteins for experiments such as SDS-PAGE which require denatured proteins.

An interesting and novel approach to solving the problems of solubilisation is to radically redesign the protein to render it soluble by substituting the exterior lipid

facing hydrophobic residues of the protein for hydrophilic ones^{182,183}. If successful, the resulting soluble protein can be studied without the need for maintaining the solubility and stability of the protein with detergents. This technique, however, is yet to be deployed on a wide scale.

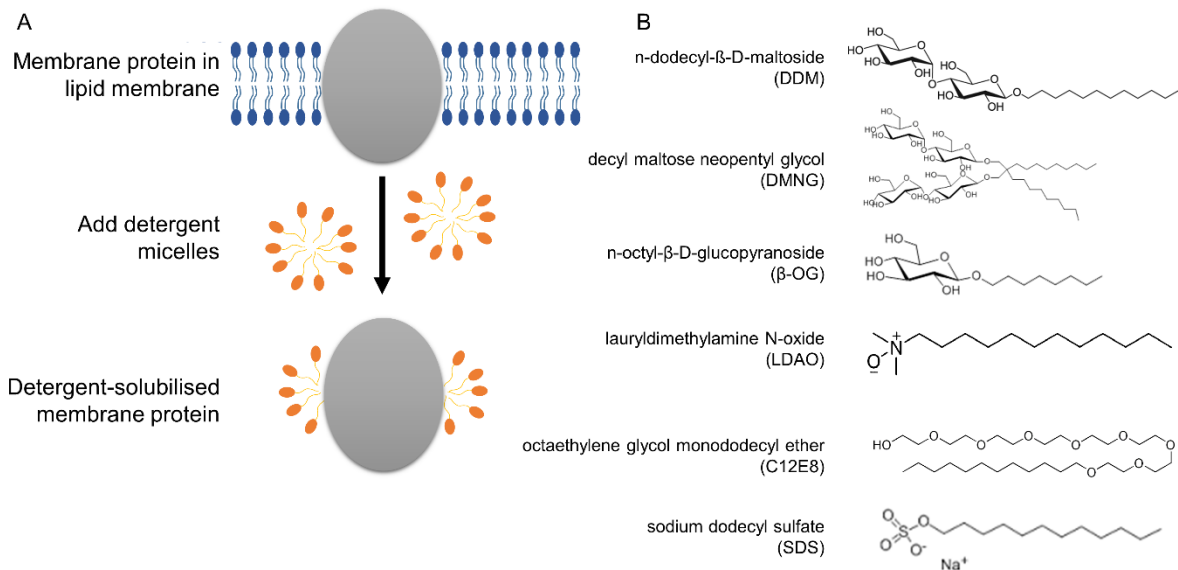


Figure 1.11 Detergent solubilisation of membrane proteins.

A) Detergent solubilisation of membrane proteins involves adding a detergent at a concentration above its CMC. Membrane proteins are then spontaneously extracted from their lipid membrane environment and coated in detergent. **B)** A range of detergent structures commonly used for membrane protein solubilisation. The maltosides DDM and DMNG, the glucoside β -OG, the zwitterionic LDAO, and the non-ionic C12E8 are shown along with SDS, an ionic detergent used regularly to denature protein, for comparison.

1.6.4 Purification

Once sufficient protein has been produced and solubilised, the purification of the protein is the next hurdle to overcome. The low expression levels discussed above prove a problem for purification, as the target protein represents only a small proportion of the total protein extracted from cells and thus preparations

already intrinsically contain a higher proportion of contaminants that must be removed. Several proteins are known to contaminate IMAC purifications¹⁸⁴, in particular, contamination by the *E. coli* membrane protein AcrB is of concern as it has been shown to crystallise readily¹⁸⁵, at pg-quantities (thus making it undetectable on coomassie-stained SDS-PAGE gels that are routinely used to examine purified protein samples for purity) and especially in glucoside detergents that are commonly used for solubilisation of the target protein¹⁸⁶.

1.6.5 Crystallisation

Even when protein can be purified in a soluble form at an appropriate concentration structural determination can be problematic. Membrane protein crystallisation is the “gold standard” structural technique, but it is limited in several ways, starting from the fact that there are a relatively small number of commercially available screens as compared to soluble proteins. Perhaps more pertinently, the low number of solved structures means that there is a lower amount of data available on conditions that have proved to work in the past. Thus a typical experiment will explore a smaller amount of the potential crystallisation conditions space, lowering the chances of finding a condition that is a hit¹⁸⁷. This point can often be moot, though, if sufficient protein cannot be produced, as the number of trials that it is possible to set up is limited instead by the low amount of protein. This is exacerbated by the fact that there is often a need to screen the same protein in several detergents, thus spreading the sample even more thinly.

A number of new strategies have emerged to aid in the production of crystals that are of particular use for those studying membrane proteins. Robots that enable screening of crystallisation conditions in drops that use 100 nL of protein

per drop enable a precious protein sample to be screened in a larger number of conditions than, for example, use of 2 μ L drops in hanging drop trays¹⁸⁷.

Crystallisation by lipidic cubic phase (LCP, or lipidic mesophases)¹⁸⁸ enables crystallisation in a native-like membrane environment and can produce higher-quality crystals as compared to traditional crystallization from detergent solutions and has shown promise in the crystallisation of a membrane kinase¹³³, several rhodopsins^{189,190}, human β 2 adrenoceptor¹⁰, and the outer membrane transporter, BtuB¹⁹¹. A sparse screen for lipidic conditions has been developed to aid screening of ideal conditions¹⁹². This can help overcome the disruption of crystal contacts by detergent micelles, as shown by **Figure 1.12** – the larger micelles cover the entire protein surface leaving no part able to make the protein-protein interactions needed for crystal formation. Finally, once crystals have been obtained, they are often fragile, difficult to handle, and poorly-diffracting¹⁸⁷, which can require extensive optimisation, or co-crystallisation with a stabilising antibody, which can help create more crystal contacts¹⁰. However, despite these difficulties, the number of solved membrane protein crystal structures is increasing year by year, in a large part due to the ever growing knowledge within the field and development of new technology, which is comprehensively reviewed in: Moraes *et al.* (2014)¹⁵. Notwithstanding these recent advances, however, many membrane proteins have proved intractable to structural study.

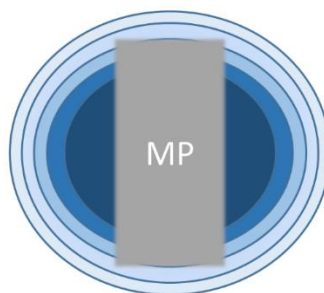


Figure 1.12 - Effect of differing micelle sizes on membrane protein visibility and crystal contacts.

A membrane protein (MP) surrounded by detergent micelles of steadily increasing size (diminishing intensity of blue). As the micelle size increases, progressively less and less of the protein surface is accessible to the solvent, and in the largest micelles the protein is completely embedded.

1.6.6 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR has limited applicability to membrane protein structural determination, with success being possible for only relatively small proteins¹⁹³, and there is generally a need to produce isotopically-labelled proteins at a high purity and high concentration – as much as 1 mM, equivalent to 50 mg/mL for a protein of 50 kDa. Solution-state NMR is made more difficult as the bound detergent gives particles of approximately double the molecular weight¹⁸¹ and so decreases the tumbling rate, impeding measurements. Solid-state NMR is an emerging technique to aid the study of membrane protein structures and can be applied to microcrystals formed in crystallography experiments that do not diffract sufficiently well¹⁹⁴. The addition of constraints derived from NMR data can be applied to data from crystallography to improve resolution, particularly useful in combination with weakly-diffracting crystals, and in one example researchers were able to improve a 3.7 Å map derived from crystallography to a precision of 0.92 Å in the membrane region¹⁹⁵.

1.6.7 Electron Microscopy

Although X-ray crystallography has so far been the dominant structural technique, recent developments in electron microscopy (EM), and especially cryo-EM, have driven a growth in solved structures year-on-year^{186,196} alongside developments such as the use of nanodiscs¹⁰⁸ or SMALPs¹⁷⁵ for protein solubilisation rather than detergents, leading to samples that are better suited to analysis by EM. While crystallography has proved a powerful tool thus far for the study of membrane proteins, EM allows the use of radically smaller amounts of protein and can enable the study of proteins within the native environment, in a variety of conformational states. Targets considered intractable for crystallographic studies can now be re-evaluated by EM with improved chance of success and with new investments in EM technology researchers can re-visit previous projects that may have failed to produce crystals. A good example of this is the AcrA-AcrB-TolC tripartite pump: the organisation of the whole complex had proved intractable by crystallography and significant discussion has taken place in the literature over the exact nature of the interactions between the subunits and their stoichiometry, but the recent structure provided by EM of the intact complex enabled visualisation of the 3:6:3 stoichiometry, and the mode of and location of interaction between AcrA and TolC¹¹⁰.

The development of direct electron detectors (DED) is a significant improvement on traditional CCD cameras due to an improved signal-to-noise and has been key to obtaining high-resolution information. Moreover, DED's are able to collect images as a series of frames, allowing the user to correct for sample movement from the electron beam and improve the quality of collected

images¹⁹⁷. The resolution of more recent EM maps is sufficient even to locate bound ligands. For example, γ -secretase had proved intractable to crystallographic studies, until 2014 when a 4.5 Å resolution EM structure was published¹⁹⁸. This was followed up by an improved study with larger dataset, increased magnification, and computationally accounting for heterogeneity in the sample that increased the resolution to 3.4 Å¹⁹⁹ and then a third study allowed visualisation of the bound inhibitor²⁰⁰. Another example of recent successful application of improvements in EM technology is the rotary ATPase family. In single particle cryo-EM studies between 2011 and 2016, the F-ATPase has improved in resolution from 32 Å to 6.2 Å^{3,201–203}, the V-ATPase from 17 Å to 6.9 Å^{204–208}, and A-ATPase from 16 Å to 6.4 Å^{29,201,209}, leading to steadily more insights in structural features and mechanism, and the determination of distinct catalytic states. An example of the improvement in the V-ATPase structure is shown in **Figure 1.13**, which shows two cross sections through the protein structure and the significant improvement in resolution seen over the course of seven years.

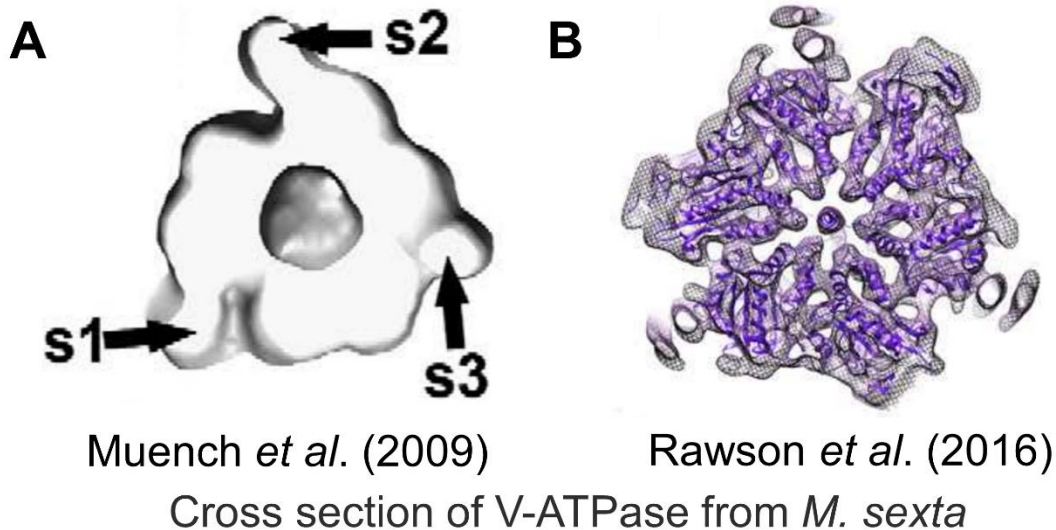


Figure 1.13 - Cross section of V-ATPase showing improvements in resolution between 2009-2016.

A) The 26 Å structure in 2009 is able to provide information on the general shape of the protein and some understanding of the symmetry and position of stators (labelled s1-s3)²¹⁰. **B)** The 8.2 Å structure in 2016 is able to show the position of individual helices in the structure. Improvements in hardware, and novel data processing techniques to account for flexibility and movement in the sample account for the improvement in resolution ²¹¹.

EM, of course, is not a panacea. EM datasets often have to be very large, with potentially over 1,000,000 particles required for analysis, necessitating thousands of micrographs to be collected over several days. The raw data requirements of these datasets are intense, as a single exposure can be 1-2 GB in size, and so an experimental dataset may represent over 7 TB of data and thus storage and processing of such data requires specialised IT infrastructure¹⁸⁶. In addition, the problems discussed above in production of sufficient protein, solubilisation and stabilisation, and purification still apply, but are somewhat lessened as smaller amounts of protein are required for EM studies, and to a certain extent protein can be “purified” during data processing as particles that are not the target protein can be discarded. Indeed, “on grid”

purification is in development with the use of Ni²⁺-doped lipids to extract polyhistidine-tagged proteins directly²¹². Overall, EM stands as an attractive structural technique for examination of membrane proteins, and will become only more attractive as developments are continuing in microscopes, detectors, and data processing²¹³.

1.7 Aims of this project

Proteins from the ArAE Family from both plants and bacteria have many vital roles, and many homologues remain uncharacterised, suggesting that further roles are yet to be elucidated. However, despite this importance no high quality structural or mechanistic data exists for any family member and much of the limited understanding of key residues and overall organisation of the proteins remain ambiguous at best. In the present study, the overall aim was to derive structural information about ALMT family members, by means of studying both the CTD and full length ALMT proteins, and bacterial family members including AaeB. This would require the acquisition of the relevant DNA, construction of expression vectors suitable for the expression systems chosen, and then optimisation of production of protein by overexpression. The targets were as follows:

1. The CTD of TaALMT1, expressed in *E. coli*.
2. Full length TaALMT1, AtALMT5, AtALMT9 transiently expressed in tobacco plants.
3. An *E. coli* protein, AaeB, expressed in *E. coli*, along with its proposed partner AaeA.

Once produced these protein targets would be used in structural and biochemical experiments. High priority studies are X-ray crystallography and electron microscopy in order to determine the 3-dimensional structure of these proteins, but other studies include investigation of complex formation and stoichiometry by crosslinking and SEC-MALLS, alongside homology modelling were also of interest, especially with the ultimate aim to resolve the open questions and matters of contention discussed above in this introduction.

Chapter 2

Materials and Methods

2

Analytical grade reagents were prepared using Milli-Q™ water (resistivity 18.2 MΩ·cm using either a Sartorius arium® mini plus (Sartorius AG), or a Milli-Q® Advantage A10 Water Purification System (Merck Millipore). Unless certified sterile by the manufacturer, solutions were sterilised by filtration through a 0.22 µm filter or autoclaving at 121°C for 20 min.

2.1 Chemicals and Reagents

Unless otherwise stated all chemicals and reagents used were obtained from Sigma-Aldrich or Fisher Scientific. Restriction enzymes were purchased from NEB.

2.2 Bacterial Growth and Expression Media

2.2.1 Media for growth

The following solutions were prepared for the growth of bacteria and expression of protein in *E. coli*, recipes were taken from Postis et al. (2013)¹¹⁶.

Table 2.1 - Lysogeny broth (LB) medium

Component	Quantity (for 1L)	Concentration
Tryptone	10 g	1% (w/v)
Yeast Extract	5 g	0.5% (w/v)
NaCl	10 g	1% (w/v)

Autoclaved

Table 2.2 - Super broth (SB) medium

Component	Quantity (for 1L)	Concentration
Tryptone	32 g	3.2% (w/v)
Yeast Extract	20 g	2% (w/v)
NaCl	10 g	1% (w/v)

Autoclaved

Table 2.3 - LB medium without salt for use in Autoinduction

Component	Quantity (for 1L)	Concentration
Tryptone	10 g	1% (w/v)
Yeast Extract	5 g	0.5% (w/v)

Autoclaved

Table 2.4 - SB medium without salt for use in Autoinduction

Component	Quantity (for 1L)	Concentration
Tryptone	32 g	3.2% (w/v)
Yeast Extract	20 g	2% (w/v)

Autoclaved

Table 2.5 - 5X M9 Salts for use in Autoinduction

Component	Quantity (for 1L)	Concentration
Na₂HPO₄	30 g	3% (w/v)
KH₂PO₄	15 g	1.5% (w/v)
NH₄Cl	5 g	0.5% (w/v)
NaCl	2.5 g	0.25% (w/v)
CaCl₂	15 mg	-

Autoclaved

Table 2.6 - 20X NSPC for use in Autoinduction

Component	Quantity (for 1L)	Concentration
Na₂HPO₄	71 g	0.5 M
KH₂PO₄	68 g	0.5 M
NH₄Cl	53.5 g	1 M
Na₂SO₄	14.2 g	0.1 M

pH adjusted to 6.7 with NaOH. Autoclaved

0.5 M MgSO₄

24.65 g MgSO₄·7H₂O, dissolved in water to give a final volume of 100 mL and autoclaved.

Table 2.7 - 50X 5052 for use in Autoinduction

Component	Quantity (for 1L)	Concentration (at 1X)
Glycerol	250 g	0.5% (w/v)
Glucose	25 g	0.05% (w/v)
α-lactose	100 g	0.2% (w/v)

Dissolved in 730 mL water, gently heated to dissolve the lactose, and sterilised by filtration.

Table 2.8 - LB for autoinduction

Component	Quantity (for 1L)	Concentration
LB without salt	930 ml	-
0.5 M MgSO₄	2 ml	1 mM
50X 5052	20 ml	1X
20X NSPC	50 ml	1X
Antibiotics	As required	

Table 2.9 - SB for autoinduction

Component	Quantity (for 1L)	Concentration
SB without salt	930 ml	-
0.5 M MgSO₄	2 ml	1 mM
50X 5052	20 ml	1X
20X NSPC	50 ml	1X
Antibiotics	As required	

Table 2.10 - M9 for autoinduction

Component	Quantity (for 1L)	Concentration
5X M9 Salts	200 ml	1X
0.5 M MgSO₄	2 ml	1 mM
50X 5052	20 ml	1X
20X NSPC	50 ml	1X
Casamino acids 5% (w/v)	40 ml	0.2%
Antibiotics	As required	

2.3 Bacterial Strains

Several *E. coli* strains were used during this project for DNA plasmid manipulation and expression of protein. Their genotypes are as follows:

OmniMAX™ 2²¹⁴: F' {*proAB+* *lacIq* *lacZΔM15* Tn10(TetR) Δ (*ccdAB*)} *mcrA* Δ (*mrr-hsdRMS-mcrBC*) Φ 80*lacZΔM15* Δ (*lacZYA-argF*) U169 *endA1* *recA1* *supE44* *thi-1* *gyrA96* *relA1* *tonA* *panD*

BL21 Star™(DE3)²¹⁵: F- *ompT* *hsdSB*(rB- mB-) *gal* *dcm* *rne131*(DE3)

BL21 Gold™(DE3)²¹⁵: F- *dcm+* *Hte* *ompT* *hsdS*(rB- mB-) *gal* λ (DE3) *endA* *Tetr*

C41(DE3)²¹⁶: F- *ompT* *gal* *dcm* *hsdSB*(rB- mB-) (DE3)

C43(DE3)²¹⁶: F- *ompT* *gal* *dcm* *hsdSB*(rB- mB-) (DE3)

Agrobacterium tumefaciens²¹⁷ strain LBA4404 rifampicin (chromosomal) and streptomycin (on the Ti plasmid)

2.3.1 Preparation of Chemically-Competent *Escherichia coli* Cells

Competent cells were prepared based upon the method of Hanahan²¹⁸ with the following modifications. The required strain was aseptically streaked from a glycerol stock onto an agar plate containing appropriate antibiotics and grown overnight at 37°C. A 5 mL LB medium starter culture was inoculated with a single colony and grown for 16 hours at 37°C with shaking (200 rpm).

Subsequently, 200 μ L of the starter culture was used to inoculate 200 mL of LB medium, and cultured at 37°C with shaking (200 rpm) until an OD₆₀₀ of 0.5 to 0.8 was reached (typically 4-8 hours). Cells were chilled on ice for 15 minutes before centrifugation at 3,500 g for 10 minutes at 4°C and the pellets

resuspended into 20 mL ice-cold Tbf1 (3 mM CH₃CO₂K, 100 mM RbCl₂ (Fluka Biochemicals), 10 mM CaCl₂·2H₂O, 50 mM MnCl₂·4H₂O, 15% (v/v) glycerol, pH 5.8 (adjusted with acetic acid)). After centrifugation (as above), the washed pellet was resuspended into 2 mL ice-cold Tbf2 (10 mM MOPS, 75 mM CaCl₂·2H₂O, 10 mM RbCl₂, 15% glycerol, pH 6.5 (adjusted with KOH)). Competent cells were divided into 50 µL aliquots snap frozen in liquid N₂ and stored at -80°C until use.

2.3.2 Heat Shock Transformation of *Escherichia coli* Cells

DNA (2 µL of a ligation reaction or 100-200 ng of plasmid DNA) was added to a freshly thawed 50 µL aliquot of chemically-competent *E. coli* cells, mixed and held at 4°C for 30 minutes. The cells were then heat shocked at 42°C for 30 seconds and returned to 4°C for 2 minutes. 100 µL LB medium was added to the transformation and incubated at 37°C for 1 hour. Transformed cells were plated out onto selective agar plates prior to overnight incubation at 37°C.

2.3.3 Plasmid Maintenance in Bacteria

Plasmids used were transformed into and maintained in One Shot OmniMAX Chemically Competent *E. coli* (Invitrogen). Colonies were grown either on solid media (Lysogeny Broth (LB) medium with 2% (w/v) agar containing appropriate antibiotic) at 37°C for 16-20 hours or liquid cultures in LB medium containing appropriate antibiotic at 37°C with shaking (200 rpm). Permanent stocks were maintained in LB medium containing 20% (v/v) sterile glycerol at -80°C in 2 mL cryo-vials snap frozen in liq. N₂. The cultures were revived by streaking onto

agar plates containing appropriate antibiotic and grown overnight until colonies appeared, which were then grown in LB medium.

2.3.4 ALMT cDNA

cDNA for AtALMT5 (AT1G68600, Clone: U14043) and AtALMT9 (AT3G18440, Clone: U24675) were obtained from the Arabidopsis Biological Resource Centre (ABRC, Ohio State University, USA). cDNA for TaALMT1 (Uniprot: Q76LB1) was synthesised by Genscript with *Xma*I and *Xho*I restriction sites flanking the gene. The primers in **Table 8.12** were used for production of vectors.

The pEAQ-HT vector and pEAQ-GFP were provided by Prof. George Lomonosoff, John Innes Centre.

2.4 Molecular Biology

2.4.1 PCR

PCR amplifications were performed using KOD Hot Start DNA Polymerase purchased from Novagen®. Typically, reactions of final volume 50 μL were set up as follows:

- 5 μL 10 \times KOD buffer (final concentration 1 \times)
- 3 μL 25 mM MgSO_4 (final concentration 1.5 mM)
- 5 μL 2 mM dNTPs (mixture) (final concentration 0.2 mM)
- 1.5 μL 10 μM Forward primer (final concentration 0.3 μM)
- 1.5 μL 10 μM Reverse primer (final concentration 0.3 μM)
- 5 μL 10 ng/ μL genomic DNA or plasmid (total amount 50 ng)
- 28 μL MilliQ H_2O
- 1 μL KOD Polymerase

PCR Protocols followed the two methods as follows:

TouchDown PCR

- 95°C - 1 min
 - 95°C - 30 s
 - 70°C - 30 s decreasing each cycle by 1°C
 - 70°C - 30 s per kb of desired product
 - 95°C - 30 s
 - 60°C - 30 s
 - 70°C - 30 s per kb of desired product
 - 70°C - 5 min
 - 4°C - continuous
-
- 10 cycles
- 30 cycles

Gradient PCR

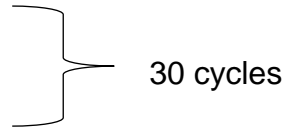
95 °C - 4 min

94 °C - 30 s

Gradient from 50 → 70 °C - 30 s

72 °C - 2 min

4 °C – continuous



The primers used for construction of vectors and mutagenesis are listed in the tables below.

Table 2.11 - List of primers used for the truncation of pTP1 and pTP2.

F and **R** refer to the forward and reverse primer, respectively.

Name	Sequence 5' -> 3'
pTP1_T-F	GAAAACCTCTGTTGCTAACACCATGGGTGGTAAAGACTCTCCGC
pTP1_T-R	GCGGAGAGTCTTTACCACCCATGGTGTAGCAACAGAGTTTTTC
pTP2_T-F	GTAAAGACTCTCCGCATATGCACAAATCTGTTC
pTP2_T-R	GAACAGATTTGTGCATATGCGGAGAGTCTTTAC

Table 2.12 - List of primers used for construction of plant expression constructs.

Suffix **C-F** and **C-R** refer to the forward primer and reverse primer respectively to construct a protein with a C-terminal His-tag. Similarly, **N-F** and **N-R** refer to the N-terminal versions. No primers were needed for ALMT1N as the gene was synthesised with appropriate flanking restriction sites.

Name	Sequence 5' -> 3'
ALMT1C-F	TAAGCAACCGGTATGGATATAGATCATCGTAGGGAATC
ALMT1C-R	TAAGCACCCGGGAGTATCACAACATCAGGCAATGG
ALMT1N-F	N/A
ALMT1N-R	N/A
ALMT5C-F	ACCGGTATGGGAGGTAAAATGGGATCAG
ALMT5C-R	CCCGGGAACCTTAGGAATCTGATCAACAGC
ALMT5N-F	CCCGGGATGGGAGGTAAAATGGGATCAG
ALMT5N-R	GCTCGAGGTTATCAAACCTTAGGAATCTGATCAACAGC
ALMT9C-F	ACCGGTATGGCGGCGAAGCAAGG
ALMT9C-R	CCCGGGTTACATCCCAAACACCTACGAATCTTC
ALMT9N-F	CCCGGGATGGCGGCGAAGCAAGG
ALMT9N-R	GCTCGAGGTTACATCCCAAACACCTACGAATCTTC

Table 2.13 - List of primers used for the amplification of bacterial genes.

F and R refer to the forward and reverse primer, respectively.

Name	Sequence 5' -> 3'
SoF	CCTAGGATGTCCGTCTTTGCCCTGC
SoR	CTGCAGAATTCGTCCCAATCGCTAAAATACATTGAAC
DvF	CCTAGGATGGACGGACGGGCATTGTGC
DvR	CTGCAGAGTACGGTACCTTGTGCGACTGC
EcF	CCTAGGATGGGTATTTTCTCCATTGCTAACC
EcR	CTGCAGAATTGATAGCCAGTTGCGTACAACATG
AaeAAaeB-F	CAGGCTAGCGTGAAAACACTAATAAGAAAATTCTCC
AaeAAaeB-R	TGCTGCAGCACTATCGGTCAACGCATGTTG
AaeA-F	CTCTCCATGGGCAAAACACTAATAAGAAAATTCTCCCGTACG
AaeA-R	CTCTGTGACACCAAACCTCACGCAGGCG
AaeAdTM-F	CTCTCCATGGGC TCCCCTGGACGCG
AaeAdTM-R	CTCTGTGACACCAAACCTCACGCAGGCG

2.4.2 Quikchange™ Site-Directed Mutagenesis

For mutagenesis of DNA sequences, primers were designed to introduce the desired base change, ensuring that the ends of the primers terminated with a pair of G or C bases in order to ensure efficient anchorage to the DNA template. The conditions were set up as for PCR, as above. After thermocycling was complete, original template DNA was digested by addition of 1 μ L *DpnI* restriction enzyme and CutSmart Buffer to 1X concentration. This mixture was incubated at 37°C for 1 h and then heat inactivated at 80°C for 20 min. A 4 μ L sample was electrophoresed on a 1% agarose gel to check for the correct product remaining after digestion, and then 1 μ L was transformed into OmniMax 2 cells, as described above in **Section 2.3.2**.

2.4.3 Cleaning of PCR Products

PCR products were cleaned using the Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacture's protocol.

2.4.4 Colony PCR

To assay the presence of the ORF of interest within transformed bacteria colony PCR was performed. A pipette tip was touched to a colony (used for plasmid purification, above) and dipped into 10 μ L of MilliQ H₂O and heated to 95°C briefly. This was then used as the "Template" in a PCR, as described above.

2.4.5 Restriction Endonuclease Digestion of DNA

DNA was digested with the appropriate restriction enzyme (New England Biolabs) using the optimal buffers and conditions recommended by the manufacturer – routinely, CutSmart Buffer at a concentration of 1X. Generally, 1 µg of plasmid DNA was digested per reaction using 5-10 units of restriction enzyme. The final volume added to any reaction was always less than 10% of the total volume. Digestions were routinely performed at 37°C for 2 hours and fragments separated by electrophoresis on an agarose gel.

2.4.6 Agarose Gel Separation of DNA

Agarose gel electrophoresis was performed to separate DNA fragments on the basis of their size. DNA gel loading dye (6x) (NEB) was added to samples at a 1:5 volume ratio prior to electrophoresis. DNA fragments were then separated on a 0.5% (w/v) SeaKem LE agarose (Lonza) TAE buffer gel containing SYBRSafe 10,000X concentrate (Life Technologies) at a 1:10,000 dilution by electrophoresis for 30 minutes at 100 V in TAE buffer.

10X TAE Buffer

Tris	48.4 g	400 mM
EDTA	2.92 g	1 mM
Acetic acid	11.4 ml	to pH 8.3

Make to 1 L with water

(This buffer was diluted to 1X before use.)

2.4.7 Agarose Gel Extraction of DNA

Bands of agarose containing DNA fragments were excised from gels using a clean scalpel blade after electrophoresis. They were subsequently purified from the agarose using the Wizard SV Gel and PCR Clean-Up System (Promega) following the manufacturer's protocol.

2.4.8 Dephosphorylation of DNA Fragments

Dephosphorylation of linearised fragments of DNA was undertaken to prevent re-circularisation. This was performed using 1 μ L Antarctic Phosphatase (NEB) in 1X Antarctic Phosphatase Buffer incubated for 15 minutes at 37°C for 5' extensions or blunt-ends and 60 minutes for 3' extensions, before inactivating at 95°C for 5 min.

2.4.9 Ligation of DNA

Previously digested linear vector DNA (25 ng) was mixed with either a threefold or tenfold molar excess of the purified DNA fragment to be inserted. Ligations were performed by two protocols: either in 1X Quick Ligation Reaction Buffer (66mM Tris-HCl, 10mM MgCl₂, 1mM DTT, 1mM ATP, 7.5% Polyethylene glycol (PEG6000), pH 7.6), with 1 μ L of Quick T4 DNA Ligase (NEB) in a total volume of 10 μ L and incubated at room temperature 5 minutes; or in T4 Ligase buffer (50mM Tris-HCl, 10mM MgCl₂, 1mM ATP, 10mM DTT, pH 7.5) with 1 μ L of T4 Ligase (NEB) in a total volume of 10 μ L incubated at 16°C for 16 h.

2.4.10 Plasmid Purification and Preparation: Mini-preparation

A single bacterial colony was used to inoculate 10 mL LB medium containing appropriate antibiotics and incubated for 16 hours at 37°C with shaking (200 rpm). Cells were harvested by centrifugation in an Eppendorf swing-bucket centrifuge at 3,500 g for 10 minutes. Purification of plasmid DNA was carried out using the PureYield™ Plasmid Miniprep Kit (Promega) following the manufacturer's protocol.

2.4.11 DNA Sequencing

DNA sequence analysis was performed by Source Bioscience (www.sourcebioscience.com). Plasmids were sent at 100 ng/μl, primers at 3.2 pmol/μL, and PCR products at 10 ng/μL per 1000 basepairs. The sequence of the primers is given in **Table 2.14** - Primers for sequencing

Table 2.14 - Primers for sequencing

Primer	Sequence 5' -> 3'
T7F	TAATACGACTCACTATAGGG
T7R	GCTAGTTATTGCTCAGCGG
M13F	TGTAAAACGACGGCCAGT
M13R	CAGGAAACAGCTATGACC
pTACF	TTGACAATTAATCATCGGC
pTACR	CGCCAGCTGGCGAAAGGGG

2.5 Protein Expression

2.5.1 Expression Screening in Small Scale Cultures

2 mL of LB with appropriate antibiotic was inoculated with a single colony and incubated overnight at 37°C with shaking in a shaking incubator (200 rpm). 200 uL of the resulting pre-culture was inoculated into 3mL samples of appropriate auto-induction media to be tested in a sterile 24-well deep-well plate. The plate was sealed with a breathable seal and grown in a humidified shaking incubator at 37°C for 24h after which 400 uL samples were taken to test protein expression levels.

2.5.2 Large Scale Expression

10 mL of LB with appropriate antibiotic was inoculated with a single colony and grown overnight at 37°C with shaking (200 rpm). 500 mL of medium (LB_{auto} or SB_{auto}) was added to each of eight 2.5 L flasks and each was inoculated with the pre-culture to give a final O.D. of 0.05 and grown at 37°C with shaking (200 rpm) for 24 or 28h. 1 mL samples were taken to test expression levels, and the remainder was centrifuged at 14,000 g for 15 minutes and cell pellets were stored at -80°C for later use.

2.5.3 Membrane Preparation

Cell pellets stored at -80°C were thawed on ice, and re-suspended with Resuspension Buffer (20 mM Tris pH 8.0, 1 mM DTT, 1x Complete Protease Inhibitors (Roche)) at a ratio of 6 mL buffer per g of cell pellet. The suspension was homogenised with an Ultra-Turrax (IKA) before lysis by passage through a cell disrupter (Constant Systems) at 30 kpsi. The lysate was centrifuged at 14,000 *g* for 45 minutes in an Optima Ultracentrifuge (Beckman) to clear cell debris before the supernatant was centrifuged at 100,000 *g* to collect membranes. These were re-suspended in 20 mM Tris pH 8.0 before membranes were again collected at 100,000 *g*. The final membrane pellet was resuspended in a minimal volume of 20 mM Tris pH 8.0 and frozen as beads in liquid nitrogen before storage at -80°C. All steps were performed on ice.

2.5.4 Detergent Screening for Membrane Protein Solubilisation

Parallel samples of 1 µg of membrane protein was combined with appropriate test concentrations of detergents (e.g. 0.5%, 1.0%, 1.5%, 2.0% (w/v) for each of DM, DDM, and so on) in a buffer containing 20mM Tris, 10% glycerol, 150mM NaCl and made up to 200 µL. Samples were incubated with mixing at 4°C for 2 hours and 50 µL samples were taken to show a “Total” fraction. The remaining 150 µL was then centrifuged at 100,000 *g* for 1 hour in a TLA100 rotor (Beckman). 50 µL of the supernatant was then taken as the “Soluble” fraction, taking care to not disturb the pellet. Samples were then analysed by SDS-PAGE and Western blotting.

2.5.5 Purification of Proteins by Immobilised Metal Affinity Chromatography

Solubilised protein was incubated with HisPur™ Cobalt Resin (Thermo) at a ratio of 1 mL of resin per 10 mg of solubilised target protein overnight at 4°C in a buffer containing 10 mM of imidazole. The resin was collected using a disposable Gravity Flow Column (Pierce) with sufficient capacity to collect the resin volume, allowing the buffer solution to run through the column and be collected. The resin was then washed with 10x volumes of buffer containing 10 mM imidazole as an initial wash, followed by identical washes containing increasing concentrations of imidazole until contaminants had been removed. Finally, target protein was eluted with buffer containing 300 mM imidazole in steps of 0.5-1 mL. Effectiveness of purification was examined by SDS-PAGE, Coomassie staining, and Western blotting.

2.5.6 Protein Estimation by the BCA Method

Protein concentration was determined using the Pierce BCA Assay Kit (Life Technologies), combining BCA Reagent A with 4% CuSO₄ in a 50:1 ratio. Bovine serum albumin (BSA; Sigma) was used to generate a standard curve (0-10 µg) and all samples were analysed in duplicate. All samples were analysed at 550 nm using a plate reader (ICN Biomedical Titerktek).

2.5.7 Protein Estimation by Absorbance at A280

Protein concentrations were also determined by measurement of absorbance at 280 nm. 1 µL samples were analysed using the "A280" mode of a Denovix

DS11+ Spectrophotometer (Generon) using molar extinction coefficients calculated by the ExPASy ProtParam website²¹⁹.

2.5.8 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis

Proteins were separated on the basis of their molecular weight by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) (Laemmli, 1970) using the Bio-Rad Mini Protean apparatus (Bio-Rad Laboratories, Ltd). Protein samples (routinely 10 µg of cell lysate or membrane fraction) were solubilised in a 5:1 ratio protein:buffer in SDS-PAGE Sample Buffer (0.15 M Tris, 2.5% (w/v) SDS, 20% (v/v) glycerol, 1 mM DTT, 0.1% (w/v) Pyronin Y) by incubation at 37°C for 30 minutes. Samples were then loaded onto SDS-polyacrylamide gels, either precast (4–15% Mini-PROTEAN® TGX™ Precast Gel (BioRad)) or handmade, prepared as shown in **Table 2.15** below. Proteins were separated electrophoretically at 100V for between 60 and 90 minutes in SDS-PAGE Running Buffer (25 mM Tris, 192 mM glycine, 0.1 % (w/v) SDS).

Table 2.15 - Composition of 12% SDS-PAGE gels

Component	12% Separating gel	4% Stacking gel
MilliQ water	3.3 ml	3.025 ml
Acrylamide-Bis acrylamide ratio 37.5:1 30% (w/v) solution	4 ml	0.625 ml
Separating buffer (1.5 M Tris-HCl, pH 8.8)	2.5 ml	-
Stacking Buffer (0.5 M Tris-HCl, pH 6.8)	-	1.25 ml
10% (w/v) SDS	100 µl	100 µl
10% (w/v) ammonium persulfate (APS)	100 µl	100 µl
Tetramethylethylenediamine (TEMED)	5 µl	5 µl

2.5.9 SDS-Polyacrylamide Gel Coomassie Staining

Gels were routinely stained with a colloidal Coomassie protein stain. After electrophoresis the gel was removed from the apparatus and fixed for 2 hours in Fixing Solution (25% (v/v) propan-2-ol, 10% (v/v) acetic acid). The gel was then covered in Stain 1 (0.025% (w/v) G250 Coomassie Brilliant Blue, 10% propan-2-ol, 10% acetic acid) for 2 hours before replacement with Stain 2 (0.0025% (w/v) G250 Coomassie Brilliant Blue, 10% propan-2-ol, 10% acetic acid) for 2 hours. The gel was then de-stained with several changes of De-stain (10% acetic acid) until background colour had diminished fully, typically between 2 hours and overnight.

2.5.10 SDS-Polyacrylamide Gel Silver Staining

As a more sensitive alternative to Coomassie staining, gels were stained using the Pierce™ Silver Stain Kit (ThermoFisher Scientific) following the manufacturer's recommendations. The silver stain has sensitivity down to 0.25 ng per band, while the typical lower limit for Coomassie Brilliant Blue G250 is around 30 ng per band.

2.5.11 Dot Blotting and Western Blotting of Proteins

For dot blotting, protein samples were mixed with SDS-PAGE Sample Buffer (see above in **Section 2.5.8**) and 2 µL samples were dotted onto a nitrocellulose membrane. To serve as a positive control, samples of known amounts of His-tagged tobacco etch virus (TEV) protease were also dotted onto the membrane. When samples were dry, proteins were detected using specific antibodies. Non-specific binding of antibodies was blocked by incubating the

membrane in 10 mL of 3 % (w/v) BSA powder in Tris-buffered saline with Tween 20 (TBST; 25 mM Tris, 0.15 M NaCl, pH 7.6, 0.2% (v/v) Tween 20) for 1 hour at room temperature or overnight at 4°C with gentle mixing. The blot was then incubated with 2 µl of anti-His antibody, for 1 hour at room temperature with gentle mixing (Horseradish peroxidase-conjugated monoclonal anti-His₆ (R&D Systems, MAB050H) at 1:5,000 dilution). After washing 3 times for 5 minutes in TBST detection of antigen/antibody interaction performed using SuperSignal® West Pico Chemiluminescent Substrate Kit (Pierce), following manufacturer's instructions. The membrane was then photographed in a G:Box Chemi (Syngene).

For western blotting the following protocol was observed. After separation by electrophoresis proteins were transferred from the polyacrylamide gel onto nitrocellulose membrane using a Trans-Blot Turbo™ Transfer System (Bio-Rad). Transfer was achieved by two separate methods. The first used hand-assembled stacks: before transfer, the nitrocellulose membrane was equilibrated by soaking in Transfer Buffer (25 mM Tris, 192 mM glycine, 20 % (v/v) methanol). For transfer, the membrane and gel were placed between sheets of pre-soaked Whatman™ 3mm filter paper, orienting the membrane beneath the gel, at the anode. After compressing with a roller to remove any air bubbles, the electrotransfer of proteins was performed at 1 A for 30 min. The second method used Trans-Blot Turbo™ Nitrocellulose Transfer Packs (Bio-Rad), used as per manufacturer's instructions. Electrotransfer was performed at 1.3 A for 7 minutes per gel. After transfer was complete, the membrane was incubated in 3% BSA in TBST, and treated as described above, using the same washing and antibody.

2.5.12 Densitometry Protein Estimation of Coomassie-stained SDS-PAGE gels, or Western blots

Samples were run on an SDS-PAGE and either Coomassie stained or visualised by Western blotting. Images were captured using a G:Box (SynGene) and the picture analysed by comparing the relative density of the eluted protein band density against that of known protein standards using ImageJ software.

2.6 Blue Native Polyacrylamide Gel Electrophoresis

Purified samples of protein were separated by blue native PAGE using 10X NativePAGE™ Sample Buffer (BioRad) added at a ratio of 1:10 prior to electrophoresis. The gel tank was filled with 10X NativePAGE™ Running Buffer (BioRad) diluted 1:10 with MilliQ water. Proteins were analyzed by blue native (BN) electrophoresis on 4-20% Mini-PROTEAN® TGX™ Precast Gels (BioRad). Proteins were then transferred onto nitrocellulose membrane as in **Section 2.5.11** or silver stained as in **Section 2.5.10**.

2.7 Glutaraldehyde Cross Linking

Purified protein (~0.1 mg/mL) was mixed with glutaraldehyde at a range of concentrations (0 mM – 100mM) and incubated for up to 1 hour at 4°C and the reaction stopped by adding SDS-PAGE sample buffer. Samples were run on a 4-20% (w/v) acrylamide SDS-PAGE gel and analysed by immunoblotting and Coomassie staining.

2.8 Expression of ALMTs in Plants

Expression of ALMTs in plants was based on the method of Sparkes *et al.*²²⁰ with modifications as follows.

2.8.1 Plant Growth Conditions

Tobacco (*Nicotiana benthamiana*) seeds were sown directly onto freshly watered compost (J Arthur Bower's Topsoil) and grown in greenhouses maintained at 20°C with a 16h day length for 5 weeks. After infiltration, plants were maintained under the same conditions for the required number of days. A portion of plants were grown to maturity and seeds were collected and stored at 15°C and 12% relative humidity in the dark.

2.8.2 Transformation of agrobacteria

1 µg of plasmid DNA was added to 50 µL of agrobacteria (strain LBA4404) and gently mixed by flicking. The tube was incubated on ice for 5 minutes and then flash frozen in liq. N₂ before thawing at 37°C. 1 mL of YEP media was added, and then the cells were shaken at 28°C for 2 hours before plating onto YEP-agar plates containing 100 µg/mL kanamycin, 50 µg/mL streptomycin, and 100 µg/mL rifampicin. Rifampicin is light sensitive so stocks were stored in the dark at -20 and plates were freshly made. Plates were incubated at 28°C for 28 hours before storage at 4°C in the dark before later use.

2.8.3 Agrobacteria tumefaciens cultures

A single well isolated colony was picked from a plate and incubated with 50 mL YEP medium containing 100 µg/mL kanamycin, 50 µg/mL streptomycin, and 100 µg/mL rifampicin and grown with shaking at 28°C for 24 hours.

2.8.4 Agrobacteria tumefaciens infiltration

The agrobacteria were collected by centrifugation at 4000 *g* for 10 minutes, pellets were then resuspended in 50 mL Infiltration Medium (50 mM MES pH 5.6, 2 mM Na₃PO₄, 0.5% glucose 100 μM acetosyringone (4'-hydroxy-3',5'-dimethoxyacetophenone)) to wash the cells. Then centrifugation was repeated and re-suspended to the required O.D.₆₀₀. Agrobacteria suspensions were then infiltrated into the underside (abaxial) of tobacco leaves by nicking the leaf surface lightly and pressing a 5 mL syringe (without needle) onto the resultant small hole, with a finger on the other side of the leaf to provide support.

2.8.5 Protein extraction from tobacco leaves

Leaves were harvested, main ribs were removed and leaf material was ground in a pestle and mortar under liq. N₂ followed by protein extraction in Plant Extraction Buffer (20mM Tris pH 8.0, 10% (v/v) glycerol, 100 mM NaCl, 25% (w/v) sucrose, and 1x cOMplete protease inhibitors (Roche)) at a 1:1 g:mL ratio of leaf mass to buffer volume. Lysates were incubated on ice for 20 minutes and vortexed before centrifugation at 600 *g* for 5 minutes to pellet cell debris. The supernatant was then spun down at 100,000 *g* for 1 hour at 4°C to collect membranes. Membranes were then re-suspended in 20 mM Tris pH 8.0 and frozen as beads in liq. N₂. Protein expression was assessed by SDS-PAGE and Western blotting as in **Sections 2.5.9** and **2.5.11**.

2.8.6 Imaging of GFP in leaf

Whole infiltrated leaves were excised from plants and imaged in a G:BOX Chemi XX6 (Syngene) directly on the black using blue LED illumination and a filter for GFP fluorescence using the “Fluorescent Gel” present mode. Images were saved as .tif files.

2.9 Measurement of Protein Stability in Varying Buffer Conditions

Screening buffer conditions to optimise protein stability was undertaken using the Optim (Avacta) (now called the UNit from Unchained Labs). 5 μ L of purified protein was mixed with 5 μ L of each buffer condition to be screened and loaded into the supplied capillaries, which were then sealed into the machine.

Temperature runs were performed in 1°C steps from 15-95°C measuring both static light scattering at 266 nm and fluorescence at 250–720 nm, from which the barycentric mean was computed and reported. The data were plotted against temperature, along with the derivative of the fluorescence data to aid location of the midpoint of the unfolding curve. Data were exported to produce further plots in Microsoft Excel.

2.10 Protein Structural Homology Modelling with Phyre2

The sequence of AaeA was submitted to the Phyre2 server (www.sbg.bio.ic.ac.uk/~phyre/) using Intensive Mode¹⁵⁹. This produced a model based on the published structures for EmrA, AcrA, MacA, ZneB, MexA, and HlyD, with areas not covered by these structures modelled *ab initio*. The model

produced placed the predicted TM helix directly adjacent to the main body of the protein. Therefore, this was manually adjusted in Chimera to lie parallel to the main axis of the protein, as expected. This model was exported for the production of figures and analysis. The conservation scores were computed by the ConSurf webserver (<http://consurf.tau.ac.il/2016/>), and these were applied to the modelled structure in Pymol replacing the crystallography B-factors with a conservation score. This conservation score was displayed in Pymol as a surface. The surface charge was calculated using the PDB2PQR server²²¹ (http://nbc222.ucsd.edu/pdb2pqr_2.1.1/). The resulting electrostatic surface was displayed in Pymol using the APBS plugin tool. RLS motifs from other MFP proteins were located manually from their sequence, and aligned using clustalW. Molecular Phylogenetic analysis was performed using MEGA7²²². The evolutionary history was investigated using the Maximum Likelihood method based on the JTT matrix-based model²²³. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 19 amino acid sequences, including four SMR-type, two ABC-type, three MFS-type, three RND-type, alongside four identified by Saier *et al.* as PET-type³⁶ and two of unknown family.

2.11 Circular Dichroism Spectroscopy

Purified protein was loaded into 1 mm cuvettes at a concentration of ~0.2-0.4 mg/mL and measured in a Jasco J715 (Jasco Analytical Instruments). Two repeat spectra were captured between 180-260 nm at 20°C in addition to a buffer blank that was subtracted from the experimental spectra. For the temperature ramp the temperature was raised in 1°C steps then held for 60 seconds before acquisition of a spectrum. Data was exported to Microsoft Excel for calculations before analysis using the algorithms available at the Dichroweb server (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>) using SELCON3, CONTIN and K2D with the SMP180 reference set in order to calculate the proportions of secondary structure present.

2.12 Size Exclusion Chromatography-Multi Angle Laser Light Scattering (SEC-MALLS) analysis of the oligomeric states of membrane proteins

SEC-MALLS analysis ^{224,225} was performed using a DAWN HELEOS II (Wyatt Technology) after separation on either a Superdex 200 10/300 column (GE Healthcare) or a WTC-MP030S5 column (Wyatt Technology), driven by a Prominence HPLC system (Shimadzu). Data analysis was undertaken using the Astra software using the Conjugate Method analysis to compute protein MW.

2.13 Electron Microscopy and Single Particle Analysis

Negative stain grids were prepared by applying 3 µL of purified protein solution (10-100 µg/mL) onto a carbon-coated copper grid that had previously been charged using a PELCO easiGlow glow-discharge unit for 30 seconds. Excess solution was blotted with Whatman paper and the grid was then stained with 3

μL 1% w/v uranyl acetate for 1 minute, before blotting the excess. The staining step was repeated and the final drying step was done under a desk lamp. Grids were then stored until imaging. Grids were imaged using a Technai T12 microscope fitted with a LaB6 source operating at 120 kV. Micrographs were recorded on a 2000 \times 2000 Gatan CCD camera with 30,000X magnification or 50,000 magnification with between 0.5 and 2 μm defocus. Particles were picked by a combination of semi-automated swarm picking and hand-picking using the Boxer program in EMAN2²²⁶. Particles were centred, masked, and classified in IMAGIC-5²²⁷, resulting in a crude set of reference images, which were used for multireference alignment. The classes produced were significantly improved, and representative ones were used for a further stage of multireference alignment. Once a stable alignment had been achieved, poorly aligning particles and those that produced poor classes were removed to generate a new data set. Further classification was then carried out on the “best” particle stack

2.14 Screening for Crystallography

One MRC Plate 96 well 2 Drop UV Crystallization Plate (Molecular Dimensions) was used for each crystallisation screening experiment. Each well was filled with 60 μL of crystallisation buffer using a multi channel pipette. Sitting drop crystallisation trials were then set up using a NT8 robot via the RockMaker software (Formulatrix). Purified samples of AaeA were concentrated to ~ 10 mg/mL and 100 nL protein was mixed with either 100 nL or 200 nL of well reservoir solution before being placed into the top well and the plate was sealed with a ClearVue Sheet (Molecular Dimensions). The plates were stored at 20°C in a RockImager 1000 (Formulatrix), which performed automated imaging of the drops over several weeks.

Chapter 3

Production of the C-terminal Domain of TaLMT1

3.1 Introduction

The C-terminal domain (CTD) of ALMTs has been shown to be critical for regulation of channel activity and has been a subject of investigation⁸³.

However, no high-quality structural data exists and this makes biochemical characterisation and mutagenesis experiments both difficult to design and their results challenging to interpret. Thus, there is a pressing need for such structural characterisation, both of the CTD and also of the full length channels. To this end, constructs were designed to express the CTD in isolation, in the anticipation that the isolated soluble regulatory domain would prove experimentally amenable. The aim was to produce sufficiently high quantities of soluble, folded, monodisperse protein and then utilise this in structural and biochemical investigations including X-ray crystallography, binding studies, and mutagenesis then use these experiments to derive an understanding of the structure and mechanism of the regulatory CTD of ALMTs.

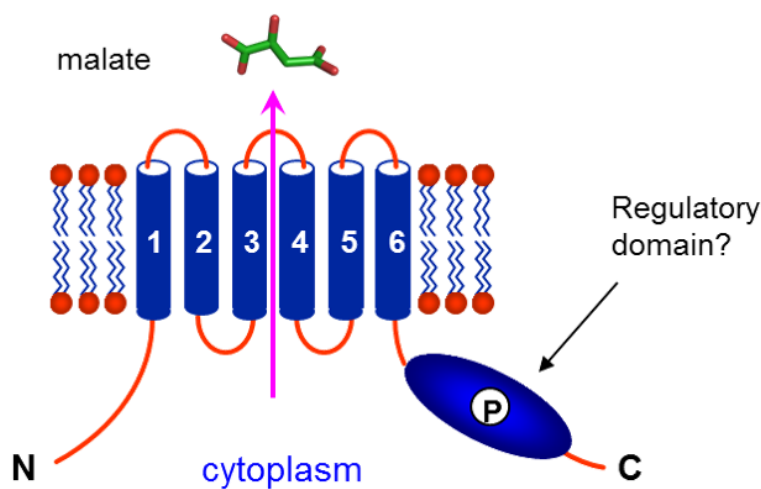


Figure 3.1 - Overall model of the domain organisation of ALMT proteins. Prepared by Steve Baldwin.

3.2 Production of the CTD of TaALMT1

3.2.1 Vector construction - pTP1 and pTP2

The first aim of this investigation was to express the CTD of *TaALMT1* in sufficient amounts for biochemical and biophysical investigations. To this end, a codon-optimised sequence for the CTD of *TaALMT1* (residues 219-433, shown in orange in

Figure 3.2B) beginning after the last predicted transmembrane segment and excluding the final 26 residues, which were predicted to be unstructured, was synthesised by Genscript and provided in the vector pL51. This was then subcloned from this vector and ligated into a pET28(b) expression vector as described in the following section. As it has been shown that the placement of affinity tags onto proteins can affect expression levels¹⁵¹, it was decided to produce two variants, with each having the affinity tag placed at opposing ends. pET28(b) also allows for a thrombin cleavage site between the N-terminal tag and the target protein, enabling later cleavage to return an untagged protein. The construct containing a C-terminal tag was designated **pTP1**, and the construct with N-terminal tag and thrombin cleavage site was designated **pTP2**.

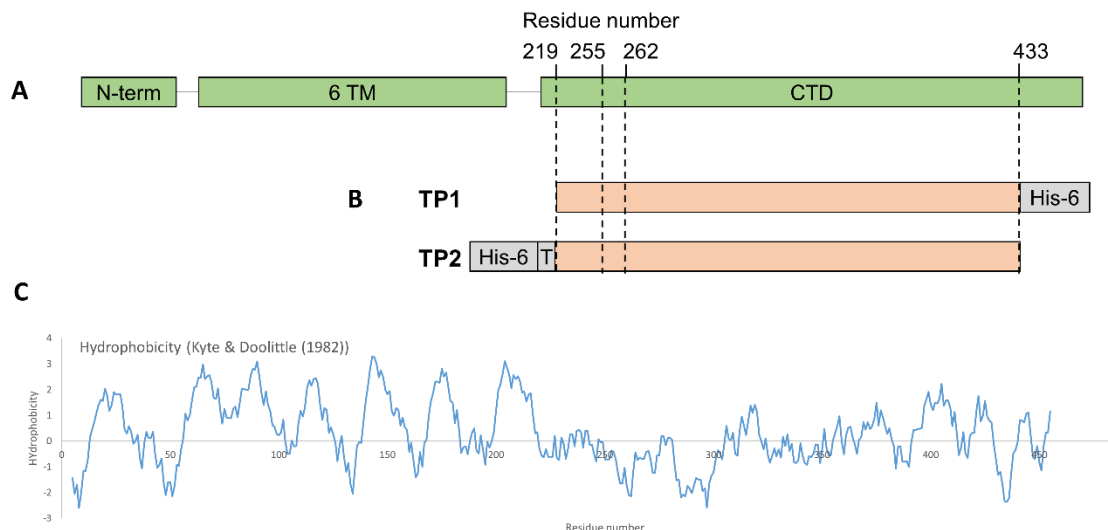


Figure 3.2 – TP1 and TP2 vectors

A) Schematic to show overall domain architecture of TaALMT1, consisting of N-terminal helix, six transmembrane helices, and a large C-terminal domain. **B)** Schematic of constructs TP1 and TP2, with C-terminal and N-terminal hexahistidine tags (**His-6**), respectively. TP2 also incorporated a thrombin cleavage site (**T**) between the tag and the coding region. It can be seen that the constructs encode a sequence that begins just after the final TM helix. **C)** Hydrophobicity across residues in TaALMT1. Graphs created from ProtParam (<http://web.expasy.org/protparam>)

By selection of restriction sites used for subcloning into the expression vector pET28(b) the two constructs were generated. Use of the restriction sites *NcoI* and *XhoI* yielded **pTP1**, an example of the digests is shown in **Figure 3.3**. By contrast, in the construction of **pTP2**, pET28(b) was digested at the sites *NheI* and *SalI*, while *NheI* is present within the coding sequence for the CTD and so could not be used to digest the insert. Thus the enzyme *AvrII* was used as it produces a compatible end with *NheI*. Restriction enzymes were initially tested on a small scale (~500 ng plasmid) for activity, and then digestions were scaled

up (~3 µg plasmid) and bands were excised from a 0.5% agarose gel and purified.

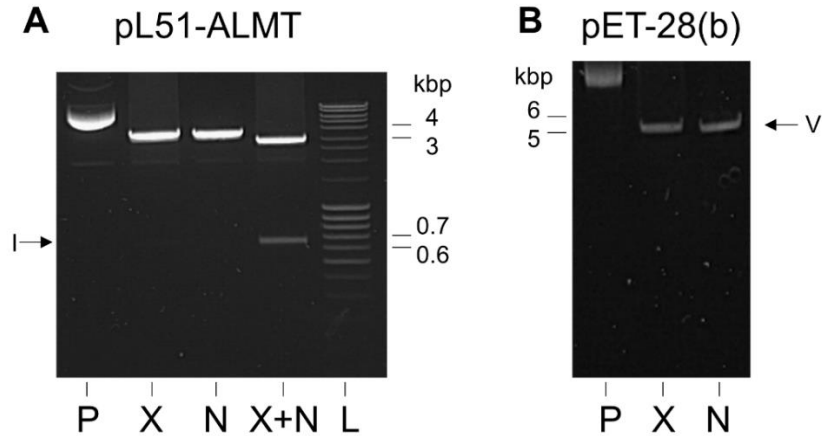


Figure 3.3 – Digests to show pTP1 construction.

A) pL51-ALMT, undigested (P), digested with XhoI (X), digested with NcoI (N), double digested with XhoI and NcoI (X+N), and DNA size marker. Insert containing ORF at approximately 700 bp indicated by arrow on the left. **B)** pET28(b), undigested (P), digested with XhoI (X), and digested with NcoI (N). Linearised vector band indicated by arrow on the right.

Purified fragments were combined and ligated with T4 DNA Ligase overnight.

After ligation of the CDS into the expression vector and isolation of the plasmid

from OmniMax2 cells, the produced constructs were digested with restriction

enzymes to assess whether cloning had been successful. Digestion of **pTP1**

plasmid singly with *NcoI* and *XhoI* yielded a band of approximately 6 kbp,

consistent with that expected for the recombinant product, while double

digestion yielded a fragment corresponding to the expected insert size of 700

bp (**Figure 3.4A**). Similarly, double-digestion of **pTP2** with *NheI* and *SalI*

produced two bands, including one at the expected size of the insert, 700 bp

(**Figure 3.4B**). The sequence of the insert plus flanking regions was then

confirmed by DNA sequencing for both constructs in both directions using the T7F and T7R sequencing sites provided by pET28(b).

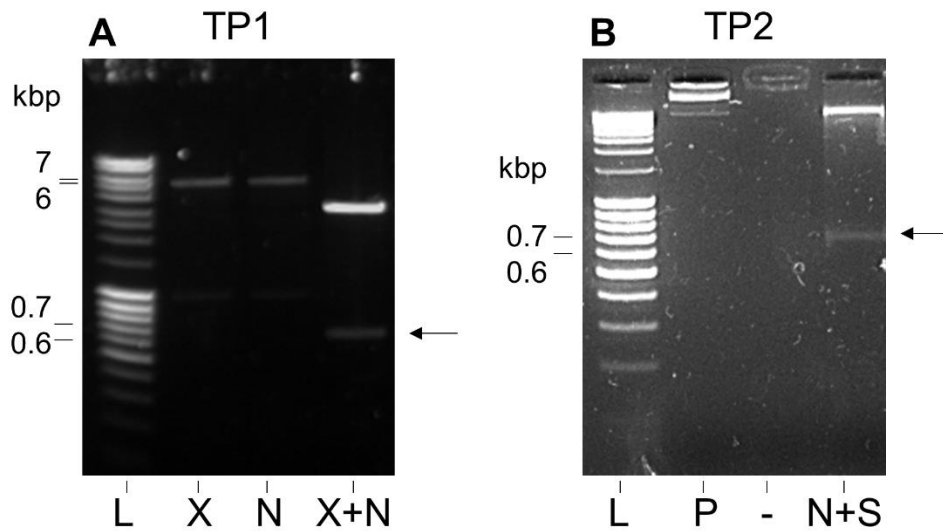


Figure 3.4 - Restriction Digests of pTP1 and pTP2 to conform correct cloning.

A) DNA size maker (L), and TP1 digested with XhoI (X), NcoI (N), or double-digested with XhoI and NcoI (X+N). Bands are seen at approximately 6000 bp for the single digests- the correct size for the full plasmid, and approximately 700 bp for the double digest – the size of the inserted ORF (indicated with an arrow). **B)** undigested TP2 (P), and TP2 digested with NheI and SmaI (N+S), which reveals a band at approximately 700 bp as expected for excision of the inserted ORF.

3.2.2 Expression trials from pTP1 and pTP2

3.2.2.1 Initial expression trial

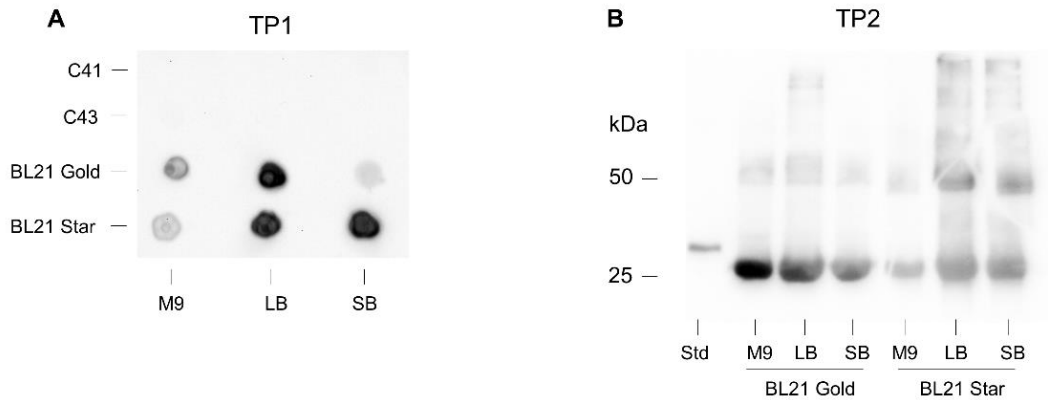


Figure 3.5 – Dot blot and Western blot of His-tagged protein expressed from pTP1

A) Dot blot of initial expression test of pTP1 in 12 conditions, 24h after autoinduction. 3 μ L of each sample was applied to nitrocellulose membrane and the blot was stained for the presence of the His₆-tag. **B)** Western blot of the six best conditions for TP2.

Once the expression constructs were confirmed by sequencing, an examination of the optimal expression conditions was conducted, starting on a 3 ml scale on a range of four *E. coli* expression strains (C41(DE3), C43(DE3), BL21-Gold (DE3), and BL21 Star (DE3)) and three media (M9, LB, and SB). Initial testing of protein expression following autoinduction for 24 h at 37°C by dot blotting cellular lysate (**Figure 3.5A**) showed that the most successful three conditions for expressing both TP1 and TP2 were BL21 Gold in LB medium and BL21 Star in either LB or SB medium. These conditions were taken forward for tests on a 50 ml scale to further optimise conditions and ensure that expression levels were maintained upon scale-up. Blots such as **Figure 3.5B** showed good expression levels, and the bands seen on the Western blot ran at the expected size of approximately 25 kDa, with a second His-reactive band at around 50

kDa, perhaps giving early evidence of multimerisation. Quantification of the bands seen in the Western blot, by densitometry, show approximately 10-30 mg protein expressed per L of culture.

3.2.2.2 Solubility Screen

In addition to assessing the conditions for their levels of total protein produced, the material was checked for solubility. This was done by two methods, firstly, centrifugation at 14,000 *g* and secondly, filtration through a 0.2 µm membrane in order to remove any particulate matter from the sample. As seen in **Figure 3.6A**, the material produced was largely insoluble, despite the initial levels of protein expression seeming promising.

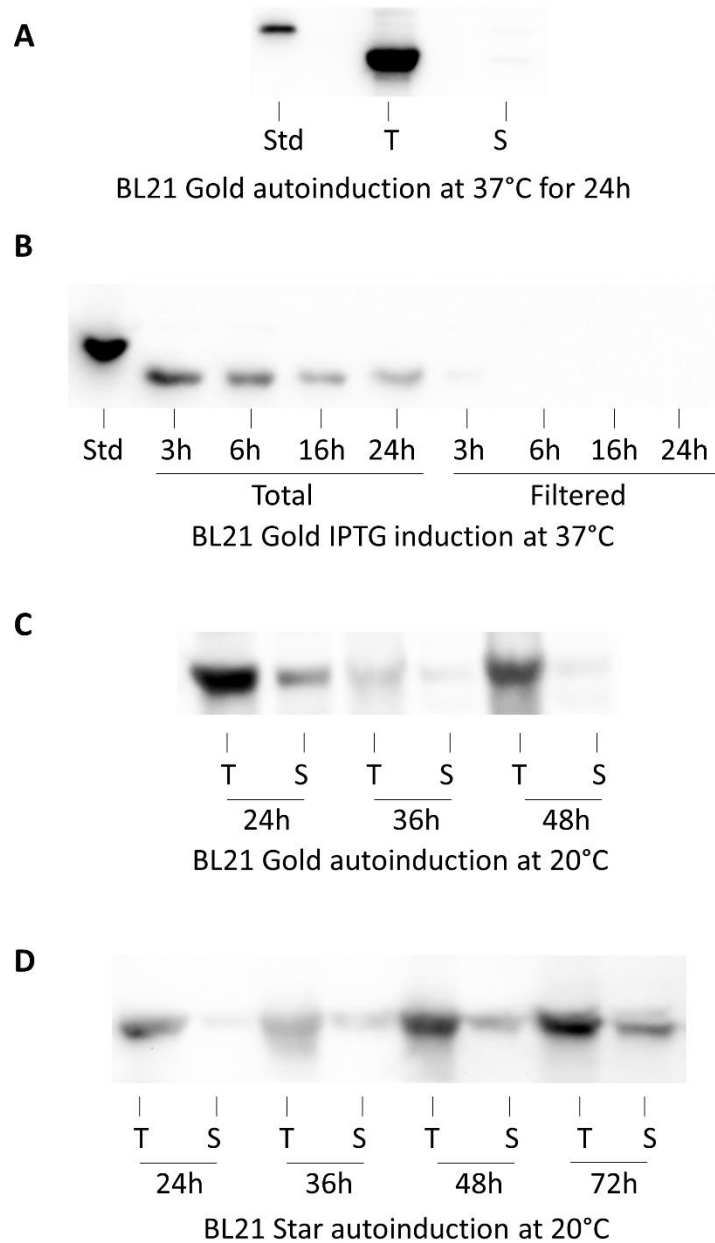


Figure 3.6 - Selection of Western blots comparing total and soluble fractions from TP1 expressed over a range of time-points in either BL21 Gold or BL21 Star in LB media.

A) Std = 125ng. While there is a significant signal in the total (T) fraction, almost no signal is seen from the soluble (S) fraction. **B)** Std = 125ng. IPTG expression does not result in an improvement in solubility. **C)** and **D)** expression at a lower temperature (20°C) may result in a slight improvement in solubility from protein expressed in both BL21 Gold and BL21 Star.

At this stage, the screening of conditions was widened to encompass a range of variables. **Table 3.1** summarises the conditions assessed for both **TP1** and **TP2** but, briefly, these included media, induction type, and cell type, as previously, but also a wider range of time-points were taken. As seen in a representative example given in **Figure 3.6B**, this was largely unproductive, with almost all conditions failing to show any signal in their soluble or filtered fractions.

An additional variable that was changed was a reduction in temperature of the expression cultures, to 20°C and 25°C; this reduction in temperature was hoped to be effective by slowing the rate of translation and thus allowing the cells to have more time to correctly fold the protein, producing natively-folded protein, albeit with a lower yield. **Figure 3.6C** and **Figure 3.6D** show that this reduction in temperature seemed somewhat effective, with a larger proportion of the produced material falling in the soluble fraction for some conditions.

Table 3.1 – Table listing expression conditions attempted for TP1 and TP2

Strain	Media	Auto/IPTG	Time (hours)	Temperature (C)
C41	M9	Autoinduction	24	37
	LB			
	SB			
C43	M9	Autoinduction	24	37
	LB			
	SB			
BL21 Gold	M9	Autoinduction	24	37
	LB			
	SB			
	LB	Autoinduction	48	37
			60	
	LB	IPTG	3	37
			6	
			16	
			24	
	LB	IPTG	4	20
			7	
			21	
			26	
	LB	Autoinduction	24	20
34				
48				
BL21 Star	M9	Autoinduction	24	37
	LB	Autoinduction	24	37
			48	
			60	
	SB	Autoinduction	24	37
			48	
			60	
	LB	IPTG	3	37
			6	
			16	
			24	
	SB	IPTG	3	37
			6	
			16	
			24	
	LB	IPTG	4	20
			7	
			21	
			26	
LB	Autoinduction	24	20	
		34		
		48		

3.2.2.3 Lysis screen

Alongside this screening, an assessment was also made as to the most effective way to extract protein from the cells, either by chemical lysis with detergents to disrupt the membrane, or by use of the cell disruptor using mechanical force to lyse cells. Equivalent samples were lysed using each method and protein signals were compared via a dot blot. The signal seen from the sample lysed by the cell disruptor was greater as seen in **Figure 3.7** and thus this was chosen as the method of choice for future protein extraction.

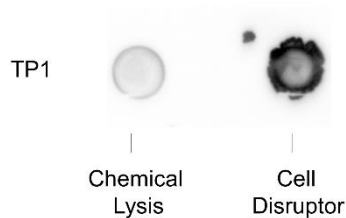


Figure 3.7 - Dot blot assessing the effectiveness of chemical lysis with Triton against lysis in a cell disruptor.

The solubility of the two protein fragments expressed from **pTP1** and **pTP2** was low, thus far. This was despite the screening of a large range of expression conditions. It was suspected that the constructs themselves were the source of the insolubility, and as such a redesign of the constructs was considered.

3.3 Production of a Truncated Construct for the CTD

3.3.1 Mutagenesis to Truncate pTP1 and pTP2

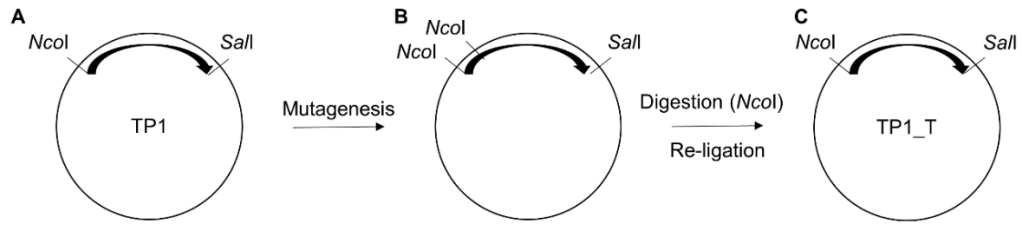


Figure 3.8 Overview of mutagenesis strategy with pTP1 as an example.

Mutagenesis is used to insert an additional NcoI site within the coding sequence (**B**). This product was then digested with NcoI, purified, and re-ligated to form an expression vector with a truncated coding region (**C**).

The constructs encoded by **pTP1** and **pTP2** both begin very soon after the final predicted TM helix of ALMT1, and so it was suspected the N-terminal part of the constructs could be part of a membrane embedded region, causing the aggregation seen in **Section 3.2.2.2**. Thus, PCR-based QuikChange mutagenesis was employed to add a restriction site to the coding region (NcoI for **pTP1** and NheI for **pTP2**) at sites identified near the 5' end of the coding sequence, as shown in **Figure 3.8**. This enabled the plasmid to be cut by restriction digestion to remove a short segment of the coding sequence, before re-ligation of the compatible sites completed the truncation of the construct and production of two new plasmids, designated **pTP1_T** and **pTP2_T**. This removed 36 or 45 residues, respectively and ensured that expression started in a hydrophilic part of the CTD, shown in **Figure 3.9**.

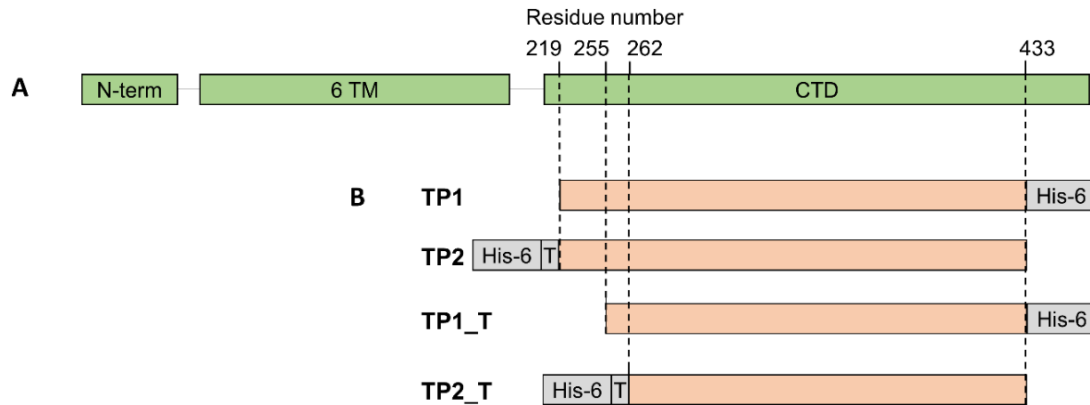


Figure 3.9 -Truncated vectors overview.

A) Schematic to show overall domain architecture of TaALMT1, consisting of N-terminal helix, six transmembrane helices, and a large C-terminal domain. **B)** Schematic of constructs, including the truncated versions TP1_T and TP2_T, with C-terminal and N-terminal His-tags, respectively.

A gradient PCR was performed between 55°C and 65°C for each mutagenesis (see **Figure 3.10**). A product of size ~7000bp was seen as expected in the gel. After digestion with DpnI to remove the original template DNA (which would be methylated as it was prepared from *E. coli*, and thus is a substrate for the enzyme, unlike the newly produced DNA), the DNA was purified transformed into OmniMax2 to allow isolation of sufficient plasmid via miniprep.

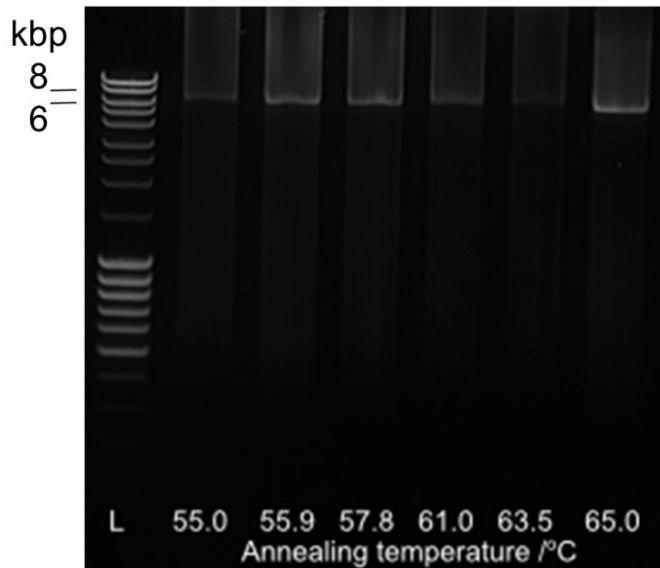


Figure 3.10 - Characterisation of products produced by QuikChange mutagenesis.

Samples were separated by electrophoresis on a 1% agarose gel of products after gradient QuikChange PCR to add an NcoI restriction site to the coding region of pTP1. All temperatures produced equivalent results. Approximate size markers are indicated on the left.

These mutagenised plasmids were then digested at the inserted restriction site, purified, and re-ligated together with T4 ligase, and isolated from OmniMax2 cells. After isolation, the final products were sequenced and this confirmed that the truncation was successful for both **pTP1_T** and **pTP2_T**.

3.3.2 Expression of Truncated Constructs

Similarly to **TP1** and **TP2**, a range of culture conditions were tested to optimise expression of the truncated constructs. BL21(DE3) Gold and BL21(DE3) Star were approximately equivalent in expression with LB being the best medium, as shown in **Figure 3.11A**. The six samples from BL21 Gold and BL21 Star were run on an SDS-PAGE gel and assessed by western blotting. The size was approximately 20 kDa for **TP1_T**, as expected, with **TP2_T** being slightly

smaller (data not shown) confirming that truncation had been successful. Again, LB media appeared to give the best expression for each cell type. Expression tests on **TP1_T** show that the protein expresses well, to similar levels seen for **TP1** and **TP2**. Expression levels seen in **Figure 3.11B** are at approximately 10 mg protein per L of culture, as quantified by densitometry.

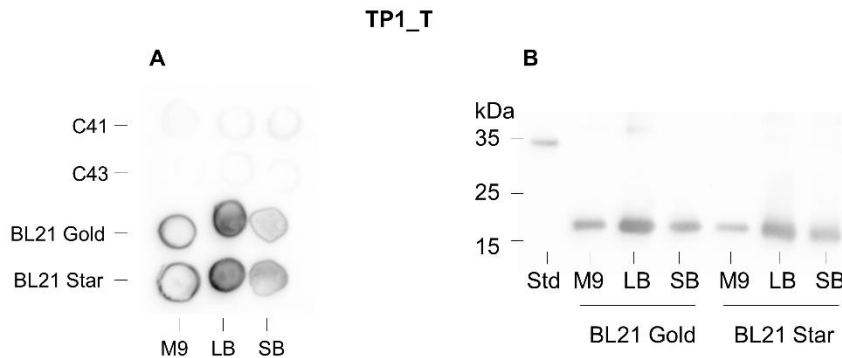


Figure 3.11 –Blots to assess truncated TP1_T

A) dot blot of 2 μ L samples of small scale expression trial of four strains and three media for TP1_T expressed by 24h of autoinduction n a 3 ml scale. BL21 Gold and BL21 Star show the best levels of expression. **B)** Western blot to assess expression of TP1_T in BL21 Gold and BL21 Star using 24h autoinduction 10 μ L of each sample was loaded and samples were resolved by SDS-PAGE, transferred to nitrocellulose and then stained for the presence of the His6 tag. Std = TeV-His6 standard, 125 ng.

The next step was to assess the solubility of the expressed protein, as before. BL21Gold and BL21 Star cells were used to express the constructs using autoinduction and IPTG induction over a range of time points between 5 h and 48 h and at both 37°C and 25°C detailed in **Table 3.2**. Cells were lysed in the cell disruptor and split into three fractions: firstly, a total fraction; secondly, a soluble fraction which was centrifuged at 14,000 g before taking the supernatant; and thirdly, a sample that was filtered through a 0.2 μ m membrane. It is clear from **Figure 3.12** (results representative for all conditions) that the truncated constructs were not markedly more soluble than the originals.

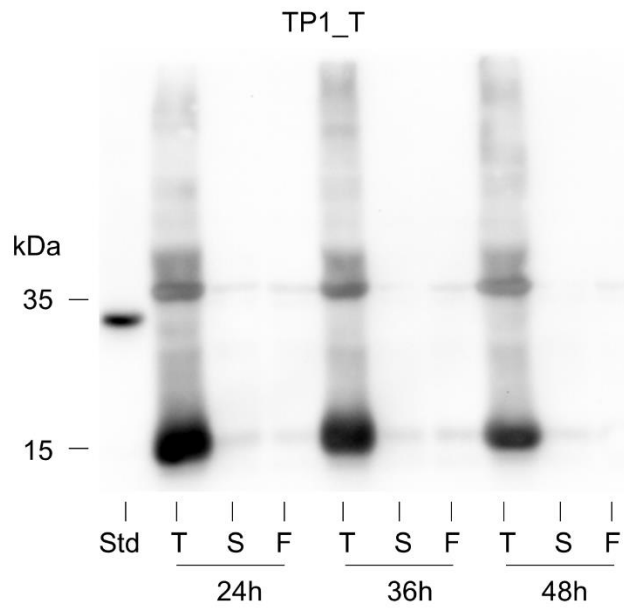


Figure 3.12 - Western blot to assess expression of TP1_T in BL21 Star using autoinduction for three time points.

10 μ L of each sample was loaded and samples were resolved by SDS-PAGE, transferred to nitrocellulose and then stained for the presence of the His6 tag. T=total, S=centrifuged at 14,000 g, F=filtered through 0.2 μ m filter. A large amount of expression is seen, but a low proportion is seen to be soluble after centrifugation or filtration. Approximate position of MW markers is indicated on the left.

Table 3.2 - Table listing expression conditions attempted for TP1_T and TP2_T

Strain	Media	Auto/IPTG	Time (hr)	Temperature (C)
C41	M9	auto	24	37
	LB			
	SB			
C43	M9	auto	24	37
	LB			
	SB			
BL21 Gold	M9	auto	24	37
	LB			
	SB			
BL21 Star	M9	auto	24	37
	LB			
	SB			
BL21 Gold	LB	IPTG	5	37
			9	
			20	
			30	
BL21 Star	LB	IPTG	5	37
			9	
			20	
			30	
BL21 Gold	LB	auto	24	37
			36	
			48	
C41	LB	auto	24	25
C43				
BL21 Gold				
BL21 Star				

3.3.3 Testing additives to aid solubility

A concurrent project in the lab that was also attempting to resolve a solubility problem while expressing an isolated protein domain found that adding Triton X-100, EDTA, and DTT after cell lysis helped to stabilise the protein produced. Thus, this was attempted for TP1_T, and as seen in **Figure 3.13A** this was able to solubilise some of the protein (by densitometry, approximately 20%). This presented a problem though as DTT and EDTA are incompatible with the standard IMAC resins used for purification of His-tagged proteins. Thus a follow-up solubility test was attempted to assess whether all of these additives were strictly required. If the EDTA and DTT were removed, or the Triton concentration was decreased, the solubility of TP1_T was sharply decreased, as seen in **Figure 3.13B**. So, it was concluded that these components are strictly required in order to aid solubility.

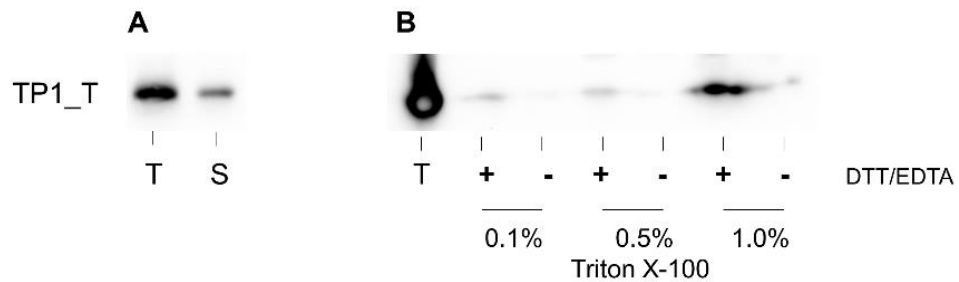


Figure 3.13 – Assessment of the effects of additives to the solubility of TP1_T.

A) Western blot of Total (T) and Soluble (S) fraction of TP1_T when assessing additives to aid solubilisation. Addition of 1% Triton X-100, 1 mM DTT and 5 mM EDTA enables partial solubilisation of TP1_T. **B)** Western blot to assess whether any of the additives can be omitted from the solubilisation buffer. Solubilisation is unsuccessful if either EDTA and DTT are omitted from the solubilisation buffer, or the level of Triton X-100 is decreased.

3.3.3.1 Attempted purification of the solubilised truncated CTD

Conditions had been found to somewhat enhance the solubility of the produced protein. The next step was to perform a purification to assess how viable it would be to produce purified soluble protein, whether this would be possible and whether it would be viable on a large enough scale. As it was necessary to include DTT and EDTA in the protein solution, cOmplete His-Tag Purification Resin (Roche) was used, as this is resistant to reducing agents and chelators.

Each of **TP1**, **TP2**, **TP1_T**, and **TP2_T** were purified using IMAC, binding protein to the resin for 1-2 hours and then washing with concentrations of imidazole (generally between 10 mM and 50 mM) before eluting the protein using 500 mM imidazole. In general, purifications showed some promise, an example is shown in **Figure 3.14** for **TP1_T**. The target protein generally bound to the resin and eluted in the appropriate fractions, and although the purity needed improvement there was some potential shown if enough soluble protein could be produced, as there is scope for significantly more washing with imidazole before elution.

Unfortunately, after purification and upon dialysis to remove the imidazole from the eluted fractions, the protein began to precipitate. This indicated that the protein was still not stable, and though the buffer conditions had seemingly improved solubility, the protein was still not in a suitable state for further studies.

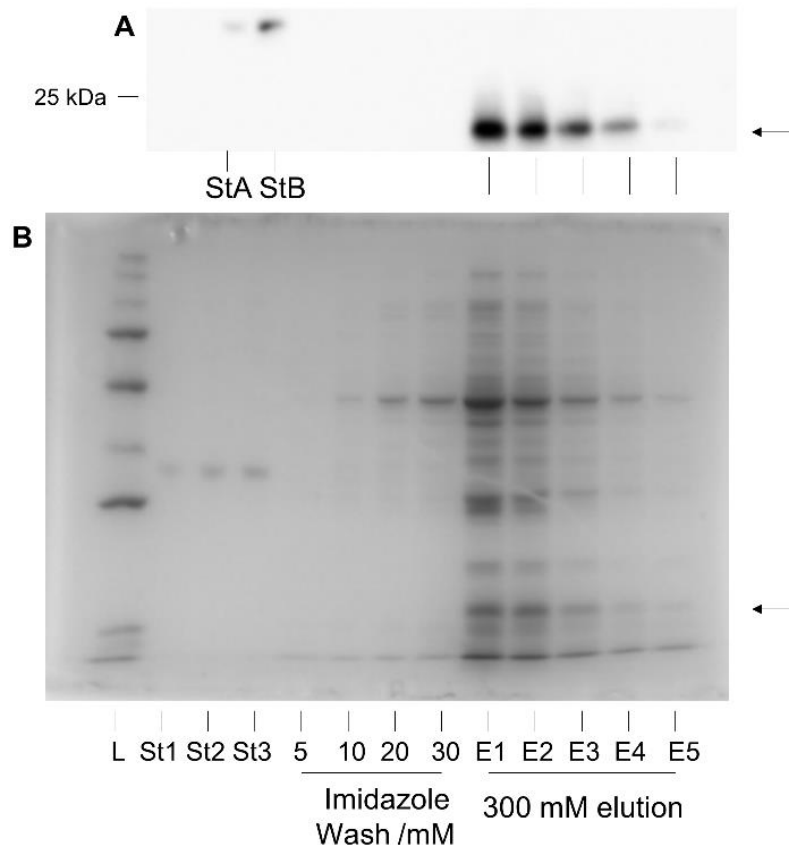


Figure 3.14 - A purification of a sample of TP1_T containing triton, EDTA, DTT.

A) Western blot for the His-tag, StA = 50 ng of His-tagged protein standard, StB = 100 ng of His-tagged protein standard. Other lanes are as marked on the gel below. **B)** Coomassie stained SDS-PAGE gel of purification fractions (20 μ L). Molecular weight maker (L), 36 ng protein standard (St1), 52 ng protein standard (St2), 112 ng protein standard (St3), washes of increasing imidazole concentration (5-30), and elutions at 300mM imidazole(E1-E5). Position of TP1_T indicated by arrows. Although TP1_T correctly binds to and is eluted from the Ni-NTA resin, a lot of contamination remains and more purification is required.

3.3.4 Denaturation and Refolding

As a final attempt to produce the CTD in a stable, soluble form refolding of the inclusion bodies was attempted following previously published methods^{138,139}. Briefly, inclusion bodies were prepared by pelleting in a centrifuge and washing with detergent and 1M NaCl, before the pellet was incubated at room temperature for 1h with 6M guanidinium chloride (GdnCl), to solubilise and denature the protein. Any insoluble material was removed by centrifugation at 100,000 *g* and then the supernatant was incubated with cOmpete Ni IMAC resin overnight. The protein was then refolded on the column by washing with buffer not containing GdnCl. However, upon buffer exchange a sandy precipitate formed in the column, indicating that refolding was unsuccessful. Protein could not be eluted from the column with 300 mM imidazole, as seen in **Figure 3.15A**. The column was re-washed with buffer containing 6M GdnCl plus 300mM imidazole to clean the column, and this material was dialysed to remove the GdnCl and imidazole. This dialysis again resulted in the crashing out of a sandy precipitate (**Figure 3.16**) and both the soluble and a sample of the re-dissolved precipitate were assessed by a dot blot. In **Figure 3.15B** it can be seen that all the **TP1** is in the insoluble precipitated material washed out of the column by the denaturing elution buffer and thus **TP1** is insoluble and cannot be refolded.

At this point the prospect of producing the material required did not seem promising so this line of enquiry was abandoned.

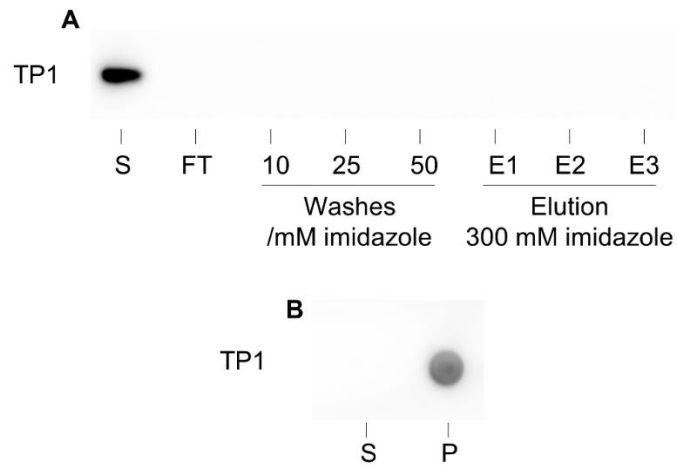


Figure 3.15 – Assessment of refolding TP1.

A) Western blot of fractions of a purification of TP1 refolded on the column. No target protein is seen in the washes or the elution fractions. **B)** Dot blot assessing the presence of His-tagged TP1 in the supernatant (**S**) and precipitate (**P**) from re-folded TP1. It can be seen that TP1 has not correctly refolded.

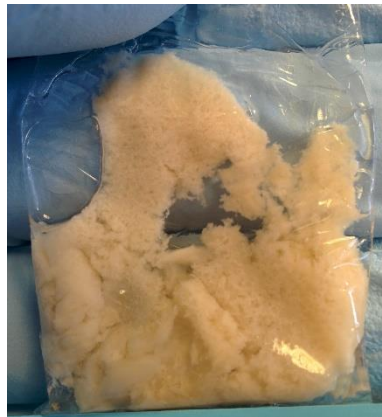


Figure 3.16 - Photograph of re-folded TP1 after dialysis. All material precipitates.

3.4 Discussion and Conclusions

Production of constructs to express the CTD of TaALMT1 (residues 219-433) with a C-terminal or N-terminal hexahistidine tag for **TP1** and **TP2**, respectively, was successful and protein could be expressed in significant amounts: above 10 mg protein per L of culture. Similarly, expression from constructs encoding a truncated CTD (residues 255-433 for **TP1_T** and 262-433 for **TP2_T**) lacking a region predicted to be disordered was successful. Unfortunately, the CTD had consistently proved to be impossible to isolate in a soluble form despite attempting a large variety of conditions. An explanation as to why is presented in analysis by Dreyer *et al.*⁴⁵ published after the conception of these experiments, summarised in **Figure 3.17**. Their bioinformatics study suggested that along with the 6 TM helices in the N-terminal half of the protein, there are potentially two further TM helices within the CTD. The position of these helices is indicated in red between **Figure 3.17B** and **Figure 3.17C** and matches well with more-hydrophobic stretches of the TaALMT1 sequence. If there really are two TM helices within the constructs expressed within this chapter then that would explain the extreme difficulty in obtaining soluble protein, and the observation that the detergent Triton X-100 was able to help solubility somewhat.

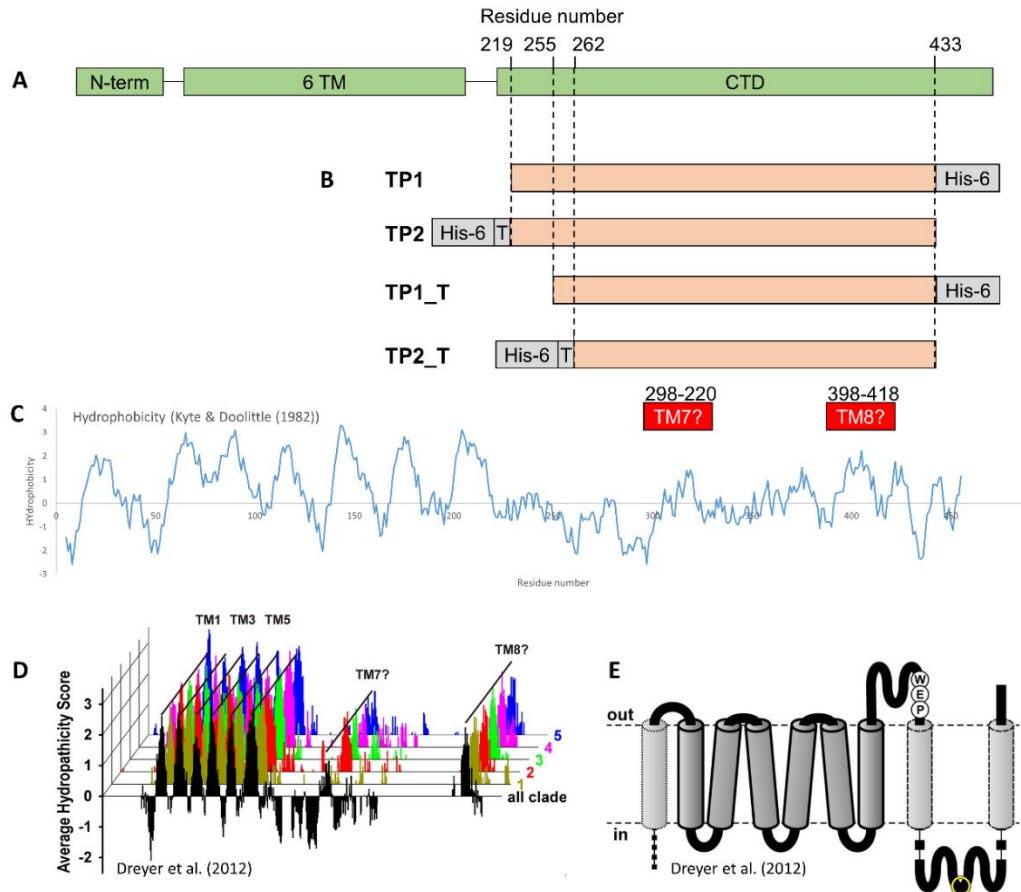


Figure 3.17 – Putative TM helices in the CTD of TaALMT1.

A) Overall schematic of TaALMT1 at the outset of this investigation. **B)** Schematic to show position of constructs expressed in this chapter. The location of two putative TM helices is shown below in red. All constructs contain these regions. **C)** Hydrophobicity analysis of TaALMT1 with position of two putative extra TM helices above in red. The location of these regions matches with regions identified as more-hydrophobic. **D)** Hydrophobicity analysis by Dreyer *et al.*⁴⁵ over 300 ALMT/QUAC family members identifying two extra putative TM regions within the CTD. **E)** Revised topology model as presented by Dreyer *et al.* containing two putative transmembrane helices in the previously-though soluble CTD⁴⁵.

The sequences identified as putative TM7 and TM8 are given in **Figure 3.18**. It can be seen that these sequences contain a high proportion of hydrophobic residues, as expected. Although there are charged residues within these sequences it is possible for these to have a role within a TM helix. For example, they could be involved in formation of a salt bridge, as seen between TM1 and

TM2 of AtALMT9⁹³. Indeed, for a channel protein there must be a hydrophilic cavity through the membrane that is solvent-accessible and thus such hydrophilic residues are entirely plausible.

TM7	W	S	Q	Y	Q	G	K	L	G	T	L	C	R	Q	C	A	S	S	M	E	A	L	A	S	Y	V	I
TM8	T	A	L	L	Q	V	M	H	V	A	V	T	A	T	L	L	A	D	L	V	D	R	V	K	E	I	A

Figure 3.18 - Sequences of two putative TM helices identified by Dreyer *et al.*⁴⁵

Hydrophobic residues highlighted in blue, charged residues are highlighted in gold and polar residues in white.

It remains to be proved whether or not the two additional TM helices predicted by Dreyer *et al.* truly are embedded within the membrane, but the hydrophobicity of the regions – which is consistent across all clades – and the strong conservation of the sequence in this region also reported by Dreyer *et al.* supports the hypothesis. The data presented here showing that the isolated CTD is not a soluble domain, even with truncation, are also consistent with this. Indeed, the findings explain the lack of success in producing soluble protein reported here and had the analysis been available before the conception of this project the above strategy of expressing the isolated CTD would not have been undertaken.

As production of the CTD in a stable, soluble form had proved unsuccessful and bioinformatics analyses by other groups had indicated that the CTD was probably not a soluble domain as we had initially expected it was decided that this approach would not produce fruitful results and as such it was decided to move on and attempt to express full-length ALMTs.

Chapter 4

Expression of ALMTs in Tobacco

4.1 Introduction

Given that production of correctly folded soluble protein from the constructs created in the previous chapter to express the C-Terminal Domain of TaALMT1 had proved impossible, an approach to express the full-length ALMTs was devised. Successful expression of these proteins and production in quantities suitable for structural studies would enable the study of ALMTs towards a detailed understanding of their structure and mechanism. Membrane proteins represent a set of high-value targets for structural studies. However, relatively few membrane protein crystal structures have been solved at present: only 604 membrane proteins are found from a total of 37,855 distinct protein structures recorded in the Protein Data Bank²²⁸ on 21 March 2016 (<http://www.rcsb.org/pdb/home/home.do>), and of these only 7 plant membrane proteins are represented. Thus, membrane proteins – and especially plant membrane proteins – represent an under-explored set of targets. ALMTs play a large variety of roles in plants, are ubiquitously expressed in all plants, and have no detailed structural data available for any family member nor for any homologue that could be used for *in silico* modelling of the rest of the family members. Thus, they present an ideal target for expression with the aim of structural studies.

4.1.1 Choice of plant expression and the pEAQ-HT vector

Expression of eukaryotic proteins for structural studies is often challenging, especially in *E. coli*, a favoured system due to its relatively low cost and its well established experimental protocols. Prokaryotic expression hosts lack the more sophisticated machinery for post-translational modification that is necessary for successful expression of functional eukaryotic membrane proteins – especially their long hydrophobic segments²²⁹. To overcome these challenges, a variety of eukaryotic expression systems exist, including yeast (both *Saccharomyces cerevisiae*²³⁰ and *Pichia pastoris*²³¹), insect (via the baculovirus system²³²), and mammalian²³³. More recently, a system for expression of proteins in plants has been developed utilising a viral-based cassette for high translational efficiency – the pEAQ-HT vector for transient expression via agro-infiltration^{132,234}.

Of the varying options available, expression *in planta* is an attractive route for successful production of plant proteins for structural studies. A plant-based expression system should be able to undertake all necessary steps including: expression, folding, posttranslational modification, trafficking to the correct intracellular compartment and insertion into the membrane, as these will be occurring in an environment very similar to the native host. Recently, differences have also been shown in the results obtained from plant-based and animal-based expression, with animal cells not producing functional protein⁷⁵.

Transient expression allows for the production of high amounts of recombinant proteins in a matter of days without the need to produce transgenic plants, and for several constructs and conditions to be screened relatively rapidly. Transient expression is also useful when expressing proteins that can have deleterious effects that would otherwise affect growth of the plants, as expression does not

occur until the plants are sufficiently grown. This can be achieved by using *Agrobacterium tumefaciens* carrying an expression plasmid, such as pEAQ-HT¹³² with an ORF of interest inserted. For this method, the intercellular space within the plant leaves is flooded with a suspension of bacteria in a process known as agro-infiltration, using a syringe (for small scale experiments) or by vacuum infiltration (if larger amounts of material are required)²²⁰.

As shown in **Figure 4.1**, the pEAQ-HT viral-based vector contains a modified 5'-untranslated region (UTR) and 3'-UTR from Cowpea mosaic virus (CPMV) RNA-2 to flank the target sequence and drive expression. These flank the polylinker region that is used for insertion of the target gene in frame with either an N- or C-terminal His-tag, which permits identification of the target protein by Western blotting and allows for purification via IMAC. A key feature is the P19 suppressor of gene silencing from Tomato bushy stunt virus, which stabilises the mRNA to enable high levels of expression. This system has been used to produce viral proteins and viral-like particles²³⁵, as well as active human and plant enzymes^{236,237} and a similar CPMV-based system has been used to produce human antibodies with properties indistinguishable from those produced by mammalian expression systems with milligram-level yields²³⁸.

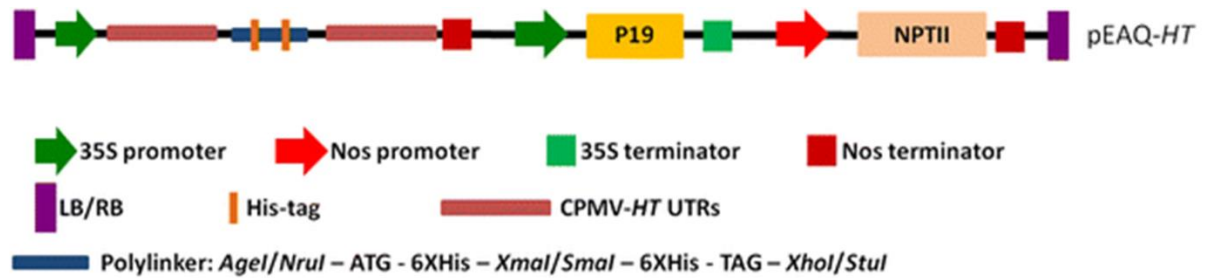


Figure 4.1 - Schematic view of pEAQ-HT sequence, highlighting key elements.

The pEAQ-HT vector contains the CPMV-*HT* expression cassette, the P19 suppressor of gene silencing, and the NPTII kanamycin resistance gene, and allows for the addition of an N- or C-terminal hexahistidine tag. Adapted from Peyret and Lomonossoff (2013)¹³².

4.1.2 Proteins chosen from ALMT family for expression

Protein targets for the present study were selected firstly from those available from the Arabidopsis Biological Resource Centre (ABRC, <https://abrc.osu.edu/>). Two cDNA clones were available from *A. thaliana*: ALMT5 and ALMT9. In addition, wheat ALMT1, the founding member of the ALMT family, was synthesised by Genscript (<http://genscript.com>) and inserted into the pUC57 cloning vector. This gave a total of three genes (as shown boxed in **Figure 4.2**): TaALMT1, plasma membrane localised and AI-responsive; AtALMT9, vacuolar-localised and AI-insensitive; and AtALMT5 of unknown function. Structural studies on these three proteins is of interest to find differences that explain their different functions, and perhaps the structural knowledge can help to suggest a functional role for AtALMT5 which can be tested in future studies, as well as other uncharacterised family members in future.

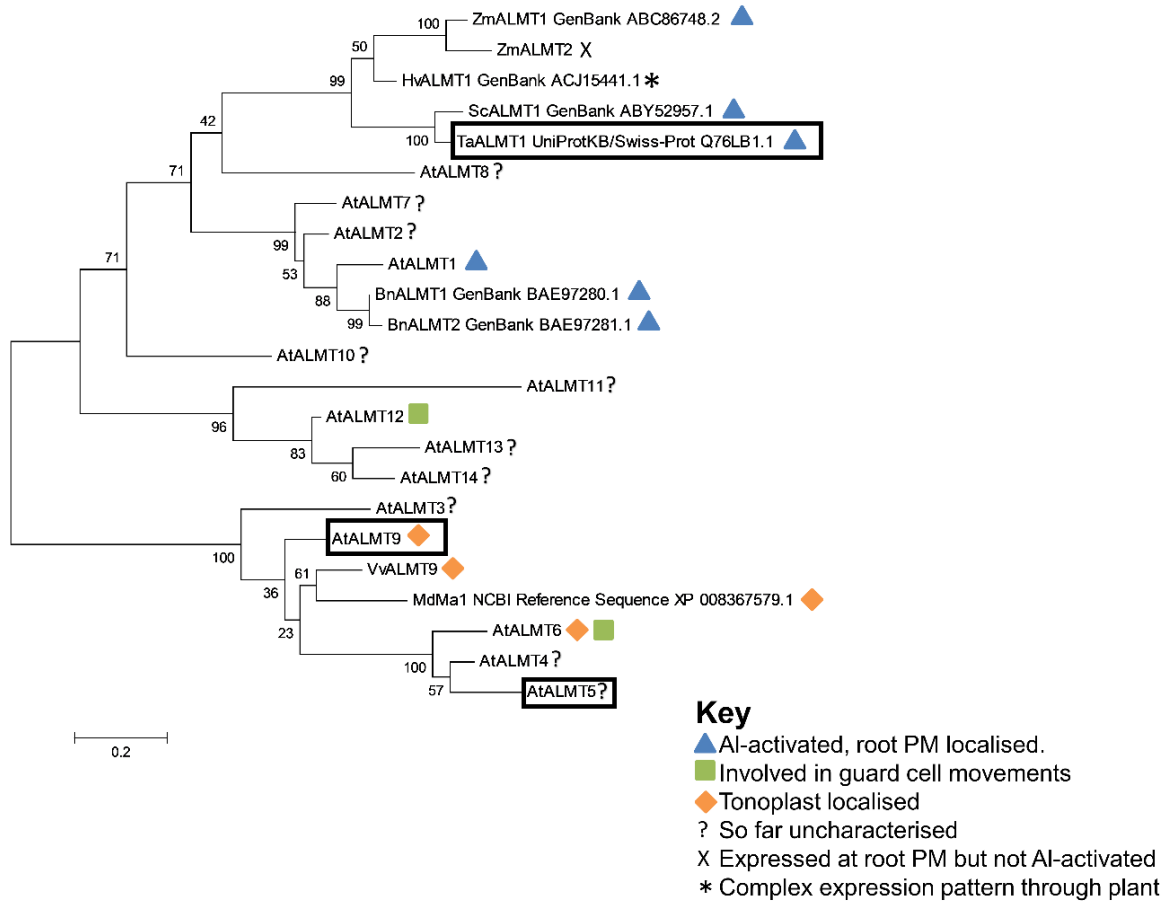


Figure 4.2 Phylogenetic tree of ALMTs with known functions plus family members from Arabidopsis with unknown function. Proteins studied herein are boxed.

Numbers on nodes denote likelihood of correctness. Tree made using MEGA6 software, sequences were aligned by the inbuilt MUSCLE functionality, and constructed with the Maximum Likelihood method and 500 iterations of bootstrapping. Figure adapted from Palmer *et al.*³⁷

4.2 Cloning

In order to produce plasmid vectors for the transient expression of ALMTs in tobacco leaves the coding sequences for three genes of interest were obtained. Firstly, *AtALMT5* and *AtALMT9* obtained from Arabidopsis Biological Resource Center (ABRC), were transformed into chemically competent OmniMax2 *E. coli* by the heat shock method, and then the plasmids were isolated for use in the present study. Secondly, *TaALMT1* was synthesised by Genscript with the addition of flanking restriction sites to allow insertion into pEAQ-HT at the sites which enable construction of N-terminally tagged protein (i.e. *Xma*I/*Xho*I) and inserted into the standard pUC57 cloning vector. The pEAQ-HT expression vector was obtained from Prof. George Lomonosoff, John Innes Centre.

4.2.1 Primer design

The pEAQ-HT vector is designed to enable production of proteins with either an N- or C-terminal His tag. By use of the *Xma*I and *Xho*I restriction sites a protein fusion is produced with an N-terminal His tag; conversely, use of the *Age*I and *Xma*I restriction sites generates a C-terminally tagged protein. It was decided to make both versions for each gene, as the placement of tags on proteins has been shown to affect their expression and function¹⁶⁷ and thus production of both options maximises the chance of a well-expressing functional protein. However, the sequence for *Xho*I (CTCGAG) can be found within the coding sequence for both *ALMT5* and *ALMT9*, necessitating that another restriction site be chosen for digestion of the ORF-containing insert fragments. *Psp*XI was chosen; this enzyme recognises the sequence VCTCGAGB (where V = A, C, or G; and B = C, G, or T), which is not found within the coding sequence of either *AtALMT5* or *AtALMT9* and when cut forms an overhang compatible with *Xho*I.

4.2.2 Vector construction

The open reading frames (ORFs) encoding the three target proteins were amplified from the corresponding purchased plasmids using primers detailed **Chapter 2** by means of Gradient PCR to assess a range of annealing temperatures between 60-70°C for each pair of primers. An example is shown in **Figure 4.3**. Small differences were seen in contaminant bands, but the major band in each case was of the expected size. As no strong effect was seen, annealing temperatures were set at 65°C. The resultant PCR products were then electrophoresed on a 1% agarose gel, and each amplification was shown to correspond approximately to the predicted size for the desired fragments from the ORFs (exact sizes 1329 bp for TaALMT1, 1623 bp for AtALMT5, 1809 bp for AtALMT9) as shown in **Figure 4.4**.

The products were then purified and ligated into *Stu*I-digested pCR-Blunt vector (Invitrogen) before transformation into OmniMax2 competent *E. coli* grown on LB-agar plates containing both kanamycin and IPTG which selects for growth of only colonies that contain the pCR-Blunt vector that contains an insert.

Plasmids were isolated from bacterial cultures grown from these colonies.

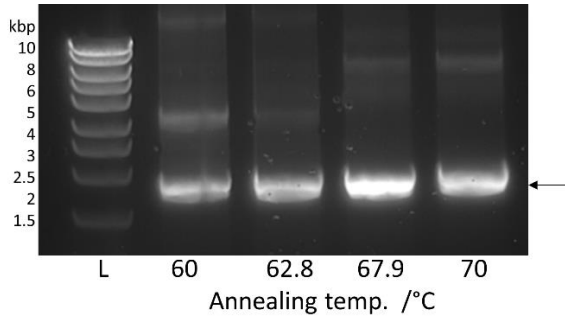


Figure 4.3 - Gradient PCR with a range of annealing temperatures to amplify AtALMT9N.

Size expected is approximately 1809 bp and this is the major band seen in each case.

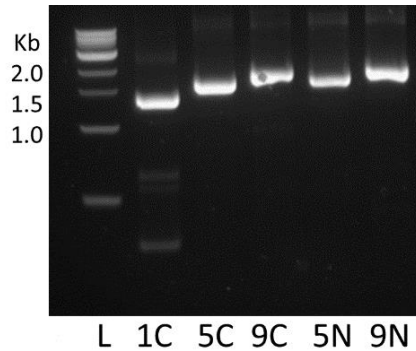


Figure 4.4 PCRs of TaALMT1C (1C), AtALMT5N (5C), AtALMT9C (9C), AtALMT5N (5N), AtALMT9N (9N).

A 1% agarose gel for PCR products. Bands are seen at approximately correct sizes (1329, 1623, 1809, 1623, 1809 bp respectively), position of known DNA MW markers are shown on left (L).

The ORFs were then digested out of their pCR-Blunt cloning vectors using appropriate restriction sites (*AgeI* and *XmaI* for C-terminally tagged constructs, *XmaI* and *PspXI* for N-terminally tagged constructs); an example is shown in **Figure 4.5**. In parallel, samples of pEAQ-HT were digested with the corresponding enzymes. Fragments were purified from a 0.5% agarose gel, and ligated together overnight at 16°C with T4 Ligase before transformation into

Omnimax2 competent *E. coli* and grown on LB-agar plates with kanamycin selection.

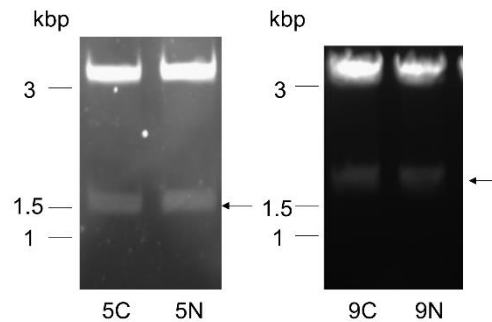


Figure 4.5 -Digests of Blunt Vectors

1% Agarose gels of pCR-Blunt-AtALMT5C digested with *AgeI* and *XmaI* (**5C**) and pCR-Blunt-AtALMT5N digested with *XmaI* and *XhoI* (**5N**) and pCR-Blunt-AtALMT9C digested with *AgeI* and *XmaI* (**9C**) and pCR-Blunt-AtALMT9N digested with *XmaI* and *XhoI* (**9N**). Samples show bands corresponding to vector (expected approximately 3519 bp) and insert indicated by an arrow (ALMT5 1623 bp and ALMT9 1809 bp).

Colonies were picked and grown in LB medium with kanamycin selection overnight then plasmids were isolated and the final constructs were assessed for the presence of an insert by double restriction digestion. Constructs encoding C-terminally tagged proteins were digested with *AgeI* and *XmaI*, while constructs encoding N-terminally tagged proteins were digested with *XmaI* and *XhoI*. Digests of final constructs are shown in **Figure 4.6** and **Figure 4.7**. An additional test for the presence of insert was by the use of Colony PCR, using the primers initially used for PCR an exemplar of the results for a successful Colony PCR is shown in **Figure 4.8**.

The final plasmids were confirmed by sequencing.

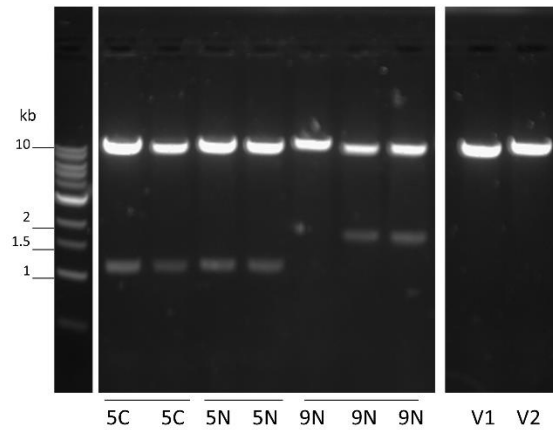


Figure 4.6 – 1% Agarose gel of digested final constructs for plant expression.

AtALMT5 with C terminal His tag (**5C**) doubly digested with *Age*I and *Xma*I, AtALML5 and AtALMT9 with N-terminal His tag (**5N**, **9N**) doubly digested with *Xma*I and *Xho*I. Bands for the insert were found in six lanes at approximately the expected size (ALMT5:1623 bp, ALMT9: 1809 bp) Also shown is the empty vector, pEAQ-HT digested singly with *Age*I (**V1**) and with *Xho*I (**V2**) to confirm enzyme activity. The gap between lanes 9N and V1 shows where empty lanes were cropped from the picture.

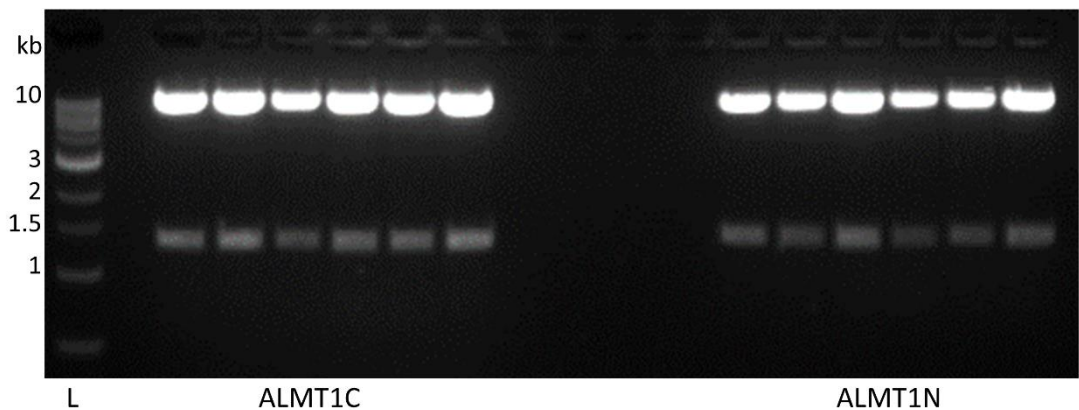


Figure 4.7 - 1% Agarose gel showing doubly digested ALMT1C and ALMT1N.

Six samples of each plasmid were prepared and samples of each were digested with either *Age*I and *Xma*I or *Xma*I and *Xho*I respectively. Bands are seen at the expected sizes in all lanes (~10,000 and ~1400 bp)

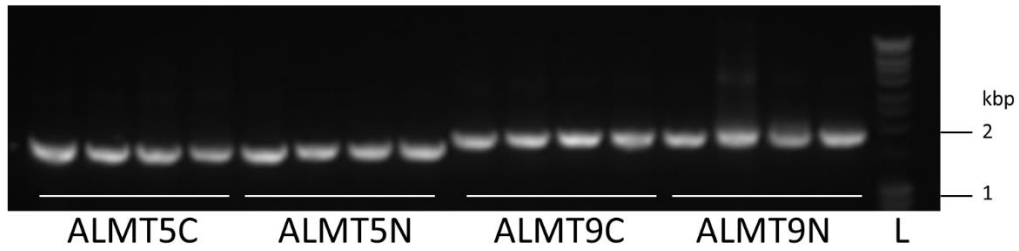


Figure 4.8 - 1% Agarose gel of 5 μ L samples of colony PCR to confirm the presence of inserts in quadruplicate.

Bands were seen at approximately the expected size for each construct (**ALMT5** ~1623, **ALMT9** ~1809) suggestive of correct incorporation of insert into the expression plasmid.

4.3 GFP expression in tobacco as a proof of principle

Before proceeding with expression of the plant ALMTs, a proof of principle trial was performed with a pEAQ-HT expression vector containing a sequence encoding GFP (also obtained from Prof. George Lomonosoff) in order to confirm that protein expression could be obtained using the system in the experimental conditions available. The following experiment was designed with reference to the method of Sparkes *et al.*²²⁰. The GFP construct was transformed into *Agrobacterium tumefaciens* LBA4404 by freeze-thaw, and cultures were grown and infiltrated into the leaves of tobacco (*Nicotiana benthamiana*) five weeks after sowing, at O.D.₆₀₀ = 0.01, 0.05, or 0.1 and left to express protein for 5 or 10 days. Control leaves were infiltrated with the Infiltration Medium only, or left to grow undisturbed.

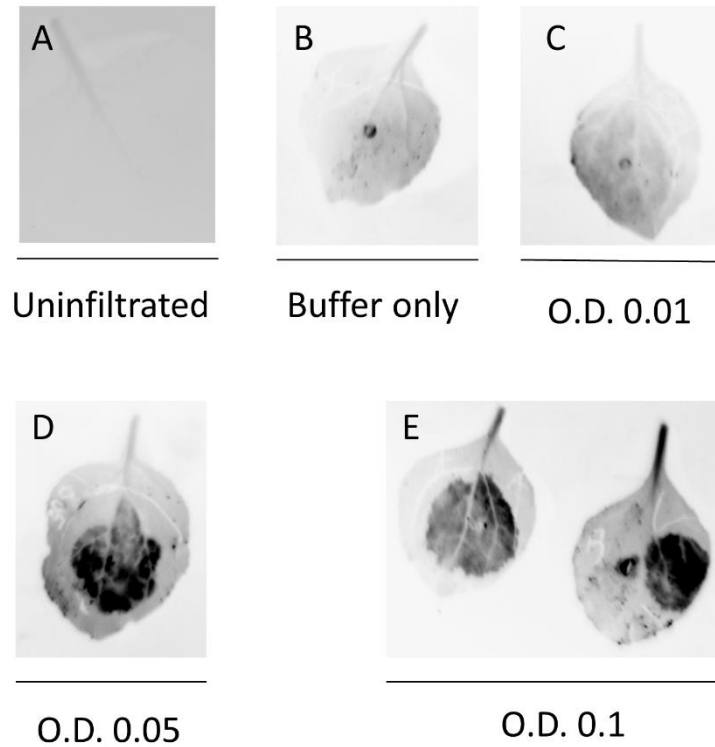


Figure 4.9 - Fluorescence image of a set of leaves infiltrated with *Agrobacterium tumefaciens* LBA4404 transformed with pEAQ-GFP.

Control leaves that were un-infiltrated (A) and infiltrated with Infiltration Medium only (B) were imaged alongside leaves infiltrated with transformed *Agrobacterium* at a range of optical densities (O.D.₆₀₀) from 0.01 (C), to 0.05 (D) to 0.1 (E) after 5 days of expression. It can clearly be seen that strong GFP expression occurs in leaves infiltrated with O.D.₆₀₀ >0.05.

Imaging of whole leaves using the fluorescence gel imaging mode of a G:Box Chemi XX6 (Syngene) was used to visually confirm the presence of expressed GFP in the leaves. No signal was seen from leaves that were not infiltrated; leaves infiltrated with Infiltration Medium only showed some autofluorescence at the site of infiltration (the small dark circle in the centre of the “Buffer only” leaf in **Figure 4.9**). By contrast, the leaves infiltrated with *agrobacterium* transformed with pEAQ-GFP showed fluorescence under illumination with the blue LED

module (470 nm), with an increase in signal intensity as the O.D.₆₀₀ of the infiltrated solution was increased from 0.01 to 0.05, but the signal was saturated above 0.05. Thus, an O.D. of 0.05 was chosen to be ideal for expression of GFP in tobacco leaves. The same relative level of expression was seen in leaves left to express for 5 days and for 10 days.

Having visually confirmed expression of GFP within the leaves, the next step was to develop a protocol for extraction of the protein from the leaves. Briefly, leaves were ground under liquid nitrogen in a chilled pestle and mortar and the resulting powder was incubated with Extraction Buffer (20 mM Tris pH 8.0, 100 mM NaCl, 10% glycerol, 5 mM EDTA, 1x protease inhibitors) before centrifugation at 100,000 g to pellet cell debris and insoluble material. The supernatant was decanted and visualised under blue light illumination, as done for the whole leaves. Fluorescence was compared against a protein sample extracted in the same manner from control leaves that were not infiltrated with *Agrobacteria*. As seen in **Figure 4.10**, the tube containing protein extracted from GFP-expressing leaves fluoresced strongly, in contrast to the control tube that contained extracted protein from un-inoculated leaves.

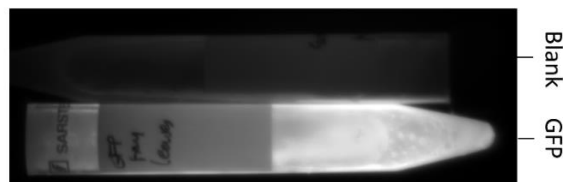


Figure 4.10 - Fluorescence image of two tubes containing protein extracted from a set of leaves infiltrated with Infiltration Medium (Blank), or *Agrobacterium tumefaciens* LBA4404 transformed with pEAQ-GFP (GFP).

To further assess the extracted GFP, the material was analysed by SDS-PAGE. 20 μ L samples of protein extracted from un-infiltrated leaves, Infiltration Medium-only leaves, and the leaves expressing GFP were loaded and run on 14-20% gradient gels and then imaged via the fluorescent gel imaging function of the G:Box Chemi XX6. As seen in **Figure 4.11**, below, a single band was found in the GFP-containing lane at approximately 30 kDa, confirming the presence of GFP.

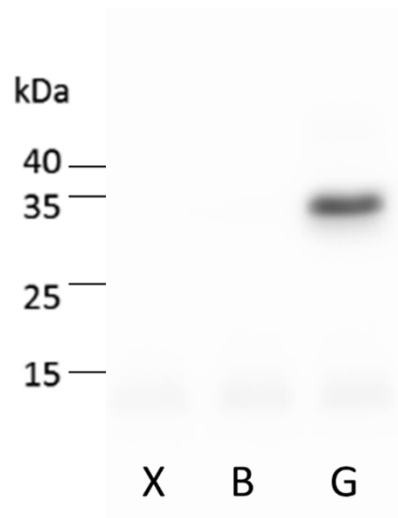


Figure 4.11 - Fluorescence image of SDS-PAGE gel of extracted samples from leaves.

Un-infiltrated control, **X**; buffer-only control, **B**; pEAQ-GFP-infiltrated sample, **G**. Markers on the left indicate approximate positions of molecular weight standards. A clear single band is shown in the lane containing the extracted protein from leaves infiltrated with GFP, at approximately 30 kDa as expected.

4.4 Transient expression of His-tagged ALMTs via agrobacteria-mediated agroinfiltration in tobacco

Having successfully expressed GFP in tobacco by agrobacterium-mediated transient expression and extracted the protein from the leaves, the focus moved onto using the system to express ALMTs with the vectors prepared in **Section 4.2**. Agrobacteria were transformed with the pEAQ-ALMT constructs as above for GFP. Simultaneously, agrobacteria were also transformed with pEAQ-GFP to provide a positive control to ensure expression was occurring consistently. Experimental conditions were maintained from **Section 4.3** as they had been shown to work for GFP. The objectives for this were three-fold. Firstly, to assess whether full length ALMTs could be expressed using the pEAQ-HT expression system in tobacco; secondly, to assess whether it was possible to produce protein in sufficient amounts to be suitable for structural studies and on what scale plants would need to be grown to do so; and thirdly, to extract, solubilise, and purify the proteins that could be produced with the intention of use in structural studies.

During the expression of full length ALMTs, a positive control was always run with the pEAQ-GFP construct as before. Expression was seen in these leaves, whether the GFP-expressing leaves were in separate plants, in separate leaves in the same plant as the ALMT, or whether co-infiltrated into different sides of the same leaf, as seen below in **Figure 4.12**. Leaves were visualised to confirm expression of GFP as in **Section 4.3**, before proceeding with protein extraction. Through each experiment the positive control leaves always showed expression.

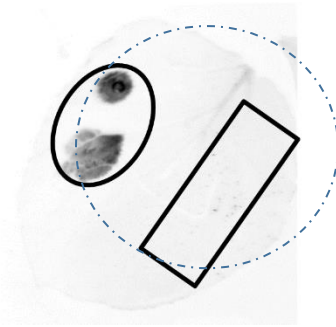


Figure 4.12 – Fluorescence image of a tobacco leaf infiltrated with two pEAQ constructs.

GFP in left hand half, indicated by the oval, imaged in G:Box XX6 with blue LEDs. Right hand side infiltrated with N-terminally His tagged TaALMT5 indicated by the rectangular box. This confirms transient expression in the leaves for GFP, suggesting the experimental conditions were conducive to transient protein expression. The dashed circle shows the approximate outline of the leaf surface.

4.4.1 An initial expression trial of AtALMT5C

pEAQ-AtALMT5C was the first construct confirmed by sequencing, so this was used for an initial test of the system. Duplicate transformations were performed into *Agrobacterium*, and a fresh transformation of pEAQ-GFP was also undertaken to provide a positive control. As previously for GFP, suspensions of transformed *Agrobacterium* were infiltrated into the leaves at $O.D_{600} = 0.01, 0.05,$ and 0.1 . After 5 and 10 days of expression, leaves were removed from the plants and protein was extracted in a similar manner as for the extraction of GFP previously, with the following modifications in light of ALMTs being membrane proteins. Membrane proteins are often separated from cellular debris and soluble proteins by a two-part centrifugation protocol¹¹⁶. After extraction protein samples were centrifuged first at $14,000\ g$ to pellet cellular debris and then the supernatant was centrifuged at $100,000\ g$ to collect the membrane fraction that should contain the protein of interest. In this initial test, three separate samples of leaves infiltrated at different $O.D_{600}$ values were

extracted in duplicate, for a total of six samples. Each of these was extracted and the fractions were assessed for the presence of His-tagged ALMT5C by dot blotting, while the membrane pellets were loaded onto a SDS-PAGE gel and assessed by Western blotting. As seen in **Figure 4.13**, expression was seen in each sample, although the protein of interest was mostly contained in the initial cell debris pellet rather than the cell membranes fraction as expected. One explanation for the target protein being found in the cell debris pellet is that it has been expressed in insoluble inclusion bodies, or otherwise this can be a sign that the fractionation buffer is not optimal, or cell lysis is not complete. This will be explored in the following section.

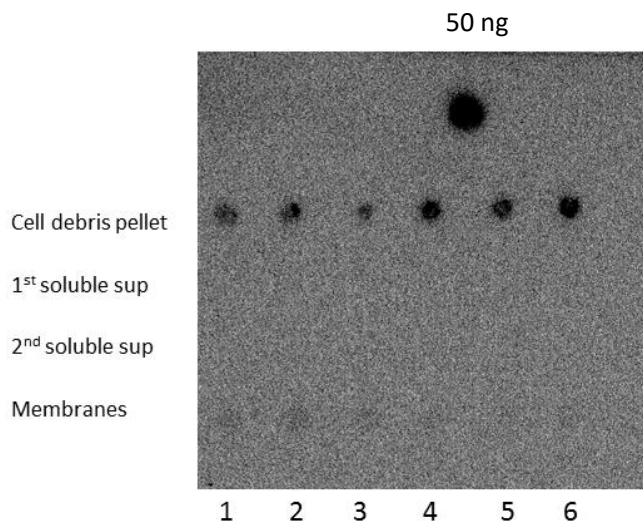


Figure 4.13 - Dot blot of two pairs of pellet/supernatant fractions extracted from each of six samples of leaves infiltrated with ALMT5C after 5 days of expression.

Samples infiltrated with Agrobacteria in duplicate at three O.D.₆₀₀ **1,2** = 0.01, **3,4** = 0.05, **5,6** = 0.1. Most of the protein of interest is seen in the first pellet containing the cell debris, and only a small amount of signal is seen in the membrane fraction (the pellet collected after centrifugation at 100,000 g). 50 ng of His-tagged protein standard was included as a positive control (**50 ng**).

The membrane pellets, which were expected to contain the protein of interest, showed only a small amount of signal in the western blot as the protein was mostly contained in the first pellets (**Figure 4.14**); however, some signal was seen in two lanes, corresponding to an infiltration at O.D.₆₀₀ 0.05 and at 0.1. Interestingly, the bands were seen at a lower MW than expected – although the molecular weight markers washed off the membrane so these could not be used for determination of the MW of the visible bands. However, the protein standard used (TEV-His) has a MW of ~27 kDa so the MW of the bands in lanes 4 and 6 can be estimated at MW 30-40 kDa. This is significantly smaller than the expected MW of AtALMT5C, which is 61 kDa. One explanation for this could be that as a membrane protein ALMT5C runs at a lower MW than expected, as is seen for many membrane proteins¹¹⁶, alternatively this could be evidence either of a contaminant protein from tobacco that cross reacts with the anti-His antibody, or evidence that ALMT5C is degrading. Given that the bands are only present in two samples it seems unlikely to be a cross-reacting native protein and so is most likely ALMT5C running anomalously low, or a cleaved product of ALMT5C.

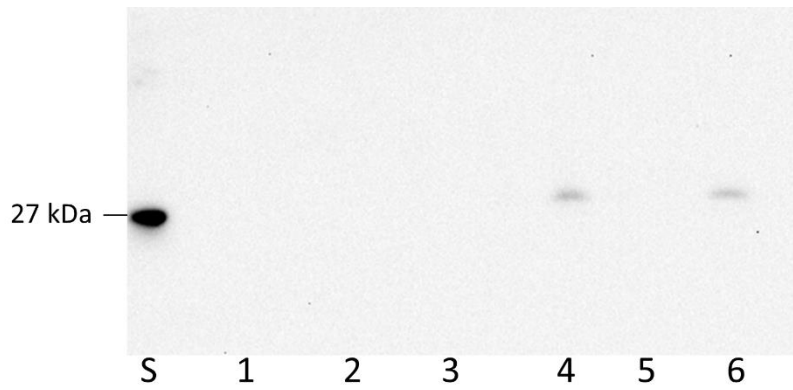


Figure 4.14 – Western blot of membranes prepared from six leaf samples inoculated with varying O.D. of Agrobacteria and expressing AtALMT5C. Two biological replicates were taken for each O.D.₆₀₀ tested.

Lanes are **S**, 50 ng TEV-His standard. **(1, 2)** O.D. = 0.01, **(3, 4)** O.D. = 0.05, **(5, 6)** O.D. = 0.1. Signal is seen in lanes 4 and 6 but at a lower height than expected (~30 kDa rather than ~61 kDa).

4.4.2 Improving fractionation with a high-density sucrose buffer

In order to move forward with protein extraction, as the protein was seen in the cell debris pellet rather than the final membrane pellet, a modification was made to the extraction buffer as it was suspected that cellular lysis was not complete and the buffer used for extraction could be optimised. Following the methodology of Abas and Luschnig²³⁹ for improved fractionation of plant membrane proteins, sucrose was added to a final concentration of 25% in order to increase the density of the buffer and prevent the membrane proteins from sedimenting in the initial spin to clear the cell debris. Furthermore, a change was made to the protocol: the plant material was passed through a cell disruptor at 30 kpsi in order to homogenise the material and fully lyse the cells, enabling all the protein to be extracted from the cell. This should also ensure the membrane protein can be separated from the pelleted material (such as the cell wall) in the initial 14,000 *g* spin.

After the first trial of ALMT expression, a second attempt was set up with ALMT1N as the protein of choice. This time, three time-points were chosen, 4, 8, and 12 days, again with the same three values for O.D.₆₀₀ of inoculation. The results from the 8-day time-point are shown in **Figure 4.15**. The Coomassie stained gel shows an improved extraction of total protein, and more of the target protein is visible in the Western blot. However, despite this improvement, expression levels as determined from the Western blot are low, measured by densitometry to be on the order of 5 μ g per leaf. Promisingly, all three conditions at this time point show some level of expression, and that material is specifically detected by the anti-His antibody and of approximately the right size (~60 kDa), and visible in the membrane fraction.

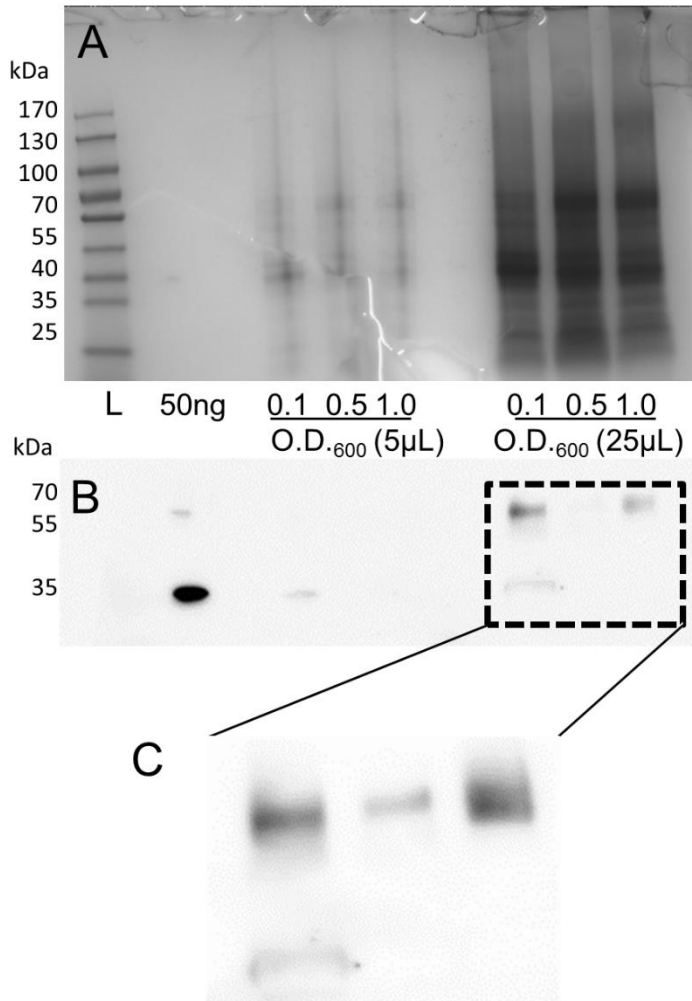


Figure 4.15 - Samples of protein extracted from leaves infiltrated with three O.D.₆₀₀ of Agrobacteria harbouring pEAQ-TaALMT1N harvested after 8 days of expression

A) Coomassie and **B)**, Western blot of the three samples, loaded in duplicate with either 5 µL or 25 µL per lane. **C)** shows a zoomed view of the section of the blot outlined in B for a clearer view of the intensities of the bands.

4.4.3 Improving expression conditions

As the levels of expression seen so far were relatively low (μg levels per leaf) a review of the literature was conducted to assess the conditions used to generate high yields of protein in previous experiments. A total of eleven studies were reviewed and these are summarised in **Table 4.1**.

Table 4.1 - Summary of conditions used for expressing high yields of protein from CPMV- and pEAQ-based expression vectors from recent papers

	Age of plants /weeks	Days of expression	O.D. of agrobacteria at infiltration	Reference
	3-5	4-6	0.3	240
	4-6	1-6	1.0	241
	8	4-10	1.0	242
	3	6	1.0	243
	4-8	14	2.0	244
	3	6	0.13-1.2	234
	4	6-8	1.2	236
	3-4	3-10	0.4	245
	6	6	0.5	238
	3	6	1.2	237
	-	5-6	1.2	246
Range	3-8	1-14	0.1-2.0	

Based on this set of papers, the conditions used in previous studies match up well with the initial experiments in this study for both the age of the plants and for the length of expression following infiltration, but the O.D.₆₀₀ of infiltrated Agrobacteria used thus far in this study was around an order of magnitude below the average used for previous studies. Based on this summary, the following conditions were chosen for a larger screen of constructs and conditions, shown in **Table 4.2**.

Table 4.2 - Revised conditions for screening ALMT expression in tobacco

Plant age	5 weeks after sowing (approx. 4 weeks after germination)
Days of expression	5 days, 10 days
O.D.₆₀₀	0.1, 0.5, and 1.0

This set of three O.D.s and two time-points gave a set of six conditions per expression trial, allowing a pair of plants to be infiltrated per condition, which gives sufficient leaf material to extract protein from while still permitting one set of plants (i.e. 12 tobacco plants) to grow together in a single tray in the greenhouses. Several further trials with all six constructs were attempted in order to attempt to find the best combination of conditions to produce the maximum yield of protein. A full list of conditions for all trials assessed in this chapter is found in **Table 4.3**. Additionally, when simply testing for expression, samples were lysed and spun to pellet cell debris, then whole protein extracts were run on SDS-PAGE gels in order to assess relative expression, and standard amounts of His-tagged protein was loaded to calculate absolute values of protein per band.

Table 4.3 –Full list of conditions tested for ALMT expression trials.

	Plant age /weeks	O.D. ₆₀₀	Days of expression	Construct
Initial screen	5	0.01	5	AtALMT5C
	5	0.05	5	
	5	0.1	5	
	5	0.01	10	AtALMT5C
	5	0.05	10	
	5	0.1	10	
Further testing	4	0.1	5	AtALMT5C
	4	0.5	5	
	4	1.0	5	
	5	0.1	4	AtALMT1N
	5	0.5	4	
	5	1.0	4	
	5	0.1	8	AtALMT1N
	5	0.5	8	
	5	1.0	8	
	5	0.1	12	AtALMT1N
	5	0.5	12	
	5	1.0	12	
	5	0.1	5	AtALMT5C
	5	0.5	5	
	5	1.0	5	
	5	0.1	5	AtALMT5N
	5	0.5	5	
	5	1.0	5	
	5	0.1	5	AtALMT9C
	5	0.5	5	
	5	1.0	5	

	Plant age /weeks	O.D. ₆₀₀	Days of expression	Construct
Further testing	5	0.1	5	AtALMT9N
	5	0.5	5	
	5	1.0	5	
	5	0.1	5	TaALMT1C
	5	0.5	5	
	5	1.0	5	
	5	0.1	5	TaALMT1N
	5	0.5	5	
	5	1.0	5	
	5	0.1	10	AtALMT5C
	5	0.5	10	
	5	1.0	10	
	5	0.1	10	AtALMT5N
	5	0.5	10	
	5	1.0	10	
	5	0.1	10	AtALMT9C
	5	0.5	10	
	5	1.0	10	
	5	0.1	10	AtALMT9N
	5	0.5	10	
	5	1.0	10	
	5	0.1	10	TaALMT1C
	5	0.5	10	
	5	1.0	10	
5	0.1	10	TaALMT1N	
5	0.5	10		
5	1.0	10		
5	1.0	10	AtALMT5N	
Final purification	5	1.0	5	AtALMT5N

An example of an expression trial is shown in **Figure 4.16**. Samples tend to smear, likely due to the samples containing whole leaf extract, and therefore being a heavily loaded lane, due to the low expression levels requiring more total protein to be loaded than is optimal (the protein may also be glycosylated, which would also cause smearing). Nonetheless, it is possible to determine that there is at least some expression of each protein tested and generally a longer expression produces a greater amount of protein. Quantification of expression levels is impossible as the bands are too smeared to compare accurately to the standards, but it is clear that the best-expressing samples have approximately the same order of magnitude of protein as the bands of standards. Working on a basis of 20 ng of target protein in 20 μ L of sample per well this extrapolates to \sim 1 μ g of target protein per g of leaf. This level of expression would be at the lower limit of feasibility for later studies, as 1 kg of leaf material would be needed for 1 mg of target protein, which would require a rational number of plants.

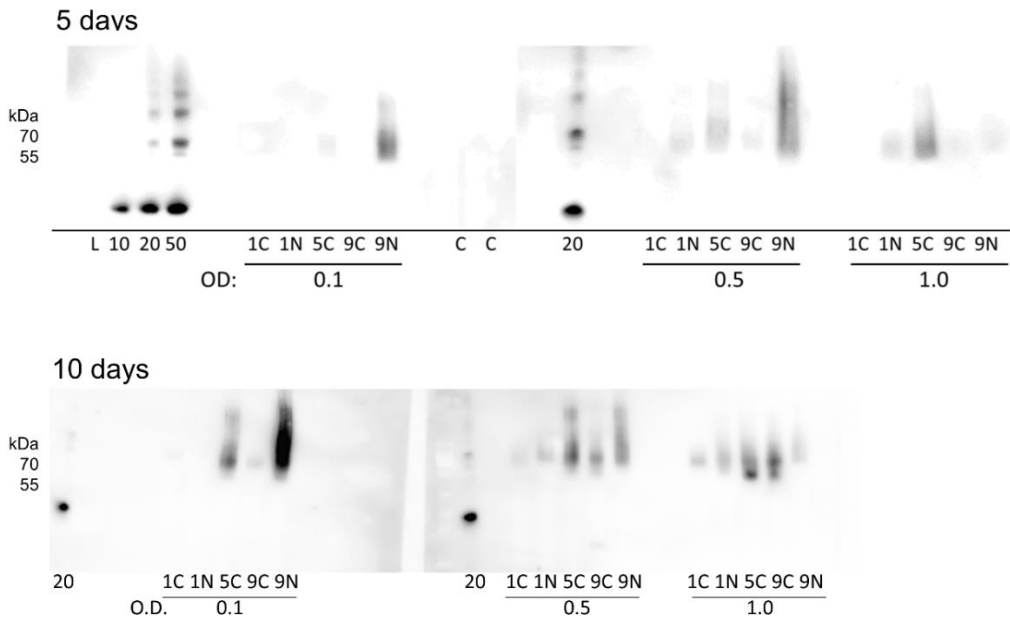


Figure 4.16 - Western blot of total protein extracted from tobacco leaves expressing TaALMT1C, TaALMT1N, AtALMT5C, AtALMT9C, AtALMT9N.

Leaves inoculated with agrobacteria at O.D.₆₀₀ = **0.01**, **0.05**, and **0.1** containing plasmids expressing TaALMT1C (**1C**), TaALMT1N (**1N**), AtALMT5C (**5C**), AtALMT9C (**9C**), AtALMT9N (**9N**). Leaves were collected after **5 days**, or **10 days** of expression. Some expression seen in a variety of conditions, with 10 days of expression producing better results than 5 days. Standards of 10, 20, and 50 ng of His-tagged TEV protease were loaded to aid quantification and act as a positive control. Lane containing MW marker indicated with L, and approximate MW markers identified to left.

4.4.4 Improving ALMT extraction

Detergents are required to solubilise membrane proteins once they are extracted from the cells, but they may also prove useful for extracting membrane proteins from the rest of the cell in the first place. It was suspected that a single extraction with buffer alone was not sufficient to extract all the target protein and so to assess the extent that detergents can aid extraction of ALMTs and ensure that the maximum amount of protein is extracted a re-extraction of the 14,000 g pellet was attempted. This second extraction contained the detergent Triton X-100 at 1% (w/v). Samples of ALMT-expressing leaf were extracted with PBS after grinding under liq. N₂ and passage through the cell disruptor as before and then centrifuged to pellet debris; the pellets were then re-extracted with detergent-containing buffer and samples of extracted protein were assessed by Western blotting.

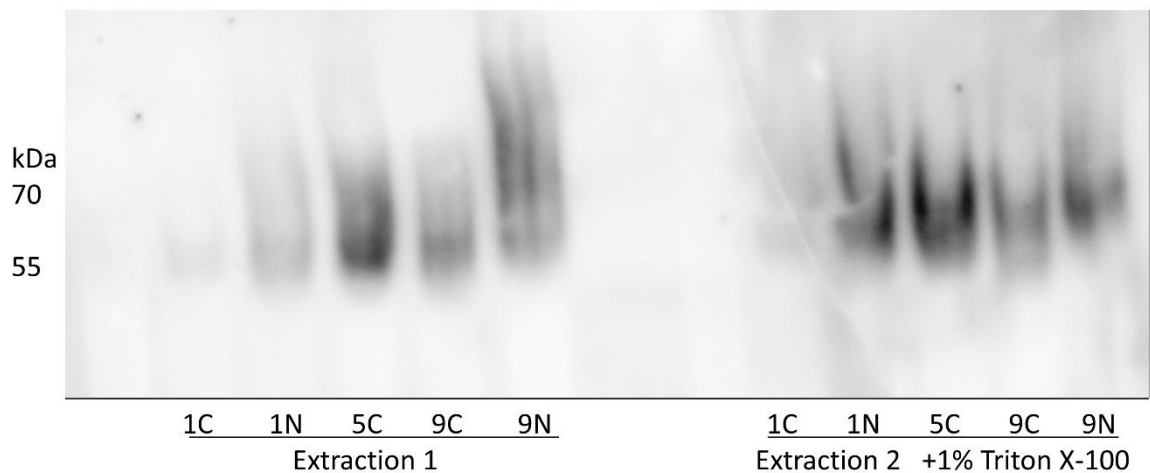


Figure 4.17 - Western blot of 20 μ L of samples extracted from tobacco leaves expressing TaALMT1C (1C), TaALMT1N (1N), AtALMT5C (5C), AtALMT9C (9C), AtALMT9N (9N) Extraction first with buffer and secondly with buffer containing 1% Triton X-100.

Figure 4.17 shows the effect of a second extraction containing Triton X-100. The second set of samples re-extracted with Triton X-100 contain a significant proportion of the target protein. Therefore, extraction can be optimised and the presence of detergent will permit more target protein to be extracted from the leaves and was used from this point forward.

A second assessment of the extraction was undertaken to assess whether detergent-containing buffer can extract all the target protein in one extraction, or whether multiple extractions will be required even with detergent added to the buffer. Three samples of ALMT5N were extracted with PBS containing 1% Triton X-100 and then re-extracted with the same buffer (in each case 1 ml buffer per g of ground leaf material). The results are shown in **Figure 4.18** and show that the first extraction is able to extract almost all the target protein. Additionally, now that the protein is extracted reliably in a single extraction, it can be seen that an infiltration of $O.D._{600} = 1.0$ appears to be best condition for expression of ALMT5N, confirming that a higher $O.D._{600}$ of infiltration is beneficial to expression levels.

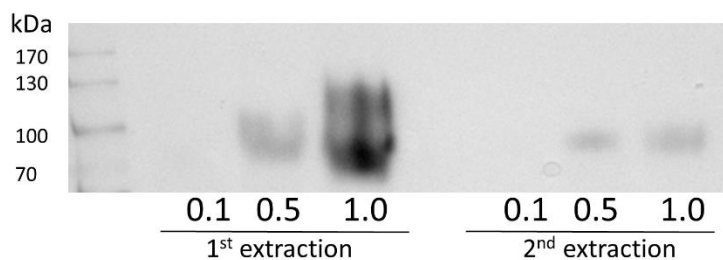


Figure 4.18 - Western blot of 20 μ L of three samples of leaf containing ALMT5N, inoculated with Agrobacteria at $O.D._{600} = 0.1, 0.5, \text{ or } 1.0$, each extracted twice with buffer containing Triton X-100.

Almost all protein is extracted in the first samples and $O.D._{600} = 1.0$ appears to be the best condition.

Thus, the best strategy appears to be using an infiltration of *Agrobacteria* at $O.D._{600} = 1.0$ and as higher densities of bacteria have not been tested as they were outside the scope of the initial screen, increasing this may prove fruitful. Moreover, it is helpful to extract protein from leaf samples with buffer containing detergent. As the target protein will require solubilisation with detergents as part of the purification in any case, this provides a useful modification to the protocol.

The next question was to quantify the amount of extracted protein from the best condition, namely, ALMT5N from leaves inoculated with *Agrobacteria* at an $O.D._{600} = 1.0$. Samples of ALMT5N from above (stored overnight at -20°C) were re-rerun on an SDS-PAGE gel and compared with His-tagged protein standards from 5-20 ng. As seen in **Figure 4.19**, no signal was seen from the lanes containing extracted ALMT5N. This could be a sign that the protein was being degraded by the action of proteases released upon homogenisation of leaf tissue. Alternatively, it could be that the protein samples were significantly more dilute than the lowest His-standard. The smallest amount of standard loaded was 5 ng and the largest volume of extracted protein loaded was 20 μL ; this gives a maximum concentration of ALMT5N of 0.25 $\text{ng}/\mu\text{L}$, equivalent to 0.25 $\mu\text{g}/\text{ml}$. As $\sim 6\text{-}8$ g of leaves were extracted for these experiments this produced approximately 6-8 ml of extracted protein i.e. 1.5-2 μg of ALMT5N. So, either the ALMT5N is being rapidly degraded, despite the presence of protease inhibitors, or it is being expressed in levels at the limit of detectability, unsuitable for structural studies. Neither of these possibilities is favourable for this study.

However, as previous gels that also contained His-tagged standards showed visible protein it seems likely that the target protein is being expressed at an appreciable level, but it is being degraded during storage.

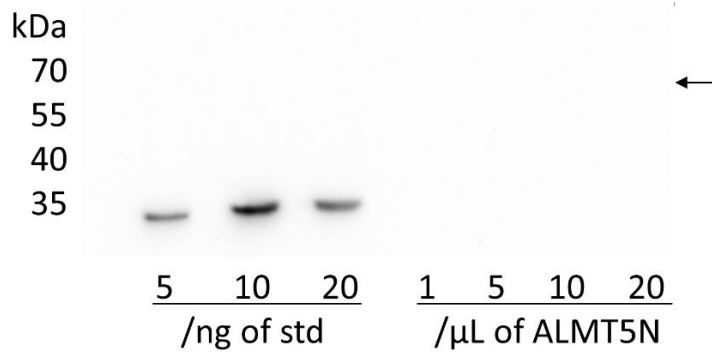


Figure 4.19 - Western blot of ALMT5N with His-standards to assess expression levels.

No signal is seen from the target protein (expected position indicated by arrow) indicating that either protein levels are below 0.25 ng/μL or that the protein has degraded.

4.4.5 Test of rapid purification of ALMT5N

Before further optimisation of conditions was attempted, given the initial relatively-low expression levels combined with the seeming instability of the extracted protein (despite the presence of protease inhibitors), an attempt was made to solubilise and purify the extracted protein immediately after extraction in order to assess firstly, whether purification was feasible, and secondly, whether purifying ALMTs would stop the apparent degradation seen in samples.

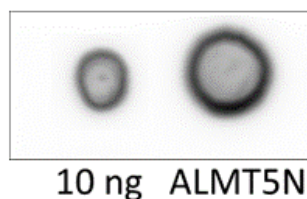


Figure 4.20 - Dot blot of protein extracted from tobacco leaves expressing AtALMT5N.

10ng of His-tagged standard and a dot of 5 μ L extracted protein. Expression of N-terminally His-tagged AtALMT5 is confirmed.

ALMT5N has shown the highest expression overall, so it was chosen for the purification test. A sample of ALMT5N-expressing leaves was ground and lysed before extraction with buffer containing 1% DDM at 4°C and centrifugation at 100,000 g in order to discover whether the material being produced was able to be solubilised. A sample of the supernatant was taken for a dot blot to confirm the presence of ALMT5N, as shown in **Figure 4.20**, then the supernatant containing solubilised proteins was bound to HisPur Cobalt resin for 2h at 4°C before washing with increasing concentrations of imidazole (10 mM, 20 mM, 40 mM) and elution with buffer containing 300 mM imidazole. 20 μ L samples of each fraction were run on SDS-PAGE gels and analysed by Western blotting and coomassie staining. This data, shown in **Figure 4.21**, did not clearly show

any of the protein of interest. More promisingly, there were very minimal contaminants in the elution fractions and 10 mM of imidazole was sufficient to wash off major contaminant bands, indicating that if target protein yields were sufficiently high then it would be possible to purify the proteins to high levels of purity in one step. However, the lack of protein present in the Western blot was the major problem.

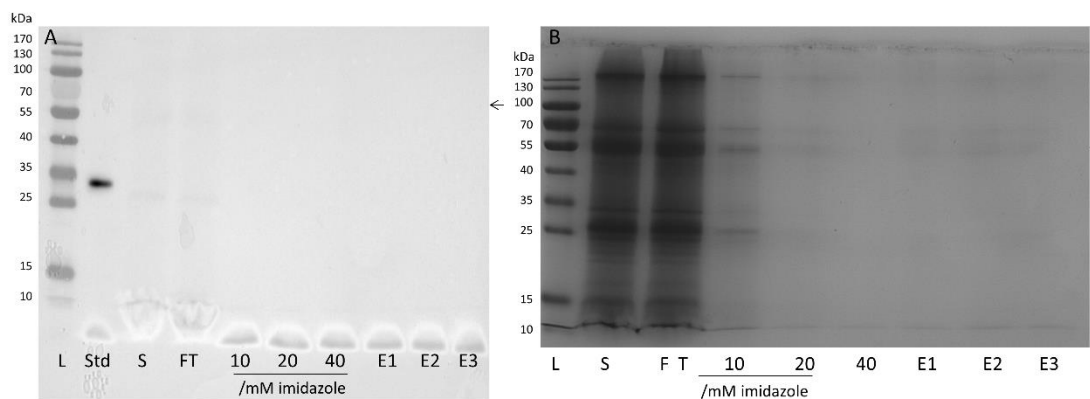


Figure 4.21 - ALMT5N purification.

Western blot (**A**), and coomassie stained (**B**) SDS-PAGE gel for a purification of ALMT5N with 20 μ L from each fraction. Washes of 10-40 mM imidazole (**10, 20, 40**) and three elutions (**E1-E3**) No signal is seen in the expected lanes for Solubilised (**S**) and Eluted (**E**) protein. 20 ng of His-tagged standard is seen in the western blot, however, indicating successful transfer and antibody binding. Arrows indicate the expected size of ALMT5N.

However, re-imaging of the Western blot was able to provide further information. Setting the camera to exclude the protein standard so that the relative brightness of that band did not obscure signal from the samples it was possible to see some signal from the purification lanes, as seen in **Figure 4.22**. Here it is possible to see that the protein has been correctly solubilised by the DDM, and although some signal is seen in the flowthrough, there is signal from the His-tagged protein in the fractions eluted from the IMAC resin. A gel was

also silver stained with the Pierce Silver Stain Kit, as this kit has a greater sensitivity than coomassie staining, but bands for ALMT5N were not seen above the background staining.

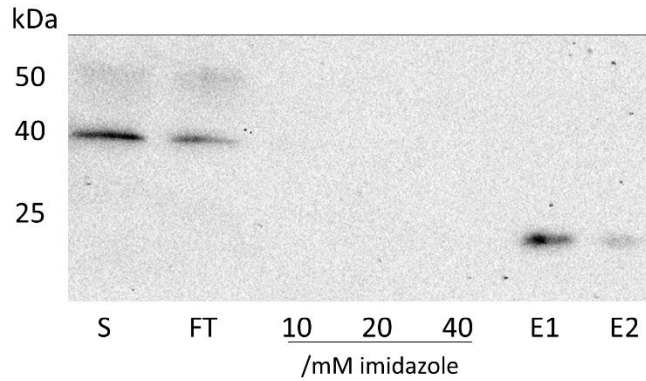


Figure 4.22 - Western blot of 20 µL of fractions of a AtALMT5N purification, without standards.

The protein is solubilised (**S**), although some material does not bind to the resin and so is seen in the flow through (**FT**). Washing with increasing imidazole (**10-40** mM) does not elute the protein until 300 mM is used, which washes the protein back off the resin (**E1, E2**). However, the protein is degraded – firstly during solubilisation the MW drops from 50 kDa to 40 kDa, and during the washing and elution of the resin, the MW decreases further to ~20 KDa.

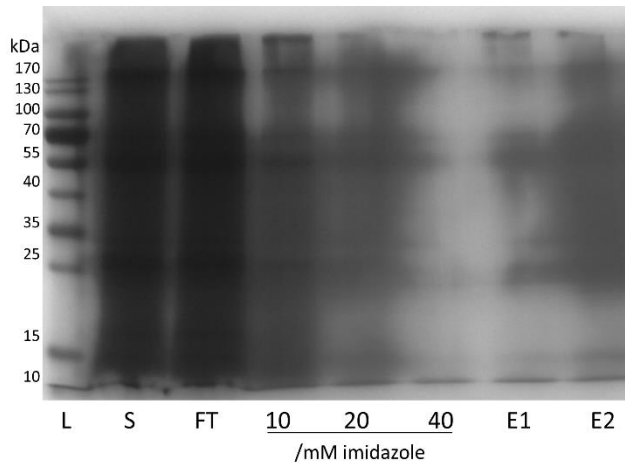


Figure 4.23 - Silver stained SDS-PAGE gel of ALMT5N purification.

MW marker (**L**), soluble fraction (**S**), flowthrough (**FT**), washes with 10-40 mM imidazole (**10, 20, 40**) and two elutions at 300 mM imidazole (**E1-2**). Amounts of protein in washes and elutions are at undetectable levels against the background signal.

The low levels of protein expression seen and especially the rapid cleavage of the expressed proteins suggests that transient agro-infiltration-mediated expression of ALMTs in tobacco is an unpromising route to production of high levels of target protein for structural studies.

In hindsight, further evidence of protein degradation is visible in the previous data. The data shown in **Figure 4.14** also indicate that there was cleavage of the protein at some point during extraction. That construct was C-terminally tagged and showed a size of ~30-35 kDa while the construct analysed in the purification trial shown in **Figure 4.22** is an N-terminally tagged construct and thus represents the “other side” of the degraded material. The MW of this fragment is initially ~40 kDa before decreasing to ~20 kDa during purification. So both the N- and C-tagged constructs are degrading. There is also a small band in **Figure 4.15C** in the first lane around 30 kDa that is detected by the antibody for the His-tag, suggesting that this band is also from the target protein and a possible sign that it was degrading in that experiment too. Also supporting this was that it was never possible to return to samples extracted the previous day in order to, for example, re-run gels as it was not possible to detect any signal from the Western blot.

4.5 Discussion and Conclusion

In this chapter, the expression of three ALMT proteins was investigated using a transient expression system in tobacco. This system was chosen to provide an environment that is as close to the native plant environment as possible and maximise the likelihood that protein would be produced in a correctly-folded and active form. Constructs were created by inserting the ORFs for each gene of interest into the pEAQ-HT vector in two positions in order to produce constructs that would be tagged at either their C- or N-terminus with an oligohistidine affinity tag. These six constructs were then transformed into *Agrobacterium* and infiltrated into leaves. Conditions for plant growth, protein expression, and protein extraction were screened, developed, and optimised. A purification showed that it was possible to produce solubilised and purified protein.

However, despite testing a range of expression conditions, expression levels of all six constructs was consistently low, producing nanograms-to-micrograms of protein, insufficient for extensive biochemical or structural studies without significant scale up. Additionally, and more problematically, the protein produced appeared to be unstable, with signal on Western blots not being detectable if samples were stored overnight. When attempting an extraction followed directly by purification it was possible to observe the cleavage and breakdown of the protein products with ALMT5N reduced from 50 kDa to 40 kDa followed by a second drop to 30 kDa.

The use of transient expression in tobacco appears not to be a promising route to the production of these plant membrane proteins for structural studies. This may be because the protein targets in question are anion channels, and high levels of expression are impossible without causing cell death as their

expression leads to making the membrane of the cell leaky due to the presence of these channels. Indeed, necrosis of the infiltrated leaves was observed during the course of many experiments. However, this was not tracked comprehensively enough to draw firm conclusions, but generally there was more necrosis observed at higher infiltrations of agrobacteria. This may reflect a higher toxicity due to the infiltration itself, or the fact that higher infiltration caused expression of target protein, and thus toxicity from overexpression.

While membrane proteins in general are known to be difficult targets for overexpression, anion channels perhaps represent an even more difficult target due to this fact. Channel activity would lead to loss of anions (in this case malate, an important compound in carbon metabolism and also chloride and nitrate, which have been shown to be substrates for ALMTs⁴⁸) and this would lead to depolarisation. These factors would especially drive the plant cell to attempt to degrade the ALMTs rather than suffer the deleterious effects of overexpression, and if uncontrolled cell death was occurring then this could also lead to the release of proteases that would degrade the target protein. Indeed, black patches were observed on the leaves, especially those infiltrated with higher concentrations of Agrobacteria, or those left to express for a longer period of time, consistent with necrosis and degradation within the leaf. This is on top of the usual potential problems with overexpression such as the risk of saturating the host quality control, posttranslational-modification, and trafficking systems.

There is some promise in these experiments, however. Each construct could be expressed, to some level and when solubilisation was attempted 1% DDM was sufficient to solubilise ALMT5N. Indeed, the final trial showed that it was

possible to purify the degraded ALMT5N by IMAC suggesting that if an approach was found that optimised the expression levels and prevented protein degradation then it would be possible to both solubilise and purify the resultant proteins.

Given that this *in planta* expression approach has not been successful, future studies (which are out of the scope of this current study) could utilise one of several options. Eukaryotic expression systems are developed for yeast, insect, and mammalian cells. For example, it may be convenient to try to exploit the similarity of yeast to plants, and express ALMTs in either *S. cerevisiae* or *P. pastoris* these two model organisms have been widely used for protein production and can be considered a proxy system for plants. Additionally, a larger range of protocols, vectors, and tools have been developed for use with these systems making them a more convenient approach. The shorter time period and better controlled expression of these systems may also help with expression. Cells can be harvested before extensive cell death occurs, and protein can be extracted before the action of proteases can degrade the protein of interest. Alternatively, success has been seen with expression of plant proteins in insect cells using the baculovirus system^{247,248}. Although this approach is more expensive and technically demanding, it may be suitable and the system has proved successful at generating crystal structures deposited in the PDB²⁴⁹. These may present future avenues of work, but have not been explored in the present study.

Chapter 5

Expression and Purification of Bacterial ArAE Exporters

5.1 Molecular biology

5.1.1 Identification of candidate genes

In an attempt to find more experimentally amenable targets for structure function investigations, three bacterial candidates from the ArAE family, with a close identify to the wheat channel, *TaALMT1* were initially selected from the genomic DNA stocks available in the lab. These three genes were SO_0154 (Uniprot:Q8EKE1) from *Shewanella oneidensis* (**So**), DVU_0077 (Uniprot:Q72FY6) from *Desulfovibrio vulgaris* (**Dv**), and AaeB (Uniprot:P46481) from *E. coli* (**Ec**), listed in **Table 5.1**.

These genes were selected based on their sequence identity to the transmembrane region of *TaALMT1*. When aligned using MUSCLE via the EMBL-EBI website^{250,251}, **So**, **Dv**, **Ec** showed 20, 22, and 17% identity to *TaALMT1*, respectively (Table 5.1). However, when looking at the predicted transmembrane regions specifically (taken as the first 180 residues of the bacterial proteins and the first 225 residues of the plant protein) the sequence identity rises to 30, 24, and 25% respectively, for identity, and 62, 70, 62% for similarity. Likewise, the bacterial proteins are predicted to have similar topology: **So** and **Dv** are predicted to have a 6 TM topology by Transmembrane Hidden Markov Modelling¹³ similar to the transmembrane domain of ALMT family members, and **Ec** is predicted to have an internally-repeated 2x 6 TM topology (Error! Reference source not found.). The *E. coli* gene has been characterised in vivo already and so provides a good target for study, while the two other genes are 6-TM and thus closer in structure to the ALMTs.

As such, it was reasoned that these proteins may provide good models for understanding the transmembrane domain of ALMT proteins, while their study could prove easier because they are bacterial proteins.

Table 5.1 – Information on genes identified as homologous to ALMT family members

Designation	Gene name	Uniprot ID	Predicted topology	Length /bp	% Sequence Identity (TaALMT1)
Ec	AaeB	P46481	2x (6 TM)	1,965	17
So	SO_0154	Q8EKE1	6 TM	1,122	20
Dv	DVU_0077	Q72FY6	6 TM	1,080	22

5.1.2 Design and Preparation of a Suitable Expression Vector

5.1.2.1 Design of pTTQ18-TP

The first step after identification of candidate genes was to prepare a suitable vector to insert the ORFs of interest into. The base vector chosen was pTTQ18, a vector that has proved successful for expression of bacterial membrane proteins due to its expression of the *LacI^q* gene, which represses any ‘leaky’ expression that could otherwise retard bacterial growth, and strong *tac* promoter for high levels of expression after induction²⁵². The expression vector, designated pTTQ18-TP, was constructed from two fragments of existing vectors used in the lab, a schematic is shown in **Figure 5.1**. The purpose of this was to introduce an improved cleavage site, swapping the thrombin cleavage site in the original vector for a more efficient HRV-3C site.

The first vector used was “pUC-T4-Lysozyme+0294 frag-0476 frag”. Three sequences were extracted from this: a ribosome binding site, a start codon, and a pair of restriction sites (*NheI* and *SbfI*) to be used for subcloning the gene of

interest between. Also included was a copy of the CDS for lysozyme, which was used for a previous project, between the NheI and SbfI restriction sites. The construct was designed for insertion in frame with the Start codon and HRV3C cleavage site and His-tag using the PstI enzyme rather than SbfI. PstI recognises the sequence CGTACG, which is found within the recognition site for SbfI (CCGTACGG). However, SbfI could be used (if, for example, the PstI recognition sequence was found within the ORF of interest) by design of PCR primers with one fewer base or two additional bases in order to put the sequence back in frame.

After insertion of the ORFs of interest, the lysozyme gene is removed, allowing for an easy check that digestion had proceeded as expected by looking for the excised band on the gel (as seen in **Figure 5.4A**). In summary, this fragment provided a cassette into which any CDS of interest could be inserted into, ensuring that the sequence would be in frame and correctly positioned with respect to the RBS and start codon, as seen in the top half of **Figure 5.1**.

Secondly, the existing expression vector “pBPT0294-CVH” provided the main pTTQ18 backbone as seen in the bottom half of **Figure 5.1**. This fragment added a HRV-3C cleavage site and His-tag followed by a stop codon onto the end of the CDS inserted between NheI and SbfI to enable purification of the target protein by IMAC, and later cleavage of the tag. It also contains the required sequences for effective replication of the vector, and for lactose-inducible expression of the target gene.

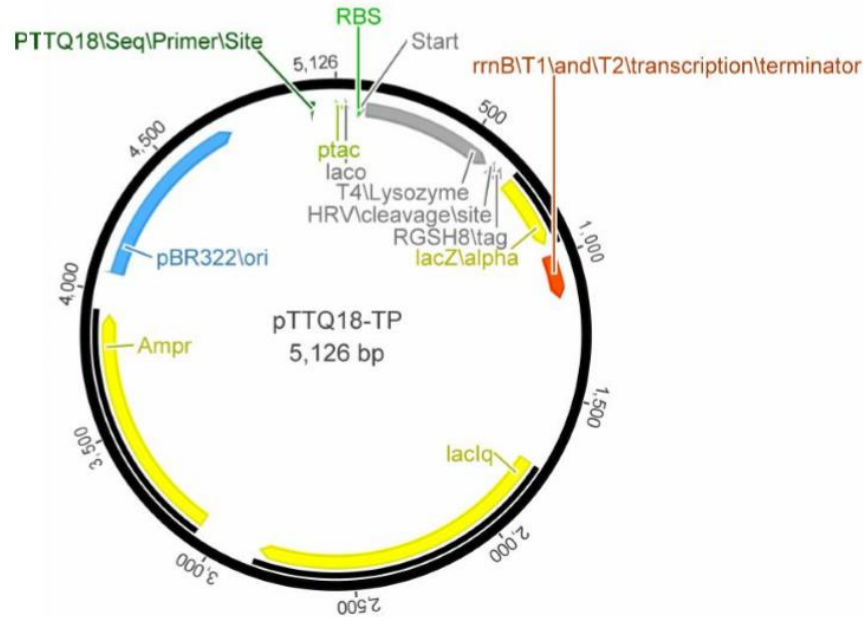


Figure 5.1 - Schematic of the pTTQ18-TP expression vector.

Fragment excised from “pUC-T4-Lysozyme+0294 frag-0476 frag” using *Xba*I and *Sbf*I sites contributes the Ribosome Binding site (**RBS**), Start Codon (**Start**), and Lysozyme gene between *Nhe*I and *Sbf*I sites as shown in the top half of the figure. Fragment excised from “pBPT0294-CVH” using *Xba*I and *Sbf*I sites contributes a C-terminal HRV-3C cleavage site (**HRV\cleavage\site**) and His₈-tag(**RGS8 tag**), followed by the stop codon (**Stop**), along with the major vector backbone comprising the Origin of Replication to allow proliferation (**Ori**), Ampicillin resistance gene for selection (**Ampr**), LacI gene for repression of expression (**LacIq**), LacZ gene for metabolism of lactose (**LacZ\alpha**). Figure produced using Geneious 8.1.

The resulting vector should result in very small levels of ‘leaky’ expression before induction due to expression of the LacI^q repressor protein, high levels of expression after induction from the Tac promoter, and a final recombinant protein product containing a His-tag for the purposes of purification and identification by western blotting, which is cleavable to provide a further purification step, or in order to produce a protein in close-to-native state.

5.1.2.2 Construction of pTTQ18-TP

To prepare the expression vector, each of the plasmids required for its construction was digested with both XbaI and SbfI in order to cut the fragments required; additionally, the vectors were cut with each of XbaI and SbfI individually to ensure that both of the enzymes were active and could cut at the recognition site. The desired bands, indicated with arrows on

Figure 5.2, were then cut from the gel and purified, before ligation together to form the final expression vector. Sequencing confirmed that the sequence was as expected.

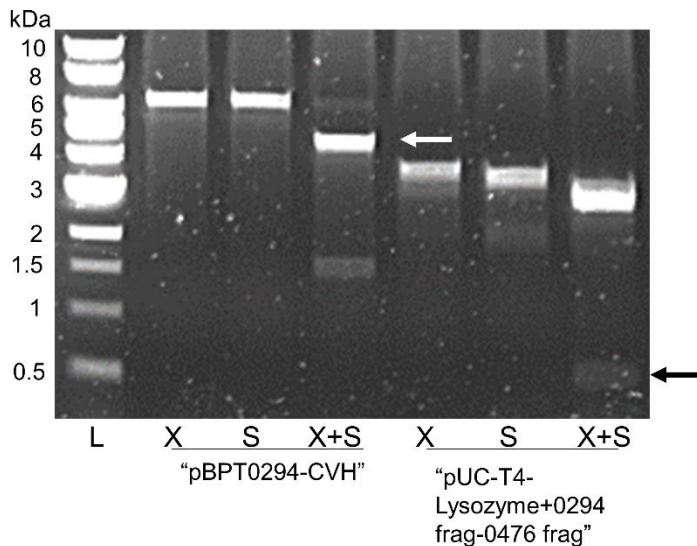


Figure 5.2 –0.5% agarose gels of digested samples of the two vectors used to create the expression plasmid used in this study.

"pBPT0294-CVH" and "pUC-T4-Lysozyme+0294 frag-0476 frag", each digested with XbaI (**X**) and SbfI (**S**) singly, and then double-digested (**X+S**). The white arrow (approximately 5 kbp) and the black arrow (approximately 500 bp) indicate the bands excised and purified.

5.1.3 Production of expression vectors

5.1.3.1 Amplification of target genes

The next step required was to amplify the open reading frames (ORFs) encoding the three target proteins from the corresponding genomic DNA via the Touchdown PCR protocol and using the primers detailed in **Chapter 2**.

Samples of each of the resultant PCR products were electrophoresed on a 0.5% agarose gel, as seen in **Figure 5.3**, with the band for **So** running at approximately 1000 bp and the band for **Ec** (AaeB) running at approximately 2000 bp, as expected. However, the **Dv** reaction produced a smear instead of a band at the expected size of approximately 1000 bp. Although further PCR optimisation by adjustment of annealing temperature and time was attempted, no improvement was seen. Accordingly, investigation into the *D. vulgaris* gene was stopped in favour of the two genes that could be successfully amplified.

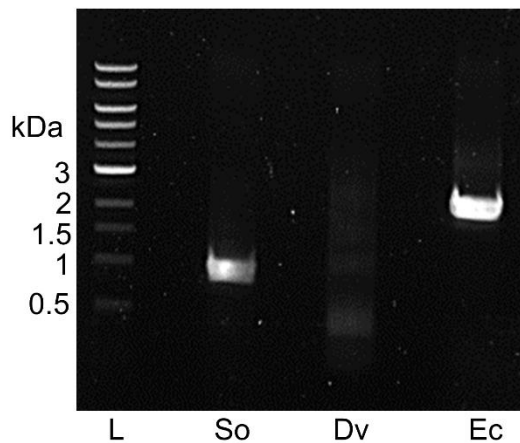


Figure 5.3 - Touchdown PCR to amplify three ORFs from *S. oneidensis* (So), *D. vulgaris* (Dv), and *E. coli* (Ec) genomic DNA.

The *S. oneidensis* and *E. coli* genes successfully amplify, with bands at the expected sizes of approximately 1000 and 2000 bp, respectively, but the reaction for the *D. vulgaris* DNA produced only a smear.

The two successful products, **So** and **Ec** (AaeB), were purified and then used to produce an 'cloning vector' by ligation into a linearised sample of the vector pCR-Blunt (Invitrogen), forming **So-Blunt** and **Ec-Blunt**. The ligation mixtures were then transformed into *E. coli* strain OminiMAX 2 cells and recombinants were then selected by plating onto LB agar containing 100 µg/mL kanamycin and 1 mM IPTG. Cloning vectors were isolated and used in the production of the expression vectors.

5.1.3.2 Insertion of ORFs into the expression vector pTTQ18-TP

In order to produce the linearised DNA fragments required for the expression vectors, pTTQ18-TP, **So-Blunt**, and **Ec-Blunt** were each digested with NheI and PstI, followed by electrophoresis on a 0.5% agarose gel, as seen in **Figure 5.4**. The digestion of pTTQ18-TP produced a band of almost 6000 bp with the removed Lysozyme gene seen at approximately 500 bp. Meanwhile, ORFs from **So-Blunt** and **Ec-Blunt** were seen at approximately 1000 bp and 2000 bp, respectively. These fragments were excised, purified, and ligated together to form the final expression vectors, designated **So-TTQ** and **Ec-TTQ**. These were confirmed to be correct by sequencing before proceeding with expression.

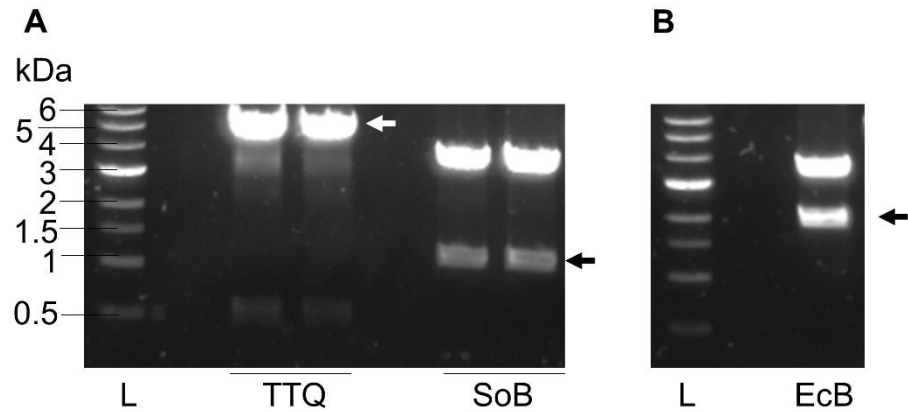


Figure 5.4 - Double digests to prepare fragments for construction of expression vectors.

Approximate DNA size markers are indicated on the left hand side and are equivalent in A and B. **A)** pTTQ18-TP digested with NheI and PstI (**TTQ**) in duplicate. The expression vector backbone taken is indicated with a white arrow at approximately 5 kbp and the excised CDS for lysozyme originally present in the vector is visible at approximately 500 bp. The *S. oneidensis* gene previously amplified and captured in PCR-Blunt was digested with NheI and PstI (**SoB**) in duplicate. The digested insert is indicated with a black arrow at approximately 1 kbp as expected. **B)** The *E. coli* gene previously amplified and captured in PCR-Blunt was digested with NheI and PstI (**EcB**) in duplicate. The digested insert is indicated with a black arrow at approximately 2 kbp as expected.

5.1.4 Expression of bacterial genes in *E. coli*

The two constructs containing genes from either *S. oneidensis* (SO_0154) or *E. coli* (AaeB) in the newly-produced vector pTTQ18-TP, designated **So**-TTQ and **Ec**-TTQ, were now transformed into expression strains of *E. coli* (C41, BL21 Gold, BL21 Star) and expression was induced on a 3 mL culture scale using autoinduction in three media (M9, LB, SB). The results are shown in **Figure 5.5**, and showed that both constructs could be expressed, with BL21 Star producing reliable expression for both constructs in either LB or SM media. M9, by contrast, produced undetectable-low levels of expression. At this stage, the expression levels of **Ec**-TTQ were significantly higher than **So**-TTQ.

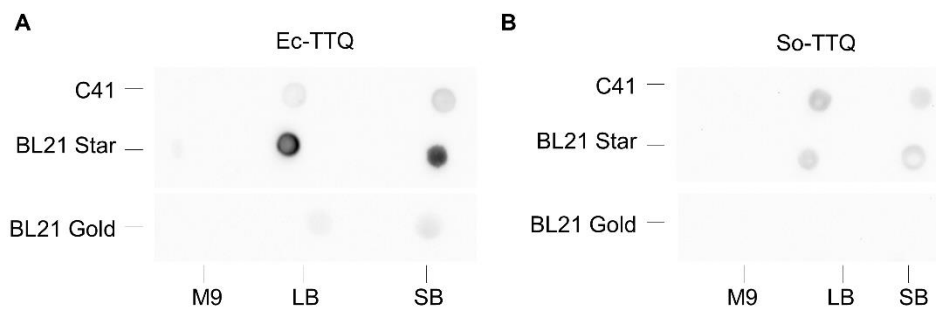


Figure 5.5 - Dot blot of 3 µL samples from small scale expression tests of two bacterial constructs.

LB and SB media appear to give the best expression for all combinations of cell lines and constructs. BL21 Star is the best expressing strain for **Ec**-TTQ, and is joint best for **So**-TTQ. By contrast, M9 media performs very poorly in all cases.

As both constructs had expressed in a range of cell and media combinations, early progress was promising. Scaling up of expression and screening conditions to optimise expression levels would be needed.

5.1.4.1 Solubilisation and purification trials for So-TTQ

So-TTQ solubilised readily in a range of detergents, such as DM, DDM, and C12E8 seen in **Figure 5.6A**. Following purification by IMAC, results looked somewhat promising as a band of expected MW was seen in the elution fractions, which were free from major contamination (**Figure 5.5B**). However, as only a small amount could be purified (resulting from low expression levels seen in the screening above) it was decided to move onto on further work with **Ec-TTQ** and experiments were not taken any further with **So-TTQ**.

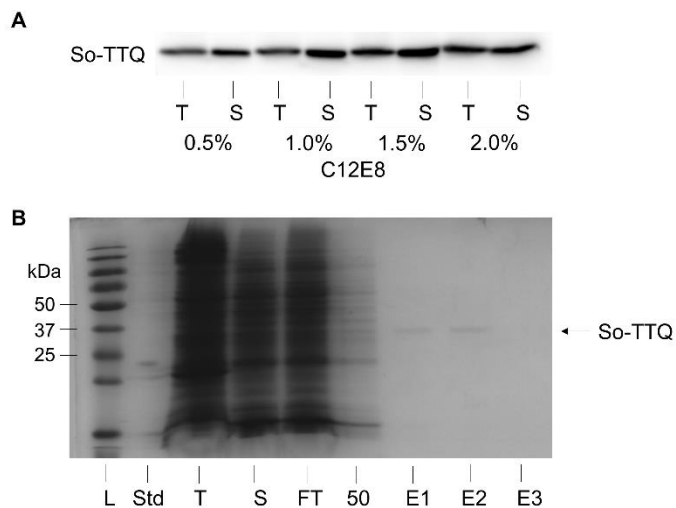


Figure 5.6 - Solubilisation and purification of So-TTQ (20 μ L).

A) Solubilisation of **So-TTQ** in the detergent C12E8, which readily and quantitatively solubilises the protein in concentrations above 0.5%. **B)** Coomassie stained SDS-PAGE gel of fractions taken from a purification of **So-TTQ** by IMAC showing molecular weight marker (**L**), 100 ng standard (**Std**) Total (**T**), Soluble (**S**), Flowthrough (**FT**), 50mM imidazole wash, and elutions with 300 mM imidazole (**E1-E3**). Bands seen in elution are of the expected size for the construct, at approximately the 37 kDa marker.

5.1.4.2 Ec-TTQ expression and solubilisation

Expression levels of EcTTQ were good and a Western blot showed a band at 70 kDa as expected (**Figure 5.7A**) giving confidence that the correct protein was being overexpressed. While culture conditions were being optimised to ensure the maximum level of expression was achieved, an assessment of solubility was undertaken. However, when a range of detergents were screened (including DM, DDM, LDAO, C12E8, β -OG, DMNG, LMNG), only very limited solubilisation was seen, with less than 10% of the material solubilised, and most often no solubilisation occurring (**Figure 5.7B**). The protein was correctly inserted into the membrane, as evidenced by it separating into the (100,000 *g*) membrane fraction upon centrifugation rather than the 14,000 *g* fraction containing aggregates, but it was impossible to extract it from the membrane and solubilise it in detergent with high yields.

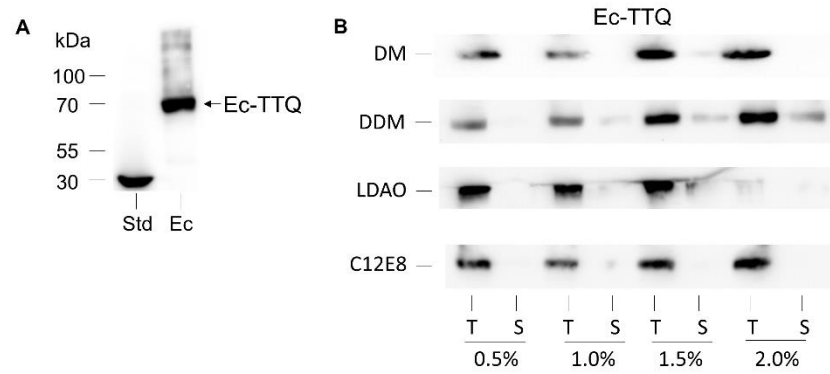


Figure 5.7 - Solubilisation of Ec-TTQ (20 µL).

A) Western Blot of membranes prepared from BL21 Gold expressing **Ec-TTQ (Ec)**. 125 ng of His-tagged standard (**Std**) and A band is seen at approximately 70 kDa as expected. **B)** Western blot showing the results of a detergent screen for **Ec-TTQ** with four detergents. Although the Total (**T**) fraction is consistent, relatively little is seen in the Soluble (**S**) fraction. The best results are seen with DDM, where some signal is seen in the soluble lanes for concentrations above 1%. However, only a small proportion is soluble.

A literature search revealed that the gene expressed from **Ec-TTQ (AaeB)** is natively expressed from an operon containing a second, smaller protein (called AaeA). Both proteins had been shown to be required to confer physiological resistance to aromatic acids³⁸, and it was hypothesised that they formed a complex. Therefore, it was suspected that both proteins were required *in vitro*, and AaeA was required to enable AaeB to be isolated in a stable, soluble form. Thus, it was decided to prepare a construct to express both AaeA and AaeB in tandem.

5.2 Construction of vectors for expression of AaeAB, AaeA, and AaeA Δ TM

In *E. coli*, both AaeB and AaeA are expressed from the same operon, and so co-expression was trialled to see if AaeA could act as a potential chaperone to improve stability. As such, PCR was performed on genomic DNA with primers designed to flank both coding sequences and amplify a fragment containing the sequences of AaeA followed by AaeB. This was then inserted into the pTTQ18-TP expression vector using the restriction sites *NheI* and *PstI*, as seen for the constructs designed in the preceding section of this chapter (A schematic can be seen in **Figure 5.8A**).

In addition to its potential role alongside AaeB, AaeA also presents an interesting target in itself. Although it has been shown to be vital, no studies have reported on its exact function, and from its sequence it has been identified to be a member of the Membrane Fusion Protein (MFP) family often found co-expressed in operons containing inner membrane transporters. AaeA is, therefore, of interest to study in its own right. Thus, a construct was designed to insert the DNA for AaeA into pET28(b) with the intention of expressing it for biochemical studies in a C-terminally His-tagged form (**Figure 5.8B**).

Analysis of the primary sequence of AaeA predicted that it was likely to contain a N-terminal transmembrane helix (as seen in **Figure 1.8A**) (approximately residues 10-30), and previous structural studies of MFPs had seen success by expressing soluble, truncated constructs^{106,253–255}. Thus, in an effort to produce a constructs that would express as soluble protein, a forward primer was designed to amplify the cDNA excluding the first 96 bases (equivalent to 32

amino acids). Thus, AaeA Δ TM represents a construct to express residues 33-310 of AaeA with the addition of a C-terminal His-tag (**Figure 5.8C**).

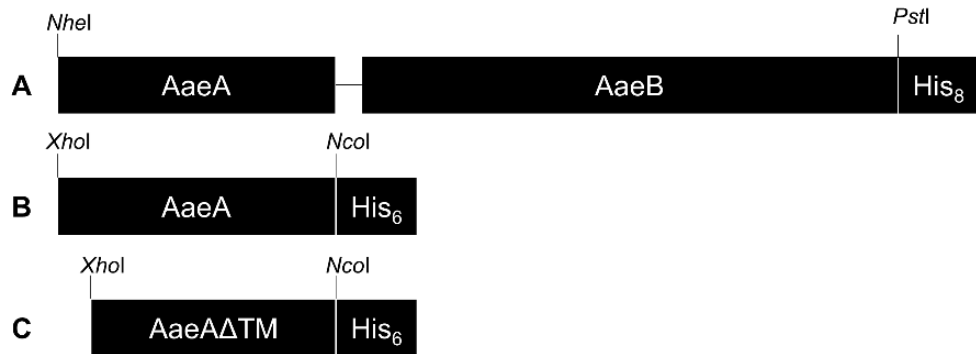


Figure 5.8 - Schematic to show constructs.

A) AaeAB. The DNA is taken from the operon and inserted into the vector between *NheI* and *PstI* restriction sites, resulting in transcription of both genes. AaeA is expressed before AaeB, which has a C-terminal His-tag. **B)** AaeA. The DNA for the full gene is inserted between *XhoI* and *NcoI* sites in pET28(b) to express a C-terminally His-tagged protein. **C)** The DNA coding for a construct with residues truncated from the N-terminus is inserted into pET28(b) as for AaeA.

In order to produce the desired expression vectors, Touchdown PCR was performed, as described previously, to amplify the required CDSs. These PCR products were purified and then directly digested with the appropriate restriction enzymes (*NheI* and *PstI* in the case of AaeA-AaeB, and *XhoI* with *NcoI* in the case of AaeA and AaeA Δ TM), purified, and then inserted into a purified linearised vector digested with the same enzymes. **Figure 5.9** shows this, with the PCR product inserts shown in **Figure 5.9A** and the linearised vector in **Figure 5.9B** (which was pTP1 made in **Chapter 3**. The previously inserted CDS for the CTD of TaALMT1 was seen cut out of the vector in **Figure 5.9B**).

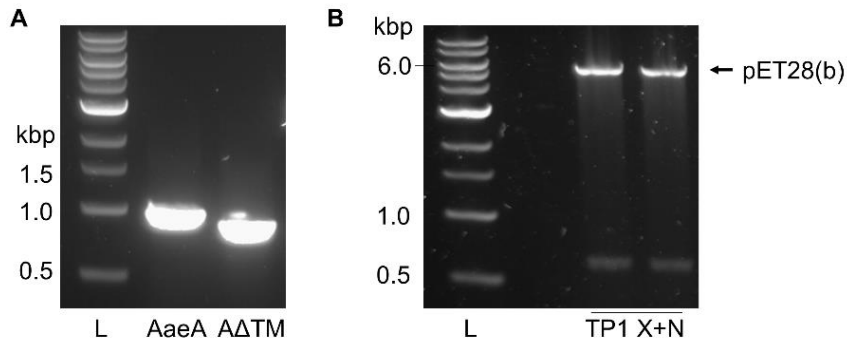


Figure 5.9 – Vector construction.

A) PCR to amplify AaeA and AaeAΔTM, bands are seen at ~1000 bp and ~800 bp as expected. **B)** Double digest in duplicate of TP1 constructed in **Chapter 3**. Vector backbone used for expression vector construction indicated by arrow on the right.

These digested fragments were ligated together using T4 DNA ligase overnight, and transformed into Omnimax 2 cells. Once sufficient plasmid had been isolated a small sample was tested by re-digestion, as seen, for example in **Figure 5.10** for AaeA, and then the vector was sequenced to confirm that they had been constructed correctly, before moving on to expression trials.

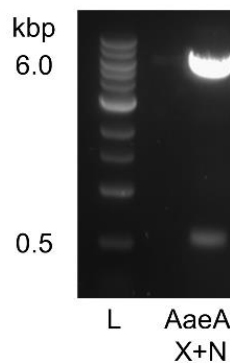


Figure 5.10 - Digest to test the expression vector AaeA in pET28(b) on a 1% agarose-TAE gel.

Double digestion with XhoI and NcoI leads to a band at approximately 6000 bp, representing the vector backbone, and a band at approximately 500 bp, representing the coding sequence cleaved in half.

5.2.1 Production of AaeB co-expressed with AaeA

Once the three plasmids for expression of AaeA-AaeB, AaeA, and AaeA Δ TM had been constructed and verified by sequencing, conditions for their expression were screened. Although AaeB had been trialled already when expressed alone, it was decided to start the process again for the new construct as co-expression with AaeA could require significantly different conditions. Optimisation of production of AaeB was thusly the first priority.

5.2.1.1 Optimisation of expression conditions for AaeA-AaeB.

After producing the AaeA-AaeB construct the next stage was to screen for the optimal overexpression conditions. To achieve this AaeA and AaeB were co-expressed by autoinduction in four *E. coli* strains (C41, C43, BL21 Gold, BL21 Star) with three media (M9, LB, SB) over four time-points (12, 24, 36, 72h) resulting in a total of 48 conditions. These cultures, grown on a 3 mL scale, had total protein extracted and AaeB-specific expression was assessed by dot-blotting. Although this method is nor quantitative, it can be seen from **Figure 5.11**, the best culture conditions were using LB media at the 48 h time-point, with any of the four cell types, and C41 over 72 h. These conditions were later further examined by Western blotting.

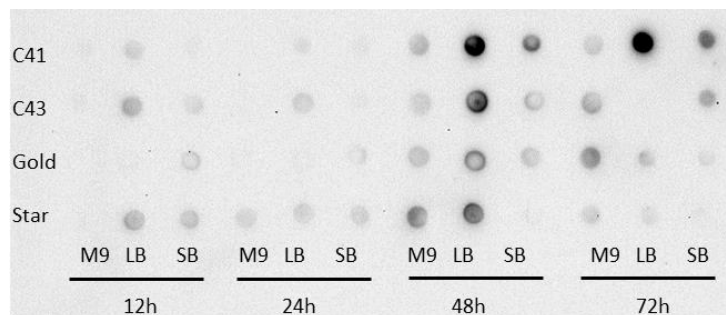


Figure 5.11 - Dot blot assessing culture conditions for AaeA-AaeB.

3 μ L of lysate was loaded onto membrane before blocking, antibody binding and visualisation.

5.2.1.2 Assessment of IPTG- vs Auto-induction for optimised yield

After this assessment, IPTG was also tested as an induction method to assess whether it could produce a greater yield than autoinduction. BL21 Gold cells in LB media were chosen for this experiment and IPTG induction was assessed at four time-points after induction (2, 3, 4, 6h), which occurred after the cultures reached an OD₆₀₀ of 0.8. The growth curves measured for these cultures can be seen in **Figure 5.12**, with exponential growth as expected up to the point of induction, followed by a slower, linear growth while the cells were expressing protein.

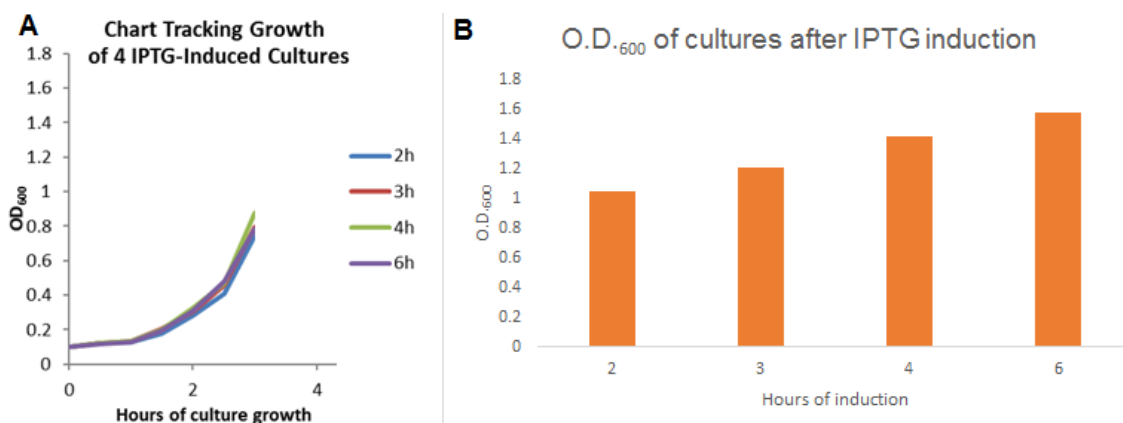


Figure 5.12 - Graph monitoring optical density of bacterial cultures expressing AaeB, co expressed with AaeA, after IPTG-induction.

A) Each culture grew with an exponential curve until 1 mM IPTG was added at 3h. **B)** Cultures continue growth during 2, 3, 4, or 6 h of induction.

These conditions were compared by Western blotting extracted protein against parallel cultures expressing AaeB (co-expressed alongside AaeA) by autoinduction and collected after 24h and 48h. The cells were lysed in a cell disruptor and total membranes were collected by centrifugation at 100,000 g and then analysed by SDS-PAGE and Western blotting. The results, seen in **Figure 5.13**, show that there is not a large difference between each of the

conditions, with 2 to 3 h of IPTG-induced expression providing approximately double the expression level as compared with 24 h of autoinduction. Either approach is equally feasible for large-scale expression, with IPTG producing slightly higher yields in a shorter time, but autoinduction requiring less hands-on experimenter time, and cheaper reagents.

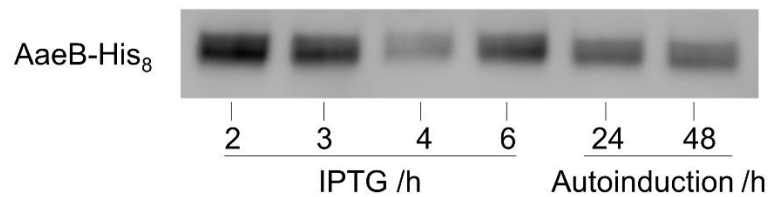


Figure 5.13 - Western blot of 20 μ L of protein extracted from six cultures of BL21 Gold expressing AaeB.

IPTG expression for 2-3 hours produces the most intense signal and thus the highest expression, although the levels seen are not greatly above those seen in autoinduced cultures collected at 24 or 48 hours.

5.2.1.3 Solubilisation of AaeB co-expressed with AaeA

As sufficient quantities of AaeB has been expressed to perform initial experiments on, the next step was to determine whether it was possible to solubilise AaeB in detergent now that it was co-expressed alongside AaeA. A range of detergents were tested between 0.5% and 2.0% and results can be seen in **Figure 5.14**. All detergents tested, except for OMNG, were able to solubilise AaeB when added at approximately 1% or above. This is expected for a well-folded membrane protein and so it appeared that co-expression of AaeA with AaeB had resulted in an improvement in stability of AaeB.

Interestingly, the blots showed higher-MW bands in addition to the bands at approximately 70 kDa (the expected size of AaeB). For example, **Figure 5.14B**

shows in the top DM-solubilised sample bands at approximately the markers for 100 kDa (which corresponds roughly to the MW of AaeA + AaeB), and 130 kDa (which corresponds to the molecular weight of a dimer of AaeB). This is an initial sign of potential multimerisation occurring, although as membrane proteins often run anomalously so it is hard to say with precision.

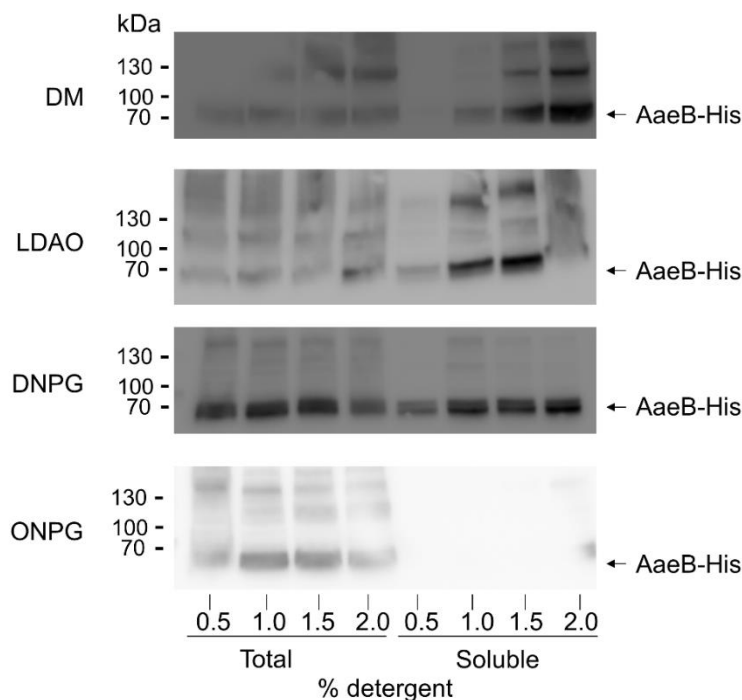


Figure 5.14 - Western blots showing that AaeB can only be solubilised in detergent when co-expressed with AaeA.

When co-expressed with AaeA, AaeB is readily solubilised by the detergents DM, LDAO, DNPG. Unusually, ONPG is not effective at solubilisation of AaeB. There are some higher-MW bands seen on the blot, perhaps evidence of multimerisation.

5.2.1.4 Purification of AaeB

After optimising the overexpression conditions for the AaeA-AaeB co-expression and ensuring it could be solubilised, the next stage was to purify the protein using its His-tag. The initial purification trial saw promising levels of

protein purified (visible by Coomassie staining without need for concentration), with only a small amount of lower-MW contaminants, and AaeA co-purifying along with AaeB, as seen in **Figure 5.15** and **Figure 5.17A-B**. AaeA is clearly interacting with AaeB in order to be co-purified alongside it, and it can be seen that over the course of multiple elution steps the proportion of AaeA decreased. AaeA and AaeB were each identified by mass spectrometry.

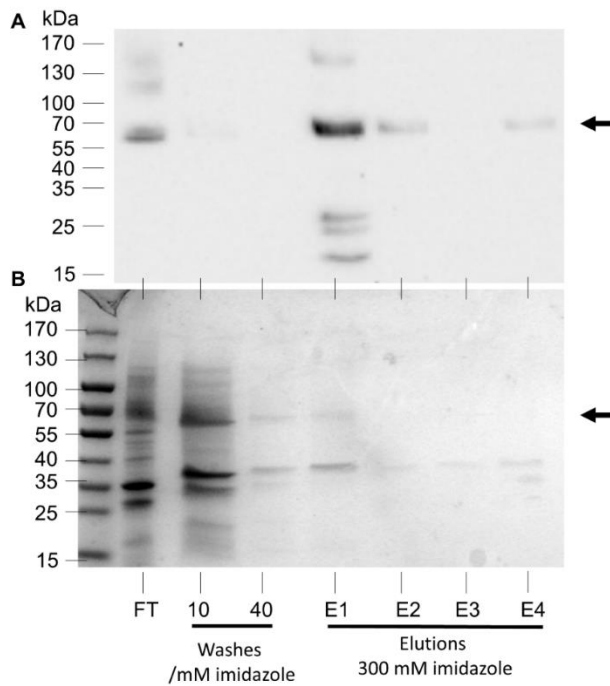


Figure 5.15 - 20 μ L samples from each fraction of a purification of AaeB.

A) Western blot, and **B)** Coomassie stained SDS-PAGE gel. Flowthrough (FT), 10 mM and 40 mM imidazole washes, and four elution fractions (E1-4) were run. The position of AaeB is indicated by arrows on the right. The washes remove a significant amount of contaminating bands

Analysis of the bands by Mass Spectrometry identified the smaller band as AaeA and the larger band as AaeB, as shown in **Figure 5.16**.

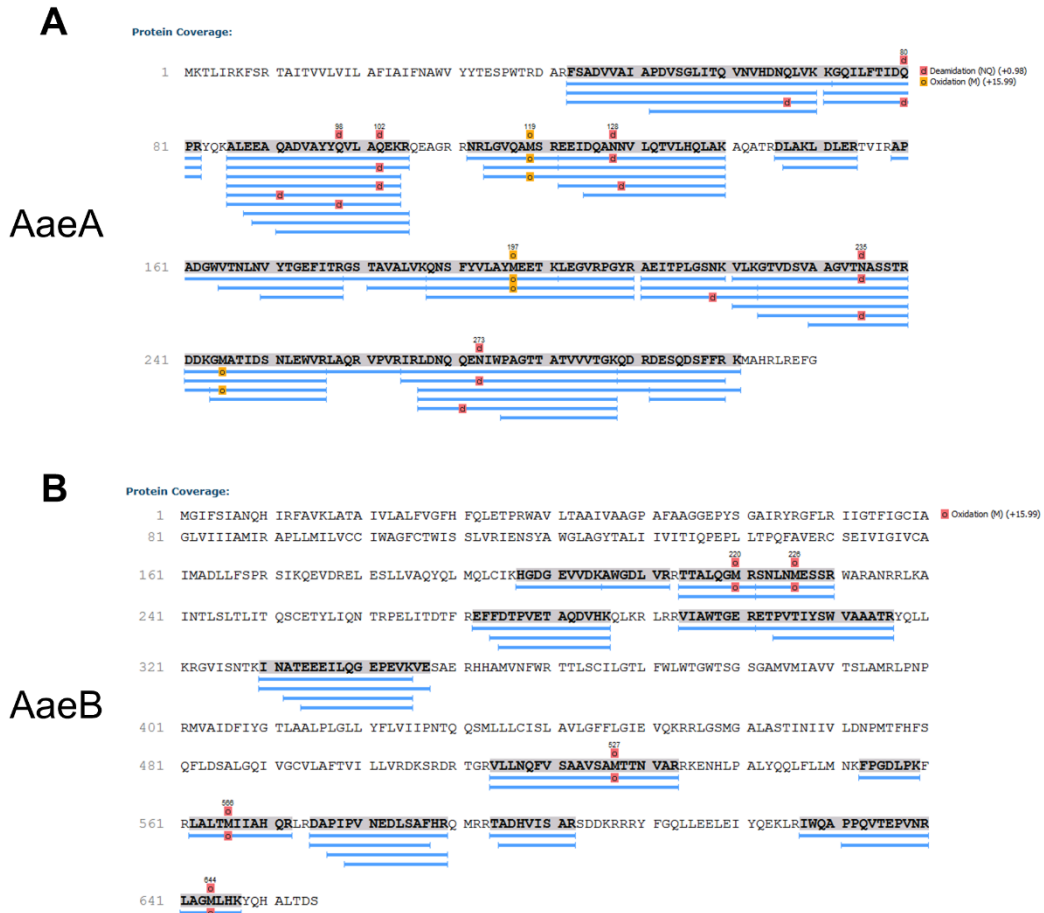


Figure 5.16 - Peptides identified in mass spectrometry analysis of bands in purified samples of AaeB co-expressed with AaeA.

A) The peptides identified in the smaller band identify this protein as AaeA. Position of deamidation identified with a red **d**, while oxidation is highlighted with an orange **o**. **B)** The peptides identified in the larger band identify this protein as AaeB. Position of oxidation identified with a red **o**.

An observation that supports AaeA having a stabilising “chaperone” type role is that when later elution fractions (that had less, or no, AaeA present) were stored overnight at 4°C, the AaeB sample precipitated. Conversely, the first elution fractions, which likely have a more optimal ratio of AaeA:AaeB, were stable over the course of several days storage.

In an attempt to enrich the purified protein samples with AaeA, membranes containing His-tagged AaeA (prepared in **Section 5.5.2**) were mixed with

membranes containing His-tagged AaeB and then purified. The elution lanes in **Figure 5.17C** show an increase in the amount of AaeA purified alongside AaeB, as expected because AaeA is also His-tagged.

This purified samples of AaeB, and AaeB enriched with AaeA, were used for later experiments in the following chapter. It had been shown that AaeB could now be expressed, solubilised, and purified in amounts that, while still relatively low in the context of crystallographic studies, could be used for EM and other biochemical studies. In an attempt to produce a significant amount of AaeB, a 30 L fermentation was undertaken by Mr David Sharples. Unfortunately, no expression of the target protein was detected, so expression in 500 ml flasks was continued.

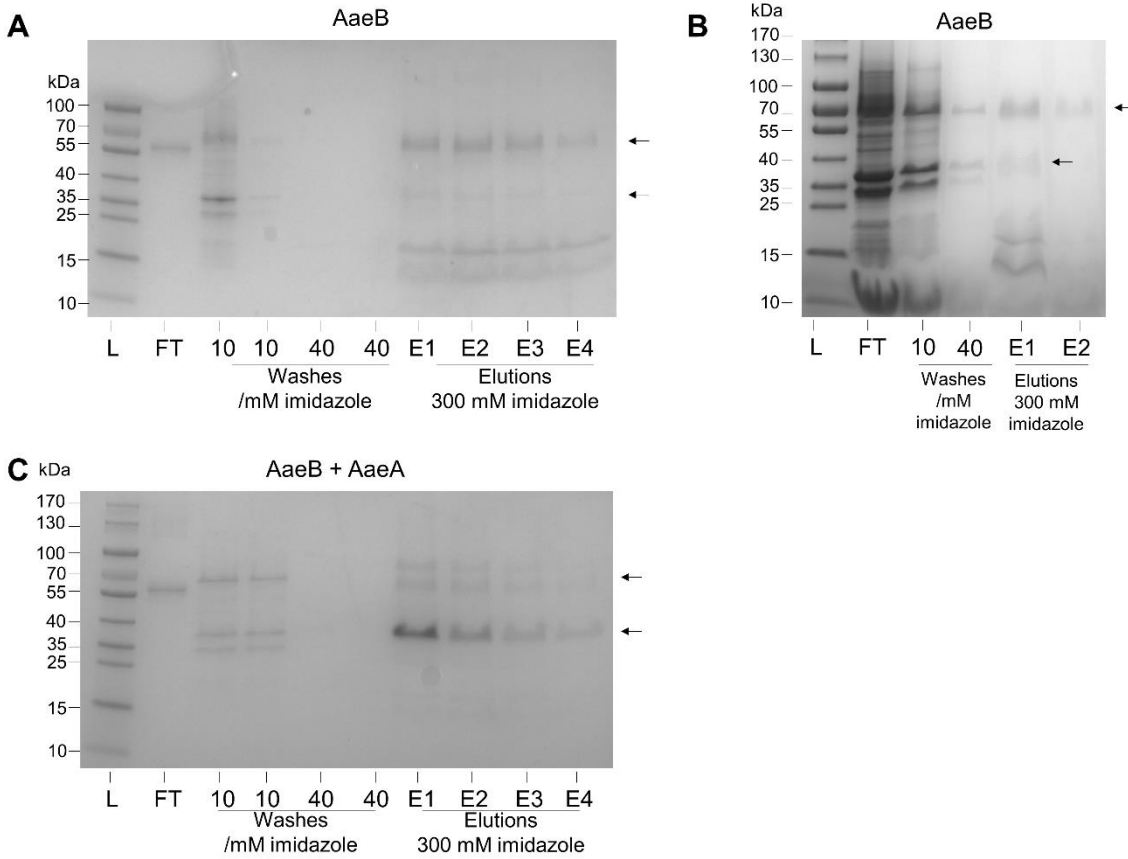


Figure 5.17 - Coomassie stained samples of three purifications of AaeB, run on SDS-PAGE gels.

A) AaeB co-expressed with AaeA purified by IMAC with 10 mM and 40 mM imidazole wash steps followed by elution at 300 mM imidazole (E1-4). **B)** A second purification of AaeB runs at approximately 70 kDa when eluted, with AaeA also present at approximately 35 kDa (both indicated by arrows). A second elution step continues to elute more AaeB, but there is no AaeA left on the resin. **C)** AaeB co-expressed with AaeA was mixed with membranes containing AaeA before purification in the same way as **A)**. It can be seen that the ratio of AaeA:AaeB in the eluted protein is increased.

5.2.2 Expression of AaeA and AaeA Δ TM

5.2.2.1 Initial expression conditions screening for AaeA and AaeA Δ TM

An initial screen of expression conditions for both AaeA and AaeA Δ TM was performed on a 3 ml scale with four strains (C41, C34, BL21 Gold, and BL21 Star) and three media (M9, LB, SB). **Figure 5.18** shows the results, with BL21 Gold and BL21 Star consistently outperforming C41 and C43. AaeA expressed well independently of the media, but AaeA Δ TM performed best in SB medium using BL21 Star cells.

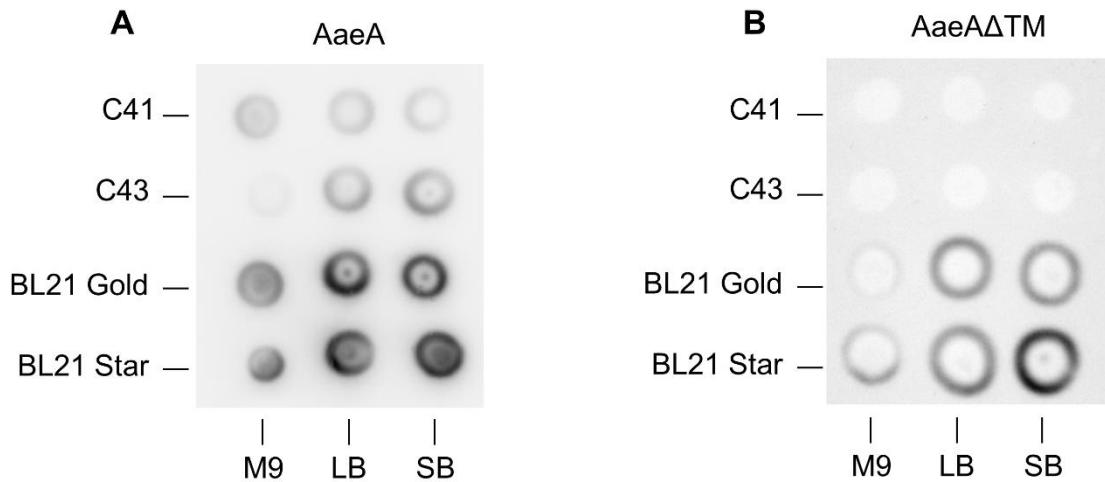


Figure 5.18 - Dot blot to assess initial expression trial for AaeA and AaeA Δ TM.

A) For AaeA, each medium performs equally well compared to each other, but best with BL21 Gold and BL21 Star cells. **B)** For AaeA Δ TM, BL21 Gold and BL21 Star are again the best cell types, with LB and SB performing as the best media. BL21 Star in SB medium appears to be the best condition.

Before further optimisation of expression was attempted, the localisation and solubility of the protein produced was assessed.

5.2.2.2 Assessing the membrane localisation and solubility of AaeA

AaeA had proved to express well in a number of conditions. The next step was to assess whether the predicted N-terminal hydrophobic helix is a transmembrane helix (as shown in **Figure 1.8A**); the hydrophobic helix could also be a hydrophobic core, around which the rest of the protein folds. The first experiment was to assess into which fraction AaeA would fall when membranes were prepared from the *E. coli* expressing AaeA.

Cells were lysed and the cell debris was removed by centrifugation at 14,000 *g*, after which the supernatant was spun at 100,000 *g* in order to collect the membrane fraction. If AaeA is a membrane protein it would be expected to be found in the final pellet of membrane proteins, whereas if it was a soluble protein it would be localised in the final supernatant, while misfolded and aggregated protein would be present in the initial pellet used to remove the cell debris. Each of these fractions was assessed by Western blotting (**Figure 5.19**) and signal was seen in the final pellet, indicating that AaeA was associated with the membrane fraction.

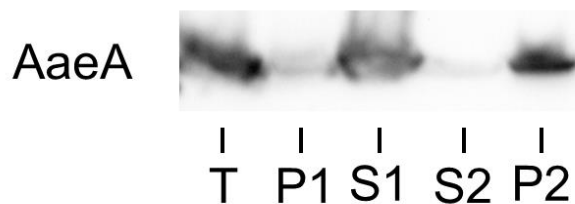


Figure 5.19 - Western blot of a fractionation of cells to locate AaeA.

20 μ L of each sample was loaded. Signal is seen in the total protein extract (**T**). After centrifugation at 14,000 *g*, some signal is seen in the pellet containing aggregates (**P1**), but most signal is seen in the supernatant (**S1**). After centrifugation of the supernatant at 100,000 *g* the protein is all seen in the pellet (**P2**), indicating that AaeA is part of the membrane fraction.

AaeA was shown to associate with the membrane fraction, as expected. The next stage was to assess whether AaeA was an integral membrane protein, or merely associated with the membrane – for example, as part of a complex with AaeB. Samples of membrane containing AaeA were re-suspended in buffer containing increasing concentrations of NaCl up to 1 M, or 50 mM sodium carbonate pH 11.5, which are both treatments that should extract AaeA into the soluble fraction if it is merely a membrane-associated protein. As can be seen in **Figure 5.20A**, after washing with salts the majority of AaeA was still present in the pellet, and so AaeA was not washed off the membranes and solubilised by these treatments. This provides strong evidence that AaeA is an integral membrane protein, as predicted.

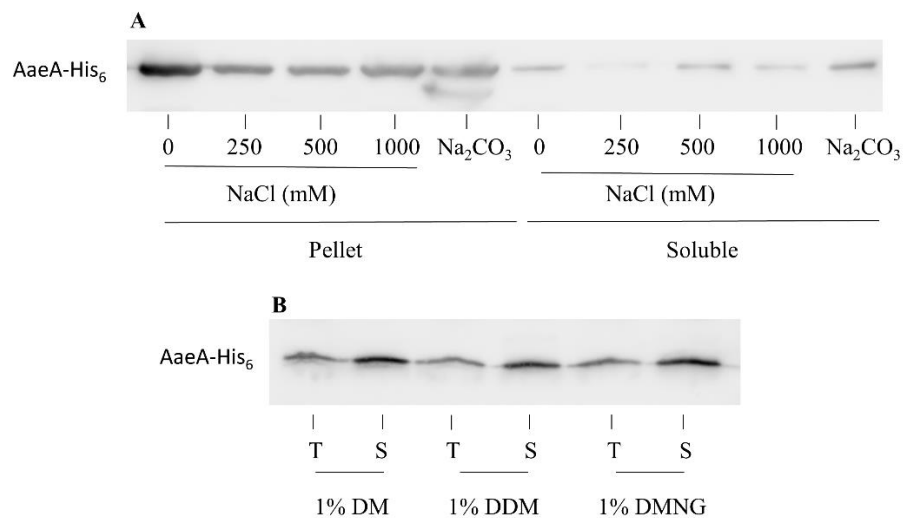


Figure 5.20 - Western blot of attempted solubilisation of AaeA.

20 μ L of each sample was loaded. **A)** NaCl and sodium carbonate washes. Increasing salt concentrations, or washing with sodium carbonate leads to a slight increase in soluble AaeA, but the majority of the material remains in the pellet, indicating that AaeA is truly an integral membrane protein. **B)** Western blot showing AaeA solubilisation by detergents. A concentration of 1% DM, DDM, and DMNG all solubilise AaeA quantitatively.

After confirming that AaeA is an integral membrane protein, the next requirement was to solubilise it in detergent. A range of detergents was trialled and **Figure 5.20B** shows that AaeA was readily made soluble by use of a 2 h incubation with 1% of either DM and DDM, commonly used maltoside detergents, or DMNG, a favoured detergent for crystallographic studies^{256,257}.

5.2.2.3 Purification of AaeA

After establishing that AaeA was an integral membrane protein that could be solubilised in detergent, a purification was performed. DDM-solubilised membranes containing AaeA were subjected to IMAC purification in order to isolate soluble, purified AaeA. As seen in **Figure 5.21**, AaeA could be isolated in high yield, with only minor contaminating bands. All bands were sent for identification by mass spectrometry and each was identified as AaeA. The higher-MW band at 70 kDa is likely a dimer of AaeA, while the lower-MW species are degradation products. The largest of these is also visible on the Western blot, indicating the His-tag is intact and thus cleavage is from the N-terminus, while the lowest bands are not detected by Western blotting, and so they must have a C-terminal deletion, possibly in addition to the N-terminal deletion. AaeA had been successfully purified, and the final step was to produce a sufficient amount to enable biochemical and structural studies to be performed.

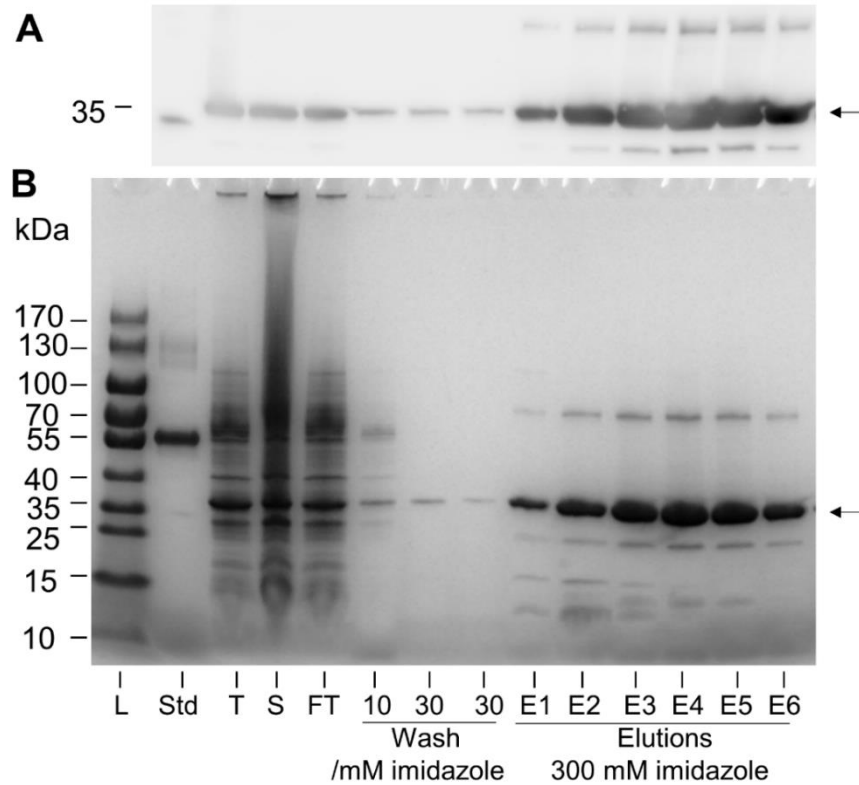


Figure 5.21 - Purification of AaeA.

A) Western with 20 ng protein standard (**Std**) and **B)** coomassie stained SDS-PAGE gel with 2 μ g protein standard (**Std**). Molecular weight markers are indicated on the left (**L**), Total (**T**), Soluble (**S**), Flowthrough (**FT**), washes of 10 mM (**10**) and 30 mM (**30**) imidazole, followed by elution in 300 mM imidazole (**E1-6**). Position of AaeA is indicated with arrows to the right. All bands were identified by mass spectrometry to be AaeA.

5.2.2.4 Production of sufficient AaeA for structural and biochemical studies

While AaeA has been shown to be membrane localised, it is readily solubilised by detergents, and successfully subject to purification by IMAC. Therefore, the next step was to ensure sufficient material could be made for biochemical investigations. Initial screening had proved promising and so to confirm the best expression conditions, the cultures were scaled up to 500 mL using BL21 Gold and LB medium, and eight conditions were assessed: autoinduction for 18, 24,

30, 48 h and IPTG induction for 2, 4, 6, 8 h. Total protein was extracted from these cultures and samples were assessed by Western blotting. **Figure 5.22** shows that yield increased over time for both IPTG induction up until 6 h, and autoinduction, with the largest increase between 18 h and 24h and smaller increases beyond. Autoinduction produced the highest yields, in the range of approximately 1-4 mg/mL, as determined by densitometry with reference to the protein standard run in the first lane. In addition, the protein runs as a clear band at approximately 35 kDa (the expected size).

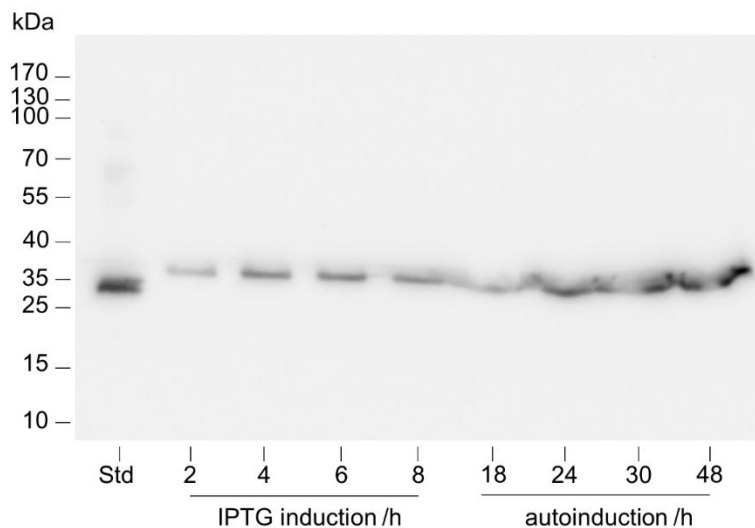


Figure 5.22 - Western blot of several samples of AaeA expressed in BL21 Gold using LB medium.

25 ng protein standard (**Std**), run alongside 20 μ L samples of four time-points of IPTG-induced (**2-8 h**) cultures and four time-points of autoinduced (**18-48 h**) cultures. Longer expression times after IPTG induction leads to more protein produced, up to a limit of 6 h, while higher overall yields are seen across autoinduced cultures. Positions of approximate protein molecular weight standards are indicated to the left.

Expression at the levels seen here is sufficient to produce enough protein (in the mg-range) for the anticipated structural and biochemical studies. In order to produce a large batch of AaeA it was decided to undertake a fermentation on a 30 L scale. For this the conditions chosen were the cell type BL21 Star in LB medium, using autoinduction for a period of 24h. This was carried out by Mr David Sharples in the fermentation facility. From the cell paste produced (approximately 500 g) membranes were prepared. These were then used for the studies that follow.

5.2.2.5 Assessing the membrane localisation and solubility of AaeA Δ TM

Having ensured that AaeA could be produced in a sufficient amount and solubilised reliably, the focus moved onto ensuring the same for AaeA Δ TM.

Having removed the putative TM helix that anchors AaeA in the membrane, it was expected that AaeA Δ TM would be soluble without the need for detergents, and that it would be found in the soluble fraction after lysing cells.

Total protein from cells expressing AaeA Δ TM was spun first at 14,000 *g* and then at 100,000 *g* and the supernatant and pellet from each step was assessed for the presence of AaeA Δ TM by dot blotting. As seen in **Figure 5.23**, the protein of interest was found in the cell debris fraction rather than the soluble fraction. This is indicative that the protein is likely to be in inclusion bodies, as seen in **Chapter 3**.

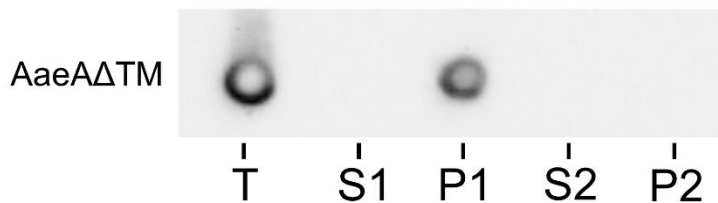


Figure 5.23 - Dot blot of supernatant and pellets from centrifugation of AaeA Δ TM.

Signal is seen in the total protein extract (**T**). After centrifugation at 14,000 *g*, all signal is seen in the pellet (**P1**), indicating that the material was part of the cell debris and AaeA Δ TM is not soluble.

Refolding of the inclusion bodies was attempted, as this has previously been shown to be an effective approach to production of proteins in *E. coli*²⁵⁸. The pelleted material was washed with 1 M NaCl and 2% Triton X-100, before

centrifugation and then the pellet was resuspended in 6 M GdnHCl to unfold the protein. After a 30-minute incubation the sample was spun to remove any aggregated material, and each fraction was assessed for the presence of the target protein by dot blot. As seen in **Figure 5.24**, salt and Triton X-100 do not solubilise AaeA Δ TM, and the protein remains in the pellet, however 6 M GdnHCl leads to retention of the target protein in the supernatant. This supernatant was taken and bound to cobalt resin overnight.

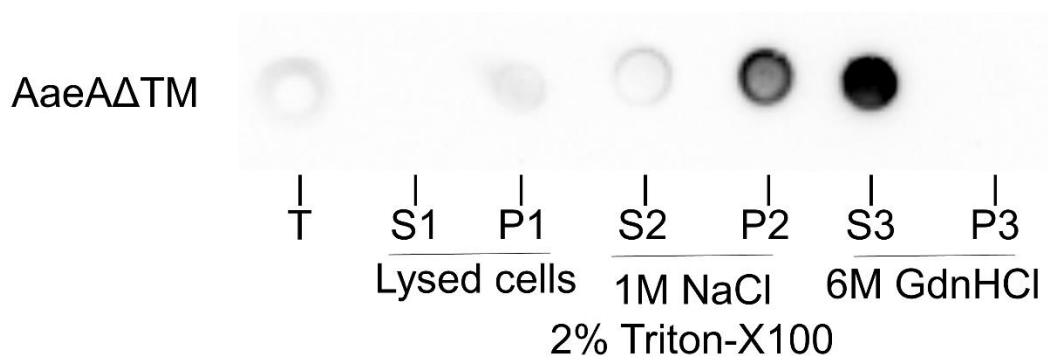


Figure 5.24 - Dot blot to monitor solubilisation of inclusion bodies of AaeA Δ TM.

Material seen in the total extracted protein (**T**) is found in the pellet after centrifugation at 14,000 g (**P1**) and after washing with 1M NaCl and 2% Triton X-100 (**P2**), but some material does seem to be solubilised by this washing, as a small fraction is seen in the discarded supernatant (**S2**). However, after incubation with 6 M GdnHCl, AaeA Δ TM is then present entirely in the supernatant (**S3**) and none remains in the pellet (**P3**).

When the protein was eluted from the resin using 300 mM imidazole buffer containing no GdnHCl, then dialysed to remove the imidazole, the protein once again precipitated out of solution, indicating refolding was not successful.

An attempt was made to express AaeA Δ TM at lower temperatures (both 30°C and 22°C) to ensure protein synthesis proceeded at a slower rate and give more time for the bacterial machinery to fold and chaperone the AaeA Δ TM, but this did not produce an improvement in solubility.

The hydrophobic helix could be hypothesised to be a hydrophobic core of the protein, around which the rest of the protein folds. This would be supported by the catastrophic lack of solubility when expressing AaeA Δ TM. However, the data presented **Section 5.5.2.2** indicate that AaeA is an integral membrane protein, and similar MFP proteins have been shown to have a transmembrane helix at the N-terminus, and thus this helix is most likely to be a legitimate transmembrane helix. In this case the most rational explanation for the complete lack of solubility of AaeA Δ TM as compared to AaeA (which expressed at similarly high levels but was readily solubilised by detergents and seems well behaved) is that the truncation leaves an unstable N-terminal end, perhaps with part of its structure unanchored or an incomplete fold. A better approach for attempting to produce a soluble truncated version of AaeA would have been to perform serial truncations, removing one residue at a time over a range of 10-20 residues after the predicted transmembrane helix. Screening all of these would then have provided more opportunities to identify a high-expressing, soluble, stable construct.

As the truncation of AaeA with the intention of forming a soluble protein had not proved successful, this line of investigation was abandoned in favour of experiments on full length AaeA and AaeB.

5.3 Summary and Conclusions

This chapter presents the work done to establish a protocol for the production and purification of bacterial ArAE family members. The first stage was construction of an expression vector to be used for bacterial membrane proteins based on pTTQ18, which has previously been used to express bacterial membrane proteins as it allows tight control of expression, minimising potential toxicity. The newly-created vector, designated pTTQ18-TP, has an improved cleavage site to allow removal of the His-tag and a cloning site that allows simple use of the restriction sites NheI and PstI to insert any bacterial ORF in frame with the RBS, start codon, and C-terminal fusion tag.

The ORFs for two bacterial genes, AaeB from *E. coli* and SO_0154 (Uniprot:Q8EKE1) from *S. oneidensis*, were cloned into this vector and expression was tested. Although some promise was seen with So, expression levels were consistently low and as such AaeB was prioritised. However, although AaeB expressed well, it was not stable and could not be readily solubilised by detergents. As such, the genes for AaeA and AaeB were both inserted into the expression vector, as this is how they are expressed from the native operon. This resulted in the production of AaeB that could be solubilised in a range of detergents, confirming that co-expression had stabilised AaeB.

In addition, a vector to express the partner to AaeB, AaeA, was produced, as this protein has also not been subject to study and although it has been shown to be vital, no studies have reported on its exact function. This protein was characterised as an integral membrane protein by assessing its localisation when fractionating *E. coli* cells and washing the membrane fraction with salt and Na₂CO₃ pH 11.5, and assessing the effect of detergents. The expression

conditions were optimised to produce mg-level quantities of AaeA and a protocol was devised for solubilisation and purification of AaeA.

A final construct was created to express a truncated version of AaeA, designated AaeA Δ TM, which was intended to express a version of AaeA lacking the putative transmembrane helix. This truncated construct, however, was not soluble when expressed in *E. coli* and could not be refolded from inclusion bodies.

In summary, an expression vector was created from two pre-existing vectors in order to express bacterial membrane proteins. Two proteins were cloned into this vector and expressed from it, with strategies devised for their solubilisation and purification. The first in vitro evidence of an interaction between AaeA and AaeB has been shown, with AaeA vital for AaeB stability (which cannot be detergent-solubilised in the absence of AaeA), perhaps operating in a “chaperone”-type manner. In addition, a strategy to express, solubilise, and purify AaeA has been developed, and its status as an integral membrane protein has been confirmed. Sufficient AaeA and AaeB has been produced to proceed with biochemical, biophysical, and structural studies.

Chapter 6

Biochemical and Structural Investigations of AaeA and AaeB

In the previous chapter both AaeA and AaeB from *E. coli* were expressed and purified in sufficient amounts to undertake a range of biochemical and structural experiments, which are detailed below.

6.1 *In silico* homology modelling of AaeA and bioinformatics analysis

The Phyre2 webserver was used to create a model of AaeA using “Intensive Mode”, which constructs a model of the protein based on the folds of homologous proteins that have already had their structures solved, and then attempts to model any remaining sections of the protein using *ab initio* techniques¹⁵⁹. The six proteins selected by the Phyre2 algorithm, due to significant sequence similarity, and used to produce the model were EmrA, AcrA, MacA, ZneB, MexA, and HlyD. These are all bacterial MFP proteins, as expected, and they were able to be used as a template for all but 67 residues. The initial 30 residues, which are predicted to be a membrane anchor region were not part of this template, as all MFP structures solved thus far have either been constructs that have been truncated to remove this section or this region has not been resolved in the crystal structure; the secondary structure prediction is modelled with high confidence in this region, however, as seen in **Figure 6.1**. In the model produced, this helix was modelled lying close to the rest of the protein and so it was manually adjusted using Chimera to be oriented along the plane of the molecule at an angle more appropriate for a transmembrane helix that needs to be embedded in the inner membrane (seen

in **Figure 6.2A**). The localisation of this helix in the membrane is supported by data in **Chapter 5** that showed that AaeA was found in the membrane fraction of *E. coli* and could be solubilised by addition of detergents.



Figure 6.1 - Secondary structure and disorder prediction by Phyre2 for AaeA.

The algorithm predicts an N-terminal sequence containing a transmembrane helix, followed by alternating regions of β -sheet and α -helix, with high confidence for secondary structure for almost the entire protein.

The model produced by Phyre2, visible in **Figure 6.2A** and also in four rotations in **Figure 6.4B**, has 3 domains after the hydrophobic N-terminal helix: a β -barrel domain, a lipoly domain, and an α -hairpin domain, each connected by flexible linkers. And as can be seen from the N-to-C rainbow colouring in the cartoon models, the β -barrel and lipoyl domains are constructed from β -sheets using

residues from both ends of the sequence. These domains are anchored to the periplasmic face of the inner membrane and project outwards into the periplasm. The model for AaeA lacks the membrane proximal domain (MPD) found in some other MFPs, notably RND- and ABC-type transporters that contain significant periplasmic loops¹⁰⁷. The MPD is predicted to act as a 'spacer' and also to interact with periplasmic loops of the inner membrane transporter by helping to load substrates from the periplasm. AaeA lacks this domain most likely because AaeB is predicted to extend only minimally into the periplasm, with the majority of the non-membrane-embedded regions of the protein being located in the cytoplasm and only small loops on the periplasmic face. Indeed, as the substrate for AaeAB is a metabolic product produced in the cytoplasm there is no need for the pump to load substrate in the periplasm, unlike in the case of multidrug transporters. Instead, the main site of interaction for AaeA and AaeB may be within the membrane, as suggested for MFS-type transporters which do not contain any significant soluble domains¹⁰⁷, and for the complex AcrA-AcrB-AcrZ, which has an additional small protein, AcrZ, located in the inner membrane that modulates the activity of the pump²⁵⁹.

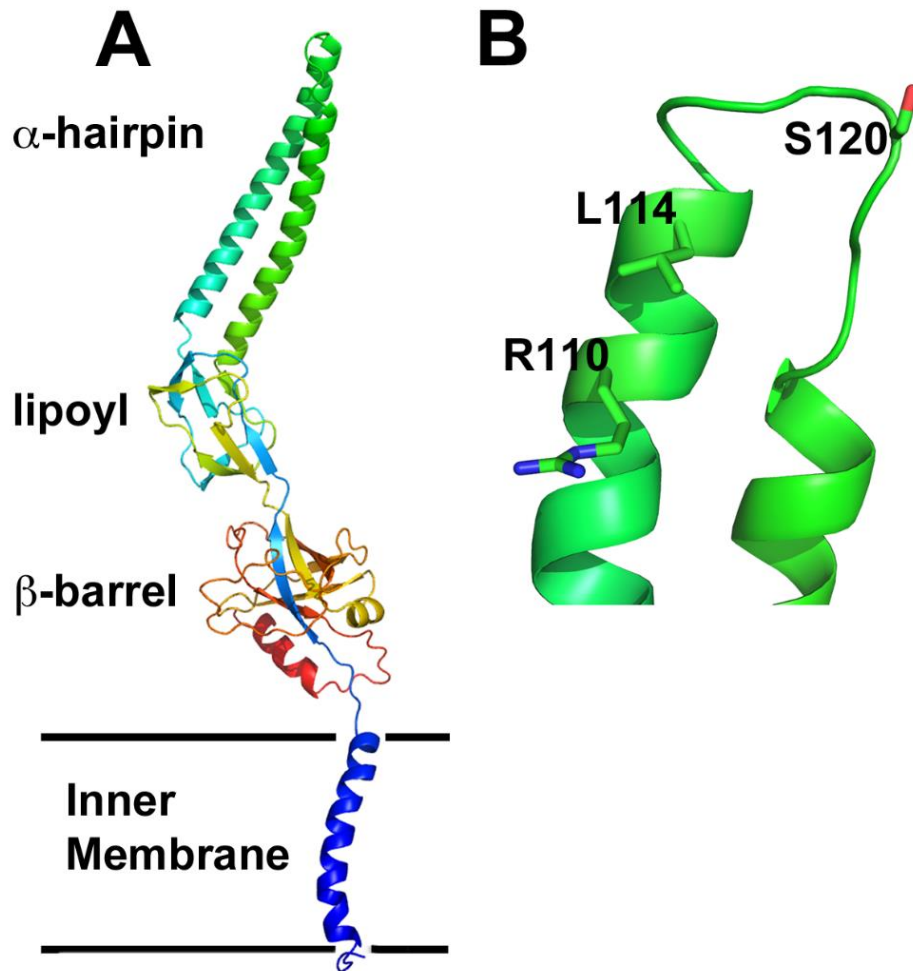


Figure 6.2 - Model of AaeA produced by Phyre2.

The position of the inner membrane helix was adjusted manually in Chimera. **A)** AaeA forms an extended structure projecting into the periplasm, consisting of a hydrophobic helix to anchor it into the inner membrane, a β -barrel domain, a lipoyl domain, and an α -hairpin domain. **B)** The residues of the RLS motif found among all MFP proteins is present at the tip of the helical α -hairpin.

6.1.1 The RLS Motif

In addition, an RLS motif (canonically, **R-X₃-L-X₆-S**) is present in AaeA and found at the tip of the α -hairpin, although with a single residue deletion between L114 and S120 as compared to the canonical sequence. This motif is found conserved throughout all MFPs and is proposed to be the key interaction between MFPs and TolC when forming tripartite complexes^{260–262}. As the AaeA-AaeB pump has not been shown to have a requirement for TolC³⁸, perhaps this deletion causes a rearrangement in the location of the residues required to form an interaction and so disrupts binding. To investigate this, the sequences at this RLS-motif region for several MFPs that form complexes with inner membrane proteins were aligned, shown in **Figure 6.3**. Several MFPs expressed in operons with PET family transporters have this single deletion conserved, and two MFPs that are not expressed in operons with any inner membrane transporter (YiaV, YibH) also show this deletion. This may support the hypothesis that that PETs do not need to use TolC to export substrate out of the cell and instead pump into the periplasmic space. As the known substrates for AaeB are organic acids such as pHBA³⁸ that are part of normal metabolism, it may be sufficient to merely remove them from the cytoplasm when they build up to toxic levels, after which it is possible for them to be taken up again and used productively.

	Starting residue		Corresponding IM Transporter	Family
Ec AaeA	110	R R N R L - G V Q A M S	Ec AaeB (YhcP)	PETs
Ec YdhJ	107	R R R H L - S Q N F I S	Ec YdhK	PETs
Sm FuaB	110	R D N A L - G N L V A S	Sm FuaA	PETs
Ec SdsR (YjcR)	139	R T E P L L K E G F V S	Ec SdsQP (YjcQ, MdtO)	PETs
Ec YiaV	102	R V D R L - M A D I V T	unknown	-
Ec YibH	102	R V D R L - Q A D L M T	unknown	-
Ec MacA	131	R Q Q R L A Q T K A V S	Ec MacA	ABC
Ec HlyD	216	R L T S L I K E Q F S T	Ec HlyB	ABC
Ec YbhG	133	R Q Q G L W K S R T I S	Ec YbhFRS	ABC
Ec EmrK	138	R R V P L A K Q G V I S	Ec EmrY	MFS
Aa EmrA	145	K L V K L D W E R Y K S	Aa EmrB	MFS
Pa MexA	119	R Y K L L V A D Q A V S	Pa MexB	RND
Ec AcrA	128	R Y Q K L L G T Q Y I S	Ec AcrB	RND
Ec YhiI	137	R S R S L A Q R G A I S	Ec YhiJ	unknown

Figure 6.3 - RLS motif alignment for several MFP proteins from several families.

MFPs from operons containing PET transporters, except for SdsR, have a deletion between the Leu and Ser of the canonical motif. Two MFPs found in operons without any inner membrane transporter also contain the same deletion, along with a S→T mutation that is also seen in HlyD.

6.1.2 Conservation and Surface Charge

The model produced by Phyre2 was analysed further by assessing the conservation of residues and also the surface charge. AaeA seems to have a well-conserved and a less-conserved face, as seen in **Figure 6.4B**, and there are a number of highly-conserved residues on the β -barrel domain which are perhaps important for interaction with AaeB. The outside face of the protein shows a number of poorly conserved residues; as these are facing the solvent rather than interacting with any other protein surface there is unlikely to be any evolutionary pressure to prevent modification of these residues. Conversely, the inner face of the β -barrel and lipoyl domains appear to have much more strongly conserved residues, reflecting their likely role in interaction with substrates and AaeB.

The analysis of surface charge showed that opposing sides of AaeA are positive and negative, as seen in **Figure 6.4C**. This may provide a mechanism for side-by-side multimerisation, with a negative “right-face” having electrostatic affinity for the positive “left-face”, allowing the AaeA monomers to assemble into a funnel shape as seen for other MFPs. The inside face of the α -hairpin, which may be expected to form a barrel for the substrate to travel along, has an overall neutral charge, with some regions of positive charge, which would interact favourably with a deprotonated acidic substrate such as pHBA.

Interestingly, the tip of the membrane helix is positively charged and as this is located in the cytoplasm: this is in agreement with the “positive inside” rule for membrane proteins²⁶³.

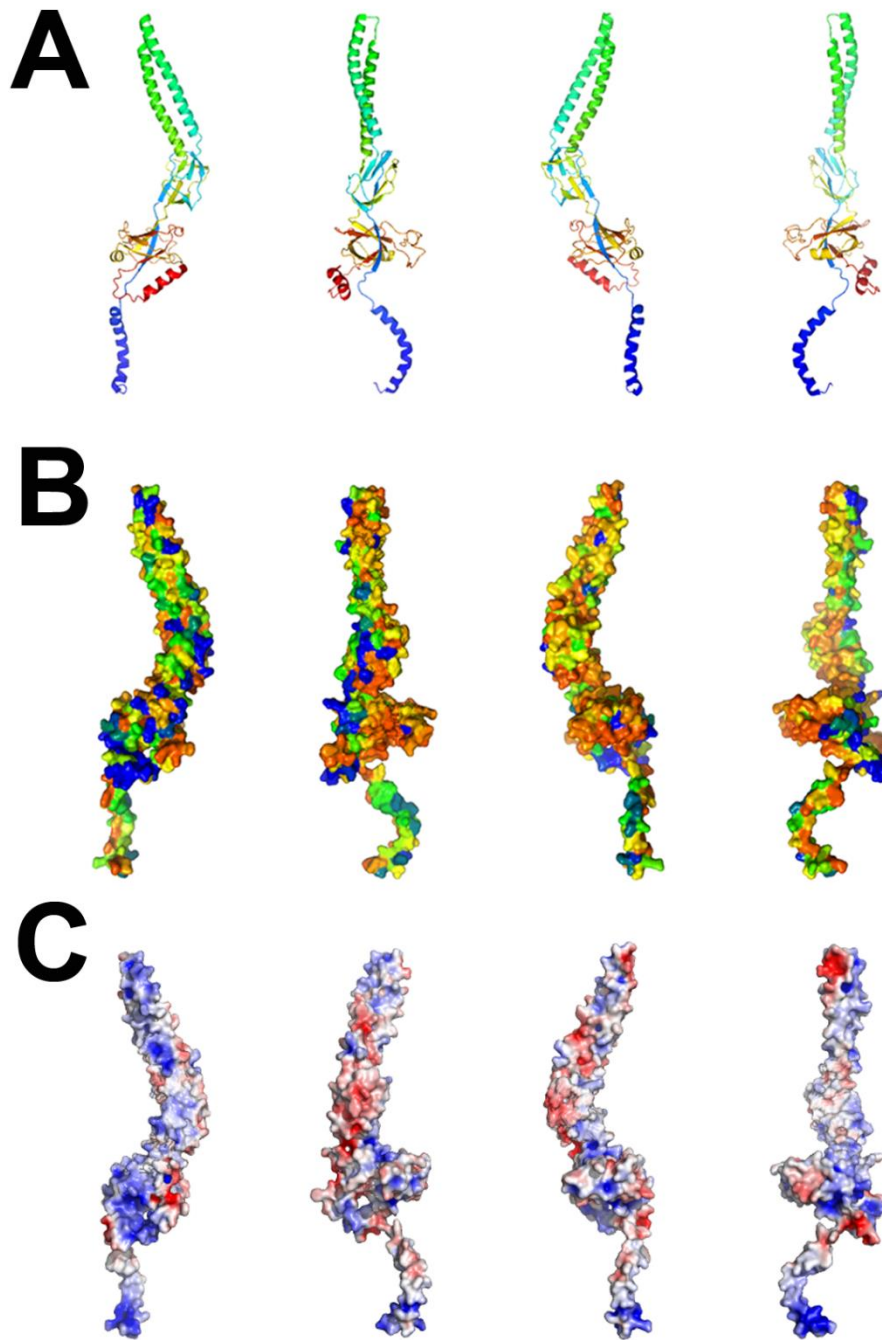


Figure 6.4 – Examination of the AaeA model.

A) Four 90 degree rotations (in order: left, outside, right, inside) of the AaeA model in cartoon representation coloured with a rainbow from N to C, and **B)** surface representation coloured as to the conservation score as calculated by Consurf²⁶⁴ in order from lowest to highest blue, green, yellow, orange, red. The protein displays a less strongly conserved left hand side than right hand side, and shows a number of highly conserved residues on the β -barrel, likely to be important to form interactions with the inner membrane transporter. **C)** surface charge coloured from blue (positive) to red (negative), showing a positively-charged left hand side and a positively charged right hand side, and while the outer side is charged, the inner face is largely neutral.

6.2 Examination of buffer conditions for optimal stability of AaeA.

Since the ultimate goal for the project was to undertake structural studies, especially crystallography, conditions for stability of AaeA were optimised. This was achieved using the Optim (Avacta) (now called the UNit from Unchained Laboratories)¹⁸⁰. The capillary system allowed 48 samples to be assayed in parallel on a very small scale (5 μ L per condition) and this allowed the following conditions to be examined: pH from 3.5 to 9.5, NaCl concentration from 0 to 1500 mM, and glycerol from 0% to 35%.

Two measurements were taken in parallel. Firstly, intrinsic fluorescence to assess protein unfolding, reported as the barycentric mean (BCM). As the protein unfolds, fluorescent side chains are exposed to a different local environment, for example by bringing residues from the core of the protein and exposing them to the solvent, and thus changing their fluorescent signal. In this case, protein unfolding leads to a decrease in BCM of fluorescence, and by taking the derivative of the fluorescence curve it is possible to identify the midpoint of the unfolding transition, and calculate the melting temperature (T_m). An example can be seen in **Figure 6.5A**, where the vertical orange line marks the midpoint of the unfolding transition. Secondly, static light scattering (SLS) was measured at 266 nm to assess aggregation. As the protein unfolds it aggregates, and this forms larger particles which scatter more light. By identifying the start of this aggregation transition as light scattering increases sharply an Aggregation Temperature (T_{agg}) can be identified. An example can be seen in **Figure 6.5B**, where the vertical orange line marks the beginning of aggregation.

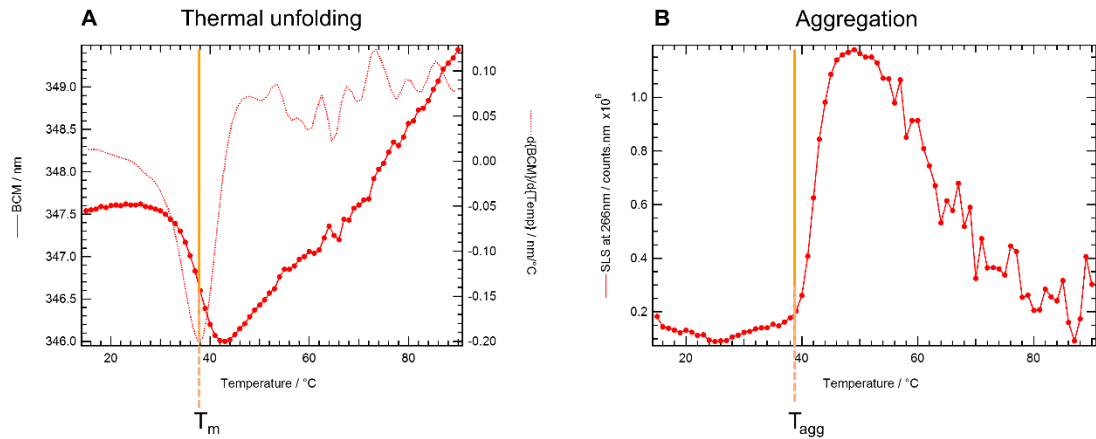


Figure 6.5 - Example graphs produced from Optim data.

A) Thermal unfolding curve. Solid line shows the BCM and a decrease as temperature is increased. Dashed line shows the derivative and the T_m can be identified at the inflection point (orange line). **B)** Graph of SLS, a sharp increase is seen at T_{agg}, indicated by the orange line.

The values measured for T_m and T_{agg} over the conditions assayed were plotted against their respective condition, as shown in **Figure 6.6**. In summary, stability as measured by thermal melting and aggregation was not strongly affected by the pH, with perhaps evidence of increased stability at pH 5.5, but no obvious and consistent trend over the pH range assessed (**Figure 6.6A** and **Figure 6.6D**). Increasing salt concentration promoted stability, but not significantly, with only a few degrees' difference between 0 mM and 1 M NaCl (**Figure 6.6B** and **Figure 6.6E**). The strongest stabilising effect came from glycerol, with increasing concentrations strongly stabilising AaeA (**Figure 6.6C** and **Figure 6.6F**). Thus, the buffer chosen for continued work with AaeA would now contain 30% (w/v) glycerol, where feasible, in order to provide maximum stability.

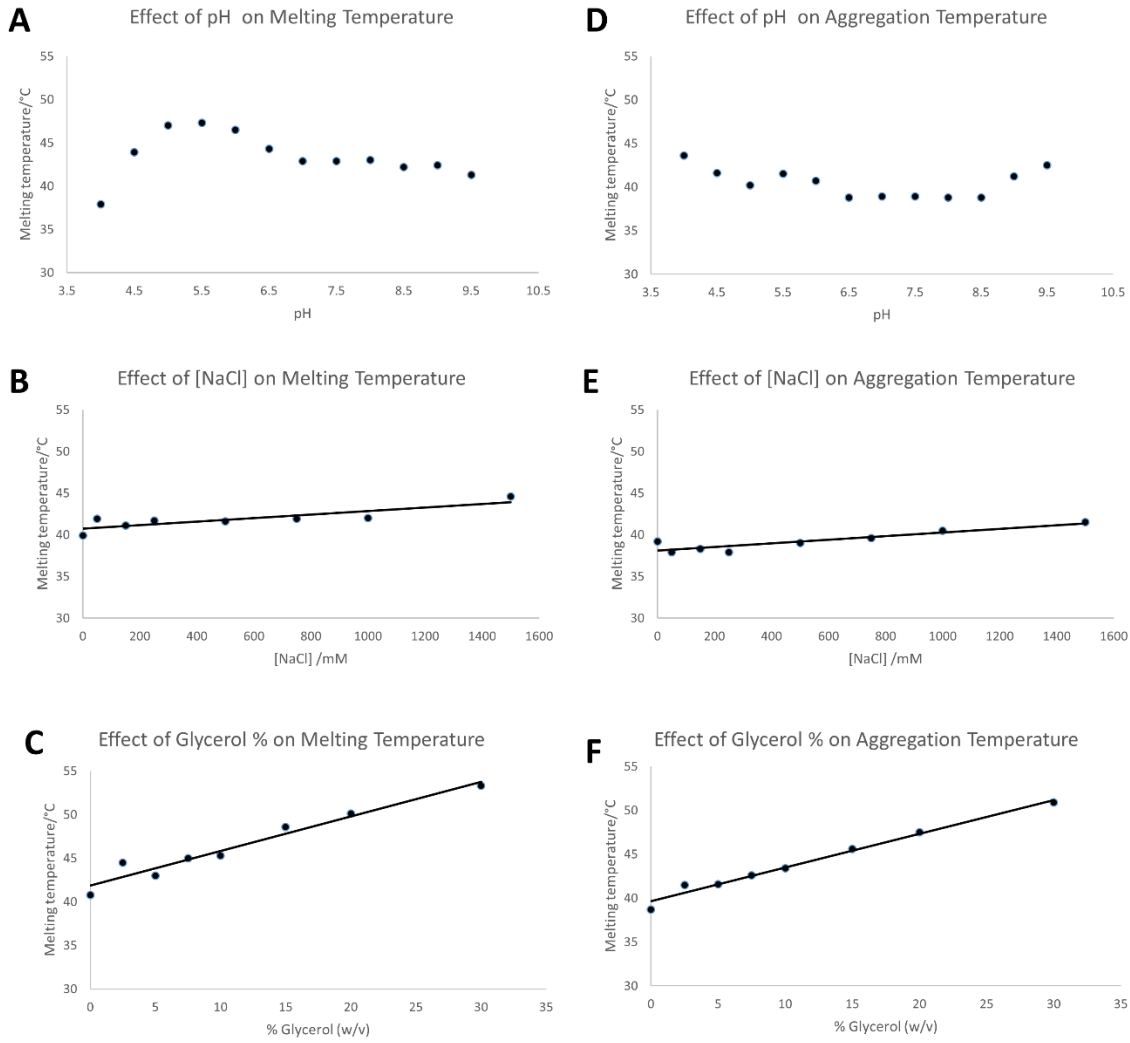


Figure 6.6 - Summary graphs from Optim reporting T_m (A-C) and T_{agg} (D-F) over a range of pH (A, D), NaCl concentration (B, E), and Glycerol % (C, F) in the buffer.

Protein stability is relatively insensitive to pH. NaCl has a small stabilising effect. A strong stabilising effect was seen when increasing glycerol.

6.3 Circular Dichroism measurement of AaeA

With the *in silico* model, some validation was required. Circular dichroism (CD) in the far-UV region of the spectrum can be used to provide information on the secondary structure composition, as α -helices and β -sheets each have distinctive CD spectra based on the regular arrangement of their backbone peptide bonds. As such, CD was employed in order to generate an estimate of the relative proportions of secondary structural elements and see if it is in agreement with the model produced by Phyre2. CD would also confirm that the protein was folded, and provide evidence that the produced material was of good quality.

Two repeat spectra were captured between 180-260 nm at 20°C in a buffer containing, 20 mM HEPES pH 8.0, 100 mM NaCl, 25% (w/v) glycerol, and 0.05% DDM, plus a spectrum of the buffer alone to be used as a blank. These spectra are presented in **Figure 6.7A**. When the buffer blank was subtracted from the measured spectrum and repeats averaged, this produced a spectrum shown in **Figure 6.7B**. It was not possible to assign secondary structure from this data when using analysis by the SELCON3, CONTIN, or K2D methods as the errors between measured data and algorithm's models were too large.

There is evidence of the typical "double bump" trough centred around 215 nm, indicating α -helical content, but the signal below 190 nm seems to drop off sharply, indicating that the conditions were not optimal. The buffer contains chloride ions, glycerol, and detergent, all of which absorb in the UV region and thus worsen the signal-to-noise ratio for the measurement. An improved choice of buffer conditions may be able to enhance this, which is explored below.

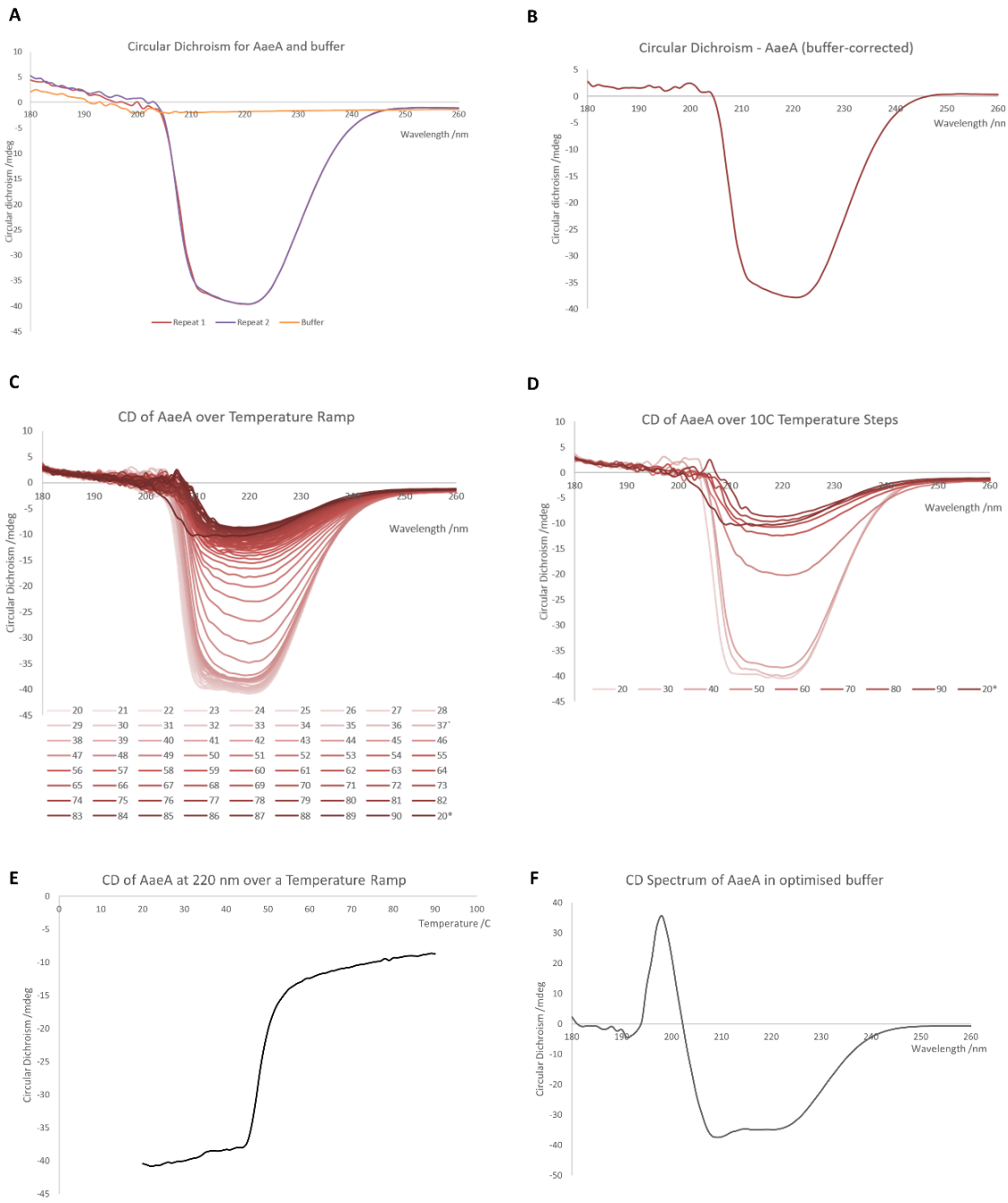


Figure 6.7 – CD analysis of AaeA.

(contd. overleaf)

A) Raw CD spectra recorded for a sample of AaeA (in duplicate) and a buffer sample used as a blank. The repeats overlay almost exactly apart from below 200 nm, where they diverge, indicating a less-reliable region with lower signal to noise. **B)** CD spectrum for AaeA, corrected using the buffer blank and averaged from two repeats. The sharply negative signal in the region between ~200-240 nm is indicative of α -helical secondary structural elements, but CD signal is not seen in the region below 200 nm. **C)** CD spectra from a sample of AaeA measured over a temperature ramp from 20-90°C in 1°C steps. The curve indicating α -helical content seen in the 20°C sample steadily disappears as the temperature is raised. **D)** Data from previous graph in 10°C steps. The collapse in structured signal is clearly seen between the 40°C and 60°C measurements. **E)** CD signal measured at 220 nm over a temperature ramp from 20-90°C. A sharp transition is seen at approximately 48°C, indicative of cooperative unfolding as would be expected of a structure containing a coiled-coil, as predicted. **F)** CD spectrum for AaeA in an improved buffer with a lowered concentration of chloride and glycerol. The characteristic α -helical “double-bump” is seen and signal below 200 nm is much better defined.

6.3.1 Thermal Melt

In addition to the static spectrum, a thermal melt was then taken in 1°C steps from 20-90°C. This showed that as temperature increased, the intensity of the circular dichroism in the spectrum dropped towards zero, indicating a loss of structure (**Figure 6.7C**). This is evidence that the apparent structure seen in the initial spectrum is real, as the thermal ramp would be expected to unfold the protein, losing structure, and therefore losing signal. A final measurement back at 20°C after the melting curve showed an almost identical spectrum to the 90°C, indicating that the structure had been irreversibly lost, likely by aggregation as the protein unfolded, or potentially because a specific chaperone is required for re-folding. The spectra presented in **Figure 6.7D**, in 10°C steps, shows more clearly the sharp transition occurring between the 40°C and 60°C measurements. An analysis of the CD intensity at 220 nm over the range of temperatures in **Figure 6.7E** shows a sharp transition centred at approximately 48°C, indicative of a defined α -helical structure that cooperatively

unfolds, as has been seen in previous studies^{265,266}. This is likely to be the coiled-coil found in the predicted structure, which would be expected to show strong cooperative unfolding. Additionally, the melting temperature derived from this experiment agrees strongly with the melting temperatures measured in **Section 6.2**, particularly as **Figure 6.6C** estimated a T_m of approximately 50°C for these buffer conditions.

6.3.2 Optimised buffer

As glycerol, chloride, and detergent may absorb UV, creating too high a signal reducing the available information, the buffer was altered for a second measurement. Although the presence of detergent is non-negotiable for a membrane protein, the other components could be lowered or removed. The glycerol content was lowered to 10%, which was judged to still provide some stabilisation and 100 mM NaCl was replaced with 500 mM NaF, which removed the chloride entirely but increased the ionic strength of the buffer to help counteract the drop in glycerol.

As can be seen in **Figure 6.7F** the CD signal at lower wavelengths is improved by the changed buffer conditions with a much more well defined shape.

Analyses of this spectrum with three algorithms (SELCON3, CONTIN, and K2D) with the reference set SMP180, designed for use with membrane proteins showed that the measured proportion of α -helical, β -sheet, and coil were each roughly a third, as predicted by the Phyre2 model. The calculated predictions are summarised in **Figure 6.8** and although the calculated values for both α -helix and β -sheet are lower than the model predicts, this can be explained by the fact that the AaeA measured here is free in detergent solution, which may promote a more 'loose' structure, and is not in complex with AaeB which could

have the effect of promoting a particular secondary structure. Conversely, the structures that the model is based on are crystal structures which will necessarily be in a more rigid conformation. Previous studies on MFPs have measured a degree of conformational flexibility, also¹⁴⁵. These data, therefore, provide support the Phyre2 model. While CD is able to determine the proportions of each secondary structure present in the solution as a whole, it cannot determine which specific residues are involved in forming those secondary structure elements; as such, it cannot be inferred that the model presented here, or any other model, is correct without further evidence, but these data support, rather than rule out the model's prediction.

A

	alpha helix	beta sheet	other/unordered
Predicted by Model	31	32	37
Average of CD analyses	26	24	50
CONTIN	28	23	49
SELCON3	19	31	51
K2D	32	18	49

B

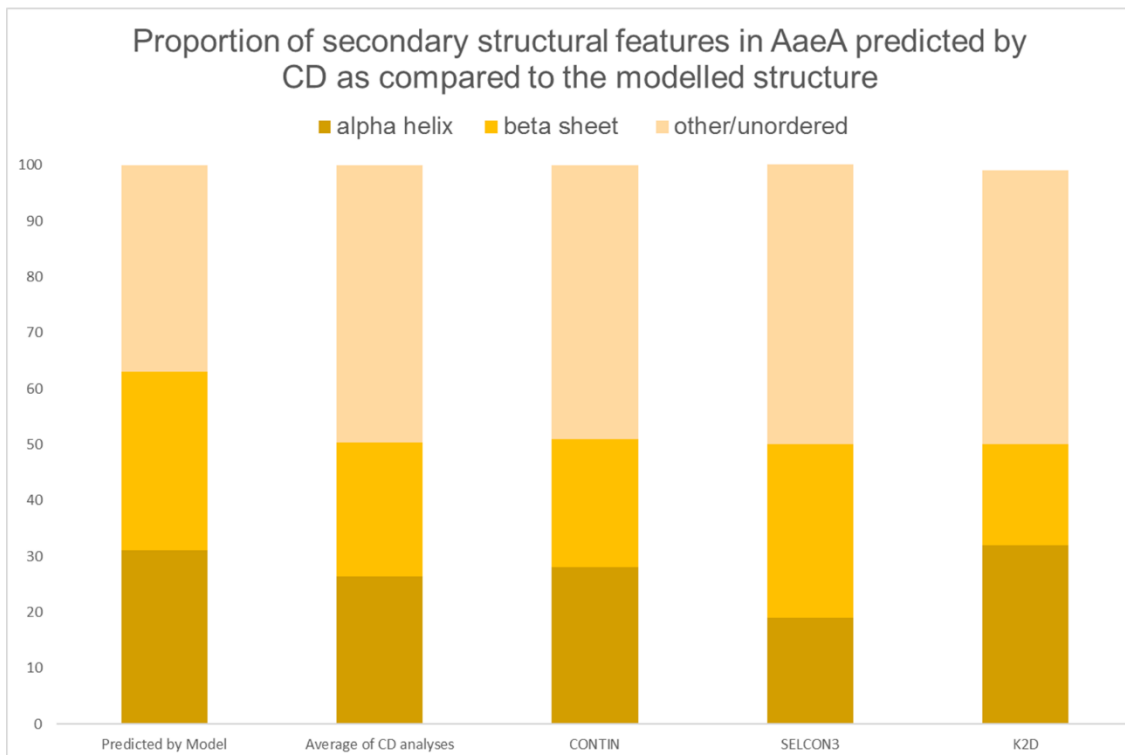


Figure 6.8 - Secondary structure prediction by CD as compared to Phyre2 Model.

A) Summary of the output of three algorithms used to process the spectrum shown in **Figure 6.7F**. **B)** stacked bar charts to show the relative proportions of α -helical and β -sheet found in the Phyre2 model presented in this chapter against the relative proportions measured by CD.

6.4 X-ray Crystallography of AaeA

As large amounts of AaeA could be produced, solubilised and purified, the material was taken and used to set up trays of protein in crystallography screening buffers. Six commercial screens were selected, which are formulated to be suitable for both soluble and membrane protein crystallography. The six screens were MemGold, MemPlus/MemStart, Morpheus, ProPlex (all Molecular Dimensions), Crystal Screen, and Peg/Ion (both Hampton Research).

Purified AaeA solubilised in DNPG with cholesterol hemisuccinate (CHS) (shown to help stabilise membrane proteins for crystallography²⁶⁷) was taken and concentrated to approximately 12 mg/mL and used to set up parallel drops for each of the 96 conditions in each screen, for a total of 1252 drops. Drop 1 in each well was made from a mixture of protein:well solution at a 1:1 ratio, while Drop 2 was made from a 2:1 ratio, meaning that it had a higher effective concentration of protein, but a lower concentration of precipitant. These trays were stored and imaged immediately and then again at 1, 2, 3, 5, 8, 13, 21 and 34 days. Normal brightfield (visible light) images were taken alongside both Ultraviolet Two-Photon Excited Fluorescence (UV-TPEF) and Second Harmonic Generation (SHG) images. UV-TPEF excited any protein crystals present, as protein crystals will absorb UV light, meaning any protein crystals present in the drop can be more easily identified, and SHG identifies chiral crystals. Both of these help to differentiate between crystals that are formed of protein, and those that are formed from buffer components, or detergents. It should be noted, however, that not all protein crystal space groups are chiral and so a negative hit does not definitely rule out a crystal as being protein, and that some non-protein molecules can crystallise in chiral space groups, especially, for example,

sulfate salts. These two imaging techniques, however, form another tool useful to help identify protein crystals.

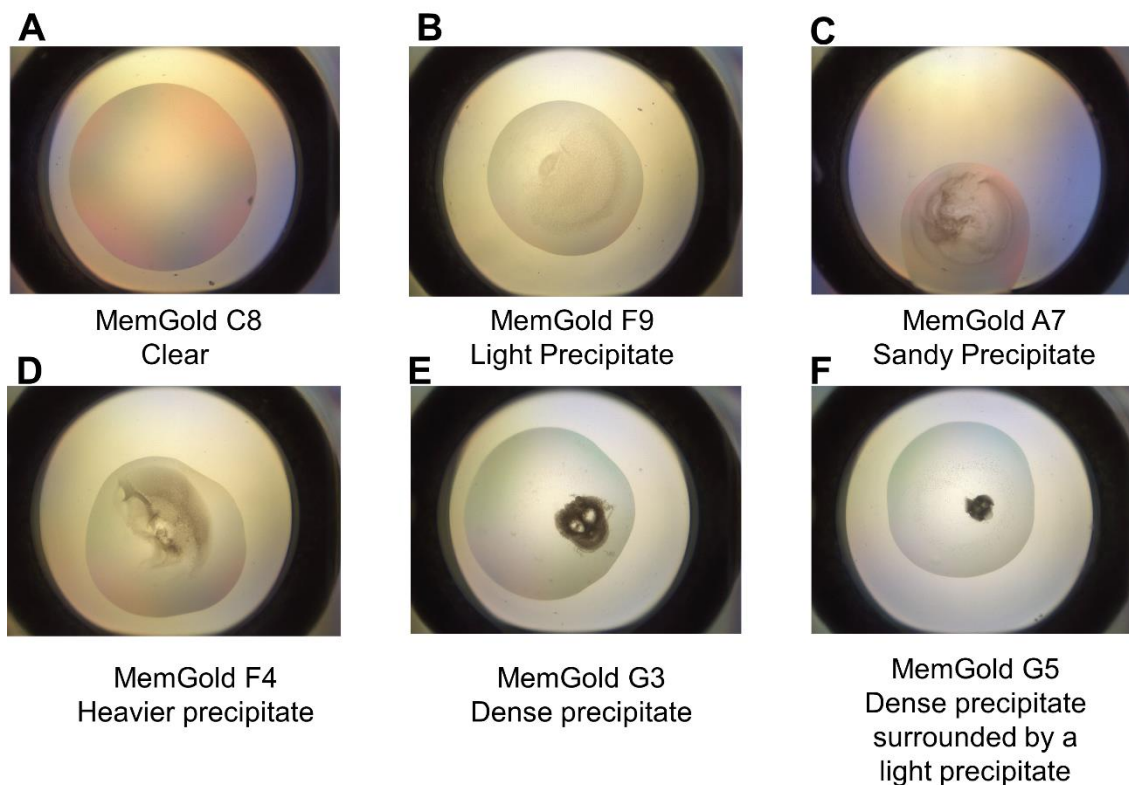


Figure 6.9 - A range of images of representative results from screening conditions for growth of crystals.

Wells ranged from completely clean (A), through a range of levels of precipitation (B-D) up to a heavy, dense precipitate (E), and some wells contained a mix of levels of precipitation (F).

A representative selection of results seen in the wells of the crystal screen is shown in **Figure 6.9**: broadly speaking observation of the images at 34 days, and on the intervening days, showed that some wells contained completely clear drops, while some had heavy precipitation, with a range of drops in between these extremes. This is indicative that the concentration of protein present in the drops is in the correct range, and the concentration used in

similar studies of other MFPs has indeed been at approximately this concentration^{145,254}.

A number of wells showed evidence of some “microcrystals”, perhaps as large as 10 µm across. A selection of these is shown in **Figure 6.10**. Crystals of this size are at the lower limit of feasibility in terms of data collection on a microfocus synchrotron beamline. Moreover, crystals of this size are difficult to fish from the drop into a loop, can be unstable, and often give poor data. As such the microcrystals seen here are unlikely to give good structural data and further optimisation is needed before any further steps of data collection are taken.

Looking at the examples given in **Figure 6.10** more closely it is possible to draw some further observations. The SHG signal seen in **Figure 6.10C** is initially promising, but when the corresponding region of the UV-TPEF image is examined there appears to be a decreased signal, with a haze of protein seen through the drop accompanied by a dark spot. This indicates that the crystal present is likely to be a salt crystal from the well buffer solution, as it does not absorb UV. By contrast, **Figure 6.10D** shows a small crystal next to a ring structure, this is present as both a brighter spot in the UV-TPEF and a clear signal in the SHG image. This is promising and indicative of a true protein crystal. Nevertheless, the size of the crystal and its location within a dust-like precipitate make it difficult to see, difficult to isolate, and it is unlikely to prove amenable to study. In all, 25 conditions showed some signs of potential hits and these conditions are summarised in **Table 6.1**. The size of potential protein crystals that could be visualised by eye and with use of UV-TPEF and SHG

imaging was consistently small and generally obscured by precipitation within the drop.

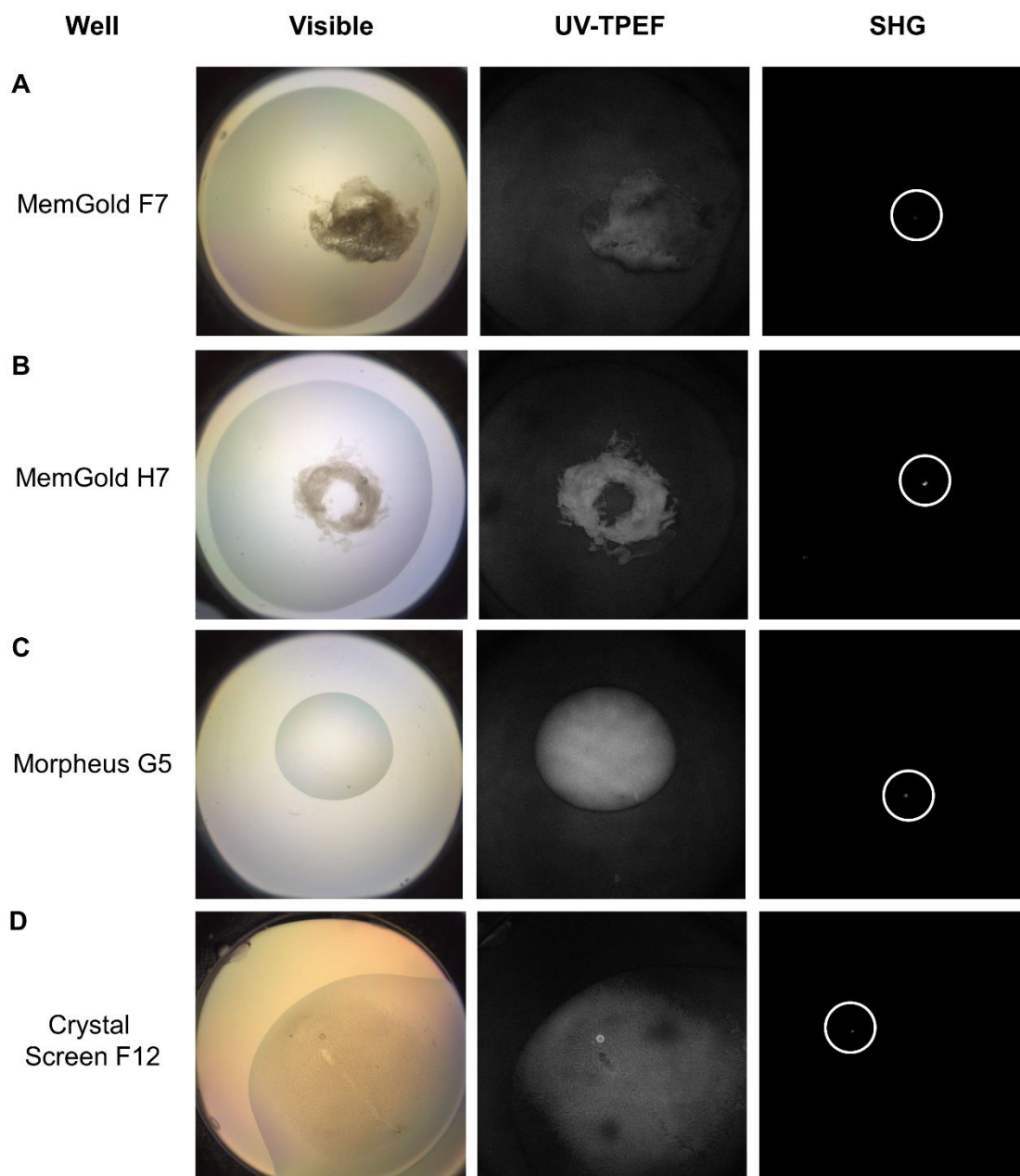


Figure 6.10 – Representative examples of potential hits from crystal screening.

The small area of signal is ringed in the SHG images for clarity, and putative crystals are in the equivalent position in the Visible and UV-TPEF images. All potential hits contain only very small crystals, giving weak signal and are generally not clearly visible under normal illumination. Drop size was 0.2 μ L

Table 6.1 - Table of conditions within the wells of putative hits in crystal screen.

The characters before the hyphen denotes the well (e.g. B8) and the number after denotes either drop 1 (1:1 protein:buffer ratio) or drop 2 (2:1 protein:buffer ratio) using that condition. No clear pattern of conditions can be drawn out of these data.

CrystalScreen		MemStartMemsys	
B8-1	0.2 M (NH ₄) ₂ SO ₄ (Salt) 0.1 M Na Acet 4.6 pH (Buffer) 25 %w/v PEG 4K (Precipitant)	A8-1	1 M MgSO ₄ (Precipitant) 0.1 M ADA 6.5 pH (Buffer) 0.1 M LiSO ₄ (Salt)
F12-2	0.1 M HEPES 7.5 pH (Buffer) 4.3 M NaCl (Precipitant)	G12-1	0.1 M MOPS 7 pH (Buffer) 12 %w/v PEG 4K (Precipitant)
MemGold		A9-2	0.1 M NH ₄ H ₂ PO ₄ 6.5 pH (Buffer)
H10-1	0.002 M ZnSO ₄ (Salt) 0.08 M HEPES 7 pH (Buffer) 25 %v/v Jeff ED-2001 (Precipitant)	ProPlex	
D11-2	0.2 M Na Acet 6.8 pH (Buffer) 8.8 %w/v PEG 6K (Precipitant)	C5-1	0.1 M TRIS 8 pH (Buffer) 20 %w/v PEG 4K (Polymer)
D12-2	0.2 M KCl (Salt) 0.1 M MES 6.5 pH (Buffer) 18 %w/v PEG 6K (Precipitant)	E10-1	0.1 M TRIS 8.5 pH (Buffer) 12 %w/v PEG 8K (Polymer) 0.2 M (NH ₄) ₂ SO ₄ (Salt)
PEG/Ions		F11-1	0.1 M Na Acet 5 pH (Buffer) 1.5 M (NH ₄) ₂ SO ₄ (Salt)
H11-1	1 %w/v Tryptone (Additive) 0.05 M HEPES 7 pH (Buffer) 12 %w/v PEG 3350 (Polymer)	G10-1	1.6 M Na K Phos 6.5 pH (Buffer)
G12-2	16 %w/v PEG 3350 (Polymer) 0.1 M CBTP 3.4 pH (Buffer)	H4-1	1.4 M Na ₂ Malon 6 pH (Buffer)
H12-2	1 %w/v Tryptone (Additive) 0.05 M HEPES 7 pH (Buffer) 20 %w/v PEG 3350 (Polymer)	E10-2	0.1 M Na Acet 4.5 pH (Buffer) 8 %w/v PEG 8K (Polymer) 0.1 M Mg Acet (Salt)
Morpheus		E11-2	0.1 M TRIS 8.5 pH (Buffer) 12 %w/v PEG 8K (Polymer) 0.2 M (NH ₄) ₂ SO ₄ (Salt)
F4-1	37.5 %w/v M1K3350 (Precipitant) 0.1 M MB1 6.5 pH (Buffer) 0.12 M Monosacc (Complex ingredient)	G12-2	0.1 M HEPES 7 pH (Buffer) 1 M Na ₃ Cit (Salt)
G5-1	30 %w/v 550M_20K (Precipitant) 0.1 M MB2 7.5 pH (Buffer) 0.1 M MCA (Complex ingredient)	H6-2	0.1 M MES 6.5 pH (Buffer) 1.6 M MgSO ₄ (Salt)
		H8-2	0.1 M Na Cacod 6 pH (Buffer) 25 %v/v MPD (Organic (non-volatile)) 0.05 M Ca Acet (Salt)
		H9-2	0.1 M Imidazole 7 pH (Buffer) 50 %v/v MPD (Organic (non-volatile))

The conditions present in the wells of potential hits (summarised in **Table 6.1**) appear to show no particular pattern of conditions that are represented. The pH ranges from 3.4 to 8.5, perhaps reflecting the broad range of stability seen in

Section 6.2 over a range of pH values. Similarly, the buffers and additives present within the wells covers a broad range of chemical species, and although it can be seen that some compounds re-occur in several of the wells, such as various PEGs, or sulfate salts, this is more likely to be a reflection of the fact that they are present in many of the wells rather than that they are specifically good at enhancing the crystallisation of AaeA. Nevertheless, the conditions listed in **Table 6.1** are a good starting point for further optimisation of crystallisation conditions and further screen could be designed to explore a range of concentrations of the various PEGs, additives, or salts that are highlighted. Additionally, the small crystals seen in the conditions above can be used for microseeding experiments, whereby the tiny fragments of crystals are added into new crystallisation drops to provide a nucleation site for future crystal growth. However, due to time limits it was not possible to perform a comprehensive optimisation process and proceed through to potential data collection on any crystals that could be grown and so this line of enquiry was put on hold.

6.5 Electron Microscopy of AaeB

Although there was not enough purified AaeB sufficient for crystallography, there was enough to perform analysis by EM. Negative stain EM provides a way to study the structure of AaeB using only μg -amounts of protein in a timescale of only a few days and can provide information on complex size and shape, and would potentially afford the possibility of assembling a low-resolution model. EM negative stain grids were prepared using samples of purified protein, and then imaged. 100 micrographs were acquired and from these 13,135 particles were picked for further analysis. The EM micrographs showed some patches of aggregated and clumped material, but a generally monodisperse and homogenous sample (seen in **Figure 6.11**), which indicates that the EM data from this sample should be reliable and is promising for further study.

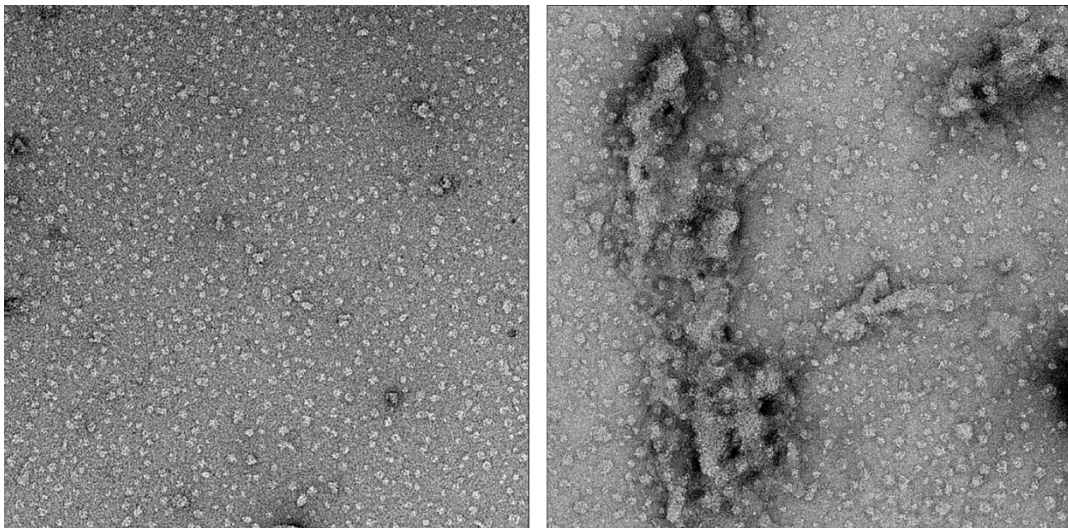


Figure 6.11 - A pair of representative micrographs at 30,000X magnification from AaeB negative stain EM grids.

6.5.1 EM classes

The 13,135 particles picked from the micrographs were taken and used to create class averages, in order to improve signal-to-noise. Several interesting classes are seen from the processed data generated from the first EM imaging experiment. It is possible to pick out classes with four regions of electron density (**Figure 6.12E, G, H, I, J, M**): putatively dimers of AaeB, with four sets of six helices forming a central pore, seen in **Figure 6.12**. In addition, there are classes that appear to be dimers of the four-lobed density (**Figure 6.12B, D, F, K**). This could be a true tetramer of AaeB, or may be an artefact, with the particles merely sat next to each other on the grid. However, the fact that the putative dimer-of-dimers is pulled out in the classes indicates it was a relatively frequent occurrence. It should also be noted that the possibility of these classes being a monomer was eliminated as a protein of 70 kDa would not be visible by negative stain EM.

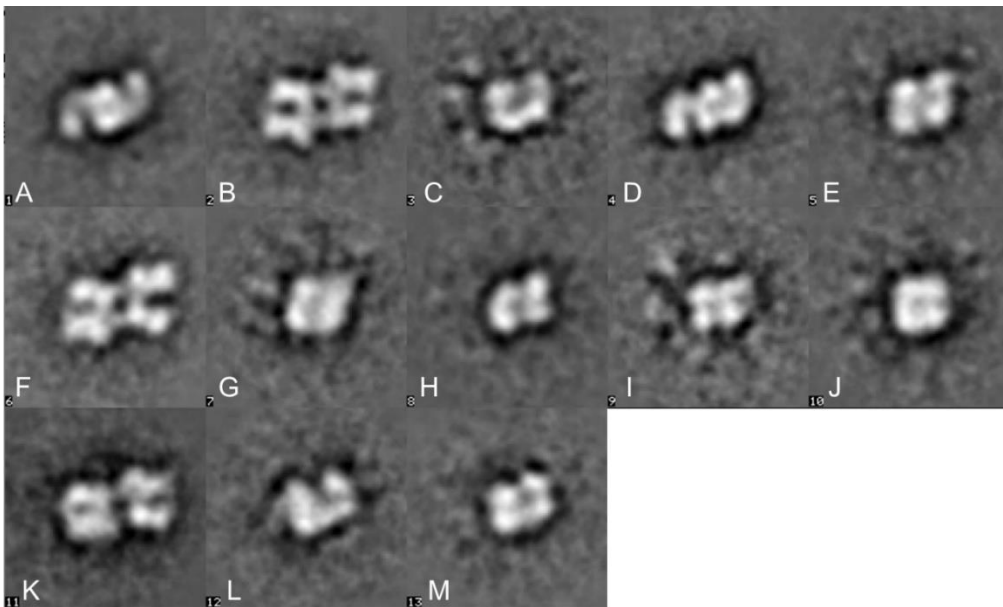


Figure 6.12 - EM class averages of AaeB imaged at 30,000X magnification.

Classes are seen containing four distinct regions of density in an approximate diamond formation.

A second set of micrographs were acquired at 50,000X magnification in an attempt to acquire higher-resolution data. Class averages from this higher-magnification data are shown in **Figure 6.13A**, and a number of the classes show a similar four-lobed structure as seen previously (selected in **Figure 6.13B**), although many of the classes were very circular and formless, which may be expected from a solution containing detergent micelles. The additional magnification seemed not to provide any significant greater understanding. Moreover, the background noise looks to contain strong features, indicative of a poorly aligned particle which is focussed on the background leading to a build-up of noise. The images looked to contain dominant features consistent with a dominant frequency in the image, likely caused by problems with the camera. As such this data is not reliable.

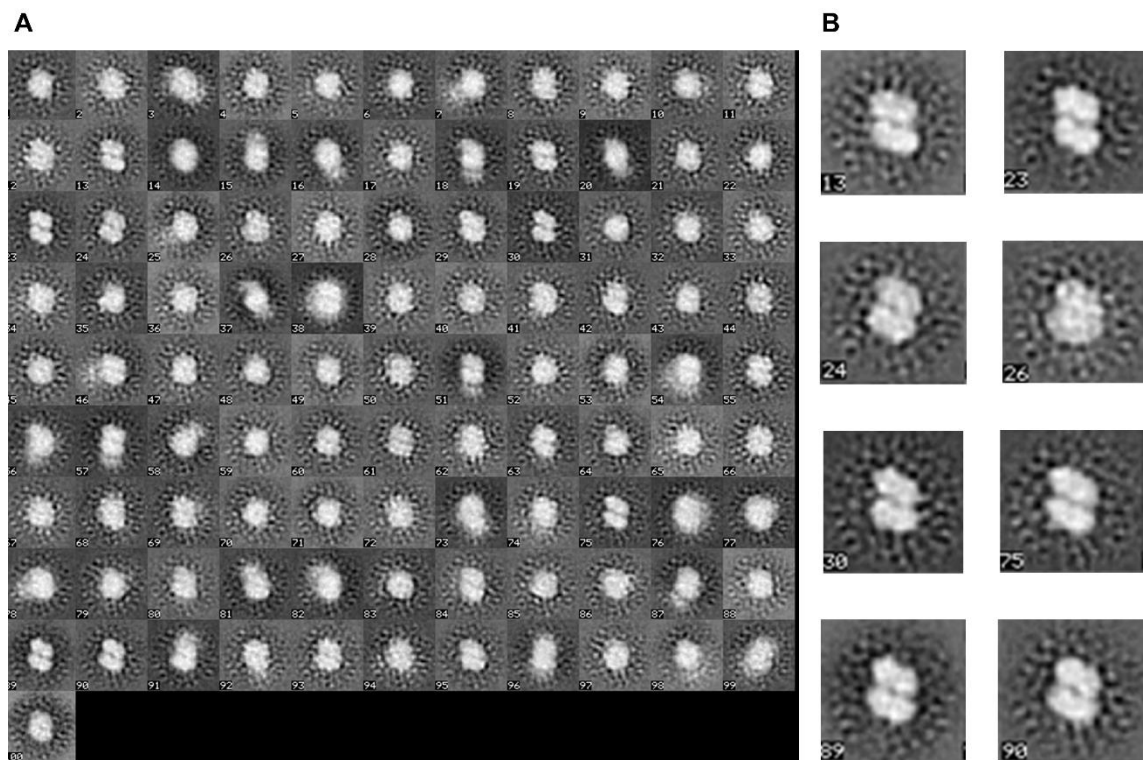


Figure 6.13 – A) EM class averages from AaeB taken at 50,000X magnification. B) Selected classes that show a more distinct morphology.

A second sample of purified AaeB was prepared as before from an independent purification and imaged in order to assess the reproducibility of the results seen. 114 micrographs were taken and from these 13,349 particles were picked. This time, as seen in **Figure 6.14**, significant heterogeneity was seen in the sample with a range of particle sizes and with a significantly noisier background.

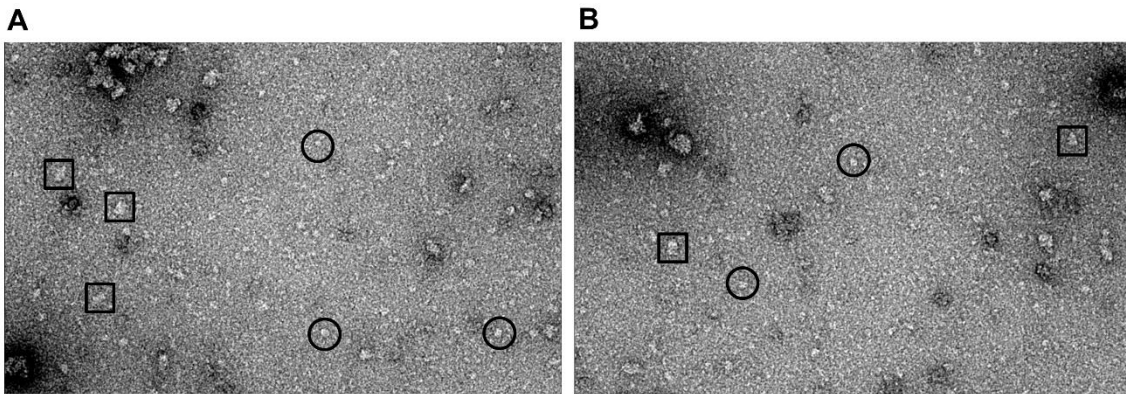


Figure 6.14 - Example micrographs of AaeB negative stain EM grids imaged at 30,000x magnification.

Larger particles indicated by boxes, smaller particles indicated in circles.

When classes were made from particles picked from this second set of data, shown in **Figure 6.15**, the background looked noisier, as expected. However, distinct classes could be picked from the data. Speculatively, these can be rationalised as a side view and a top view of a complex of AaeB, with the top view showing threefold symmetry and three pores, suggesting a trimer. The second sample of AaeB contained a higher proportion of AaeA, so perhaps AaeA promotes AaeB to form trimers rather than dimers. Alternatively, AaeB could remain as a dimer and AaeA may form a hexamer around AaeB, giving the hexagonal shape seen. While distinct classes have been generated by this

EM investigation there is not enough data to produce a full 3D model and the final model is likely be influenced by the starting model chosen.

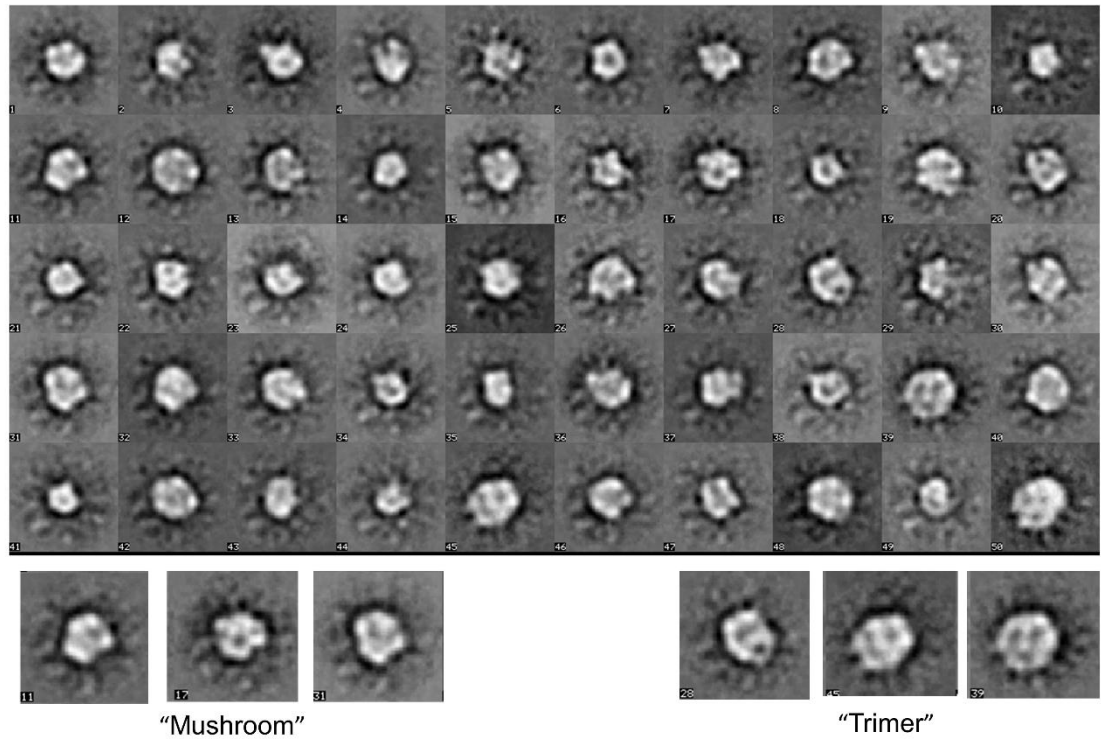


Figure 6.15 - Second set of EM class averages from an independently prepared sample.

Below are selected classes. Firstly, a set that have the appearance of a "mushroom" which is perhaps a side view of AaeB with a narrow membrane-embedded region and wider cytoplasmic domain. Secondly, a set that have the appearance of a trimer, with an approximately-hexagonal outline and three

6.5.2 Gold labelling

In order to assess the oligomeric state, AaeB was labelled with Ni-NTA-nanogold, which has affinity for the His-tag²⁶⁸ and has previously been used to locate the position of a subunit in the photosystem 2 complex²⁶⁹ in EM micrographs. As seen in **Figure 6.16**, much of the protein is found in large aggregated clumps, and the rest of the material is seen in mostly monomeric form, although there is some evidence for rare dimers in **Figure 6.16C-D**. A sample of Ni-NTA-nanogold with a smaller diameter of 1.8 nm was also used, and as seen in **Figure 6.16E-F**, a similar pattern of monomers was seen, although the nanogold appeared less homogenous. It is noteworthy that the micrographs shown in **Figure 6.16A-D**, in which the EM grids were washed to remove excess gold particles before staining, show the clearest background of all the samples. This suggests that a washing step is important to produce high quality data.

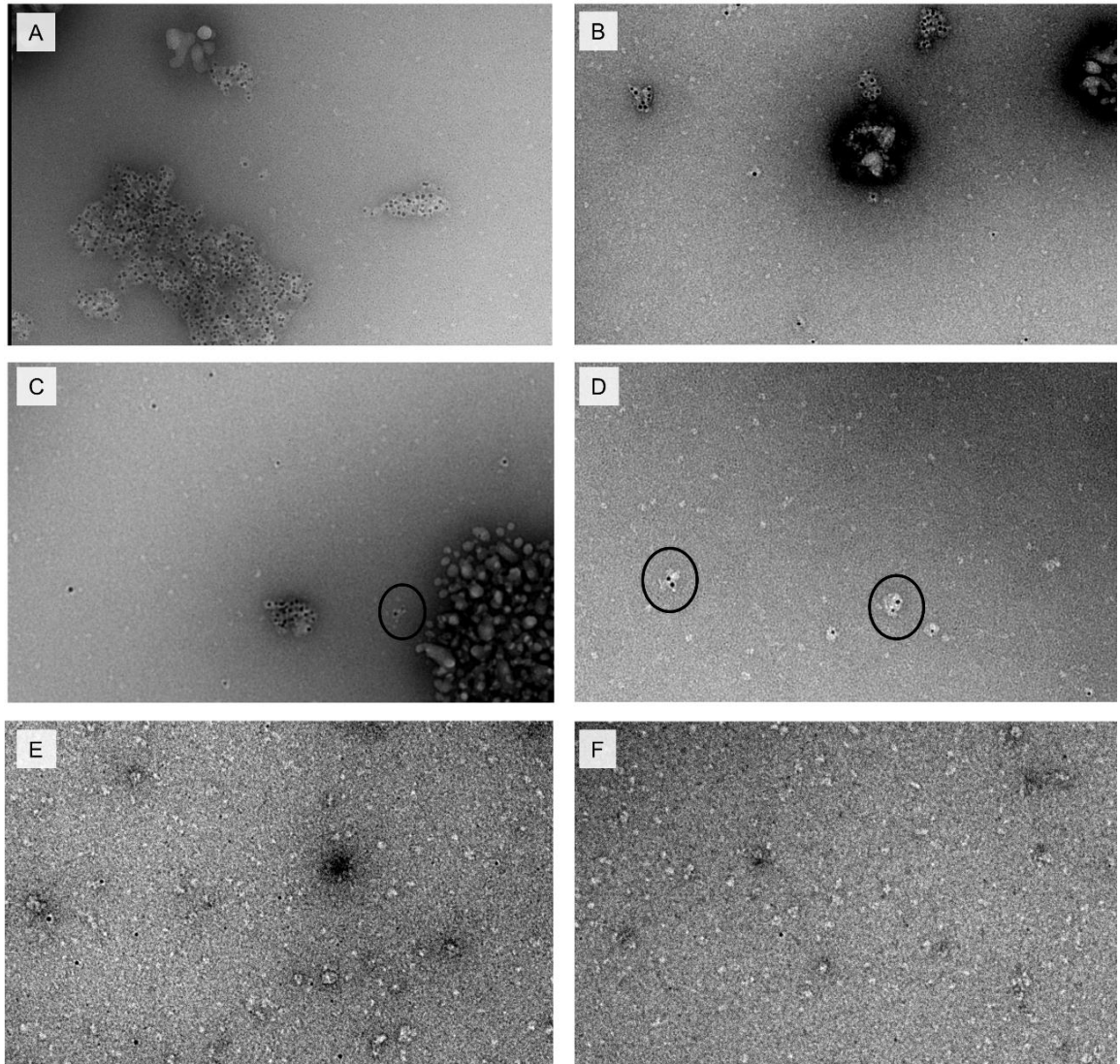


Figure 6.16 – EM negative stain images of samples of AaeB labelled with gold.

A-D) 5nm Ni-NTA-nanogold and **E-F)** 1.8nm Ni-NTA-nanogold. **A)** large aggregated regions are seen. **B)** smaller aggregates of AaeB are also seen. **C-D)** the majority of particles seen on the micrographs are associated with only one nanogold particle, rare particles are seen with two nanogold particles (circled), which are approximately double in size and so putative dimers.

Overall, the examination of AaeB by EM suggested two potential structures: firstly, four-lobed protein particles, rationalised as putative dimers of AaeB, with each monomer providing two regions of density, each corresponding to a set of 6 TM helices; and secondly, a trimer or dimer-with-hexamer structure. The Au

labelling showed that dimers were relatively rare, with most monodisperse AaeB having only one associated Au label. However, it must be noted that the position of the His tags and bulk of the Au label may result in only one His tag being labelled through steric clashes. A further problem maybe that the Au labelling could force apart multimers which has resulted in the large scale aggregation. The monodispersity of the sample on the grid could be improved to reduce the clumping seen in some micrographs, perhaps by adjusting the buffer used. In addition, the most-likely improvement that could be made is to reduce the background on the grids; this is achievable by concentrating the protein sample initially before a large-fold dilution in dH₂O directly before application onto the grid to dilute any interfering buffer components. Due to time constraints it was not possible to perform extensive optimisation, although the data presented here indicate that washing after sample application helps to reduce the background signal.

6.6 Studies on the oligomeric state of AaeA and AaeB

Purified AaeB, solubilised in DDM or DNPG was used to investigate the potential complex formed by AaeA and AaeB, multimerisation, and structure. When the purified protein was concentrated, an additional band was seen on the Coomassie stained SDS-PAGE gels. Seen as the highest band in **Figure 6.17**, a band at around the 130 kDa molecular weight marker was identified by mass spectrometry as being AaeB, therefore, this is evidence of a dimer of AaeB, which persists even in SDS.

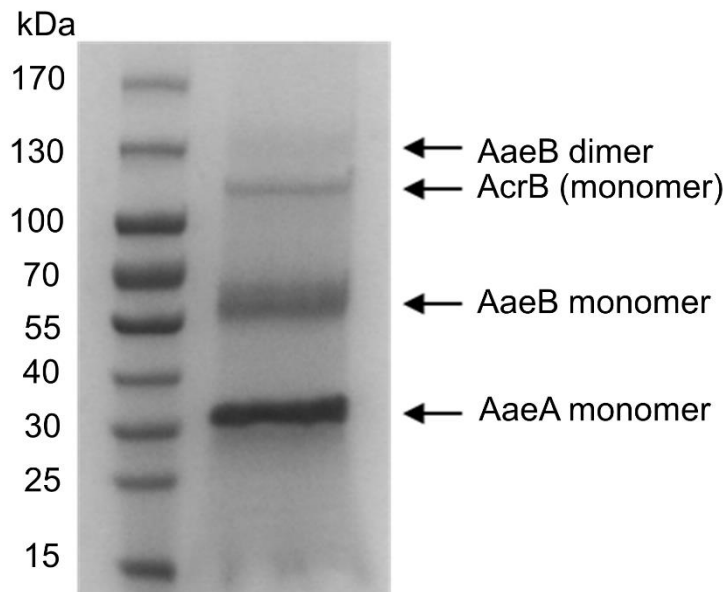


Figure 6.17 - Coomassie stained SDS-PAGE gel of 20 μ L of IMAC-purified AaeB.

Bands can be seen for AaeA, which co-purifies with AaeB; AcrB, a common contaminant during IMAC; and a dimer of AaeB. All bands were identified by mass spectrometry, as seen previously.

A similar tight dimer running on SDS-PAGE has been reported for the inner membrane protein MacB²⁷⁰, which has been shown to be a dimer in later studies²⁷¹. An example of where a similar co-purification is seen in *E. coli* is with

the AcrA-AcrB-AcrZ complex, which are shown to co-purify together²⁵⁹. No outer membrane proteins (OMP) were seen co-purifying with AaeB in this study, giving further evidence to the hypothesis that AaeAB does not form a complex with an outer membrane channel, although it is possible that if there was such an interaction the level of expression of the OMP may merely be too low to be detected. This co-purification with an OMP has been seen, for example, in an analogous tripartite system MacAB-OmpF with the MacAB pump (in inner membrane components), which co-purified with OmpF (the cognate outer membrane component)²⁷⁰.

In addition to this evidence of dimerization of AaeB, purifications of AaeA were also seen to contain bands corresponding to dimers of AaeA (as discussed in the previous chapter). This is initial evidence that AaeA and AaeB each form homomeric assemblies, in addition to the heteromeric interaction between the two proteins. However, this leaves many possibilities for the functional state of the AaeA-AaeB complex including dimers, trimers, dimers-of-dimers, trimers-of-dimers, and so on.

6.6.1 Crosslinking of AaeB

To further investigate the apparent multimeric state seen on gels, chemical crosslinking was investigated as a tool to shed light on the multimeric state of AaeB in solution. The homo-bifunctional cross-linker glutaraldehyde is widely used for crosslinking experiments because it rapidly reacts with amine groups at around neutral pH and generates stable covalent bonds²⁷². Purified protein was incubated with glutaraldehyde at a range of concentrations (5 mM-100 mM) and crosslinking was monitored by SDS-PAGE and Western blotting.

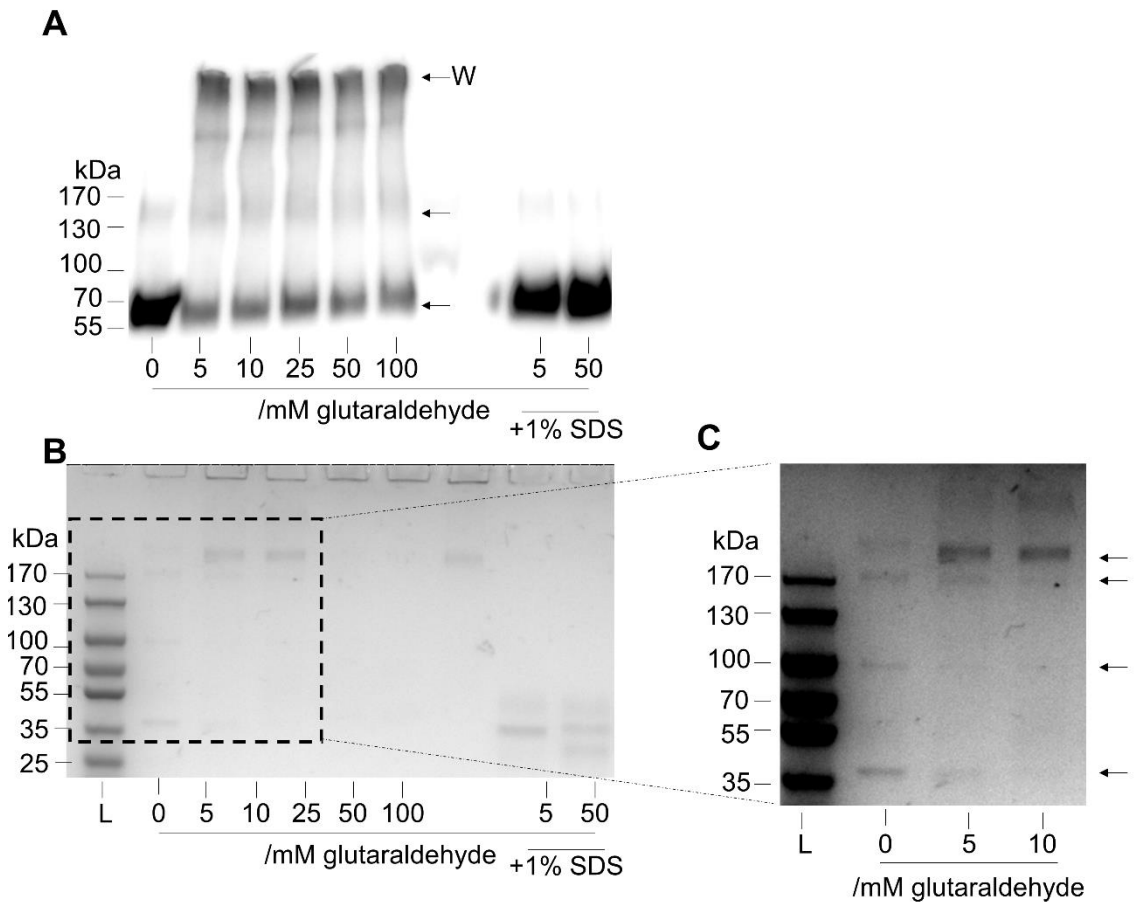


Figure 6.18 – Crosslinking of AaeB.

A) Western blot of 20 μ L samples of crosslinked samples of purified AaeB. Arrows indicate the approximate position of a monomer and dimer of AaeB, with potential evidence of a band significantly above the size of the MW ladder. The well is indicated by arrow with **W**. **B)** Coomassie-stained SDS-PAGE gel of the same, and **C)** Blown up and contrast-adjusted section of the coomassie gel. Four distinct bands can be seen initially, marked with arrows to the right, but upon addition of glutaraldehyde the largest band at >170 kDa becomes the major band, becoming the only band at 10 mM glutaraldehyde. This band is in the region expected for a trimer of AaeB. No crosslinking is seen in the samples denatured by SDS, confirming that crosslinking requires higher order structure to be intact.

The Western blot of crosslinked samples in **Figure 6.18A** shows adding glutaraldehyde led to a diminishment in the amount of protein in the band at the height of an AaeB monomer (70 kDa), and an increase in higher-MW species. A

faint band is seen in the non-crosslinked lane at the height expected of a dimer (140 kDa) and this increases in intensity when glutaraldehyde is added, along with a smear of higher-MW species that continue up to the height of the wells, including evidence of a band at a size of a putative trimer (224 kDa).

Importantly, when the protein sample was first denatured with 1% SDS, no crosslinking was seen, indicating that the crosslinking is dependent on correct protein secondary structure. The Coomassie stained gel in **Figure 6.18C** shows a band at >170 kDa, approximately in the region expected of a AaeB trimer (224 kDa). Some doubt is cast on the Coomassie results, however, as the bands seen in the gel appear to match poorly to the expected sizes of AaeB, running lower than expected, and lower than seen in other gels. The bands shown in the Western blot are closer to the expected sizes and identification is aided by the specificity of the anti-His-antibody.

The crosslinking was repeated on a freshly-purified sample of AaeB (which had a greater proportion of AaeA co-purified along with it) using a lower range of concentrations of glutaraldehyde, shown in **Figure 6.19**, and both the Coomassie stained SDS-PAGE gel and Western blot showed both monomers and dimers but no significant higher bands, and when concentrations of glutaraldehyde of 5 mM or higher were employed the protein aggregated and was found in the well.

These crosslinking data provide contrasting evidence for AaeB forming a multimeric complex, either in the form of a dimer, or a trimer, but do not appear to show any crosslinking between AaeA and AaeB.

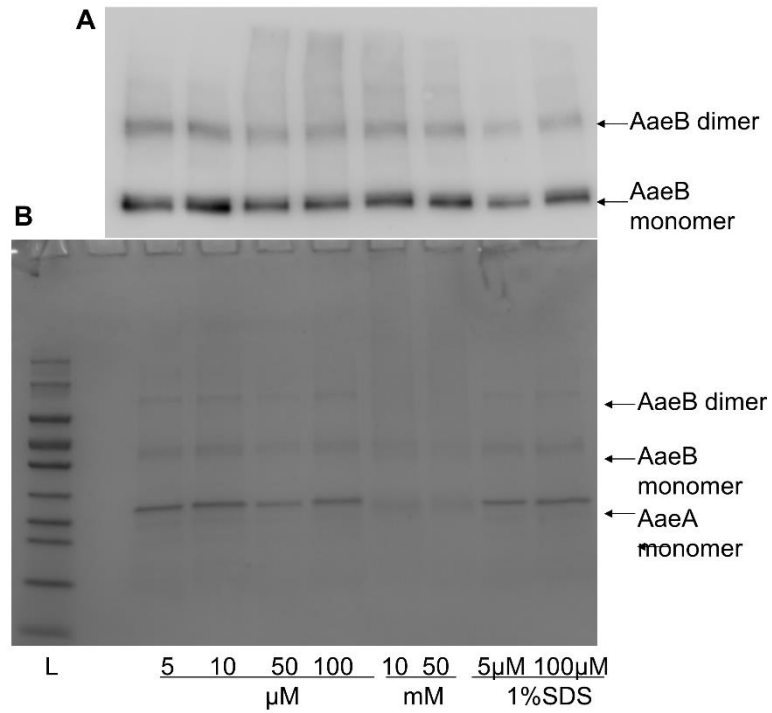


Figure 6.19 - Crosslinking of AaeB.

A) Western blot of 20 μ L sample of crosslinked samples of AaeB. A monomer and dimer size band is seen, but only limited evidence is seen of larger-MW species. **B)** Coomassie-stained SDS-PAGE gel of crosslinked samples of purified AaeB. Also seen is AaeA, which co-purifies with AaeB.

6.6.2 Native PAGE Gel of AaeB

A second approach to investigating the multimeric state of AaeB is a native PAGE gel. As native gels do not contain SDS they can assess complexes that form between proteins. Evidence is seen in **Figure 6.20** for both a monomer and dimer in the first lane, plus higher-MW species which tend towards aggregates as salt concentration increases. Without addition of NaCl, smeared bands are seen at approximately 70 kDa and from 150 kDa up to around 200 kDa (although this value is above the level of the molecular weight markers) and these sizes are approximately consistent with the presence of monomers, dimers, and trimers of AaeB. In 250 mM NaCl the protein runs almost entirely in a single band at the approximate size of 200 kDa, a putative trimer. At higher salt concentrations, the protein appears as high-MW aggregates. These data also provide potential evidence for the putative trimer, as they again contain an apparent band at >170 kDa.

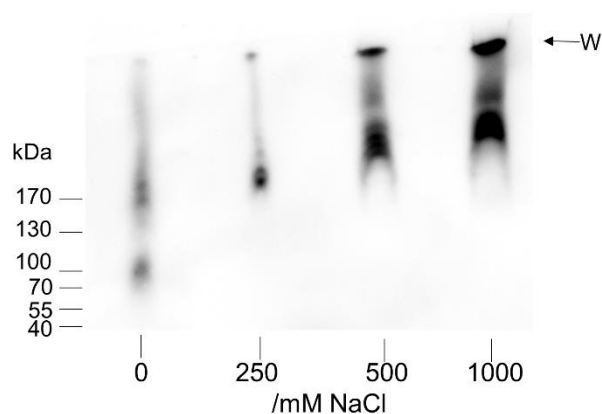


Figure 6.20 – Western blot of a Native PAGE gel.

20 μ L of purified AaeB was run in native sample buffer containing varying concentrations of NaCl from 0 – 1000 mM. Increasing salt promoted the formation of higher-MW species. Position of the well shown labelled **W** and MW markers shown on the left.

6.6.3 Crosslinking of AaeA

AaeA is likely to form functional multimers similarly to other MFPs and so this was tested by crosslinking. Samples of purified AaeA solubilised with DDM were treated with freshly-defrosted glutaraldehyde at concentrations up to 250 μM and, when assessed by Western blotting, bands are seen at molecular weights approximately equal to the expected sizes for monomer (34 kDa), dimer (69 kDa), trimer (103 kDa), tetramer (138 kDa), pentamer (172 kDa), and hexamer (206 kDa), as seen in **Figure 6.21**.

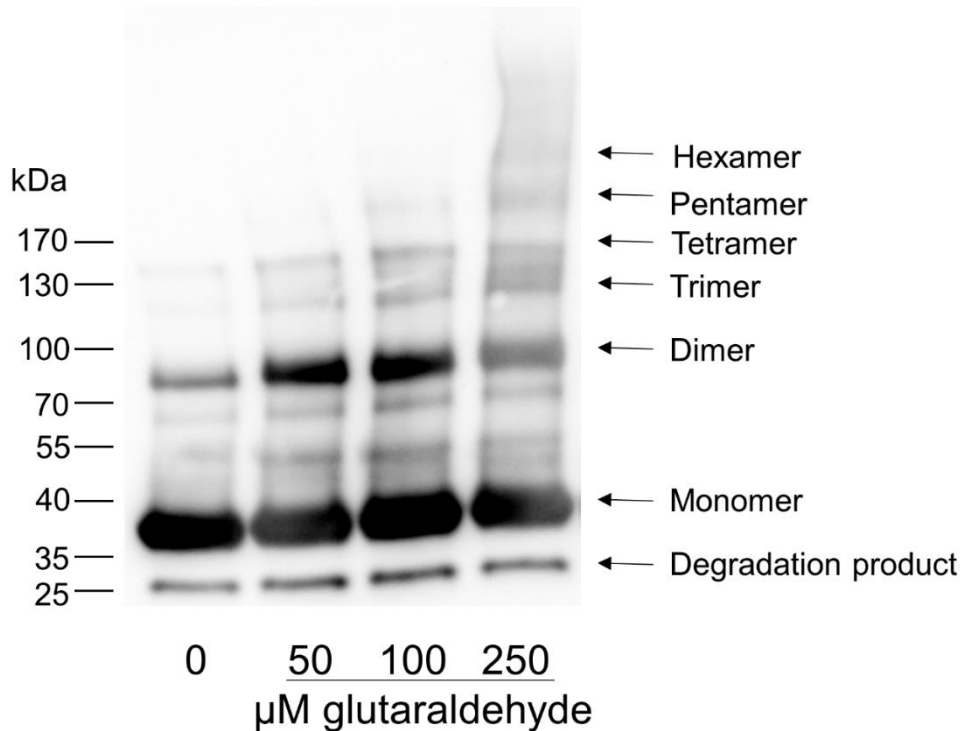


Figure 6.21 - Western blot of 20 μL samples of purified AaeA crosslinked with glutaraldehyde for 15 minutes.

As crosslinking is initiated an increase in the intensity of the dimer, trimer, and tetramer bands is seen, when the glutaraldehyde concentration is increased to 100 μM a band corresponding to the pentamer appears, and at 250 μM there is evidence for a hexamer along with smearing of larger-MW species all the way up to the size of the well.

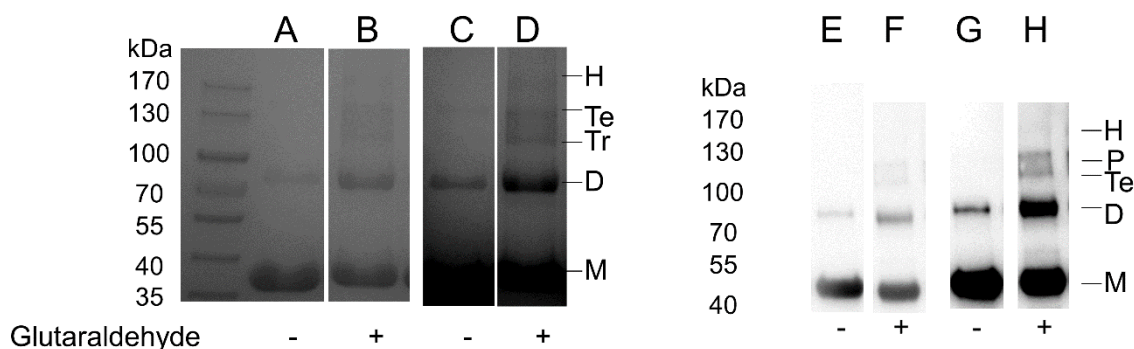


Figure 6.22 - Crosslinking of purified DDM-solubilised AaeA

A) B) Coomassie-stained SDS-PAGE gel of 20 μ L samples of AaeA alone, and after glutaraldehyde treatment; which have been contrast-adjusted to give: **C) and D).** **E) F)** Western blot of the same, which have been contrast-adjusted to give: **G) and H).** Bands are seen at sizes corresponding to monomer (**M**), dimer (**D**), trimer (**Tr**), tetramer (**Te**), pentamer (**P**), and hexamer (**H**)

Glutaraldehyde crosslinking was again employed on a second freshly-purified sample of AaeA at a concentration of 100 μ M. When assessed by Coomassie staining and Western blotting, evidence was seen in **Figure 6.22** of monomers becoming higher oligomers up to hexamer size. Additionally, each of these bands was identified by mass spectrometry to contain AaeA. This is congruent with evidence that other MFPs are hexameric in their functional state¹⁰⁷, and thus it is proposed that AaeA is a hexamer *in vivo* and the strength of the dimer band indicates AaeA is likely to dimerise before these dimers assemble into larger multimers.

6.7 Assessment of the Oligomeric State of AaeB by SEC-MALLS

Size Exclusion Chromatography-Multi Angle Laser Light Scattering (SEC-MALLS) provides a method for calculation of the molecular weight of solutes present in solution, by measurement of the UV absorbance, change in refractive index, and light scattering of a sample compared to a reference buffer. This enables measurement of particle size and molecular weight of macromolecules in solution and as such means that determination of complex stoichiometry is possible. Unlike conventional SEC, which is also used for molecular weight estimation by reference to known standards, SEC-MALLS is an absolute technique and measures molecular weight directly. Furthermore, SEC-MALLS is able to distinguish between the protein (which absorbs UV) and the detergent surrounding it (which does not), and thus can determine the size of the protein specifically. The ability of SEC-MALLS to assess the oligomeric state of any species makes it suited to studying AaeA and AaeB complex formation and stoichiometry; AaeA is co-purified with AaeB, so samples of this purified material were taken to assess oligomerisation and interaction between the two proteins.

For this specific case, this analysis is complicated by the fact that the size of AaeB (74 kDa) is approximately equal to the size of a dimer of AaeA (70 kDa), this difference is approximately 5%, the expected error in SEC-MALLS measurements as quoted by the manufacturer. To give an example, this means that it is possible to rationally assign three separate identities to a peak of approximately 145 kDa: with AaeA₄ (140 kDa), AaeA₂AaeB (144 kDa), or AaeB₂

(148 kDa) all falling at approximately that value. With this caveat, the data are presented below.

Before each SEC-MALLS experiment, a sample of the protein standard carbonic anhydrase was run to ensure the equipment was correctly calibrated before proceeding with the experiment. An example of one of these runs is shown in **Figure 6.23**; in **panel A**, the three traces for Light Scattering (LS, red), UV absorbance (UV, green), and dynamic refractive index (dRI, blue) overlay with the peaks aligned, which they should if the instrument is operating correctly. Additionally, **panel B** shows that when the region of the peak is analysed a consistent MW is calculated across the whole peak, and that the protein is running as a monodisperse species of a single MW. The calculated MW of the protein species is ~ 29 kDa, as expected. These initial tests ensured that the equipment was correctly calibrated and aligned.

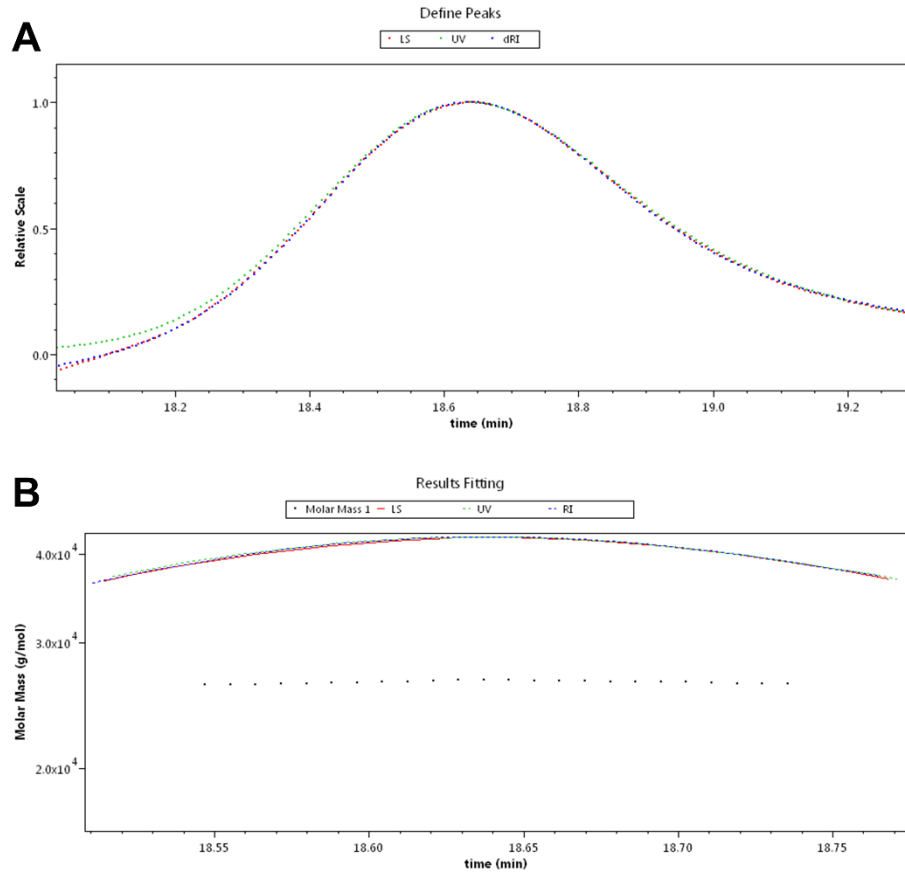


Figure 6.23 - SEC-MALLS analysis of carbonic anhydrase standard to assess calibration.

Light scattering is shown in red line, dRI shown in dotted blue line, and UV shown in dotted green. The traces for light scattering (LS), UV absorption (UV), and refractive index (dRI) overlay at the peak (**A**), and the calculated MW is consistent across the peak (**B**) indicating that the equipment is correctly calibrated and the protein standard is running as a single monodisperse species across the column.

6.7.1 SEC-MALLS analysis of AaeB

AaeB solubilised in DDM and purified by IMAC (co-purified with AaeA) was concentrated to approximately 1 mg/mL and 50 μ L was injected into the SEC-MALLS apparatus and run over a Superdex S200 10/300 column (GE Healthcare), producing the chromatogram seen in **Figure 6.24A**. A peak was seen at 7 minutes that had a shoulder at 8 minutes, and a third smaller peak present at just after 10 minutes. Although these peaks were not well resolved, they were analysed and the MW of the species present were calculated as 158, 117, and 2.48 kDa, respectively (3 s.f.). This can be seen in **Figure 6.24B**, with MWs superimposed as black lines onto the chromatograms. The first two peaks were in the expected range for oligomers containing AaeA and AaeB, but the final peak was clearly too low to be a protein species of interest, and thus was discarded. SEC-MALLS is also able to calculate the amount of detergent present in particles within each selected peak by finding the mass of protein (which absorbs UV light, unlike detergent) and the total mass of both protein and detergent (from light scattering). Peaks 1 and 2 also have approximate 50% detergent and 50% protein (Seen as Protein Fraction in **Table 6.2**), which is the expected ratio for membrane proteins solubilised in detergent, indicating that these peaks correspond to species of AaeA. The data corresponding to these peaks is summarised in **Table 6.2**. There are two potential assignments for Peak 1: AaeA₂AaeB (144 kDa), or AaeB₂ (148 kDa). Peak 2 may be assigned as the heterodimer AaeAAaeB (109 kDa) or a homotrimer AaeA₃ (105 kDa).

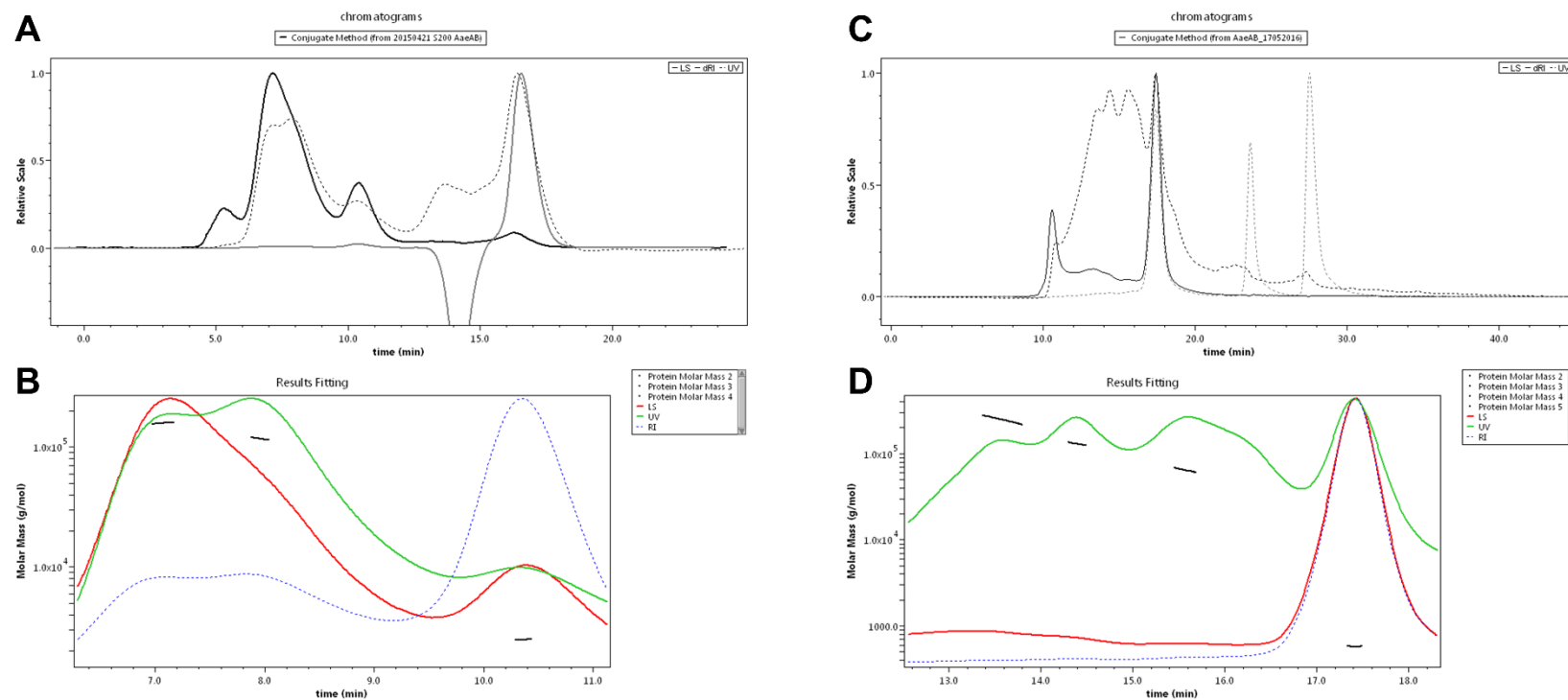


Figure 6.24 - SEC-MALLS analysis of purified AaeB.

Light scattering is shown in solid black or red line, dRI shown in solid grey or dotted blue line, and UV shown in dotted black or dotted green. **A)** chromatograms and **B)** peaks with calculated MWs for AaeB run on a Superdex S200 10/300 column. **C)** chromatograms and **D)** peaks with calculated MWs for AaeB run on a WTC-MP030S5 column. Calculated MWs (indicated by the horizontal black lines overlaying the peaks in **B** and **D**) are listed in **Table 6.2** and **Table 6.3**. In both cases the peaks are relatively poorly resolved from one another.

Table 6.2 - Summary of SEC-MALLS results for AaeB on S200 column

	Peak 1	Peak 2	Peak 3
MW /kDa	158 ($\pm 0.436\%$)	117 ($\pm 0.457\%$)	2.48 ($\pm 0.364\%$)
Micelle MW /kDa	173 ($\pm 0.928\%$)	118 ($\pm 1.008\%$)	45 ($\pm 0.514\%$)
Protein fraction %	47.7 ($\pm 0.079\%$)	49.9 ($\pm 0.077\%$)	5.2 ($\pm 0.116\%$)

As the peaks were poorly resolved, a second run was performed on a newly-purified sample of AaeB, using the WTC-MP030S5 column (Wyatt Technology) which is designed for improved resolution when separating membrane proteins in detergent. The chromatogram produced can be seen in **Figure 6.24C**, and several peaks could again be identified. Four peaks were selected for analysis, indicated in **Figure 6.24D**. These peaks were calculated to contain macromolecules with MW of 245, 129, 64 and 0.6 kDa, respectively. Again, the final peak was too low to be relevant. However, the peak of 64 kDa is approximately the size of a monomer of AaeB, and thus the three other peaks were deemed likely to be appropriate species. The data corresponding to these peaks is summarised in Table 6.3. Peak 1 could be assigned as AaeB₃ (222 kDa), or AaeA₂AaeB₂ (248 kDa), or AaeA₃AaeB₂ (253 kDa). Peak 2 may represent AaeAAaeB (109 kDa) or AaeA₂AaeB (144 kDa). Peak 3 is closest to AaeB (74 kDa) or AaeA₂ (70 kDa), although as AaeA is present only as a consequence of having co-purified along with AaeB and so complexes that contain only AaeA are unlikely to be valid species.

Table 6.3 - Summary of SEC-MALLS results for AaeB on WTC-MP030S5 column

	Peak 1	Peak 2	Peak 3	Peak 4
MW /kDa	245 (±6.429%)	129 (±6.324%)	64.0 (±5.455%)	0.576 (±5.002%)
Micelle MW /kDa	157 (±17.730%)	131 (±14.068%)	106 (±10.321%)	57.6 (±7.109%)
Protein fraction %	61.0 (±0.126%)	50.0 (±0.097%)	38.0 (±0.081%)	1.0 (±0.048%)

The protein fraction in each of these peaks in both experiments was approximately half, which is expected for a detergent-solubilised membrane protein¹⁸¹ and supports the conclusion that they are relevant species, surrounded by DDM. There is evidence to support the existence of AaeB monomers (WTC-MP030S5 Peak 3, 64 kDa) and AaeB dimers (S200 Peak 1, 158 kDa), and AaeB trimers (WTC-MP030S5 Peak 1, 245 kDa), which is in agreement with bands seen on SDS-PAGE gels such as **Figure 6.17** and crosslinking. A range of peaks may also represent various AaeA-AaeB heteromers.

However, although the MWs calculated for the peaks fall within the expected range, it is important not to over-analyse the results. The peaks are, in general, poorly resolved and generally the line denoting MW is seen to increase or decrease across the peak rather than being flat, as would be expected for a single monodisperse species. What can be drawn from these data is that the species AaeA and AaeB are part of a dynamic, labile complex that does not have one clear, dominant species over the conditions of the SEC-MALLS run. It may be possible to extract the entire complex using SMALP-based purification, rather than detergent, and run this SEC-MALLS again, which would probe the native state of the complex without the disruption caused by detergent

purification. Also, perhaps crosslinking the complex may stabilise it in one conformation and allow analysis by SEC-MALLS by producing a single peak, preventing the complexes falling apart.

A third examination of the complex AaeA-AaeB was attempted with the WTC-MP030S5 column. Purified AaeA and AaeB were mixed, ensuring that there was an excess of AaeA, before running the SEC-MALLS. The chromatogram in **Figure 6.25**, and analysis summarised in Table 6.4 show that, again there seems to be a dynamic complex, with a MW in the range 112-126 kDa, which is broadly what is expected for complexes of AaeB dimers, heteromeric complexes of AaeAAaeB, or AaeA trimers or tetramers. Again, the protein molar mass lines on the “Results Fitting”, seen in **Figure 6.25B**, show a curve rather than a flat line, suggesting species of higher and lower MW are also present in the peak, and there is a labile, dynamic complex.

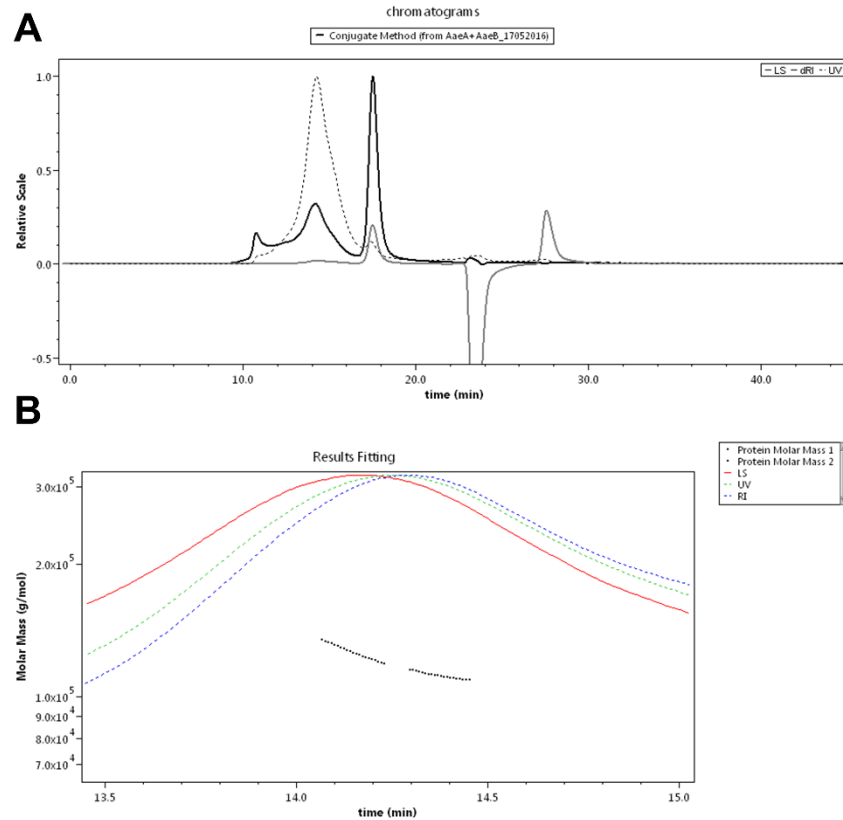


Figure 6.25 - SEC-MALLS analysis of AaeA mixed with AaeB.

Light scattering is shown in solid black or red line, dRI shown in solid grey or dotted blue line, and UV shown in dotted black or dotted green. **A)** chromatogram shows a single major peak in the UV trace with a shoulder. **B)** analysis of the peak with calculated MW overlaid.

Table 6.4 - Summary of SEC-MALLS results for AaeA plus AaeB on WTC-MP030S5 column

	Peak 1	Peak 2
MW /kDa	126 ($\pm 5.686\%$)	112 ($\pm 5.711\%$)
Micelle MW /kDa	72.8 ($\pm 16.529\%$)	72.2 ($\pm 15.618\%$)
Protein fraction %	63.3 ($\pm 0.042\%$)	60.7 ($\pm 0.041\%$)

6.7.2 SEC-MALLS analysis of AaeA

Just as was done for AaeB in **Section 6.7**, purified AaeA was assessed by SEC-MALLS. Both the S200 column (**Figure 6.26A and B**) and WTC-MP030S5 column (**Figure 6.26C and D**) were used, and the chromatograms and MW estimation is shown in **Figure 6.26**. For the S200 column, it was possible to

calculate the MW of only one peak from the chromatogram, and that peak was calculated to be 2.7 MDa, significantly larger than any oligomeric species is expected to be, while the detergent micelle was calculated to have a negative MW: clearly not physically possible.

Table 6.5 - Summary of SEC-MALLS results for AaeA on S200 column

	Peak 1
MW /kDa	2,718 (±10.801%)
Micelle MW /kDa	-107 (±262.313%)
Protein fraction %	104 (±0.019%)

Due to this, the experiment was repeated with freshly-purified AaeA using the WTC-MP030S5 column. The chromatogram (**Figure 6.26C**) showed one very broad peak, with a shoulder visible on the left hand side of the UV trace. Taking several measurements of the MW over the 15-21 minute range it was clear that the shoulder contained much higher-MW material, with a calculated MW of 427 kDa, while the main peak centred around 19.5 minutes was comprised of material in the MW range 143-173 kDa. These data are summarised in Table 6.6. Similarly to the experiment carried out on AaeB, the main peak likely represents a complex that is labile and dynamic. The MW lines overlaid in black on **Figure 6.26D** slope downwards, indicating an unstable species. The predicted maximum oligomeric state of AaeA is a hexamer, which would have a MW of approximately 209 kDa, and the values measured over the peak are a small amount below this value. If the analysis of the broad second peak were extended to the left, following the upward slope of the line it may plausibly reach 209 kDa. Interestingly, the shoulder on the peak is approximately the MW of

AaeA₁₂ and may represent a pair of hexamers associated together in a DDM micelle (or, potentially a dodecamer), although there is no evidence that this is a physiologically-relevant state.

Table 6.6 - Summary of SEC-MALLS results for AaeA on WTC-MP030S5 column

	Peak 1	Peak 2	Peak 3	Peak 4	Peak 5
MW /kDa	427 (±10.014%)	173 (±9.082%)	166 (±9.063%)	157 (±9.097%)	143 (±9.025%)
Micelle MW /kDa	340 (±24.709%)	99.0 (±26.549%)	85.8 (±28.149%)	77.2 (±29.084%)	73.6 (±28.017%)
Protein fraction %	55.7 (±0.013%)	63.6 (±0.005%)	66.0 (±0.005%)	67.1 (±0.005%)	66.0 (±0.006%)

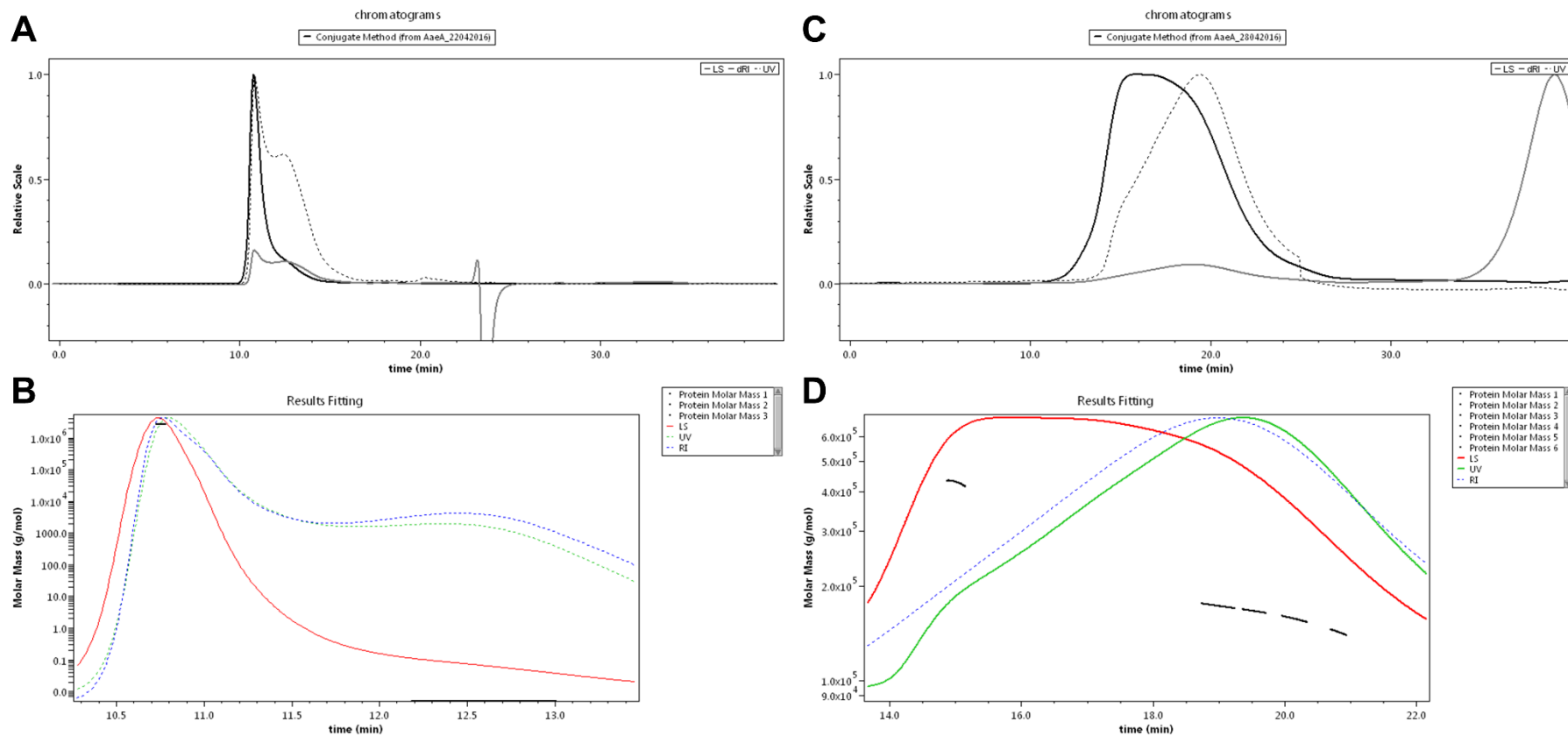


Figure 6.26 - SEC-MALLS analysis of purified AaeA.

Light scattering is shown in solid black or red line, dRI shown in solid grey or dotted blue line, and UV shown in dotted black or dotted green. A) chromatograms and B) one assigned peak with calculated MWs for AaeA run on a Superdex S200 10/300 column. C) chromatograms and D) five assigned peaks with calculated MWs for AaeA run on a WTC-MP030S5 column. Calculated MWs are listed in Table 6.5 and Table 6.6

6.8 Summary and conclusions

Samples of AaeA and AaeB, produced and purified as seen in the previous chapter were used for a variety of biochemical and structural studies. AaeA was previously shown to be vital for stability of AaeB in solution, which is in agreement with the previously shown requirement for both proteins to confer resistance *in vivo*, and purifications of AaeB showed untagged AaeA co-purifying, confirming the tight association between these proteins. These data suggest that AaeA has a stabilising, chaperone type role with AaeB.

A homology model of AaeA was constructed *in silico* based on closely related MFP family proteins, and this model contained linearly arranged α -hairpin, lipoyl, and β -barrel domains attached an inner-membrane-embedded helix, as expected for MFP members. No MPD domain was predicted in the structure, which in other MFPs interacts with the periplasmic domains of the cognate inner membrane transporters. This is to be expected as AaeA interacts with AaeB, which does not have a large periplasmic domain, with the largest globular loops being located in the cytoplasm and only minimal presence in the periplasm. This model was validated by CD, which showed approximately one third each of α -helical and β -sheet content, and thermal unfolding was undertaken to give a melting temperature of just below 50°C.

AaeAB may not form an association with TolC, unlike analogous inner-membrane transporters, and perhaps this absence of interaction is based on a single deletion between the bases of the RLS motif in AaeA that is known to be vital for interaction with TolC. Thus, the role of their associated MFPs, such as AaeA, is open to speculation with their role perhaps being key to promoting

stability (supported by the requirement of AaeA to stabilise AaeB) and regulating transport activity.

A survey of buffer conditions was undertaken to optimise the stability of AaeA. The protein was shown to be tolerant of a wide range of pH values, while increasing the salt concentration provided slightly increased stability, and increased glycerol concentration had a strong stabilising effect. This wide pH tolerance may reflect the location of AaeA in the periplasm of *E. coli*, as this is likely to experience wide variations in pH as gut bacteria must cope with stomach acid as well as a more alkaline intestine. The unfolding temperatures measured matched well with the unfolding temperature measured in the CD experiments.

AaeB was studied and evidence was shown for multimerisation by SEC-MALLS and crosslinking, with both suggesting that there is a dynamic complex present, with AaeB likely forming a homodimer or -trimer and also evidence was seen for heteromeric complexes containing AaeA and AaeB. AaeA alone similarly showed evidence of dynamic multimerisation with strong dimerization and the likely functional complex hexameric. The mechanism of this multimerisation may be explained by the finding that in the model of AaeA opposing faces of the molecule are positively and negatively charged, respectively, suggesting a side-by-side assembly based on electrostatic interactions.

EM was undertaken to investigate the structure of AaeB complexes and after analysis of over 26,000 particles evidence was found for particles that perhaps represent dimers of AaeB, in addition to particles that may be trimers of AaeB, or alternatively a dimer of AaeB complexed with a hexamer of AaeA, although

no 3D model could be constructed from the data collected. Optimisation of grids, especially washing, can be undertaken to provide clearer negative-stain images, before moving on to cryo-EM in order to generate data at a higher resolution, perhaps enough to see a tertiary structure. This is a promising approach to study the AaeA-AaeB complex as a whole in addition to homomeric complexes.

For AaeA, enough material could be purified to allow crystallography to be done. No clear crystals were formed, but some potential hits containing microcrystals that can be followed up by further optimisation to produce crystals that can lead to solution of the structure. This optimisation is out of the scope of the present study, however, as it is a lengthy process. If production of AaeB is optimised, there is hope for crystals of that too, and the complex as a whole.

In summary, the first *in vitro* experiments have been performed on AaeA and AaeB, a pair of proteins that form a complex within the inner membrane of *E. coli*. An *in silico* homology model was generated for AaeA, which proved highly plausible when considering the literature concerning related MFPs. This model was validated by CD and provides a starting point to develop further experiments to examine AaeA. In addition, a variety of buffer conditions were assayed in order to derive optimal conditions to stabilise AaeA within solution, a key step in any structural experiment. A strategy for structural examination of AaeA by crystallography and AaeB by EM has been developed in order to provide the first structural data for these proteins, or any protein of the PET family of which they are a part. Consistent within the EM data and additional data from SEC-MALLS and crosslinking on AaeA and AaeB is that they form a dynamic, labile complex, likely formed of either dimers or trimers or AaeB and

tetramers or hexamers of AaeA. This represents the first biochemical and structural data gathered for either of these two proteins, and the first step in understanding their structure and mechanism. This understanding may also be helpful in understanding the related ALMT family in plants.

Chapter 7

General Discussion and Conclusions

7.1 Strategies for overexpression of membrane proteins for structural studies

Several approaches were taken in this study in order to attempt to characterise ArAE family proteins including ALMTs, a family of membrane proteins which are found ubiquitously in plants, and their bacterial members, part of the relatively-underexplored PET family. Firstly, the C-terminal domain of wheat ALMT1 – putatively a soluble domain – was expressed in isolation; secondly, full length ALMTs (wheat ALMT1, and *Arabidopsis* ALMT5 & ALMT9) were expressed transiently in the leaves of the model plant *Nicotiana benthamiana*; and thirdly, bacterial proteins were identified from *E. coli* and *S. oneidensis* and overexpressed in *E. coli*. Usually, tens or even hundreds of milligrams of highly purified protein are needed in order to undertake detailed structural and functional studies, and, unfortunately, the overexpression of membrane proteins remains a challenge. Attempts at production of high levels of protein are often limited by poor expression levels, misfolded protein, or the formation of inclusion bodies, all of which were seen within this project.

7.1.1 The CTD of ALMT1

In order to study the isolated C-terminal domain of ALMTs, four expression constructs were made to express the C-terminal domain of wheat ALMT1 in *E. coli*. Two of these encoded the sequence starting from after the final transmembrane helix until the end of the protein, excluding the very C-terminal residues, which were predicted to be disordered. These were tagged with either a C-terminal or N-terminal His-tag. Two additional truncations were made at the

N-terminal end of this construct to remove a further region predicted to be disordered and ensure that none of the transmembrane helix was included; again these constructs had either N- or C- terminal His-tags. After extensive expression trials (encompassing bacterial strain, media, temperature, time, and type of induction) it was concluded that it was not possible to express this CTD in a soluble form. Attempts at refolding the protein from inclusion bodies were similarly not able to produce the soluble, monodisperse protein required, with a heavy precipitate forming upon re-folding. An extensive bioinformatics study on ALMTs published during this study showed that the CTD of ALMTs contain two well-conserved hydrophobic regions, that the authors predicted to be additional transmembrane helices⁴⁵. Thus, it is concluded that the CTD of ALMT1 is not a soluble domain and at least one of the hypothesised additional transmembrane helices are likely to be present within the CTD. As the CTD is not likely to be the independent, globular domain that was initially expected, this line of investigation was discontinued.

7.1.2 Full length ALMTs

The second line of investigation focussed on the full-length plant ALMT proteins. Vectors were produced for transient expression of full length ALMTs in *Nicotiana benthamiana* for three family members, TaALMT1, AtALMT5 and AtALMT9, which were produced in both C-terminally and N-terminally His-tagged forms using the pEAQ-HT vector¹³². This is the first use of transient expression of membrane proteins *in planta* for the purpose of structural studies. A proof-of-concept study showed that production of GFP in high yield was possible using the system, confirming previously-published work²³⁴ and so the

target proteins were then expressed. Production of full length ALMTs was possible, and expression of all six constructs was seen, but the protein was rapidly degraded. This perhaps may reflect the situation *in vivo*, with the proteins being rapidly turned over. In the case of ALMTs involved in stomatal opening specifically this is an appealing hypothesis, as stomatal movements need to match the diurnal cycle, evidence of which is seen in the expression patterns from microarray data (<http://www.bar.utoronto.ca/>)²⁷³. After trialling a variety of infiltration conditions and time-points for expression it was concluded that it would not be possible to extract protein from the leaves after degradation was monitored via SDS-PAGE during extraction.

7.1.3 Bacterial ArAE Proteins

The third approach concerned investigation of bacterial proteins, including AaeB from *E. coli*, which was originally identified as part of a new protein family designated PETs in 2000³⁶. Vectors were produced for expression of three bacterial proteins: firstly, two inner membrane transporters AaeB from *E. coli* and Q8EKE1 from *S. oneidensis*, which both have sequence identity to ALMTs; and secondly two constructs concerning the MFP AaeA from *E. coli*, both full-length and an N-terminal truncation of AaeA to remove the hydrophobic helix with the aim of producing a soluble protein. All four of these were expressed in *E. coli* successfully, and AaeB was shown to require co-expression with AaeA in order to be solubilised, supporting the previous physiological study that showed both components were required to form a functional complex³⁸. AaeA Δ TM was only expressed as an insoluble aggregate, despite attempts at optimisation of culture conditions (including strain, media, and temperature). The *S. oneidensis*

protein was produced and purified successfully, however, the amount of protein was low and, as such, further investigations were not continued with this. But, if expression levels could be optimised and increased then this protein shows promise for use in further experiments as a strategy for its solubilisation and purification has already been developed. Both AaeA and AaeB were expressed in sufficient amounts, solubilised by detergents, and purified successfully and so were used for a range of biochemical and structural experiments.

The finding that the solubility of AaeB requires co-expression with AaeA supports previous studies that showed a requirement for both proteins for physiological function and the pull-down of AaeA during the purification of AaeB is the first direct evidence of a direct interaction between AaeA and AaeB *in vitro*. This provides strong evidence that AaeA and AaeB form a functional complex in the inner membrane *in vivo*.

AaeA was subjected to crystallography, and some evidence of microcrystals was seen, although there was not sufficient time to follow up this discovery and fully optimise the crystallisation conditions. Nevertheless, the development of a strategy to express, solubilise, and purify AaeA to the point of being able to obtain microcrystals is a significant step on the path to obtaining high resolution structural data.

In the absence of any experimental structure, homology modelling has become an important tool in structural biology to predict secondary and tertiary structure of membrane proteins. The models that result can be used to guide experimental approaches, such as mutagenesis, that are aimed at shedding light on mechanisms of activity, for example examination of transport in the

case of channels and transporters. The first homology model for the protein AaeA was generated in this study, based on several MFPs that form complexes with inner membrane transporters from the RND, ABC, and MFS families. This model contained three of the canonical domains identified in other MFPs, alongside the transmembrane helix that anchors the protein into the inner membrane. This membrane localisation was supported by experimental data. Overexpressed AaeA segregated with the membrane fraction when produced in *E. coli*, and the protein could not be washed from the membrane into the soluble fraction with either addition of NaCl, or NaCO₃ pH 11.5, indicating that AaeA was a true integral membrane protein, not merely membrane-associated. The bioinformatics analysis of AaeA also drove the production of a construct to express a truncated version of AaeA (designated AaeA Δ TM) that aimed to exclude the N-terminal transmembrane helix. Unfortunately, (in contrast to other studies on related MFPs that also took this approach²⁵⁴) expression of this truncated protein produced only aggregated protein in inclusion bodies that could not be re-folded into a soluble form. It is possible that the exact location of the truncation was not optimal, perhaps leaving a disordered or untethered N-terminus. If so, then expression of a range of constructs with serial deletions from the N-terminus may identify a specific construct that is both well-expressed and correctly-folded. The high expression levels seen for both AaeA and AaeA Δ TM and previous success seen by other groups in expressing N-terminally-truncated MFPs^{253,254} leads to the conclusion that this approach is a promising route, although time constraints dictated that this was not possible to investigate this in the course of the present study.

The homology model and sequence analysis also presented several interesting avenues of investigation and provide new insights into the AaeA-AaeB complex and wider family members. And the validation of the model using CD spectroscopy provides confidence in the model and allows such predictions to be made using it. The finding that AaeA has opposing faces that are positive and negative, respectively, suggests a potential mode of side-by-side multimerisation and is the first description of such a mechanism for AaeA. Additionally, the conserved “RLS motif”, found within MFPs to be vital for interaction with TolC in the outer membrane¹⁰⁷, was identified in the model of AaeA at the tip of the α -hairpin with a single residue deletion, which appears to be conserved in the MFPs known to be expressed in operons with PET/AraE family inner membrane transporters. This deletion is a possible explanation for the previous observation that TolC is not required by the AaeAB transporter³⁸. As such, the homology model and sequence alignment have shown how the AaeAB exporter may have become adapted to extrusion into the periplasm rather than past the outer membrane.

Data was gathered by crosslinking and SEC-MALLS in order to investigate the complex formed by AaeA and AaeB, along with potential multimerisation of the individual subunits. The data gathered, *in toto*, show that AaeA and AaeB forms a dynamic complex, with AaeA being necessary to stabilise AaeB. Both AaeA and AaeB appear to form homodimers readily, and the most likely configuration of the final complex is a 6:3 configuration similar to AcrA-AcrB, with evidence seen for dimers and trimers of AaeB, and multimers up to hexamers of AaeA.

Overall, this study has developed a strategy for expression and purification of two bacterial genes from *S. oneidensis* and *E. coli* that have homology to

ALMTs, as well as an accessory protein that forms a complex with the *E. coli* gene. A variety of biochemical and structural approaches have been applied to investigate the *E. coli* genes, AaeA and AaeB, alongside the production of the first homology model of AaeA. This is the first *in vitro* study of any PET family member or its associated MFP as well as the first EM study, showing a possible arrangement of the complex. A schematic view of the proposed arrangement of AaeAB is shown in **Figure 7.1**. In the absence of an interaction with an outer membrane channel, the MFP AaeA may have a role as a “funnel” or “chimney” to help channel and deposit substrates further away from the inner membrane.

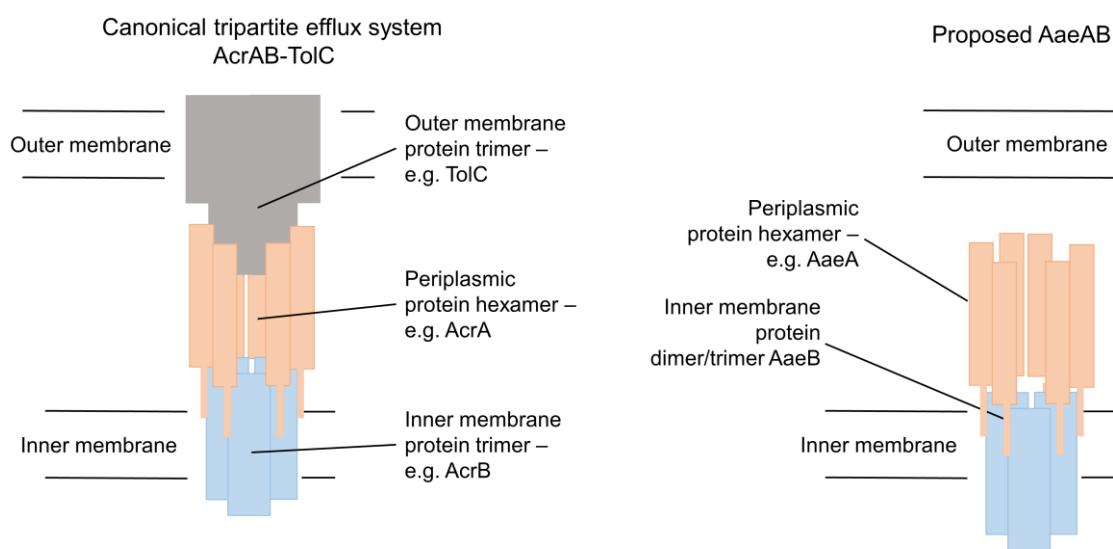


Figure 7.1 - Overall schematic proposed for the complex AaeAB as compared to AcrAB-TolC

In the canonical description of a tripartite efflux system the complex contains a trimeric inner membrane transporter connected to a trimeric outer membrane channel via a hexameric periplasmic membrane fusion protein. It is proposed that AaeAB similarly forms a complex with either dimeric or trimeric AaeB as the inner membrane component with hexameric AaeA. No connection to an outer membrane protein has thus far been demonstrated for AaeAB as so AaeA may represent a “chimney” to deposit substrate away from the inner membrane.

7.2 Future work

7.2.1 Continuation of work on AaeA-AaeB

The next aim for studies of AaeA and AaeB would be to move forward with further structural studies. Microcrystals of AaeA were seen in several crystallisation conditions, and further optimisation of the well solutions should provide larger crystals suitable for X-ray diffraction. Once obtained, a structure would be able to provide the ultimate validation of the model presented herein. However, it should be noted that the refinement of microcrystals to form well-diffracting crystals can take several months, if not years. Beyond this structure, a structure of AaeA in complex with AaeB would provide further insights into the nature of their interaction. To achieve this, optimisation is required to improve expression levels of AaeB, and material can be produced with a scale up to fermenter scale. This will allow sufficient amounts of protein to be made in order to undertake crystallographic studies, either alone, or as part of a complex with AaeA. The commercially available ProPlex screen is particularly designed for assessment of protein complexes and may be of specific interest for such a study. Additionally, once greater expression is obtained it would be worthwhile to perform screening of buffer conditions to find the optimal conditions for complex stability, as done with AaeA. Additionally, production of a soluble fragment of AaeA would also be beneficial for crystallographic studies. Unfortunately, there was insufficient time in the study presented herein to produce and screen the required range of constructs with deletions in order to find a soluble construct. However, if a soluble version of AaeA were able to be produced it would have advantages for the various biochemical and structural characterisations of interest, especially as detergents or other reagents would

not be required to solubilise and stabilise the protein in solution, crystallisation is also likely to be easier without the presence of detergents.

Beyond crystallography, more work can be done with the EM. Optimisation of grids coated with AaeB, particularly washing to improve the background would likely produce significant improvement in aligning the classes against one another, with the aim of producing a preliminary 3D model from the data. Once negative stain EM has provided this initial model, cryo-EM can be used to follow up and provide higher-resolution data. Another interesting avenue would be to examine AaeA in isolation by EM. If it forms higher-order oligomers such as hexamers, then they should be visible by negative stain. A hexamer with six-fold symmetry should be particularly clear, and gold labelling would provide further direct evidence. Additional techniques could be used following the approach of Janganan *et al.* (2011)²⁷⁴ who used crosslinking followed by both atomic force microscopy and EM to visualise the hexamer of a MFP and mass spectrometry to measure the mass of the intact complex.

Although no interaction with TolC is predicted from physiological data, a biochemical investigation would be of interest; in particular, an assessment could be made to if any evidence of binding between AaeAB and outer membrane proteins known to associate with inner membrane transporters in tripartite complexes such as TolC, MtrE, and OprM via binding studies.

Particularly of interest would be SdsP, as this protein is found in an operon with another AeAE family member (the SdsQRP operon). Highly-conserved residues identified from the model could also be investigated by mutagenesis, in order to assess their role. Mutation of residues vital for oligomerisation should abolish the interaction when mutated, for example, and this can be measured with

binding studies such as isothermal titration calorimetry or microscale thermophoresis.

7.2.1.1 An alternative to detergent

It could be beneficial to attempt solubilisation of AaeA-AaeB in styrene maleic acid co-polymer (SMA) to form styrene maleic acid co-polymer lipid particles (SMALPs) containing both AaeA and AaeB. This may allow isolation of the entire complex as a single unit by wrapping around it in the membrane, enabling analysis of the complex as a whole. Extraction of whole complexes has been shown, with for example, the entire SecYEG complex, which was co-extracted with SecA²⁷⁵. A key benefit of SMALPs over other techniques is that it extracts proteins directly from the membrane without the need for detergents that may destabilise the complex or individual polypeptides²⁷⁶. As well as this improved stability, later experiments are often improved when SMALPs are used rather than detergent. SMALP protein appear much better looking in EM compared to detergent-solubilised proteins, when tested on the related AcrB¹⁷⁵, with significantly improved background. SMALPs have also been used to obtain improved data for EPR spectroscopy²⁷⁷, CD and NMR experiments²⁷⁸. Similarly, SMALP AaeAB could be assessed by NativePAGE, SEC-MALLS, crosslinking, and EM, as seen in this study and the ability to extract and stabilise the whole complex would be extremely beneficial for study of the true native state. SMALPs do have the drawback – shared with detergents – of not being suitable for functional studies to test transport, as this experiment requires a closed system with membrane across which substrate can be transported, so they are not appropriate for every experiment, but they do provide a useful tool.

7.2.2 Study of ALMTs

Given the unpromising nature of expression *in planta* shown in **Chapter 4**, alternative expression systems are needed. Several approaches suggest themselves, the first of which is the possibility of using engineered plants that lack proteases, which would promote the stability of the expressed ALMTs^{279,280}. Alternatively, it may be possible to prevent the action of proteases with addition of a greater variety of protease inhibitors, but as inhibitors were used in the extraction buffers in this study it is not clear what additional components would prove suitable. Bacterial expression presents a lower-cost system that lends itself well to screening a large number of constructs, with the caveat that eukaryotic proteins have a low rate of success in bacterial expression systems and as such it is likely to not provide success without a very wide-ranging set of constructs to be tested. Recent developments in plant-based cell free expression systems also are an attractive proposition due to their similarity to the native system. Systems based on tobacco BY2²⁸¹ and *Arabidopsis* callus²⁸² have joined the wheatgerm system that has been in use for over 30 years²⁸³. Additionally, a benchmarking of several systems showed that HeLa- and *Leishmania*-based systems produced protein that was highly monodisperse²⁸⁴, significantly expediting structural studies, though it should be noted that these proteins were globular rather than membrane proteins. Possibly the best choice is expression in Sf9 insect cells, from which have been produced a significant proportion of membrane protein structures deposited in the PDB. Here, again, screening of as large a range of candidate genes and fusion tags as possible is the best approach to finding an eventual high-expressing, readily-solubilised protein that is amenable to structural study.

Isolation of specific eukaryotic cDNA sequences is somewhat more difficult compared to bacterial sequences, due to the presence of introns in the genomic DNA and thus the need to prepare it from mRNA, but this should not cause a major problem when compared to the challenges of membrane protein structural biology. Two additional cDNA have now become available to order from ABRC in addition to those studied in **Chapter 4** (AtALMT9 and AtALMT12), and it is also possible to order cDNA libraries for *Arabidopsis*, from which can be amplified the genes of interest. For other species, cDNA can be produced from plant material, as long as they are expressed in the tissues in question. In particular, the sequence for ALMT1 could be extracted from root tissue of plants grown in Al-containing media, and the coding sequences of ALMT6, ALMT9, and ALMT12 can be extracted from leaves, where they are known to be expressed in the stomata. Of course, another approach is to have the cDNA of interest chemically synthesised, which can be a fast and economical choice, with Genscript offering synthesis from 99 USD per gene. This approach also allows sequences to be codon-optimised, which can often prove helpful for heterologous expression by allowing faster translation, although, paradoxically, this can also lead to problems as rapid overexpression of membrane proteins can then overwhelm downstream processing machinery. In addition to starting with a larger number of sequences, it is likely to be a good idea to set up a range of vectors that have a variety of different tags for example, MBP fusion, GFP fusion, His-tag, Strep-tag, and combinations thereof. It is known that the location of type of fusion tags can affect the expression and stability of proteins and so screening several gives a greater chance of finding a construct that expresses well¹⁶⁷. A recent example in the lab

involved re-engineering an expression vector to incorporate a larger flexible linker and a GFP fusion, which led to a significant improvement in both expression levels and yield after purification and thus a larger yield of the target protein was obtained, accelerating biochemical studies.

A potential further line of enquiry would be to renew analysis of the CTD of TaALMT1. New constructs designed to encode the CTD while deleting the two predicted TM regions, either both or individually could be created, and their solubility assessed. The region between the two putative TM helices is approximately 75 residues long, and expressing this alone, presuming that this region folds as an independent soluble domain, could enable the study of it by crystallography, binding studies, and crosslinking. Production of this region from homologues such as AtALMT1, which also is AI-activated, and AtALMT6, 9, and 12, which are not, may produce comparisons that can elucidate the structural basis of differences between the proteins. Additionally, if either of the two putative TM regions can be left in the construct yet still produce soluble protein then this could provide evidence that either of the regions is not actually a TM helix, gaining further evidence for topology and architecture of ALMTs.

7.2.3 Study of Bacterial Homologues

In the absence of an experimental structure for ALMTs, acquisition of a structure of a related bacterial protein, such as AaeB, would enable insights to be drawn by homology modelling. This is of course in addition to the knowledge gained about the bacterial protein, which is important in itself. Similar to the discussion for ALMTs above, it would be advantageous to clone all of the PETs identified in *E. coli*, as well as SmFuaA, which has a known function, and other

similar proteins from other bacterial species (Prof. Stephen Baldwin identified 9 additional genes from species whose DNA was not available in the lab, for example). Once a great number of proteins has been identified, these should be fused to a variety of tags before screening all of the constructs created in order to screen to find best-expressing, as discussed above. In addition, it is worthwhile recommencing study on the *S. oneidensis* gene investigated in Chapter 5 as the expression and purification of this protein has already shown promise and further optimisation of fusion tags and culture conditions could well provide sufficient material for structural studies. Once a sufficient yield is obtained for any construct, the expression can be scaled up and produced protein can be used for crystal trials and EM. A crystal structure obtained for any individual family member is likely to provide a good template for homology modelling of any other member of the family, and proteins showing high sequence identity to ALMTs can similarly be used to guide the production of a homology model for ALMTs.

7.3 Summary

At the outset of this study the aim was to structurally and biochemically characterise ALMT family members, which have not been subject to such study thus far despite their vital biological importance in plants, and bacterial PET family members which are related to ALMTs and equally understudied. In particular, ALMTs are anion channels found throughout all plant species, and have key roles in acid soil tolerance, stomatal movements, and fruit flavour, with many more likely roles yet to be elucidated. Yet despite this clear importance, little headway has been made with discovering their three-dimensional structure, hampering the design of and interpretation of the results of any biochemical investigations. As such this study aimed to express and characterise both full-length ALMT channels and the isolated C-terminal domain of TaALMT1 and while a set of strategies was developed to produce members of this difficult family it was not possible to isolate protein for the studies intended. However, a great deal was learnt about the challenges of expressing these targets in *E. coli* and *Nicotiana benthamiana*. To overcome the problems with production of the plant proteins a shift was made to study of the related AaeB from *E. coli*, which is also poorly characterised. This work has shown the interaction between AaeA and AaeB and that they likely form a complex in vivo. Additionally, AaeA has been shown to be an integral membrane protein and a homology model has been developed, which was validated by CD spectroscopy. This was used to draw conclusions about AaeA such as identifying a potential mechanism for side-by-side homo-oligomerisation. The oligomeric state has been shown to be likely to be hexameric, and although stoichiometry of the overall AaeA-AaeB complex is difficult to determine, there

is evidence to support a 6:2 or 6:3 ratio. And finally EM analysis has shown the first structural evidence for the organisation of the complex, albeit at a low resolution. Overall this work has provided a significant development in our understanding of this bacterial family of proteins.

Chapter 8

References

1. Appelqvist, H., Wåster, P., Kågedal, K. & Öllinger, K. The lysosome: from waste bag to potential therapeutic target. *J. Mol. Cell Biol.* **5**, 214–226 (2013).
2. Staehelin, L. A. & Arntzen, C. J. Regulation of chloroplast membrane function: protein phosphorylation changes the spatial organization of membrane components. *J. Cell Biol.* **97**, 1327–1337 (1983).
3. Mannella, C. A. Structure and dynamics of the mitochondrial inner membrane cristae. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1763**, 542–548 (2006).
4. Baker, M. Making membrane proteins for structures: a trillion tiny tweaks. *Nat. Methods* **7**, 429–434 (2010).
5. Tan, S., Tan, H. T. & Chung, M. C. M. Membrane proteins and membrane proteomics. *PROTEOMICS* **8**, 3924–3932 (2008).
6. Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727–730 (2002).
7. Zwick, M., Esposito, C., Hellstern, M. & Seelig, A. Interplay between CFTR Phosphorylation, CFTR-ATPase Activity, and Anion Flux. *J. Biol. Chem.* (2016). doi:10.1074/jbc.M116.721415
8. Fox, R. K. & Muniraj, T. Pharmacologic Therapies in Gastrointestinal Diseases. *Med. Clin. North Am.* **100**, 827–850 (2016).
9. Bhardwaj, A. K. & Mohanty, P. Bacterial efflux pumps involved in multidrug resistance and their inhibitors: rejuvenating the antimicrobial chemotherapy. *Recent Patents Anti-Infect. Drug Disc.* **7**, 73–89 (2012).
10. 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. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. *Nature* **450**, 383–387 (2007).
11. Fakhoury, M. Revisiting the Serotonin Hypothesis: Implications for Major Depressive Disorders. *Mol. Neurobiol.* **53**, 2778–2786 (2016).
12. Bian, M., Zhou, M., Sun, D. & Li, C. Molecular approaches unravel the mechanism of acid soil tolerance in plants. *Crop J.* **1**, 91–104 (2013).
13. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
14. Overington, J. P., Al-Lazikani, B. & Hopkins, A. L. How many drug targets are there? *Nat. Rev. Drug Discov.* **5**, 993–996 (2006).

15. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination — The next generation. *Biochim. Biophys. Acta BBA - Biomembr.* **1838**, 78–87 (2014).
16. Lodish, H., Berk, A. & Zipursky, S. in *Molecular Cell Biol.* (W. H. Freeman, 2000). at <<http://www.ncbi.nlm.nih.gov/books/NBK21626/>>
17. Bienert, G. P., Schjoerring, J. K. & Jahn, T. P. Membrane transport of hydrogen peroxide. *Biochim. Biophys. Acta BBA - Biomembr.* **1758**, 994–1003 (2006).
18. Kasanicki, M. A., Cairns, M. T., Davies, A., Gardiner, R. M. & Baldwin, S. A. Identification and characterization of the glucose-transport protein of the bovine blood/brain barrier. *Biochem. J.* **247**, 101–108 (1987).
19. Ritzel, M. W. L., Ng, A. M. L., Yao, S. Y. M., Graham, K., Loewen, S. K., Smith, K. M., Ritzel, R. G., Mowles, D. A., Carpenter, P., Chen, X.-Z., Karpinski, E., Hyde, R. J., Baldwin, S. A., Cass, C. E. & Young, J. D. Molecular Identification and Characterization of Novel Human and Mouse Concentrative Na⁺-Nucleoside Cotransporter Proteins (hCNT3 and mCNT3) Broadly Selective for Purine and Pyrimidine Nucleosides (System cib). *J. Biol. Chem.* **276**, 2914–2927 (2001).
20. Takata, K., Matsuzaki, T. & Tajika, Y. Aquaporins: water channel proteins of the cell membrane. *Prog. Histochem. Cytochem.* **39**, 1–83 (2004).
21. Berg, J., Tymoczko, J. & Stryer, L. in *Biochemistry (Mosc.)* (WH Freeman, 2002). at <Available from: <http://www.ncbi.nlm.nih.gov/books/NBK22509/>>
22. Hilf, R. J. C. & Dutzler, R. Structure of a potentially open state of a proton-activated pentameric ligand-gated ion channel. *Nature* **457**, 115–118 (2009).
23. Sands, Z., Grottesi, A. & Sansom, M. S. P. Voltage-gated ion channels. *Curr. Biol.* **15**, R44–R47 (2005).
24. Peyronnet, R., Tran, D., Girault, T. & Frachisse, J.-M. Mechanosensitive channels: feeling tension in a world under pressure. *Front. Plant Sci.* **5**, 558 (2014).
25. Doyle, D. A. Structural changes during ion channel gating. *Trends Neurosci.* **27**, 298–302 (2004).
26. Törnroth-Horsefield, S., Wang, Y., Hedfalk, K., Johanson, U., Karlsson, M., Tajkhorshid, E., Neutze, R. & Kjellbom, P. Structural mechanism of plant aquaporin gating. *Nature* **439**, 688–694 (2006).
27. A L Olson & Pessin, and J. E. Structure, Function, and Regulation of the Mammalian Facilitative Glucose Transporter Gene Family. *Annu. Rev. Nutr.* **16**, 235–256 (1996).

28. Dean, M., Rzhetsky, A. & Allikmets, R. The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res.* **11**, 1156–1166 (2001).
29. Schep, D. G., Zhao, J. & Rubinstein, J. L. Models for the a subunits of the *Thermus thermophilus* V/A-ATPase and *Saccharomyces cerevisiae* V-ATPase enzymes by cryo-EM and evolutionary covariance. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 3245–3250 (2016).
30. Xie, H., Patching, S. G., Gallagher, M. P., Litherland, G. J., Brough, A. R., Venter, H., Yao, S. Y. M., Ng, A. M. L., Young, J. D., Herbert, R. B., Henderson, P. J. F. & Baldwin, S. A. Purification and properties of the *Escherichia coli* nucleoside transporter NupG, a paradigm for a major facilitator transporter sub-family. *Mol. Membr. Biol.* **21**, 323–336 (2004).
31. Pochini, L., Scalise, M., Galluccio, M. & Indiveri, C. Membrane transporters for the special amino acid glutamine: structure/function relationships and relevance to human health. *Front. Chem.* **2**, (2014).
32. Kotrba, P., Inui, M. & Yukawa, H. Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *J. Biosci. Bioeng.* **92**, 502–517 (2001).
33. Batista, A. P., Marreiros, B. C. & Pereira, M. M. The role of proton and sodium ions in energy transduction by respiratory complex I. *IUBMB Life* **64**, 492–498 (2012).
34. Artigas, P. & Gadsby, D. C. Ion channel-like properties of the Na⁺/K⁺ Pump. *Ann. N. Y. Acad. Sci.* **976**, 31–40 (2002).
35. Parker, J. L. & Newstead, S. Molecular basis of nitrate uptake by the plant nitrate transporter NRT1.1. *Nature* **507**, 68–72 (2014).
36. Harley, K. T. & Saier, M. H. A novel ubiquitous family of putative efflux transporters. *J. Mol. Microbiol. Biotechnol.* **2**, 195–198 (2000).
37. Palmer, A. J., Baker, A. & Muench, S. P. The varied functions of aluminium-activated malate transporters—much more than aluminium resistance. *Biochem. Soc. Trans.* **44**, 856–862 (2016).
38. Van Dyk, T. K., Templeton, L. J., Cantera, K. A., Sharpe, P. L. & Sariaslani, F. S. Characterization of the *Escherichia coli* AaeAB Efflux Pump: a Metabolic Relief Valve? *J. Bacteriol.* **186**, 7196–7204 (2004).
39. Saier, M. H., Reddy, V. S., Tamang, D. G. & Västermark, A. The transporter classification database. *Nucleic Acids Res.* **42**, D251-258 (2014).
40. Hu, Y.-B., Sosso, D., Qu, X.-Q., Chen, L.-Q., Ma, L., Chermak, D., Zhang, D.-C. & Frommer, W. B. Phylogenetic evidence for a fusion of archaeal and bacterial SemiSWEETs to form eukaryotic SWEETs and identification of

- SWEET hexose transporters in the amphibian chytrid pathogen *Batrachochytrium dendrobatidis*. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* (2016). doi:10.1096/fj.201600576R
41. Tao, Y., Cheung, L. S., Li, S., Eom, J.-S., Chen, L.-Q., Xu, Y., Perry, K., Frommer, W. B. & Feng, L. Structure of a eukaryotic SWEET transporter in a homo-trimeric complex. *Nature* **527**, 259–263 (2015).
 42. Shimada, T., Yamamoto, K. & Ishihama, A. Involvement of the leucine response transcription factor LeuO in regulation of the genes for sulfa drug efflux. *J. Bacteriol.* **191**, 4562–4571 (2009).
 43. Sasaki, T., Yamamoto, Y., Ezaki, B., Katsuhara, M., Ahn, S. J., Ryan, P. R., Delhaize, E. & Matsumoto, H. A wheat gene encoding an aluminum-activated malate transporter. *Plant J.* **37**, 645–653 (2004).
 44. Hoekenga, O. A., Maron, L. G., Piñeros, M. A., Cançado, G. M. A., Shaff, J., Kobayashi, Y., Ryan, P. R., Dong, B., Delhaize, E., Sasaki, T., Matsumoto, H., Yamamoto, Y., Koyama, H. & Kochian, L. V. AtALMT1, which encodes a malate transporter, is identified as one of several genes critical for aluminum tolerance in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **103**, 9738–9743 (2006).
 45. Dreyer, I., Gomez-Porrás, J. L., Riano-Pachon, D. M., Hedrich, R. & Geiger, D. Molecular Evolution of Slow and Quick Anion Channels (SLACs and QUACs/ALMTs). *Front. Plant Sci.* **3**, (2012).
 46. Takanashi, K., Sasaki, T., Kan, T., Saida, Y., Sugiyama, A., Yamamoto, Y. & Yazaki, K. A dicarboxylate transporter, LjALMT4, mainly expressed in nodules of *Lotus japonicus*. *Mol. Plant. Microbe Interact.* (2016). doi:10.1094/MPMI-04-16-0071-R
 47. Meyer, S., Mumm, P., Imes, D., Endler, A., Weder, B., Al-Rasheid, K. A. S., Geiger, D., Marten, I., Martinoia, E. & Hedrich, R. AtALMT12 represents an R-type anion channel required for stomatal movement in *Arabidopsis* guard cells. *Plant J.* **63**, 1054–1062 (2010).
 48. De Angeli, A., Zhang, J., Meyer, S. & Martinoia, E. AtALMT9 is a malate-activated vacuolar chloride channel required for stomatal opening in *Arabidopsis*. *Nat. Commun.* **4**, 1804 (2013).
 49. Piñeros, M. A., Cancado, G. M. A., Maron, L. G., Lyi, S. M., Menossi, M. & Kochian, L. V. Not all ALMT1-type transporters mediate aluminum-activated organic acid responses: the case of ZmALMT1 - an anion-selective transporter. *Plant J.* **53**, 352–367 (2008).
 50. Khan, S. A., Beekwilder, J., Schaart, J. G., Mumm, R., Soriano, J. M., Jacobsen, E. & Schouten, H. J. Differences in acidity of apples are

- probably mainly caused by a malic acid transporter gene on LG16. *Tree Genet. Genomes* **9**, 475–487 (2013).
51. Xu, M., Gruber, B. D., Delhaize, E., White, R. G., James, R. A., You, J., Yang, Z. & Ryan, P. R. The barley anion channel, HvALMT1, has multiple roles in guard cell physiology and grain metabolism. *Physiol. Plant.* **153**, 183–193 (2015).
 52. Ramesh, S. A., Tyerman, S. D., Xu, B., Bose, J., Kaur, S., Conn, V., Domingos, P., Ullah, S., Wege, S., Shabala, S., Feijó, J. A., Ryan, P. R. & Gillham, M. GABA signalling modulates plant growth by directly regulating the activity of plant-specific anion transporters. *Nat. Commun.* **6**, (2015).
 53. Soto-Cerda, B. J., Inostroza-Blancheteau, C., Mathias, M., Penaloza, E., Zuniga, J., Munoz, G., Rengel, Z. & Salvo-Garrido, H. Marker-assisted breeding for TaALMT1, a major gene conferring aluminium tolerance to wheat. *Biol. Plant.* **59**, 83–91 (2015).
 54. Bindraban, P. S., van der Velde, M., Ye, L., van den Berg, M., Materechera, S., Kiba, D. I., Tamene, L., Ragnarsdóttir, K. V., Jongschaap, R., Hoogmoed, M., Hoogmoed, W., van Beek, C. & van Lynden, G. Assessing the impact of soil degradation on food production. *Curr. Opin. Environ. Sustain.* **4**, 478–488 (2012).
 55. Cordell, D., Drangert, J.-O. & White, S. The story of phosphorus: Global food security and food for thought. *Glob. Environ. Change* **19**, 292–305 (2009).
 56. Sade, H., Meriga, B., Surapu, V., Gadi, J., Sunita, M. S. L., Suravajhala, P. & Kavi Kishor, P. B. Toxicity and tolerance of aluminum in plants: tailoring plants to suit to acid soils. *Biometals Int. J. Role Met. Ions Biol. Biochem. Med.* **29**, 187–210 (2016).
 57. Kochian, L. V., Piñeros, M. A., Liu, J. & Magalhaes, J. V. Plant Adaptation to Acid Soils: The Molecular Basis for Crop Aluminum Resistance. *Annu. Rev. Plant Biol.* **66**, 571–598 (2015).
 58. Kochian, L. V. Cellular Mechanisms of Aluminum Toxicity and Resistance in Plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **46**, 237–260 (1995).
 59. Baker, A., Ceasar, S. A., Palmer, A. J., Paterson, J. B., Qi, W., Muench, S. P. & Baldwin, S. A. Replace, reuse, recycle: improving the sustainable use of phosphorus by plants. *J. Exp. Bot.* **66**, 3523–3540 (2015).
 60. WANG, C., ZHAO, X.-Q., AIZAWA, T., SUNAIRI, M. & SHEN, R.-F. High Aluminum Tolerance of *Rhodotorula* sp. RS1 is Associated with Thickening of the Cell Wall Rather than Chelation of Aluminum Ions. *Pedosphere* **23**, 29–38 (2013).

61. Xia, J. X., Yamaji, N., Kasai, T. & Ma, J. A. F. Plasma membrane-localized transporter for aluminum in rice. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 18381–18385 (2010).
62. Zhou, G., Delhaize, E., Zhou, M. & Ryan, P. R. The barley MATE gene, HvAACT1, increases citrate efflux and Al³⁺ tolerance when expressed in wheat and barley. *Ann. Bot.* **112**, 603–612 (2013).
63. Magalhaes, J. V. How a microbial drug transporter became essential for crop cultivation on acid soils: aluminium tolerance conferred by the multidrug and toxic compound extrusion (MATE) family. *Ann. Bot.* **106**, 199–203 (2010).
64. Zhang, W.-H., Ryan, P. R., Sasaki, T., Yamamoto, Y., Sullivan, W. & Tyerman, S. D. Characterization of the TaALMT1 Protein as an Al³⁺-Activated Anion Channel in Transformed Tobacco (*Nicotiana tabacum* L.) Cells. *Plant Cell Physiol.* **49**, 1316–1330 (2008).
65. Ryan, P. R., Skerrett, M., Findlay, G. P., Delhaize, E. & Tyerman, S. D. Aluminum activates an anion channel in the apical cells of wheat roots. *Proc. Natl. Acad. Sci.* **94**, 6547–6552 (1997).
66. Tashiro, M., Furihata, K., Fujimoto, T., Machinami, T. & Yoshimura, E. Characterization of the malate-aluminum(III) complex using ¹H and ²⁷Al NMR spectroscopy. *Magn. Reson. Chem.* **45**, 518–521 (2007).
67. Meyer, S., De Angeli, A., Fernie, A. R. & Martinoia, E. Intra- and extra-cellular excretion of carboxylates. *Trends Plant Sci.* **15**, 40–47 (2010).
68. Delhaize, E., Taylor, P., Hocking, P. J., Simpson, R. J., Ryan, P. R. & Richardson, A. E. Transgenic barley (*Hordeum vulgare* L.) expressing the wheat aluminium resistance gene (*TaALMT1*) shows enhanced phosphorus nutrition and grain production when grown on an acid soil. *Plant Biotechnol. J.* **7**, 391–400 (2009).
69. Delhaize, E., Ryan, P. R., Hebb, D. M., Yamamoto, Y., Sasaki, T. & Matsumoto, H. Engineering high-level aluminum tolerance in barley with the ALMT1 gene. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15249–15254 (2004).
70. Delhaize, E., Ma, J. F. & Ryan, P. R. Transcriptional regulation of aluminium tolerance genes. *Trends Plant Sci.* **17**, 341–348 (2012).
71. Kobayashi, Y., Hoekenga, O. A., Itoh, H., Nakashima, M., Saito, S., Shaff, J. E., Maron, L. G., Pineros, M. A., Kochian, L. V. & Koyama, H. Characterization of AtALMT1 Expression in Aluminum-Inducible Malate Release and Its Role for Rhizotoxic Stress Tolerance in Arabidopsis. *Plant Physiol.* **145**, 843–852 (2007).

72. Sawaki, Y., Iuchi, S., Kobayashi, Y., Kobayashi, Y., Ikka, T., Sakurai, N., Fujita, M., Shinozaki, K., Shibata, D., Kobayashi, M. & Koyama, H. STOP1 regulates multiple genes that protect arabidopsis from proton and aluminum toxicities. *Plant Physiol.* **150**, 281–294 (2009).
73. Kobayashi, Y., Ohyama, Y., Kobayashi, Y., Ito, H., Iuchi, S., Fujita, M., Zhao, C.-R., Tanveer, T., Ganesan, M., Kobayashi, M. & Koyama, H. STOP2 Activates Transcription of Several Genes for Al- and Low pH-Tolerance that Are Regulated by STOP1 in Arabidopsis. *Mol. Plant* **7**, 311–322 (2014).
74. Ding, Z. J., Yan, J. Y., Xu, X. Y., Li, G. X. & Zheng, S. J. WRKY46 functions as a transcriptional repressor of ALMT1, regulating aluminum-induced malate secretion in Arabidopsis. *Plant J.* **76**, 825–835 (2013).
75. Sasaki, T., Tsuchiya, Y., Ariyoshi, M., Ryan, P. R. & Yamamoto, Y. A chimeric protein of aluminum-activated malate transporter generated from wheat and Arabidopsis shows enhanced response to trivalent cations. *Biochim. Biophys. Acta* **1858**, 1427–1435 (2016).
76. Sasaki, T., Tsuchiya, Y., Ariyoshi, M., Ryan, P. R., Furuichi, T. & Yamamoto, Y. A Domain-Based Approach for Analyzing the Function of Aluminum-Activated Malate Transporters from Wheat (*Triticum aestivum*) and Arabidopsis thaliana in *Xenopus oocytes*. *Plant Cell Physiol.* (2014). doi:10.1093/pcp/pcu143
77. Ligaba, A., Katsuhara, M., Ryan, P. R., Shibasaka, M. & Matsumoto, H. The BnALMT1 and BnALMT2 Genes from Rape Encode Aluminum-Activated Malate Transporters That Enhance the Aluminum Resistance of Plant Cells. *Plant Physiol.* **142**, 1294–1303 (2006).
78. Hoffland, E., Findenegg, G. R. & Nelemans, J. A. Solubilization of rock phosphate by rape. *Plant Soil* **113**, 155–160 (1989).
79. Fontecha, G., Silva-Navas, J., Benito, C., Mestres, M. A., Espino, F. J., Hernández-Riquer, M. V. & Gallego, F. J. Candidate gene identification of an aluminum-activated organic acid transporter gene at the Alt4 locus for aluminum tolerance in rye (*Secale cereale* L.). *TAG Theor. Appl. Genet. Theor. Angew. Genet.* **114**, 249–260 (2007).
80. Collins, N. C., Shirley, N. J., Saeed, M., Pallotta, M. & Gustafson, J. P. An ALMT1 gene cluster controlling aluminum tolerance at the Alt4 locus of rye (*Secale cereale* L.). *Genetics* **179**, 669–682 (2008).
81. Liang, C., Piñeros, M. A., Tian, J., Yao, Z., Sun, L., Liu, J., Shaff, J., Coluccio, A., Kochian, L. V. & Liao, H. Low pH, Aluminum, and Phosphorus

- Coordinately Regulate Malate Exudation through GmALMT1 to Improve Soybean Adaptation to Acid Soils. *Plant Physiol.* **161**, 1347–1361 (2013).
82. Ligaba, A., Kochian, L. & Pineros, M. Phosphorylation at S384 regulates the activity of the TaALMT1 malate transporter that underlies aluminum resistance in wheat. *Plant J.* **60**, 411–423 (2009).
83. Furuichi, T., Sasaki, T., Tsuchiya, Y., Ryan, P. R., Delhaize, E. & Yamamoto, Y. An extracellular hydrophilic carboxy-terminal domain regulates the activity of TaALMT1, the aluminum-activated malate transport protein of wheat. *Plant J.* **64**, 47–55 (2010).
84. Ligaba, A., Dreyer, I., Margaryan, A., Schneider, D. J., Kochian, L. & Piñeros, M. Functional, structural and phylogenetic analysis of domains underlying the Al sensitivity of the aluminum-activated malate/anion transporter, TaALMT1. *Plant J.* **76**, 766–780 (2013).
85. Sivaguru, M., Ezaki, B., He, Z.-H., Tong, H., Osawa, H., Baluska, F., Volkman, D. & Matsumoto, H. Aluminum-Induced Gene Expression and Protein Localization of a Cell Wall-Associated Receptor Kinase in Arabidopsis. *Plant Physiol.* **132**, 2256–2266 (2003).
86. Mumm, P., Imes, D., Martinoia, E., Al-Rasheid, K. A. S., Geiger, D., Marten, I. & Hedrich, R. C-terminus mediated voltage gating of Arabidopsis guard cell anion channel QUAC1. *Mol. Plant* (2013).
doi:10.1093/mp/sst008
87. Ligaba, A., Maron, L., Shaff, J., Kochian, L. & Piñeros, M. Maize ZmALMT2 is a root anion transporter that mediates constitutive root malate efflux. *Plant Cell Environ.* **35**, 1185–1200 (2012).
88. Wang, B. & Qiu, Y.-L. Phylogenetic distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* **16**, 299–363 (2006).
89. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
90. Kollist, H., Nuhkat, M. & Roelfsema, M. R. G. Closing gaps: linking elements that control stomatal movement. *New Phytol.* **203**, 44–62 (2014).
91. Roelfsema, M. R. G., Levchenko, V. & Hedrich, R. ABA depolarizes guard cells in intact plants, through a transient activation of R- and S-type anion channels. *Plant J. Cell Mol. Biol.* **37**, 578–588 (2004).
92. Meyer, S., Scholz-Starke, J., De Angeli, A., Kovermann, P., Burla, B., Gambale, F. & Martinoia, E. Malate transport by the vacuolar AtALMT6 channel in guard cells is subject to multiple regulation. *Plant J. Cell Mol. Biol.* **67**, 247–257 (2011).

93. Zhang, J., Baetz, U., Krügel, U., Martinoia, E. & Angeli, A. D. Identification of a Probable Pore-Forming Domain in the Multimeric Vacuolar Anion Channel AtALMT9. *Plant Physiol.* **163**, 830–843 (2013).
94. Zhang, J., Martinoia, E. & De Angeli, A. Cytosolic Nucleotides Block and Regulate the Arabidopsis Vacuolar Anion Channel AtALMT9. *J. Biol. Chem.* **289**, 25581–25589 (2014).
95. Sasaki, T., Mori, I. C., Furuichi, T., Munemasa, S., Toyooka, K., Matsuoka, K., Murata, Y. & Yamamoto, Y. Closing Plant Stomata Requires a Homolog of an Aluminum-Activated Malate Transporter. *Plant Cell Physiol.* **51**, 354–365 (2010).
96. Kovermann, P., Meyer, S., Hörtensteiner, S., Picco, C., Scholz-Starke, J., Ravera, S., Lee, Y. & Martinoia, E. The Arabidopsis vacuolar malate channel is a member of the ALMT family. *Plant J. Cell Mol. Biol.* **52**, 1169–1180 (2007).
97. Gruber, B. D., Ryan, P. R., Richardson, A. E., Tyerman, S. D., Ramesh, S., Hebb, D. M., Howitt, S. M. & Delhaize, E. HvALMT1 from barley is involved in the transport of organic anions. *J. Exp. Bot.* **61**, 1455–1467 (2010).
98. Gruber, B. D., Delhaize, E., Richardson, A. E., Roessner, U., James, R. A., Howitt, S. M. & Ryan, P. R. Characterisation of HvALMT1 function in transgenic barley plants. *Funct. Plant Biol.* **38**, 163–175 (2011).
99. Bai, Y., Dougherty, L., Li, M., Fazio, G., Cheng, L. & Xu, K. A natural mutation-led truncation in one of the two aluminum-activated malate transporter-like genes at the Ma locus is associated with low fruit acidity in apple. *Mol. Genet. Genomics* **287**, 663–678 (2012).
100. Angeli, A., Baetz, U., Francisco, R., Zhang, J., Chaves, M. M. & Regalado, A. The vacuolar channel VvALMT9 mediates malate and tartrate accumulation in berries of *Vitis vinifera*. *Planta* **238**, 283–291 (2013).
101. Palanivelu, R., Brass, L., Edlund, A. F. & Preuss, D. Pollen Tube Growth and Guidance Is Regulated by POP2, an Arabidopsis Gene that Controls GABA Levels. *Cell* **114**, 47–59 (2003).
102. Kvist, M., Hancock, V. & Klemm, P. Inactivation of Efflux Pumps Abolishes Bacterial Biofilm Formation. *Appl. Environ. Microbiol.* **74**, 7376–7382 (2008).
103. Jones, C. M., Hernández Lozada, N. J. & Pfleger, B. F. Efflux systems in bacteria and their metabolic engineering applications. *Appl. Microbiol. Biotechnol.* **99**, 9381–9393 (2015).

104. Lee, S., Nam, D., Jung, J. Y., Oh, M.-K., Sang, B.-I. & Mitchell, R. J. Identification of *Escherichia coli* biomarkers responsive to various lignin-hydrolysate compounds. *Bioresour. Technol.* **114**, 450–456 (2012).
105. Monnappa, A. K., Lee, S. & Mitchell, R. J. Sensing of plant hydrolysate-related phenolics with an *aaeXAB::luxCDABE* bioreporter strain of *Escherichia coli*. *Bioresour. Technol.* **127**, 429–434 (2013).
106. Hinchliffe, P., Greene, N. P., Paterson, N. G., Crow, A., Hughes, C. & Koronakis, V. Structure of the periplasmic adaptor protein from a major facilitator superfamily (MFS) multidrug efflux pump. *FEBS Lett.* **588**, 3147–3153 (2014).
107. Symmons, M. F., Marshall, R. L. & Bavro, V. N. Architecture and roles of periplasmic adaptor proteins in tripartite efflux assemblies. *Front. Microbiol.* **6**, (2015).
108. Daur, L., Orange, F., Taveau, J.-C., Verchère, A., Monlezun, L., Gounou, C., Marreddy, R. K. R., Picard, M., Broutin, I., Pos, K. M. & Lambert, O. Tripartite assembly of RND multidrug efflux pumps. *Nat. Commun.* **7**, 10731 (2016).
109. Zgurskaya, H. I., Yamada, Y., Tikhonova, E. B., Ge, Q. & Krishnamoorthy, G. Structural and functional diversity of bacterial membrane fusion proteins. *Biochim. Biophys. Acta* **1794**, 794–807 (2009).
110. Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., Ohene-Agyei, T., Venter, H., Chiu, W. & Luisi, B. F. Structure of the AcrAB-TolC multidrug efflux pump. *Nature* **509**, 512–515 (2014).
111. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci. Publ. Protein Soc.* **7**, 1029–1038 (1998).
112. Ravna, A. W., Sager, G., Dahl, S. G. & Sylte, I. in *Transp. Targets Drugs* (eds. Napier, S. & Bingham, M.) 15–51 (Springer Berlin Heidelberg, 2008). at <http://link.springer.com/chapter/10.1007/7355_2008_023>
113. Ward, A., Sanderson, N., O'Reilly, J., Rutherford, N. & Poolman, B. in *Membr. Transp. Pract. Approach* (ed. Baldwin, S.) (Oxford: Blackwell, 2000).
114. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 1760–1765 (2009).
115. Indiveri, C., Galluccio, M., Scalise, M. & Pochini, L. Strategies of Bacterial Over Expression of Membrane Transporters Relevant in Human Health:

- The Successful Case of the Three Members of OCTN Subfamily. *Mol. Biotechnol.* **54**, 724–736 (2013).
116. Postis, V. G. L., Rawlings, A. E., Lesiuk, A. & Baldwin, S. A. Use of *Escherichia coli* for the production and purification of membrane proteins. *Methods Mol. Biol. Clifton NJ* **998**, 33–54 (2013).
117. Ward, A., O'Reilly, J., Rutherford, N. G., Ferguson, S. M., Hoyle, C. K., Palmer, S. L., Clough, J. L., Venter, H., Xie, H., Litherland, G. J., Martin, G. E., Wood, J. M., Roberts, P. E., Groves, M. A., Liang, W. J., Steel, A., McKeown, B. J. & Henderson, P. J. Expression of prokaryotic membrane transport proteins in *Escherichia coli*. *Biochem. Soc. Trans.* **27**, 893–899 (1999).
118. Schägger, H., Link, T. A., Engel, W. D. & von Jagow, G. Isolation of the eleven protein subunits of the bc1 complex from beef heart. *Methods Enzymol.* **126**, 224–237 (1986).
119. Wagner, S., Bader, M. L., Drew, D. & de Gier, J.-W. Rationalizing membrane protein overexpression. *Trends Biotechnol.* **24**, 364–371 (2006).
120. Drew, D., Fröderberg, L., Baars, L. & de Gier, J.-W. L. Assembly and overexpression of membrane proteins in *Escherichia coli*. *Biochim. Biophys. Acta* **1610**, 3–10 (2003).
121. Schnell, D. J. & Hebert, D. N. Protein translocons: multifunctional mediators of protein translocation across membranes. *Cell* **112**, 491–505 (2003).
122. Ma, P., Varela, F., Magoch, M., Silva, A. R., Rosário, A. L., Brito, J., Oliveira, T. F., Nogly, P., Pessanha, M., Stelter, M., Kletzin, A., Henderson, P. J. F. & Archer, M. An Efficient Strategy for Small-Scale Screening and Production of Archaeal Membrane Transport Proteins in *Escherichia coli*. *PLoS ONE* **8**, (2013).
123. Wild, R., Gerasimaite, R., Jung, J.-Y., Truffault, V., Pavlovic, I., Schmidt, A., Saiardi, A., Jessen, H. J., Poirier, Y., Hothorn, M. & Mayer, A. Control of eukaryotic phosphate homeostasis by inositol polyphosphate sensor domains. *Science* **352**, 986–990 (2016).
124. Bernaudat, 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. Heterologous Expression of Membrane Proteins: Choosing the Appropriate Host. *PLOS ONE* **6**, e29191 (2011).

125. Molina, D. M., Cornvik, T., Eshaghi, S., Haeggström, J. Z., Nordlund, P. & Sabet, M. I. Engineering membrane protein overproduction in *Escherichia coli*. *Protein Sci. Publ. Protein Soc.* **17**, 673–680 (2008).
126. Kurland, C. & Gallant, J. Errors of heterologous protein expression. *Curr. Opin. Biotechnol.* **7**, 489–493 (1996).
127. Klepsch, M. M., Persson, J. O. & de Gier, J.-W. L. Consequences of the overexpression of a eukaryotic membrane protein, the human KDEL receptor, in *Escherichia coli*. *J. Mol. Biol.* **407**, 532–542 (2011).
128. Emmerstorfer, A., Wriessnegger, T., Hirz, M. & Pichler, H. Overexpression of membrane proteins from higher eukaryotes in yeasts. *Appl. Microbiol. Biotechnol.* **98**, 7671–7698 (2014).
129. Altmann, F., Staudacher, E., Wilson, I. B. H. & März, L. Insect cells as hosts for the expression of recombinant glycoproteins. *Glycoconj. J.* **16**, 109–123 (1999).
130. Zhao, Y., Bishop, B., Clay, J. E., Lu, W., Jones, M., Daenke, S., Siebold, C., Stuart, D. I., Yvonne Jones, E. & Radu Aricescu, A. Automation of large scale transient protein expression in mammalian cells. *J. Struct. Biol.* **175**, 209–215 (2011).
131. Mardanov, E. S., Kotlyarov, R. Y., Kuprianov, V. V., Stepanova, L. A., Tsybalova, L. M., Lomonosoff, G. P. & Ravin, N. V. Rapid high-yield expression of a candidate influenza vaccine based on the ectodomain of M2 protein linked to flagellin in plants using viral vectors. *BMC Biotechnol.* **15**, (2015).
132. Peyret, H. & Lomonosoff, G. P. The pEAQ vector series: the easy and quick way to produce recombinant proteins in plants. *Plant Mol. Biol.* **83**, 51–58 (2013).
133. Boland, C., Li, D., Shah, S. T. A., Haberstock, S., Dötsch, V., Bernhard, F. & Caffrey, M. Cell-free Expression and In Meso Crystallisation of an Integral Membrane Kinase for Structure Determination. *Cell. Mol. Life Sci. CMLS* **71**, 4895–4910 (2014).
134. Wu, J. J. & Swartz, J. R. High yield cell-free production of integral membrane proteins without refolding or detergents. *Biochim. Biophys. Acta* **1778**, 1237–1250 (2008).
135. RCSB PDB. at
<http://www.rcsb.org/pdb/static.do?p=general_information/pdb_statistics/index.html>

136. Schmitz, A. & Galas, D. J. The interaction of RNA polymerase and lac repressor with the lac control region. *Nucleic Acids Res.* **6**, 111–137 (1979).
137. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
138. Palmer, I. & Wingfield, P. T. Preparation and Extraction of Insoluble (Inclusion-Body) Proteins from Escherichia coli. *Curr. Protoc. Protein Sci. Editor. Board John E Coligan AI CHAPTER*, Unit-6.3 (2004).
139. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **99**, 303–310 (2005).
140. Kiefer, H. In vitro folding of alpha-helical membrane proteins. *Biochim. Biophys. Acta BBA - Biomembr.* **1610**, 57–62 (2003).
141. Rogl, H., Kosemund, K., Kühlbrandt, W. & Collinson, I. Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett.* **432**, 21–26 (1998).
142. Gorzelle, B. M., Nagy, J. K., Oxenoid, K., Lonzer, W. L., Cafiso, D. S. & Sanders, C. R. Reconstitutive Refolding of Diacylglycerol Kinase, an Integral Membrane Protein. *Biochemistry (Mosc.)* **38**, 16373–16382 (1999).
143. Osborne, A. R. & Rapoport, T. A. Protein Translocation Is Mediated by Oligomers of the SecY Complex with One SecY Copy Forming the Channel. *Cell* **129**, 97–110 (2007).
144. Long, S. B., Tao, X., Campbell, E. B. & MacKinnon, R. Atomic structure of a voltage-dependent K⁺ channel in a lipid membrane-like environment. *Nature* **450**, 376–382 (2007).
145. Waugh, D. S. Crystal structures of MBP fusion proteins. *Protein Sci. Publ. Protein Soc.* **25**, 559–571 (2016).
146. Nallamsetty, S. & Waugh, D. S. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expr. Purif.* **45**, 175–182 (2006).
147. Ware, S., Donahue, J. P., Hawiger, J. & Anderson, W. F. Structure of the fibrinogen gamma-chain integrin binding and factor XIIIa cross-linking sites obtained through carrier protein driven crystallization. *Protein Sci. Publ. Protein Soc.* **8**, 2663–2671 (1999).
148. Corsini, L., Hothorn, M., Scheffzek, K., Sattler, M. & Stier, G. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Sci. Publ. Protein Soc.* **17**, 2070–2079 (2008).

149. Spurlino, J. C., Lu, G. Y. & Quioco, F. A. The 2.3-Å resolution structure of the maltose- or maltodextrin-binding protein, a primary receptor of bacterial active transport and chemotaxis. *J. Biol. Chem.* **266**, 5202–5219 (1991).
150. Bell, M. R., Engleka, M. J., Malik, A. & Strickler, J. E. To fuse or not to fuse: What is your purpose?: Protein Fusion Technology. *Protein Sci.* **22**, 1466–1477 (2013).
151. Costa, S., Almeida, A., Castro, A. & Domingues, L. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia coli*: the novel Fh8 system. *Front. Microbiol.* **5**, (2014).
152. Nannenga, B. L. & Baneyx, F. Folding engineering strategies for efficient membrane protein production in *E. coli*. *Methods Mol. Biol. Clifton NJ* **899**, 187–202 (2012).
153. Betenbaugh, M. J., Ailor, E., Whiteley, E., Hinderliter, P. & Hsu, T. A. Chaperone and foldase coexpression in the baculovirus-insect cell expression system. *Cytotechnology* **20**, 149–159 (1996).
154. Tegel, H., Tourle, S., Ottosson, J. & Persson, A. Increased levels of recombinant human proteins with the *Escherichia coli* strain Rosetta(DE3). *Protein Expr. Purif.* **69**, 159–167 (2010).
155. Schlegel, S., Löfblom, J., Lee, C., Hjelm, A., Klepsch, M., Strous, M., Drew, D., Slotboom, D. J. & de Gier, J.-W. Optimizing membrane protein overexpression in the *Escherichia coli* strain Lemo21(DE3). *J. Mol. Biol.* **423**, 648–659 (2012).
156. Mahamad, P., Boonchird, C. & Panbangred, W. High level accumulation of soluble diphtheria toxin mutant (CRM197) with co-expression of chaperones in recombinant *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **100**, 6319–6330 (2016).
157. Butz, J. A., Niebauer, R. T. & Robinson, A. S. Co-expression of molecular chaperones does not improve the heterologous expression of mammalian G-protein coupled receptor expression in yeast. *Biotechnol. Bioeng.* **84**, 292–304 (2003).
158. 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. Crystal structure of a prokaryotic homologue of the mammalian oligopeptide[ndash]proton symporters, PepT1 and PepT2. *EMBO J.* **30**, 417–426 (2010).
159. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015).

160. Facchiano, A. M., Stiuso, P., Chiusano, M. L., Caraglia, M., Giuberti, G., Marra, M., Abbruzzese, A. & Colonna, G. Homology modelling of the human eukaryotic initiation factor 5A (eIF-5A). *Protein Eng.* **14**, 881–890 (2001).
161. Liang, J., Naveed, H., Jimenez-Morales, D., Adamian, L. & Lin, M. Computational studies of membrane proteins: Models and predictions for biological understanding. *Biochim. Biophys. Acta BBA - Biomembr.* **1818**, 927–941 (2012).
162. Chen, K.-Y. M., Sun, J., Salvo, J. S., Baker, D. & Barth, P. High-Resolution Modeling of Transmembrane Helical Protein Structures from Distant Homologues. *PLOS Comput Biol* **10**, e1003636 (2014).
163. Young, J. D., Yao, S. Y. M., Baldwin, J. M., Cass, C. E. & Baldwin, S. A. The human concentrative and equilibrative nucleoside transporter families, SLC28 and SLC29. *Mol. Aspects Med.* **34**, 529–547 (2013).
164. Baldwin, S. A., McConkey, G. A., Cass, C. E. & Young, J. D. Nucleoside transport as a potential target for chemotherapy in malaria. *Curr. Pharm. Des.* **13**, 569–580 (2007).
165. Vaziri, H., Baldwin, S. A., Baldwin, J. M., Adams, D. G., Young, J. D. & Postis, V. L. G. Use of molecular modelling to probe the mechanism of the nucleoside transporter NupG. *Mol. Membr. Biol.* **30**, 114–128 (2013).
166. Johnson, Z. L., Cheong, C.-G. & Lee, S.-Y. Crystal structure of a concentrative nucleoside transporter from *Vibrio cholerae* at 2.4 Å. *Nature* **483**, 489–493 (2012).
167. Ma, C., Hao, Z., Huysmans, G., Lesiuk, A., Bullough, P., Wang, Y., Bartlam, M., Phillips, S. E., Young, J. D., Goldman, A., Baldwin, S. A. & Postis, V. L. G. A Versatile Strategy for Production of Membrane Proteins with Diverse Topologies: Application to Investigation of Bacterial Homologues of Human Divalent Metal Ion and Nucleoside Transporters. *PLoS ONE* **10**, (2015).
168. Cava, F., Hidalgo, A. & Berenguer, J. *Thermus thermophilus* as biological model. *Extrem. Life Extreme Cond.* **13**, 213–231 (2009).
169. Deacon, S. E., Roach, P. C. J., Postis, V. L. G., Wright, G. S. A., Xia, X., Phillips, S. E. V., Knox, J. P., Henderson, P. J. F., McPherson, M. J. & Baldwin, S. A. Reliable scale-up of membrane protein over-expression by bacterial auto-induction: from microwell plates to pilot scale fermentations. *Mol. Membr. Biol.* **25**, 588–598 (2008).
170. Roach, P. C. J., Postis, V. L. G., Deacon, S. E., Wright, G. S. A., Ingram, J. C., Xia, X., McPherson, M. J. & Baldwin, S. A. Large-scale preparation of

- bacterial cell membranes by tangential flow filtration. *Mol. Membr. Biol.* **25**, 609–616 (2008).
171. Duquesne, K. & Sturgis, J. N. Membrane protein solubilization. *Methods Mol. Biol. Clifton NJ* **601**, 205–217 (2010).
172. Gao, Y., Cao, E., Julius, D. & Cheng, Y. TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action. *Nature advance online publication*, (2016).
173. Linke, D. Detergents: an overview. *Methods Enzymol.* **463**, 603–617 (2009).
174. Inagaki, S., Ghirlando, R. & Grisshammer, R. Biophysical Characterization of Membrane Proteins in Nanodiscs. *Methods San Diego Calif* **59**, 287–300 (2013).
175. Postis, V., Rawson, S., Mitchell, J. K., Lee, S. C., Parslow, R. A., Dafforn, T. R., Baldwin, S. A. & Muench, S. P. The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochim. Biophys. Acta BBA - Biomembr.* **1848**, 496–501 (2015).
176. Zoonens, M. & Popot, J.-L. Amphipols for Each Season. *J. Membr. Biol.* **247**, 759–796 (2014).
177. Eshaghi, S., Hedrén, M., Nasser, M. I. A., Hammarberg, T., Thornell, A. & Nordlund, P. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Sci. Publ. Protein Soc.* **14**, 676–683 (2005).
178. Gutmann, D. A. P., Mizohata, E., Newstead, S., Ferrandon, S., Postis, V., Xia, X., Henderson, P. J. F., van Veen, H. W. & Byrne, B. A high-throughput method for membrane protein solubility screening: the ultracentrifugation dispersity sedimentation assay. *Protein Sci. Publ. Protein Soc.* **16**, 1422–1428 (2007).
179. Postis, V. L. G., Deacon, S. E., Roach, P. C. J., Wright, G. S. A., Xia, X., Ingram, J. C., Hadden, J. M., Henderson, P. J. F., Phillips, S. E. V., McPherson, M. J. & Baldwin, S. A. A high-throughput assay of membrane protein stability. *Mol. Membr. Biol.* **25**, 617–624 (2008).
180. Platt, G., Postis, V., Hao, Z., Palmer, T. & Baldwin, S. Measuring the conformational stability of membrane proteins using the UNit. at <<https://www.unchainedlabs.com/downloads/Application-Note-Measuring-the-conformational-stability-of-membrane-proteins-using-the-UNit.pdf>>

181. Vinogradova, O., Sönnichsen, F. & Sanders, C. R. On choosing a detergent for solution NMR studies of membrane proteins. *J. Biomol. NMR* **11**, 381–386 (1998).
182. Rawlings, A. E. Membrane proteins: always an insoluble problem? *Biochem. Soc. Trans.* **44**, 790–795 (2016).
183. Perez-Aguilar, J. M. & Saven, J. G. Computational Design of Membrane Proteins. *Structure* **20**, 5–14 (2012).
184. Robichon, C., Luo, J., Causey, T. B., Benner, J. S. & Samuelson, J. C. Engineering Escherichia coli BL21(DE3) derivative strains to minimize E. coli protein contamination after purification by immobilized metal affinity chromatography. *Appl. Environ. Microbiol.* **77**, 4634–4646 (2011).
185. Glover, C. A. P., Postis, V. L. G., Charalambous, K., Tzokov, S. B., Booth, W. I., Deacon, S. E., Wallace, B. A., Baldwin, S. A. & Bullough, P. A. AcrB contamination in 2-D crystallization of membrane proteins: lessons from a sodium channel and a putative monovalent cation/proton antiporter. *J. Struct. Biol.* **176**, 419–424 (2011).
186. Thompson, R. F., Walker, M., Siebert, C. A., Muench, S. P. & Ranson, N. A. An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. *Methods San Diego Calif* **100**, 3–15 (2016).
187. Carpenter, E. P., Beis, K., Cameron, A. D. & Iwata, S. Overcoming the challenges of membrane protein crystallography. *Curr. Opin. Struct. Biol.* **18**, 581–586 (2008).
188. Liu, W. & Cherezov, V. Crystallization of Membrane Proteins in Lipidic Mesophases. *J. Vis. Exp. JoVE* (2011). doi:10.3791/2501
189. Kubicek, J., Schlesinger, R., Baeken, C., Büldt, G., Schäfer, F. & Labahn, J. Controlled In Meso Phase Crystallization – A Method for the Structural Investigation of Membrane Proteins. *PLOS ONE* **7**, e35458 (2012).
190. Pebay-Peyroula, E., Rummel, G., Rosenbusch, J. P. & Landau, E. M. X-ray Structure of Bacteriorhodopsin at 2.5 Angstroms from Microcrystals Grown in Lipidic Cubic Phases. *Science* **277**, 1676–1681 (1997).
191. Misquitta, L. V., Misquitta, Y., Cherezov, V., Slattery, O., Mohan, J. M., Hart, D., Zhalnina, M., Cramer, W. A. & Caffrey, M. Membrane Protein Crystallization in Lipidic Mesophases with Tailored Bilayers. *Structure* **12**, 2113–2124 (2004).
192. Wöhri, A. B., Johansson, L. C., Wadsten-Hindrichsen, P., Wahlgren, W. Y., Fischer, G., Horsefield, R., Katona, G., Nyblom, M., Öberg, F., Young, G., Cogdell, R. J., Fraser, N. J., Engström, S. & Neutze, R. A Lipidic-Sponge

- Phase Screen for Membrane Protein Crystallization. *Structure* **16**, 1003–1009 (2008).
193. Sanders, C. R. & Sönnichsen, F. Solution NMR of membrane proteins: practice and challenges. *Magn. Reson. Chem. MRC* **44 Spec No**, S24-40 (2006).
194. Shahid, S. A., Bardiaux, B., Franks, W. T., Krabben, L., Habeck, M., van Rossum, B.-J. & Linke, D. Membrane-protein structure determination by solid-state NMR spectroscopy of microcrystals. *Nat. Methods* **9**, 1212–1217 (2012).
195. Tang, M., Sperling, L. J., Berthold, D. A., Schwieters, C. D., Nesbitt, A. E., Nieuwkoop, A. J., Gennis, R. B. & Rienstra, C. M. High-resolution membrane protein structure by joint calculations with solid-state NMR and X-ray experimental data. *J. Biomol. NMR* **51**, 227–233 (2011).
196. Vinothkumar, K. R. Membrane protein structures without crystals, by single particle electron cryomicroscopy. *Curr. Opin. Struct. Biol.* **33**, 103–114 (2015).
197. Scheres, S. H. W. Semi-automated selection of cryo-EM particles in RELION-1.3. *J. Struct. Biol.* **189**, 114–122 (2015).
198. Lu, P., Bai, X., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S. H. W. & Shi, Y. Three-dimensional structure of human γ -secretase. *Nature* **512**, 166–170 (2014).
199. Bai, X., Yan, C., Yang, G., Lu, P., Ma, D., Sun, L., Zhou, R., Scheres, S. H. W. & Shi, Y. An atomic structure of human γ -secretase. *Nature* **525**, 212–217 (2015).
200. Bai, X., Rajendra, E., Yang, G., Shi, Y. & Scheres, S. H. W. Sampling the conformational space of the catalytic subunit of human γ -secretase. *eLife* **4**, (2015).
201. Allegretti, M., Klusch, N., Mills, D. J., Vonck, J., Kühlbrandt, W. & Davies, K. M. Horizontal membrane-intrinsic α -helices in the stator a-subunit of an F-type ATP synthase. *Nature* **521**, 237–240 (2015).
202. Baker, L. A., Watt, I. N., Runswick, M. J., Walker, J. E. & Rubinstein, J. L. Arrangement of subunits in intact mammalian mitochondrial ATP synthase determined by cryo-EM. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 11675–11680 (2012).
203. Rubinstein, J. L., Walker, J. E. & Henderson, R. Structure of the mitochondrial ATP synthase by electron cryomicroscopy. *EMBO J.* **22**, 6182–6192 (2003).

204. Benlekbir, S., Bueler, S. A. & Rubinstein, J. L. Structure of the vacuolar-type ATPase from *Saccharomyces cerevisiae* at 11-Å resolution. *Nat. Struct. Mol. Biol.* **19**, 1356–1362 (2012).
205. Rawson, S., Harrison, M. A. & Muench, S. P. Rotating with the brakes on and other unresolved features of the vacuolar ATPase. *Biochem. Soc. Trans.* **44**, 851–855 (2016).
206. Rawson, S., Phillips, C., Huss, M., Tiburcy, F., Wieczorek, H., Trinick, J., Harrison, M. A. & Muench, S. P. Structure of the vacuolar H⁺-ATPase rotary motor reveals new mechanistic insights. *Struct. Lond. Engl.* **1993** **23**, 461–471 (2015).
207. Song, C. F., Papachristos, K., Rawson, S., Huss, M., Wieczorek, H., Paci, E., Trinick, J., Harrison, M. A. & Muench, S. P. Flexibility within the rotor and stators of the vacuolar H⁺-ATPase. *PloS One* **8**, e82207 (2013).
208. Zhao, J., Benlekbir, S. & Rubinstein, J. L. Electron cryomicroscopy observation of rotational states in a eukaryotic V-ATPase. *Nature* **521**, 241–245 (2015).
209. Bickel-Sandkötter, S., Wagner, V. & Schumann, D. ATP-synthesis in archaea: Structure-function relations of the halobacterial A-ATPase. *Photosynth. Res.* **57**, 335–345 (1998).
210. Muench, S. P., Huss, M., Song, C. F., Phillips, C., Wieczorek, H., Trinick, J. & Harrison, M. A. Cryo-electron Microscopy of the Vacuolar ATPase Motor Reveals its Mechanical and Regulatory Complexity. *J. Mol. Biol.* **386**, 989–999 (2009).
211. Rawson, S., Iadanza, M. G., Ranson, N. A. & Muench, S. P. Methods to account for movement and flexibility in cryo-EM data processing. *Methods* **100**, 35–41 (2016).
212. Benjamin, C. J., Wright, K. J., Hyun, S.-H., Krynski, K., Yu, G., Bajaj, R., Guo, F., Stauffacher, C. V., Jiang, W. & Thompson, D. H. Nonfouling NTA-PEG-Based TEM Grid Coatings for Selective Capture of Histidine-Tagged Protein Targets from Cell Lysates. *Langmuir ACS J. Surf. Colloids* **32**, 551–559 (2016).
213. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
214. Grant, S. G., Jessee, J., Bloom, F. R. & Hanahan, D. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4645–4649 (1990).

215. Grodberg, J. & Dunn, J. J. ompT encodes the Escherichia coli outer membrane protease that cleaves T7 RNA polymerase during purification. *J. Bacteriol.* **170**, 1245–1253 (1988).
216. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
217. Yadav, S., Sharma, P., Srivastava, A., Desai, P. & Shrivastava, N. Strain specific Agrobacterium-mediated genetic transformation of Bacopa monnieri. *J. Genet. Eng. Biotechnol.* **12**, 89–94 (2014).
218. Hanahan, D. Studies on transformation of Escherichia coli with plasmids. *J. Mol. Biol.* **166**, 557–580 (1983).
219. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., 'everine, Wilkins, M. R., Appel, R. D. & Bairoch, A. *Protein identification and analysis tools on the ExPASy server.* (Springer, 2005). at <<http://link.springer.com/protocol/10.1385/1-59259-890-0:571>>
220. Sparkes, I. A., Runions, J., Kearns, A. & Hawes, C. Rapid, transient expression of fluorescent fusion proteins in tobacco plants and generation of stably transformed plants. *Nat. Protoc.* **1**, 2019–2025 (2006).
221. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665-667 (2004).
222. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* msw054 (2016). doi:10.1093/molbev/msw054
223. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* *CABIOS* **8**, 275–282 (1992).
224. Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73–82 (2008).
225. Folta-Stogniew, E. in *New Emerg. Proteomic Tech.* 97–112 (Humana Press, 2006). at <<http://dx.doi.org/10.1385/1-59745-026-X%3A97>>
226. Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I. & Ludtke, S. J. EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).
227. van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R. & Schatz, M. A new generation of the IMAGIC image processing system. *J. Struct. Biol.* **116**, 17–24 (1996).

228. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. The protein data bank: A computer-based archival file for macromolecular structures. *Arch. Biochem. Biophys.* **185**, 584–591 (1978).
229. Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L. & McCafferty, J. Production of soluble mammalian proteins in *Escherichia coli*: identification of protein features that correlate with successful expression. *BMC Biotechnol.* **4**, 32 (2004).
230. Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**, 661–670 (2004).
231. Ahmad, M., Hirz, M., Pichler, H. & Schwab, H. Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* **98**, 5301–5317 (2014).
232. Gómez-Sebastián, S., López-Vidal, J. & Escribano, J. M. Significant Productivity Improvement of the Baculovirus Expression Vector System by Engineering a Novel Expression Cassette. *PLoS ONE* **9**, (2014).
233. Dalton, A. C. & Barton, W. A. Over-expression of secreted proteins from mammalian cell lines. *Protein Sci. Publ. Protein Soc.* **23**, 517–525 (2014).
234. Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnol. J.* **7**, 682–693 (2009).
235. D'Aoust, M.-A., Couture, M. M.-J., Charland, N., Trépanier, S., Landry, N., Ors, F. & Vézina, L.-P. The production of hemagglutinin-based virus-like particles in plants: a rapid, efficient and safe response to pandemic influenza. *Plant Biotechnol. J.* **8**, 607–619 (2010).
236. Vardakou, M., Sainsbury, F., Rigby, N., Mulholland, F. & Lomonossoff, G. P. Expression of active recombinant human gastric lipase in *Nicotiana benthamiana* using the CPMV-HT transient expression system. *Protein Expr. Purif.* **81**, 69–74 (2012).
237. Miyamoto, K., Shimizu, T., Lin, F., Sainsbury, F., Thuenemann, E., Lomonossoff, G., Nojiri, H., Yamane, H. & Okada, K. Identification of an E-box motif responsible for the expression of jasmonic acid-induced chitinase gene *OsChia4a* in rice. *J. Plant Physiol.* **169**, 621–627 (2012).
238. Sainsbury, F., Sack, M., Stadlmann, J., Quendler, H., Fischer, R. & Lomonossoff, G. P. Rapid transient production in plants by replicating and non-replicating vectors yields high quality functional anti-HIV antibody. *PLoS One* **5**, e13976 (2010).

239. Abas, L. & Luschnig, C. Maximum yields of microsomal-type membranes from small amounts of plant material without requiring ultracentrifugation. *Anal. Biochem.* **401**, 217–227 (2010).
240. Yaeno, T., Li, H., Chaparro-Garcia, A., Schornack, S., Koshiba, S., Watanabe, S., Kigawa, T., Kamoun, S. & Shirasu, K. Phosphatidylinositol monophosphate-binding interface in the oomycete RXLR effector AVR3a is required for its stability in host cells to modulate plant immunity. *Proc. Natl. Acad. Sci.* **108**, 14682–14687 (2011).
241. Kagale, S., Uzuhashi, S., Wigness, M., Bender, T., Yang, W., Borhan, M. H. & Rozwadowski, K. TMV-Gate vectors: Gateway compatible tobacco mosaic virus based expression vectors for functional analysis of proteins. *Sci. Rep.* **2**, (2012).
242. Feller, T., Thom, P., Koch, N., Spiegel, H., Addai-Mensah, O., Fischer, R., Reimann, A., Pradel, G., Fendel, R., Schillberg, S., Scheuermayer, M. & Schinkel, H. Plant-Based Production of Recombinant Plasmodium Surface Protein Pf38 and Evaluation of its Potential as a Vaccine Candidate. *PLoS ONE* **8**, (2013).
243. Hwang, M. S., Lindenmuth, B. E., McDonald, K. A. & Falk, B. W. Bipartite and tripartite Cucumber mosaic virus-based vectors for producing the *Acidothermus cellulolyticus* endo-1,4- β -glucanase and other proteins in non-transgenic plants. *BMC Biotechnol.* **12**, 66 (2012).
244. Van der Hoorn, R. A. L., Laurent, F., Roth, R. & De Wit, P. J. G. M. Agroinfiltration Is a Versatile Tool That Facilitates Comparative Analyses of Avr9/Cf-9-Induced and Avr4/Cf-4-Induced Necrosis. *Mol. Plant. Microbe Interact.* **13**, 439–446 (2000).
245. Sainsbury, F., Saxena, P., Geisler, K., Osbourn, A. & Lomonossoff, G. P. Using a virus-derived system to manipulate plant natural product biosynthetic pathways. *Methods Enzymol.* **517**, 185–202 (2012).
246. Sainsbury, F. & Lomonossoff, G. P. Extremely High-Level and Rapid Transient Protein Production in Plants without the Use of Viral Replication. *Plant Physiol.* **148**, 1212–1218 (2008).
247. De Marcos Lousa, C., van Roermund, C. W. T., Postis, V. L. G., Dietrich, D., Kerr, I. D., Wanders, R. J. A., Baldwin, S. A., Baker, A. & Theodoulou, F. L. Intrinsic acyl-CoA thioesterase activity of a peroxisomal ATP binding cassette transporter is required for transport and metabolism of fatty acids. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 1279–1284 (2013).
248. Evans-Roberts, K. M., Mitchenall, L. A., Wall, M. K., Leroux, J., Mylne, J. S. & Maxwell, A. DNA Gyrase Is the Target for the Quinolone Drug

- Ciprofloxacin in *Arabidopsis thaliana*. *J. Biol. Chem.* **291**, 3136–3144 (2016).
249. Kozma, D., Simon, I. & Tusnády, G. E. PDBTM: Protein Data Bank of transmembrane proteins after 8 years. *Nucleic Acids Res.* **41**, D524–D529 (2013).
250. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
251. The EMBL-EBI bioinformatics web and programmatic tools framework. at <<http://nar.oxfordjournals.org/content/early/2015/04/06/nar.gkv279>>
252. Psakis, G., Saidijam, M., Shibayama, K., Polaczek, J., Bettaney, K. E., Baldwin, J. M., Baldwin, S. A., Hope, R., Essen, L.-O., Essenberg, R. C. & Henderson, P. J. F. The sodium-dependent d-glucose transport protein of *Helicobacter pylori*. *Mol. Microbiol.* **71**, 391–403 (2009).
253. Mikolosko, J., Bobyk, K., Zgurskaya, H. I. & Ghosh, P. Conformational Flexibility in the Multidrug Efflux System Protein AcrA. *Structure* **14**, 577–587 (2006).
254. Kim, J.-S., Song, S., Lee, M., Lee, S., Lee, K. & Ha, N.-C. Crystal Structure of a Soluble Fragment of the Membrane Fusion Protein HlyD in a Type I Secretion System of Gram-Negative Bacteria. *Structure* **24**, 477–485 (2016).
255. Su, C.-C., Yang, F., Long, F., Reyon, D., Routh, M. D., Kuo, D. W., Mokhtari, A. K., Van Ornam, J. D., Rabe, K. L., Hoy, J. A., Lee, Y. J., Rajashankar, K. R. & Yu, E. W. Crystal Structure of the Membrane Fusion Protein CusB from *Escherichia coli*. *J. Mol. Biol.* **393**, 342–355 (2009).
256. Chae, P. S., Rasmussen, S. G. F., Rana, R., Gotfryd, K., Chandra, R., Goren, M. A., Kruse, A. C., Nurva, S., Loland, C. J., Pierre, Y., Drew, D., Popot, J.-L., Picot, D., Fox, B. G., Guan, L., Gether, U., Byrne, B., Kobilka, B. & Gellman, S. H. Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat. Methods* **7**, 1003–1008 (2010).
257. Cho, K. H., Husri, M., Amin, A., Gotfryd, K., Lee, H. J., Go, J., Kim, J. W., Loland, C. J., Guan, L., Byrne, B. & Chae, P. S. Maltose Neopentyl Glycol-3 (MNG-3) Analogues for Membrane Protein Study. *The Analyst* **140**, 3157–3163 (2015).
258. Burgess, R. R. in *Methods Enzymol.* (ed. Richard R. Burgess and Murray P. Deutscher) **Volume 463**, 259–282 (Academic Press, 2009).
259. Hobbs, E. C., Yin, X., Paul, B. J., Astarita, J. L. & Storz, G. Conserved small protein associates with the multidrug efflux pump AcrB and

- differentially affects antibiotic resistance. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 16696–16701 (2012).
260. Lee, M., Jun, S.-Y., Yoon, B.-Y., Song, S., Lee, K. & Ha, N.-C. Membrane fusion proteins of type I secretion system and tripartite efflux pumps share a binding motif for TolC in gram-negative bacteria. *PLoS One* **7**, e40460 (2012).
261. Lee, M., Kim, H.-L., Song, S., Joo, M., Lee, S., Kim, D., Hahn, Y., Ha, N.-C. & Lee, K. The α -barrel tip region of Escherichia coli TolC homologs of Vibrio vulnificus interacts with the MacA protein to form the functional macrolide-specific efflux pump MacAB-TolC. *J. Microbiol. Seoul Korea* **51**, 154–159 (2013).
262. Xu, Y., Sim, S.-H., Song, S., Piao, S., Kim, H.-M., Jin, X. L., Lee, K. & Ha, N.-C. The tip region of the MacA alpha-hairpin is important for the binding to TolC to the Escherichia coli MacAB-TolC pump. *Biochem. Biophys. Res. Commun.* **394**, 962–965 (2010).
263. Elazar, A., Weinstein, J. J., Prilusky, J. & Fleishman, S. J. Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 10340–10345 (2016).
264. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344-350 (2016).
265. Greenfield, N. J. & Hitchcock-DeGregori, S. E. Conformational intermediates in the folding of a coiled-coil model peptide of the N-terminus of tropomyosin and alpha alpha-tropomyosin. *Protein Sci. Publ. Protein Soc.* **2**, 1263–1273 (1993).
266. Fujiwara, Y., Kurokawa, T., Takeshita, K., Kobayashi, M., Okochi, Y., Nakagawa, A. & Okamura, Y. The cytoplasmic coiled-coil mediates cooperative gating temperature sensitivity in the voltage-gated H⁺ channel Hv1. *Nat. Commun.* **3**, 816 (2012).
267. Hanson, M. A., Cherezov, V., Griffith, M. T., Roth, C. B., Jaakola, V.-P., Chien, E. Y. T., Velasquez, J., Kuhn, P. & Stevens, R. C. A specific cholesterol binding site is established by the 2.8 Å structure of the human beta2-adrenergic receptor. *Struct. Lond. Engl.* **16**, 897–905 (2008).
268. Hainfeld, J. F., Liu, W., Joshi, V. & Powell, R. D. Nickel-NTA-Nanogold Binds His-Tagged Proteins. *Microsc. Microanal.* **8**, 832–833 (2002).

269. Büchel, C., Morris, E., Orlova, E. & Barber, J. Localisation of the PsbH subunit in photosystem II: a new approach using labelling of His-tags with a Ni(2+)-NTA gold cluster and single particle analysis. *J. Mol. Biol.* **312**, 371–379 (2001).
270. Tikhonova, E. B., Devroy, V. K., Lau, S. Y. & Zgurskaya, H. I. Reconstitution of the Escherichia coli macrolide transporter: the periplasmic membrane fusion protein MacA stimulates the ATPase activity of MacB. *Mol. Microbiol.* **63**, 895–910 (2007).
271. Lin, H. T., Bavro, V. N., Barrera, N. P., Frankish, H. M., Velamakanni, S., van Veen, H. W., Robinson, C. V., Borges-Walmsley, M. I. & Walmsley, A. R. MacB ABC transporter is a dimer whose ATPase activity and macrolide-binding capacity are regulated by the membrane fusion protein MacA. *J. Biol. Chem.* **284**, 1145–1154 (2009).
272. BioTechniques - Glutaraldehyde: behavior in aqueous solution, reaction with proteins, and application to enzyme crosslinking. at <<http://www.biotechniques.com/BiotechniquesJournal/2004/November/Glut-araldehyde-behavior-in-aqueous-solution-reaction-with-proteins-and-application-to-enzyme-crosslinking/biotechniques-117448.html>>
273. Winter, D., Vinegar, B., Nahal, H., Ammar, R., Wilson, G. V. & Provart, N. J. An 'Electronic Fluorescent Pictograph' Browser for Exploring and Analyzing Large-Scale Biological Data Sets. *PLOS ONE* **2**, e718 (2007).
274. Janganan, T. K., Bavro, V. N., Zhang, L., Matak-Vinkovic, D., Barrera, N. P., Venien-Bryan, C., Robinson, C. V., Borges-Walmsley, M. I. & Walmsley, A. R. Evidence for the Assembly of a Bacterial Tripartite Multidrug Pump with a Stoichiometry of 3:6:3. *J. Biol. Chem.* **286**, 26900–26912 (2011).
275. Prabudiansyah, I., Kusters, I., Caforio, A. & Driessen, A. J. M. Characterization of the annular lipid shell of the Sec translocon. *Biochim. Biophys. Acta* **1848**, 2050–2056 (2015).
276. Dörr, J. M., Scheidelaar, S., Koorengel, M. C., Dominguez, J. J., Schäfer, M., van Walree, C. A. & Killian, J. A. The styrene–maleic acid copolymer: a versatile tool in membrane research. *Eur. Biophys. J.* **45**, 3–21 (2016).
277. Sahu, I. D., McCarrick, R. M., Troxel, K. R., Zhang, R., Smith, H. J., Dunagan, M. M., Swartz, M. S., Rajan, P. V., Kroncke, B., Sanders, C. R. & Lorigan, G. A. DEER EPR Measurements for Membrane Protein Structures via Bi-functional Spin Labels and Lipodisq Nanoparticles. *Biochemistry (Mosc.)* **52**, (2013).

278. Knowles, T. J., Finka, R., Smith, C., Lin, Y.-P., Dafforn, T. & Overduin, M. Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer. *J. Am. Chem. Soc.* **131**, 7484–7485 (2009).
279. Goulet, C., Khalf, M., Sainsbury, F., D'Aoust, M.-A. & Michaud, D. A protease activity-depleted environment for heterologous proteins migrating towards the leaf cell apoplast. *Plant Biotechnol. J.* **10**, 83–94 (2012).
280. Robert, S., Khalf, M., Goulet, M.-C., D'Aoust, M.-A., Sainsbury, F. & Michaud, D. Protection of recombinant mammalian antibodies from development-dependent proteolysis in leaves of *Nicotiana benthamiana*. *PLoS One* **8**, e70203 (2013).
281. Buntru, M., Vogel, S., Spiegel, H. & Schillberg, S. Tobacco BY-2 cell-free lysate: an alternative and highly-productive plant-based in vitro translation system. *BMC Biotechnol.* **14**, 37 (2014).
282. Murota, K., Hagiwara-Komoda, Y., Komoda, K., Onouchi, H., Ishikawa, M. & Naito, S. Arabidopsis cell-free extract, ACE, a new in vitro translation system derived from Arabidopsis callus cultures. *Plant Cell Physiol.* **52**, 1443–1453 (2011).
283. Olliver, C. L., Grobler-Rabie, A. & Boyd, C. D. In vitro translation of messenger RNA in a wheat germ extract cell-free system. *Methods Mol. Biol. Clifton NJ* **2**, 137–144 (1985).
284. Gagoski, D., Polinkovsky, M. E., Mureev, S., Kunert, A., Johnston, W., Gambin, Y. & Alexandrov, K. Performance benchmarking of four cell-free protein expression systems. *Biotechnol. Bioeng.* **113**, 292–300 (2016).