LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. 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.

Title

A structural study of AlgE, transporter of alginate from *Pseudomonas aeruginosa* PAO1

Author

Jingquan Tan

Supervisor

Martin Caffrey

Submitted to University of Dublin, Trinity College, for a Doctor of Philosophy

April 2013

TRINITY COLLEGE
2 4 MAY 2013
LIBRARY DUBLIN

Thesis 9962

Abstract

Pseudomonas aeruginosa is an opportunistic human pathogen that causes serious diseases in humans with immune system deficiencies, such as cystic fibrosis, wound infections, urinary tract infections, and blood stream infections (Bentzmann and Plésiat 2011). Biofilm formation plays an important role in the life cycle of *P. aeruginosa* to protect cells from adverse environment and facilitate the infection to hosts (Drenkard and Frederick 2002). In Membrane Structural and Functional Biology Group, we seek to control biofilm formation by reducing the secretion of one of its important components, alginate. For alginate to be secreted to the cell surface, the outer membrane protein, AlgE, is required as alginate transporter (Rehm *et al.* 1994). Thus, AlgE is chosen as the study target of this project. A recombinant gene of AlgE has been cloned and the protein over-expressed in *Escherichia coli*. It has been solubilised from the membrane in detergent solution and purified by affinity and size exclusion chromatography. Both *in meso* and *in surfo* crystallisation trials have been set up and the protein has been successfully crystallised. After optimisation, a complete X-ray diffraction data set of the crystals has been collected to the resolution of 1.9 Å. Different strategies have been designed in order to get the phase information for structure determination. By heavy atom soaking, the achieved preliminary phase information yielded a possible model for AlgE structure. However, because an *in surfo* structure of AlgE from *P. aeruginosa* PAO1, 3RBH in PDB, was published in August 2011, the phasing strategy is changed to molecular replacement. The resulted *in meso* structures share nearly the same features in transmembrane segment but shows very important difference in some loop regions. The existence of a citrate binding site inside the pore region reveals further atructura features that the previous *in surfo* structure is in a closed state and possibly reveals that the high-resolution *in meso* structure is in an inward open state. In addition, an unidentified molecule is found beside the longest extracellular loop that is proven to be crucial for alginate secretion, which may point out the direction for future research on the mechanism of alginate secretion. Details of the *in meso* structures will be discussed.

Keywords: AlgE, Biofilm, cloning, crystallogenesis, *in meso*, *in surfo*, lipid screening, MIR phasing, molecular replacement, over-expression, closed state, inward open state, X-ray crystallography

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the libarary to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions and acknowledgement.

Signed:

Acknowledgements

I can't thank Professor Martin Caffrey more, for his supervision during my years in his lab. He gave me the opportunity to work in this fantastic lab and has instructed me through the whole project ever since. He is always more than happy to have scientific discussions or to give advices on other matters, like presentation skills, writing skills and etc.

Thanks very much to Dr. Dianfan Li. He is the best source for my questions in terms of practical experimental problems. I have benefited a lot from his wide knowledge of science and outstanding patience in the past few years.

I would also like to thank Dr. David Aragão, Dr. Valerie Pye, Dr. Joseph Lyons, Ms. Nichole Hoefer, Dr. Coiline Boland and Dr. Vadim Cherezov for their hard work at synchrotrons. David and Valerie also devoted large amount of time in AlgE data processing and teaching me crystallography. Without their days and nights spent on this, the project can't progress to this far.

Thanks Dr. Jean Lee, Dr. Shiva Shankar, Dr. Syed Tasadaque Ali Shah and Dr. Joseph Lyons for synthesizing the lipids that are used in this work.

Thanks Dr. Declan Doyle for sharing his wisdom with me and strategizing the AlgE phasing work.

Thanks to all the members in the group for making the life in laboratory easy and interesting, in particular: Mr. Toufic El Arnaout, Mr. Emmanuel Nji, Dr. Abhiram Dukkipati, Ms. Margaret Ahern, Mrs Martha Motherway, Ms. Christina Doherty and Ms. Caroline Barry.

Finally, I would like to thank Science Foundation Ireland and National Institutes of Health for their funding.

Table of Contents

Abstract 2

Declaration 3

Acknowledgements 4

Table of Contents 5

List of Figures 8

List of Tables 11

Abbreviations 12

Chapter 1 Introduction 14

1.1 *Pseudomonas aeruginosa* 14

1.2 *P. aeruginosa* alginate and *algD* operon 15

1.2.1 *P. aeruginosa* biofilm and alginate 15

1.2.2 *algD* operon 19

1.2.3 AlgE 22

1.3 Project overview 29

Chapter 2 30

Cloning, expression, purification, identification 30

and characterization of AlgE 30

2.1 Chapter overview 30

2.2 Materials and instruments 30

2.3 Methods 36

2.3.1 Primer design 36

2.3.2 Media preparation and cell culture 37

2.3.3 PCR amplification 37

2.3.4 PCR product purification 38

2.3.5 TOPO cloning reaction 39

2.3.6 Transformation into cloning strain 39

2.3.7 Plasmid isolation and sequencing 40

2.3.8 Transformation into the expression strain 42

2.3.9 IPTG induction test 42

2.3.10 Scale-up cell culture 43

2.3.11 AlgE purification flowchart 44

2.3.12 Cell breakage 45

2.3.13 Membrane preparation and protein solubilisation 45

2.3.14 Cu^{2+}-IMAC Column equilibration 45

2.3.15 Detergent exchange and Cu^{2+}-IMAC purification 46

2.3.16 MALDI-TOF Mass Spectrometry 46

2.3.17 Size exclusion chromatography 47

2.3.18 Purity estimation 47

2.3.19 Heat modification test 48

2.3.20 Protein storage 48

2.4 Results and discussion 49

2.4.1 PCR amplification 49

2.4.2 Sequencing result 50

2.4.3 IPTG induction test 51

2.4.4 Cu^{2+}-IMAC chromatography for AlgE 54

2.4.5 MALDI-TOF MS identification 57

2.4.6 Size exclusion chromatography 59
2.4.7 Purity estimation for final AlgE sample 62
2.4.8 Heat modifiability of AlgE 64
2.5 Conclusions 66

Chapter 3 68

AlgE crystallisation and experimental approaches for structure determination 68

3.1 Introduction 68
3.1.1 Methods for protein structure determination 68
3.1.2 General protein crystallisation methods 69
3.1.3 *In meso* crystallisation method 70
3.1.4 Novel monoacylglycerol lipids (MAGs) 76
3.1.5 Phasing strategies 78
3.2 Materials and instruments 81
3.3 Methods 88
3.3.1 AlgE *In meso* crystallisation trials in 9.9 MAG 88
3.3.1.1 Cubic phase preparation 88
3.3.1.2 Commercial screen trial 90
3.3.1.3 Optimisation trial 94
3.3.2 AlgE *In meso* crystallisation trials in novel MAGs 94
3.3.2.1 Commercial screen trials 94
3.3.2.2 Optimisation trials 94
3.3.3 AlgE *in surfo* sitting-drop crystallisation trials 95
3.3.4 Crystal harvesting and data collection at synchrotrons 96
3.3.5 MIR phasing work for AlgE structure determination 100
3.3.5.1 Native AlgE crystals production 100
3.3.5.2 Heavy atom solution preparation 100
3.3.5.3 Colour screening experiment 102
3.3.5.4 Preliminary study on phasing power of K_2OsCl_6 and $(NH_4)_2IrCl_6$ 105
3.3.5.5 Systematic study on K_2OsCl_6's phasing power 106
3.4 Results and discussion 108
3.4.1 *In meso* crystals in 9.9 MAG 108
3.4.1.1 Initial hits from commercial screen trials 108
3.4.1.2 Crystals after optimisation 116
3.4.1.3 X-ray diffraction 120
3.4.2 *In meso* crystals in Novel MAGs 122
3.4.2.1 Initial hits from commercial screen trials 122
3.4.2.2 Crystals after optimisation 133
3.4.2.3 X-ray diffraction 137
3.4.3 AlgE *in surfo* crystallisation trials 144
3.4.4 Phasing work 144
3.4.4.1 Colour screening experiment 144
3.4.4.2 Phasing power for K_2OsCl_6 and $(NH_4)_2IrCl_6$ 146
3.5 Conclusions 150

Chapter 4 *In meso* Structure of AlgE 152
4.1 Introduction 152
4.2 Method 156
4.2.1 Molecular replacement and structure refinement 156
4.2.2 Analysis 158
4.3 Results and discussion 159
4.3.1 Molecular replacement and refinement 159
4.3.2 Overall structure 166
4.4.3 Citrate binding site 177

4.4.4 L7 and unidentified molecule 192
4.4.5 Lipid binding sites 197
4.4.6 Metal ions 203
4.5 Conclusion 208
Chapter 5 General conclusions and future work 210
References 213
Appendix 227

List of Figures

Figure 1.1 Chemical structure of alginate....16
Figure 1.2 Chemical structure of psl....17
Figure 1.3 The members of pel, algD and psl operon....18
Figure 1.4 Cartoon model for *P. aeruginosa* alginate synthesis and transportation machinery....19
Figure 1.5 GDP-mannuronic acid synthesis pathway....20
Figure 1.6 Topological model of mature AlgE based on structure determination....22
Figure 1.7 Sequence alignments among 17 AlgE or AlgE-like proteins....26
Figure 2.1 Full-length DNA sequence of AlgE gene with signal sequence from *P. aeruginosa* PAO1 Holloway strain....35
Figure 2.2 A flowchart for AlgE purification procedures....43
Figure 2.3 Result for annealing temperature gradient of PCR to find out the optimal condition for AlgE gene amplification....49
Figure 2.4 One of the six plasmids sent for sequencing, showing A128E point mutation....50
Figure 2.5 The result of IPTG induction test....51
Figure 2.6 A cartoon representation of major fragments in full-length recombinant AlgE protein....53
Figure 2.7 Cu^{2+}-IMAC chromatography for AlgE purification....54
Figure 2.8 The purity of fractions collected in the 80 mM - 130 mM range from Cu^{2+}-IMAC checked by SDS-PAGE....55
Figure 2.9 MALDI-TOF result for the protein purified by Cu2+IMAC....56
Figure 2.10 Database search result showing the protein sample is AlgE....57
Figure 2.11 Database search result showing which regions in protein sequence the identified fragments cover....57
Figure 2.12 Size exclusion chromatogram for AlgE purification....59
Figure 2.13 The purity of fractions from size exclusion chromatography checked on SDS-PAGE....60
Figure 2.14 AlgE purity estimation....62
Figure 2.15 The heat-modification behaviour of AlgE checked by SDS-PAGE....64
Figure 3.1 A cartoon representation of cubic phase system formed by water and monoolein....70
Figure 3.2 Cartoon representation of the proposed mechanism for *in meso* crystallisation process....71
Figure 3.3 The chemical structures of all monoacylglycerol lipids and detergents used in the project....76
Figure 3.4 The lipid mixing device....89
Figure 3.5 The set up of *in meso* crystallisation plate....91
Figure 3.6 The *in meso* crystallisation robot....92
Figure 3.7 Layout of the SwiisC1 MRC 2-well plate used for *in surfo* sitting-drop crystallisation trials....95
Figure 3.8 Cartoon representation of harvesting procedure from *in meso* crystallisation plate....97
Figure 3.9 Cartoon representation of heavy atom soaking procedure from *in meso* crystallisation plate....102
Figure 3.9 Example 1 of initial hits found for AlgE / 9.9 MAG *in meso* trials....109

Figure 3.10 Example 2 of initial hits found for AlgE / 9.9 MAG *in meso* trials.....110
Figure 3.11 Example 3 of initial hits found for AlgE / 9.9 MAG *in meso* trials.....111
Figure 3.12 Example 4 of initial hits found for AlgE / 9.9 MAG *in meso* trials.....112
Figure 3.13 Example 5 of initial hits found for AlgE / 9.9 MAG *in meso* trials.....113
Figure 3.14 Example 6 of initial hits found for AlgE / 9.9 MAG *in meso* trials.....114
Figure 3.15 Typical parallelogram type crystals found for AlgE / 9.9 MAG *in meso* trials 116
Figure 3.16 Typical hexagonal type crystals are found for AlgE / 9.9 MAG *in meso* trials 117
Figure 3.17 Typical needle type crystals are found for AlgE / 9.9 MAG *in meso* trials.118
Figure 3.18 Conditions of crystals found on crystallisation plates that are set up using AlgE Screen Kit 1.....119
Figure 3.19 The best diffraction data of crystals from AlgE / 9.9 MAG *in meso* trials..120
Figure 3.20 Example 1 of 2D crystals found for AlgE / 8.9 MAG *in meso* trials.....127
Figure 3.21 Example 2 of 2D crystals found for AlgE / 8.9 MAG *in meso* trials.....128
Figure 3.22 Example 1 of 2D crystals found for AlgE / 8.8 MAG *in meso* trials.....129
Figure 3.23 Example 1 of 2D crystals found for AlgE / 7.8 MAG *in meso* trials.....130
Figure 3.24 Example 1 of crystal cluster found for AlgE / 7.9 MAG *in meso* trials.....131
Figure 3.25 Typical parallelogram type crystals for AlgE / 7.8 MAG *in meso* trials using AlgE Screen Kit 1.....133
Figure 3.26 Typical "square" type crystals for AlgE / 7.8 MAG *in meso* trials using AlgE Screen Kit 1.....134
Figure 3.27 Typical rectangular type crystals for AlgE / 7.8 MAG *in meso* trials using AlgE Screen Kit 5.....135
Figure 3.28 The crystal that yielded the best and complete data set to 1.9 Å.....138
Figure 3.29 The best diffraction data of the crystal in Figure 3.28 from AlgE / 7.8 MAG *in meso* trials.....139
Figure 3.30 Suzuki-Miyaura coupling of an iodoalkene (orange) and the terminally unsaturated ester (red and black) to prepare MAG acetonide with different length of necks and tails.....140
Figure 3.31 Proton NMR profiles for two different batches of 7.8 MAG.....141
Figure 3.32 Expanded view of the boxed region in Figure 3.31.....142
Figure 3.33 View of probable AlgE β-barrel in the unit cell, along c;ab face.....147
Figure 3.34 View of layer of density, along b;ac face.....148
Figure 4.1 Overall *in surfo* structure of AlgE.....154
Figure 4.2 Illustration of ϕ and ψ angles of peptide bond that define Ramachandran plot for L- amino acids.....159
Figure 4.3 Ramachandran plots for AlgE_1.9.....161
Figure 4.4 Ramachandran plots for AlgE_2.4.....162
Figure 4.5 Ramachandran plots for AlgE_2.8.....163
Figure 4.6 Overview of AlgE_1.9 (code 4AFK in PDB).....165
Figure 4.7 Overview of AlgE_2.4 (code 4B61 in PDB).....166
Figure 4.8 Overview of AlgE_2.8 (code 4AZL in PDB).....167
Figure 4.9 A superposed view of all three *in meso* structures.....170
Figure 4.10 The hydrophobic thickness calculated by PPM server.....171
Figure 4.11 Crystal packing styles for all AlgE structure models.....173
Figure 4.12 Zoomed view in the crystal contact regions on loops L5 and L6 of AlgE *in meso* structures.....175
Figure 4.13 Position of citrate molecule found in AlgE_1.9.....177

Figure 4.14 Slab view of AlgE_1.9 reveals the position of citrate molecule.....178
Figure 4.15 The citrate molecule modelled into AlgE_1.9.....179
Figure 4.16. Slab view of AlgE_in_surfo from membrane shows no citrate molecule inside the barrel region.....180
Figure 4.17 The citrate binding site in AlgE_1.9 forming H-bonding with surrounding residues and waters.....183
Figure 4.18 The hypothetical secretion pathway of AlgE.....185
Figure 4.19 The citrate binding site in hypothetical secretion pathway in AlgE_1.9....186
Figure 4.20 A comparison between the size of citrate and size of mannuronate.....187
Figure 4.21 Clashing sites around the citrate site when AlgE_1.9 and AlgE_in_surfo are superposed.....189
Figure 4.22 Space-filling model that shows AlgE_in_surfo is possibly in a closed state.....190
Figure 4.23 Space-filling model that shows AlgE_1.9 is possibly in an inward open state.....191
Figure 4.24 Cartoon representation of L7 and unidentified molecule.....194
Figure 4.25 The electron density of the unidentified molecule.....195
Figure 4.26 Overall distributions of complete lipid molecules in each *in meso* AlgE structure model.....198
Figure 4.27 An example of 7.8 MAG (7.8 MAG-1) molecule with electron density from *in meso* AlgE_1.9 structure model.....199
Figure 4.28 7.8 MAG-1 binding site in *in meso* AlgE structure models.....200
Figure 4.29 7.8 MAG-2 lipid binding site in *in meso* AlgE structure models.....201
Figure 4.30 Ions modelled in *in meso* AlgE structure models.....203
Figure 4.30 The Cu^{2+} ions in AlgE_2.4 with surrounding environment.....206

List of Tables

Table 1.1 Amino Acid composition of mature AlgE.....24
Table 2.1 A list of materials and chemicals that have been used in this project.....29
Table 2.2 A list of all equipment that have been used in the project.....32
Table 2.3 A list of all components in 50 μL PCR mixture.....36
Table 2.4 A list of all components in TOPO Cloning reaction system.....38
Table 2.5 Summary for the yield of AlgE preps.....61
Table 3.1 Structures in Membrane Protein Data Bank (MPDB) that are solved by *in meso* method.....72
Table 3.2 A list of materials that have been used in this chapter.....80
Table 3.3 A list of instruments that have been used in this chapter.....84
Table 3.4 List of heavy atoms with working concentrations in colour screening experiment.....100
Table 3.5 Soaking conditions used for phasing power determination of K_2OsCl_6 and $(NH_4)_2IrCl_6$.....104
Table 3.6 Conditions used for systematic study on the phasing power of K_2OsCl_6.....105
Table 3.7 Summary of all initial hits in AlgE / 9.9 MAG in meso commercial trials.....107
Table 3.8 Summary of all initial hits found in different novel MAGs.....121
Table 3.9 Summary of diffraction quality for the initial hits found in all MAGs.....132
Table 3.10 Summary of data reduction statistics for three crystal forms of AlgE.....137
Table 3.11 Summary of X-ray diffraction qualities of all crystal forms.....143
Table 3.12 Results for colour screening experiment.....144
Table 3.13 Phasing power of K_2OsCl_6 and $(NH_4)_2IrCl_6$ in AlgE structure work.....145
Table 4.1 Data collection and refinement statistics for AlgE in surfo structure model, 3RBH.....153
Table 4.2 A summary of root-mean-square deviation of C_α between each of four AlgE structure models.....159
Table 4.3 Refinement and geometry statistics for AlgE *in meso* structure models.....160
Table 4.4 Summary of molecules in one asymmetric unit for all four AlgE models.....164
Table 4.5 Contacts present between citrate and protein residues or water molecules.....181
Table 4.6 Alignment of the residues shown in Figure 4.17 that have close contact with citrate molecule.....184
Table 4.7 Alignment of arginine residues found on loop L7 of AlgE protein.....192
Table 4.8 Chemicals used for the determination of all four AlgE structure models.....193
Table 4.9 List of metal ions in one asymmetric unit of all *in meso* structure models.....202
Table 4.10 Contacts present in between residues, acetate or water molecules and the four Cu^{2+} ions in AlgE2.4 structure model.....205

Abbreviations

Å	Ångström
Ala	Alanine
APS	Advanced Photon Source
Arg	Arginine
Asn	Asparagine
Asp	Aspartic acid
ATP	Adenosine Triphosphate
BCA	Bicinchoninic Acid Assay
BSA	Bovine Serum Albumin
Cys	Cysteine
CCD	Charge-Coupled Device
CMC	Critical Micelle Concentration
C-terminal	Carboxyl-terminal
DDM	n-dodecyl-β-D–maltoside
DLS	Diamond Light Source
DM	n-decyl-β-D–maltoside
DN	Deoxyribonucleic Acid
dNTPs	Deoxy-Nucleotide-Triphosphates
E. coli	*Escherichia coli*
FPLC	Fast Protein Liquid Chromatography
Gln	Glutamine
Glu	Glutamic acid
Gly	Glycine
His	Histidine
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl- β-D-Thiogalactopyranoside
Ile	Isoleucine
kDa	Kilo Daltons
LB	Luria Bertani
LDAO	N,N-Dimethyldodecylamine-N-Oxide
Leu	Leucine

Lys	Lysine
MAG	Monoacylglycerol
MALDI-TOF-MS	Matrix-Assisted Laser Desorption / Ionization Time of Flight Mass Spectrometry
Met	Methionine
MPDB	Membrane Protein Data Bank
MWCO	Molecular Weight Cut-Off
NMR	Nuclear Magnetic Resonance Spectroscopy
N-terminal	Amino-terminal
OD	Optical Density
OG	n-octyl-β-D-glucopyranoside
PAGE	Polyacrylamide Gel Electrophoresis
P. aeruginosa	*Pseudomonas aeruginosa*
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
Phe	Phenylalanine
Ser	Serine
SDS	Sodium Dodecyl Sulphate
Thr	Threonine
Tris	Tris-(Hydroxymethyl) Aminomethane
Trp	Tryptophan
UV	Ultra Violet
v/v	Volume per Volume
Vo	Void Volume
Vt	Included Volume
w/w	Weight per Weight

Chapter 1 Introduction

1.1 *Pseudomonas aeruginosa*

P. aeruginosa is a gram-negative, aerobic, rod-shaped bacterium with unipolar mobility. It can be motile by means of a single polar flagellum when colonising lung tissues, although its survival is independent of the existence of flagella (Drake and Montie 1988). It secretes a variety of pigments, including pyocyanin (blue-green), fluorescein (yellow-green and fluorescent, now also known as pyoverdin), and pyorubin (red-brown) (King *et al.* 1954). Although classified as an aerobic organism, *P. aeruginosa* is well adapted to proliferate in conditions of partial or total oxygen depletion (Palmer *et al.* 2007).

The ability to live in microaerobic or anaerobic environments is very important for some life stages of *P. aeruginosa.* For the instance, during lung infection in cystic fibrosis patients, the biofilm surrounding bacterial mucoid cells contains gel-like alginate component, which can limit the diffusion of oxygen and create a microaerobic environment (Hassett 1996, Worlitzsch *et al.* 2002, Cooper *et al.* 2003, Williams *et al.* 2007).

P. aeruginosa is an opportunistic pathogen and can initiate an infection when a weakness in the host defences is found. Infection by *P. aeruginosa* can cause urinary tract infections, respiratory system infections, dermatitis, soft tissue infections, bacteraemia, bone and joint infections, gastrointestinal infections and a variety of systemic infections, particularly in patients with severe burns, cystic fibrosis (CF), cancer and AIDS who are immunosuppressed (Todar 2011). In cystic fibrosis patients, infection in lungs by *P. aeruginosa* is one of the major causes of death (Bentzmann and Plésiat 2011). Unfortunately, here in Ireland, CF is one of the most common life-threatening inherited diseases. Approximately 1 in 19 people in Ireland are carriers of the CF gene and where two carriers parent a child together, there is a 1 in 4 chance of the baby being born with CF ((CFAI) 2011). So there is an urgent need to research and know more about *P. aeruginosa* infection in humans.

Because *P. aeruginosa's* potential threat to human health has been a big problem in hospitals, many antibiotics have been used in the fight against the bacteria. However, three major defensive mechanisms give *P. aeruginosa* the ability to survive an adverse environ-

ment. Firstly, *P. aeruginosa* has very low permeability in the outer membrane, which is only one percent of the value in *E. coli* and reduces the ability of antibiotics to enter the *P. aeruginosa* cells (Yoshimura and Nikaido 1982). Secondly, a biofilm is produced to further protect the cells from antibiotics. Although the mechanism how biofilm helps *P. aeruginosa* resist antibiotics remains unknown, it is proposed to be regulated by PvrR gene, which can convert the phenotype of biofilm from non-antibiotics-resistant to antibiotics-resistant (Drenkard and Frederick 2002). Thirdly, many complex multidrug efflux pump systems exist in *P. aeruginosa,* including the two most important ones: MexEF-OprN (Fetar *et al.* 2011) and MexAB-OprM (Esquisabel *et al.* 2011). Each pump has its own specificity and they work together to expel antibiotics from the organism in order to make the organism resistant to various antibiotics (Morita *et al.* 2001).

1.2 *P. aeruginosa* alginate and *algD* operon

1.2.1 *P. aeruginosa* biofilm and alginate

Biofilm is a mixture of various substances, such as DNA, proteins and polysaccharides, produced by microorganisms. In this media of biofilm, cells adhere to and interact with each other (Hall-Stoodley *et al.* 2004). There is a remarkable diversity in the composition of this matrix in different bacterial (Peter *et al.* 2011).

The existence of biofilm helps protect cells from a range of adverse environmental factors, like extreme pH, ultraviolet (UV) light, temperature or hydration (Hall-Stoodley *et al.* 2004). Biofilm has also been found to be involved in a wide variety of microbial infections in the body, by an estimated 80% of all infections ((NIH) 2002). Diseases linked to biofilm includes urinary tract infections, catheter infections, middle-ear infections, formation of dental plaque, gingivitis (Rogers 2008) and coating contact lenses (Imamura *et al.* 2008). Some less common but more lethal diseases include endocarditis, infections of lungs in cystic fibrosis patients, and infections of permanent indwelling devices such as joint prostheses and heart valves (Parsek and Singh 2003, Lewis 2011).

In industry, biofilms can also be detrimental. It can be erosive to metal surface (Mortensen and Conley 1994). When biofilms reside on the surface of heat exchangers, the rate of heat

transfer will be greatly reduced and then the energy will be wasted (Mortensen and Conley 1994). It is one of the major concerns in food industry, as biofilms help bacterial stick onto the surface of processing instruments (Carpentier and Cerf 1993).

However, biofilms can also be beneficial sometimes. They can be used for water cleaning, remediation of contaminated soil and groundwater, mining and nutrition maintenance in farmland because of its very good absorbing ability of various organic and inorganic substances. Some scientists have also used biofilms to prevent pathogenic bacterial from damaging crops possibly because the biofilm from some bacterial like *P. aeruginosa* and *Bacillus subtilis* contains antimicrobial agent. (Peyton and Characklis 1995, Morikawa 2006)

P. aeruginosa is not only an important opportunistic pathogen and causative agent of emerging hospital-acquired infections, but can also be considered a model organism for the study of diverse bacterial mechanisms that contribute to bacterial persistence. As mentioned in section 1.1, the capability of the bacteria to form biofilm that protects them from adverse environmental factors is one of the reasons for its resistance to many antibiotics.

As many other bacterial like *Escherichia coli, Aliivibrio fischeri, Acinetobacter sp, Bacillus subtilis* and so on, *P. aeruginosa*'s biofilm also consists of DNA, proteins, extracellular polysaccharides. There are three major types of polysaccharides in the *P. aeruginosa* biofilm: alginate (Hassett 1996), product of Polysaccharide Synthesis Locus (psl) (Byrd *et al.* 2009) and another pellicle-formation related polysaccharide (pel), whose nature or composition has not been identified (Mann and Wozniak 2011).

Alginate from *P. aeruginosa* is composed of a linear random co-polymer of (1, 4)-linked β-D-mannuronic acid and its C-5 epimer, α-L-guluronic acid (Figure 1.1). The mannuronic acid residues are variably modified with *O*-acetyl groups but the guluronic acid residues are not. The removal of *O*-acetyl group increases the amount of divalent cations that are bound to alginate but reduces the solubility of alginate (Sutherland 2001).

Unlike alginate, psl has an unusual pentarepeating unit (Figure 1.2) containing D-mannose, L-rhamnose, and D-glucose residues (Franklin *et al.* 2011). Two forms of psl can be found in *P. aeruginosa* cells: high molecular weight polymer and relatively smaller soluble form (Byrd *et al.* 2009).

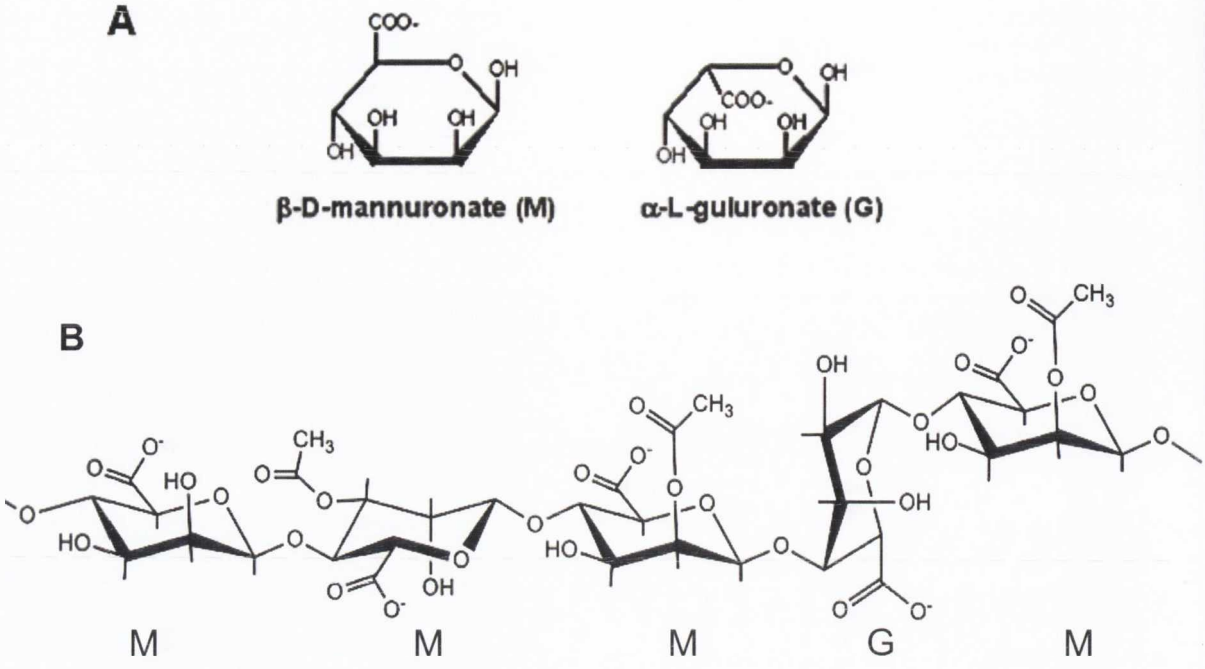


Figure 1.1 Chemical structure of alginate. In panel A, structures of each monomeric unit of alginate are shown. M and G is used to refer to mannuronate and guluronate, respectively. In panel B, the structure for a small fragment of alginate is shown. Acetylation randomly happens to the M unit but not to the G unit. Copied from (Franklin *et al.* 2011).

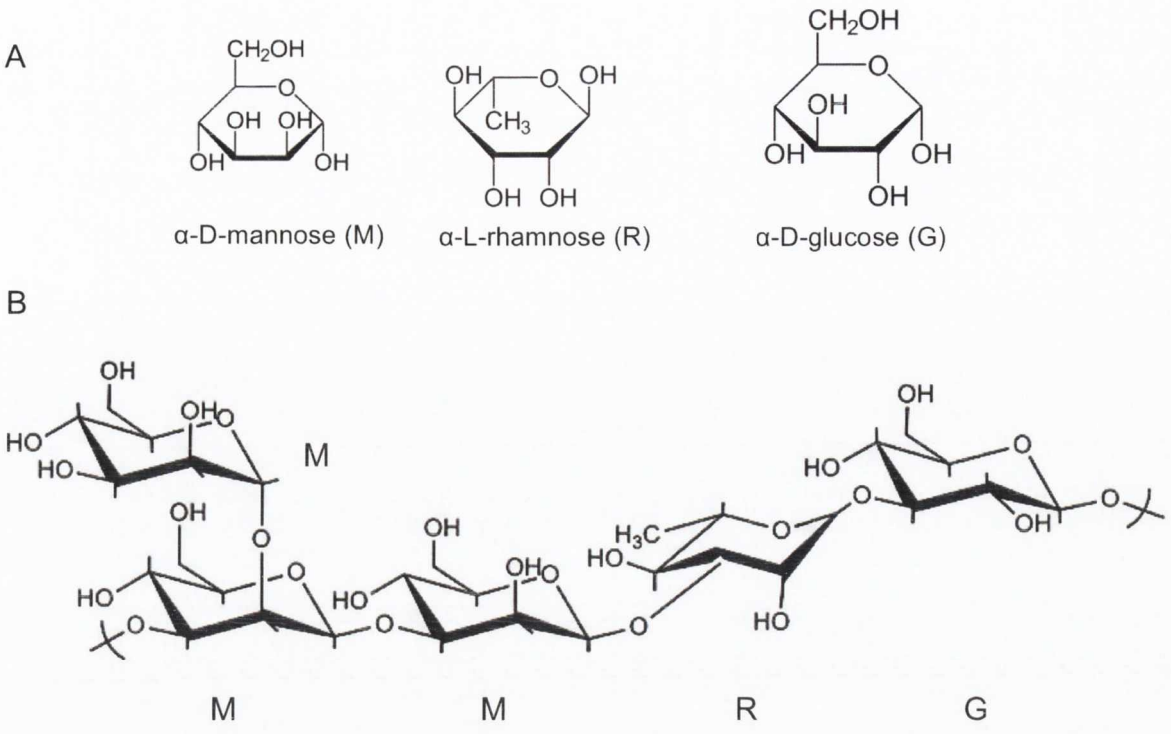


Figure 1.2 Chemical structure of psl. In panel A, structures of each monomeric unit of psl are shown. M, R and G is used to refer to mannose, rhamnose and glucose, respectively. In panel B, the structure for the pentarepeating unit of psl is copied from literature (Franklin *et al.* 2011).

Each of these polysaccharides plays an important role in the infection process of this bacterial and according to the function they can be divided into two groups: the capsular polysaccharides that provides coat and protection to the cells (alginate) and the aggregative polysaccharides that contribute to the structural integrity of biofilm and interact with other biofilm substances (psl and pel) (Mann and Wozniak 2011).

Although not required for *P. aeruginosa* biofilm development, alginate is important for the formation of three-dimensional biofilms (David *et al.* 2001) and causes mucoid phenotype. Alginate is able to absorb large quantities of water to keep moisture around cells. Some studies show that alginate protects *P. aeruginosa* from being damaged or killed by the consequences of inflammation at infection sites as it not only removes free radicals released by infected tissue (Simpson 1989), but also provides a physical and chemical barrier to the bacterium (Govan and Deretic 1996, Pier 1998, Leid *et al.* 2005). In addition, some evidence suggests that alginate could play a supplementary role in the pathogenic biofilm

formation that are related to *Pseudomonas* infection, as mucoid *P. fluorescence* strains adhere to plant root tighter than nonmucoid strains (Bianciotto *et al.* 2001).

1.2.2 *algD* operon

In *P. aeruginosa*, an 18 kb operon named *algD* operon is responsible for alginate synthesis, polymerization, modification and later secretion. *algD* operon encodes twelve genes altogether, which is shown in Figure 1.3. A cartoon representation of how these proteins might work together is shown in Figure 1.4.

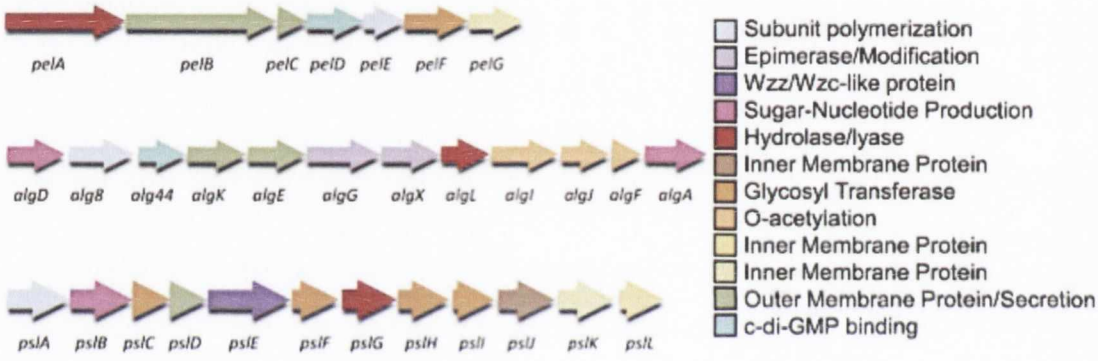


Figure 1.3 The protein members of pel, algD and psl operons. The proteins are color coded to match the functions listed on the right-hand side. Copied from (Franklin *et al.* 2011).

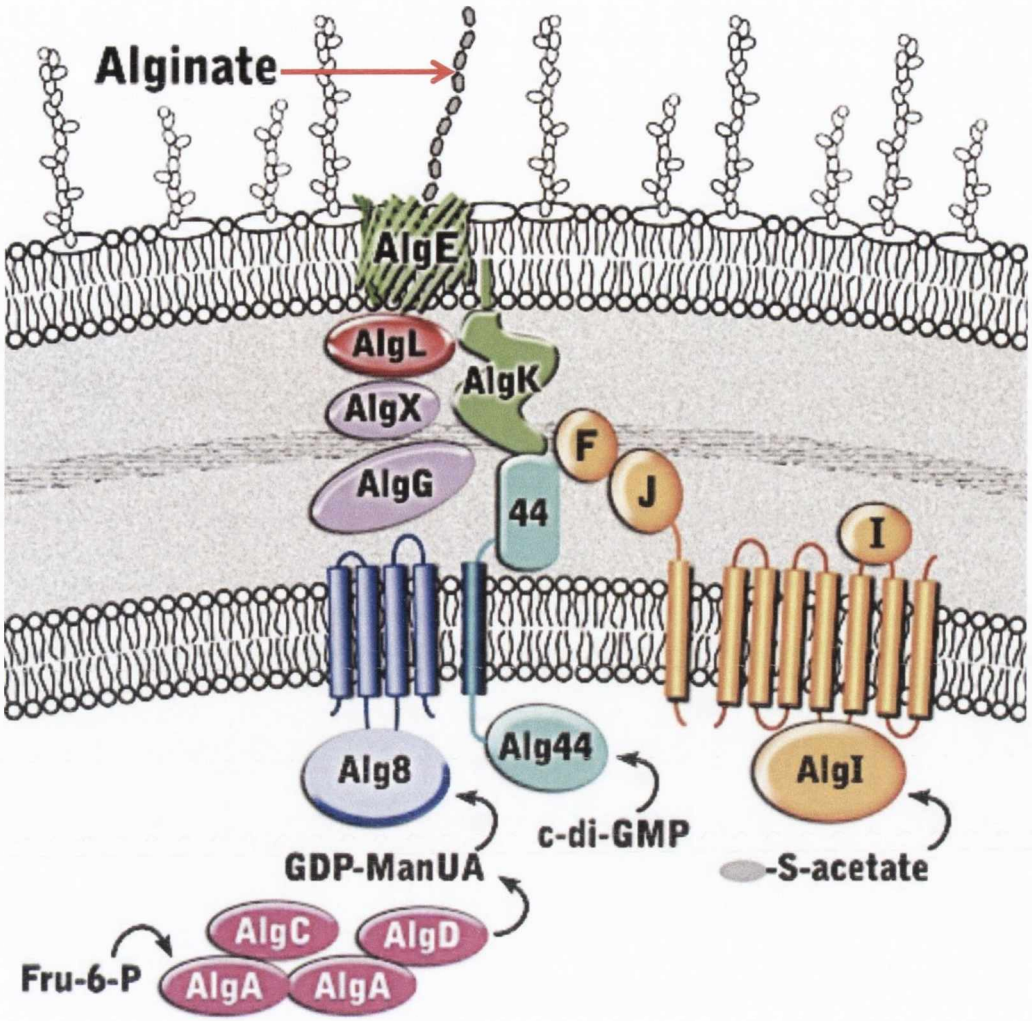


Figure 1.4 Cartoon model for *P. aeruginosa* alginate synthesis and transportation machinery. Copied from (Franklin *et al.* 2011).

AlgA and AlgD, together with AlgC, which is not included in the operon, are responsible for the synthesis of alginate precursor, GDP-mannuronic acid, in the cytoplasm, from fructose-6-phosphate (Remminghorst and Rehm 2006). Although the biosynthesis of GDP-mannuronic acid has been well characterized (Figure 1.5), very little is known about the later assembly and transportation of the full-length alginate polymer.

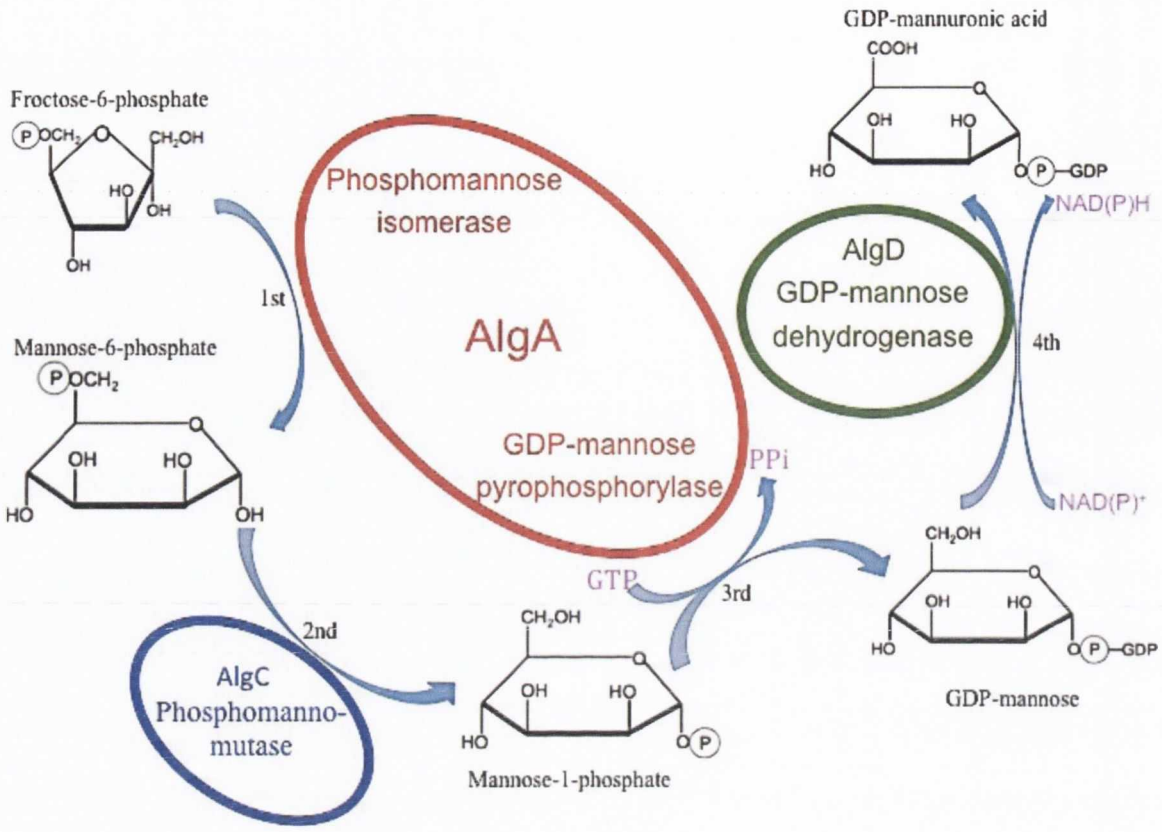


Figure 1.5 GDP-mannuronic acid synthesis pathway (Pindar and Bucke 1975). In the first step, fructose-6-phosphate is converted to mannose-6-phosphate by the phosphomannose isomerase function of AlgA. Then mannose-6-phosphate is converted to mannose-1-phosphate by phosphomannomutase, AlgC. After that, mannose-1-phosphate is transferred to GDP-mannose by the GDP-mannose pyrophosphorylase function of AlgA. In the final step, GDP-mannose is converted to GDP-mannuronic acid by GDP-mannose dehydrogenase, AlgD.

The two inner membrane proteins, Alg8 and Alg44 are required for alginate polymerization (Remminghorst and Rehm 2006, 2006). Alg8 is predicted to encode a glycosyltransferase (Remminghorst and Rehm 2006, Oglesby *et al.* 2008) and the periplasmic domain of AlgE44 shows homology to MexA, one periplasmic component of the multidrug transporter complex MexAB-OprM (Merighi *et al.* 2007). However, little is known about how Alg8 and Alg44 work together to fulfill the function of polymerization.

When polymannuronate is passed into the periplasm, random acetylation is activated by an unknown mechanism through the combined activities of AlgI, AlgJ and AlgF (Franklin and Ohman 2002) and selectively epimerized by the mannuronan C-5 epimerase AlgG to guluronate (Franklin *et al.* 1994).

In the complex machinery, AlgL is particularly interesting by having two functions involving both the degradation of alginate within the periplasm (Bakkevig *et al.* 2005) and forming a scaffold, together with AlgG, AlgK and AlgX, to pass alginate chain from inner membrane to outer membrane (Jain and Ohman 2005). The precise function of AlgX remains a mystery although it is known that the knockout strain of AlgX shows nonmucoid phenotype (Bakkevig *et al.* 2005, Jain and Ohman 2005). AlgK is another protein that people have little knowledge about, however, in recent studies, it is believed to interact with AlgE and direct it onto outer membrane. It also helps to stabilize the whole complex and make transportation of alginate possible (Jain and Ohman 1998, Keiski *et al.* 2010). AlgE is the only outer membrane protein in the machinery and reported to be responsible for alginate secretion (Rehm *et al.* 1994).

1.2.3 AlgE

AlgE is the only member located on the outer membrane fraction of all of the proteins encoded by *algD* operon.

After first being found to be highly expressed in the outer membrane of mucoid strains but not in nonmucoid strains (Grabert *et al.* 1990), only the following few studies have been carried out on this important protein. One study has confirmed its localization to outer membrane and the ability to form ion channel in lipid bilayers (Rehm *et al.* 1994). It was predicted and now has been confirmed by structure determination to form 18 membrane-spanning β-strands (Hay *et al.* 2010) as in Figure 1.6. It has a 32-amino-acids long signal peptide on the N-term, which is cleaved in cells to form the mature protein. Mutation work on AlgE gene shows that the longest loop, L7, of AlgE protein is very important in the function of alginate transport. An 18-amino-acid deletion on L7 generates a mutant that shows no transport of alginate and synthesized alginate is degraded by AlgL into free mannuronic acid and guluronic acid (Hay *et al.* 2010). In the meantime, this deletion-mutant

also affects the localization of AlgE onto the outer membrane for unknown reason. AlgE protein also has an especially long periplasmic turn, T8. T8 contains 19 residues and is proposed to interact with AlgK as part of the alginate transportation mechanism (Whitney *et al.* 2011).

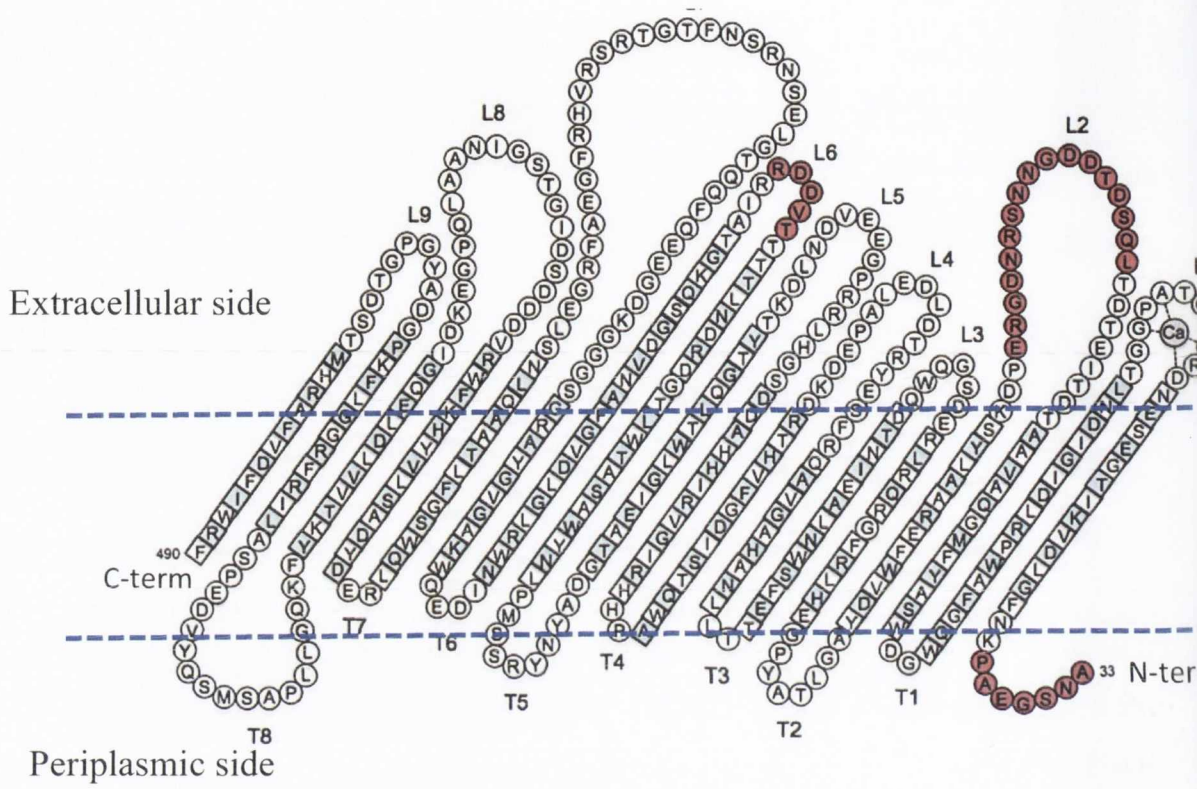


Figure 1.6 Topological model of mature AlgE based on structure determination. N-term and C-term of the protein are labelled at the ends of the sequence. Extracellular side and periplasmic side are also shown in the figure. Membrane boundaries are indicated by dotted blue line and calculated by the OPM server based on hydrophobic thicknesses: (http://opm.phar.umich.edu/server.php). The β-barrel protein consists of 18 antiparallel transmembrane β-strands, nine extracellular loops (L1 to L9), and eight periplasmic turns (T1 to T8). The red residues in the figure are disordered and can't be modelled in the *in surfo* AlgE structure (PDB ID: 3RBH). A calcium binding site is present on L1 with calcium ion co-ordinating with Asp 55, Asp57, Ala61 and Gly63. Redrawn from (Whitney *et al.* 2011).

As stated above, the first 32 residues of AlgE are cleaved off when expressed in both *P. aeruginosa* and *E. coli* and, generating a mature protein with 458 amino acid residues. The mature protein has a molecular weight of 51 kD and calculated pI of 4.6. The calculation of molecular weight and pI is done by online programme: http://web.expasy.org/compute_pi/.

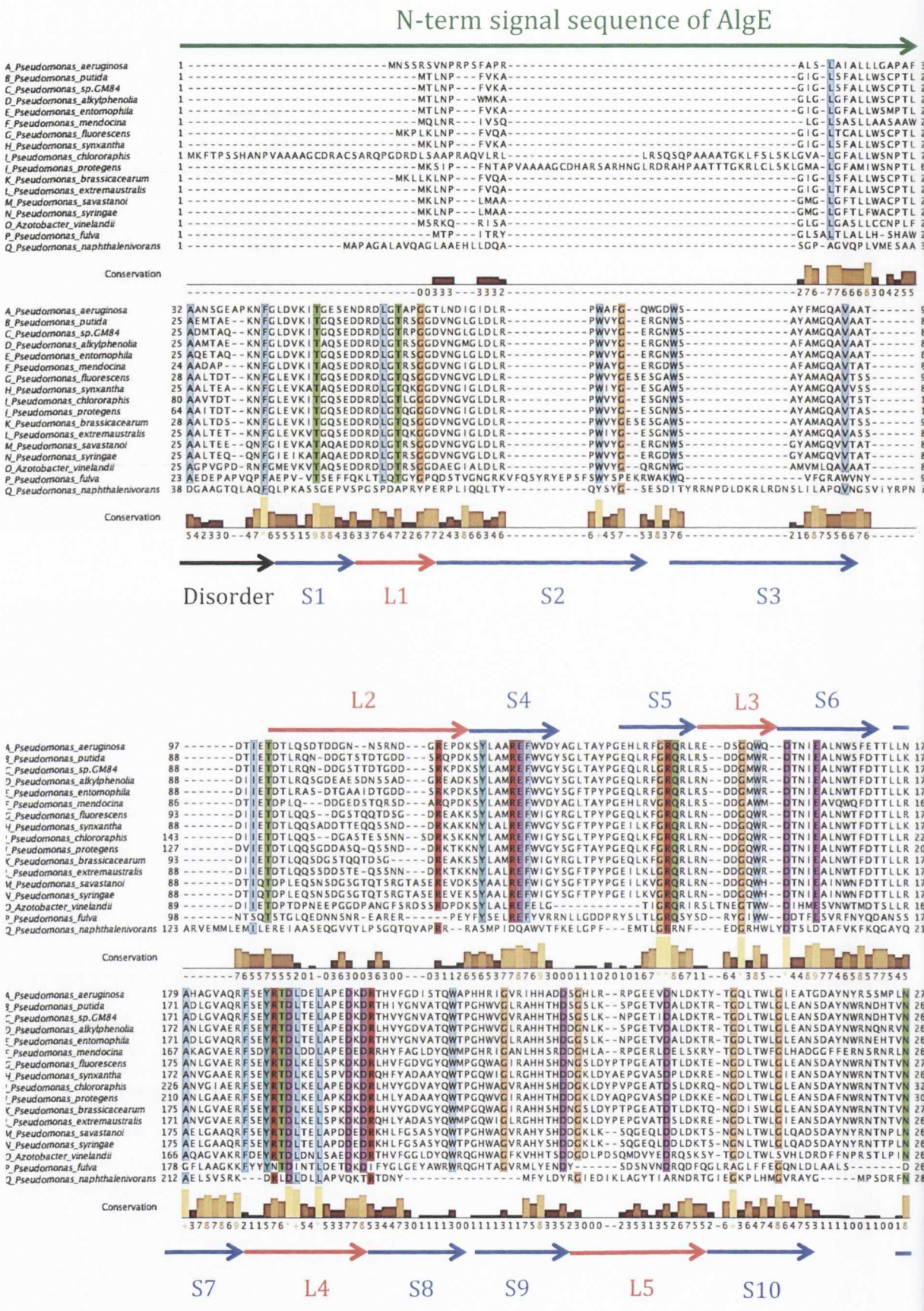
The amino acid composition of mature AlgE is listed in Table 1.1. In mature AlgE, there are 15 Trp and 15 Tyr residues but no Cys residue. Because the extinction coefficient of Trp and Tyr at 280 nm is 5,500 $M^{-1}cm^{-1}$ and 1,490 $M^{-1}cm^{-1}$, respectively (Pace *et al.* 1995), the extinction coefficient of AlgE protein at 280 nm is around 105,150 $M^{-1}cm^{-1}$. This value can be used to calculate the protein concentration at various stages. There are 4 Met residues in the AlgE sequence, which makes it 1 Met residue per 114 residues. This ratio is slightly lower than the minimum recommended level (1 Met in 100 residues) for using multi-wavelength anomalous dispersion (MAD) to obtain phase information for structure determination (Hendrickson and Ogata 1997) but still proved to be enough in the case of AlgE by recent study (Whitney *et al.* 2011). There is no Cys residue in AlgE sequence, which means site-directed mutagenesis is needed if one wants to solve the AlgE structure by taking advantage of the interaction between mercury derivatives and the sulphur on Cys residue.

Table 1.1 Amino Acid composition of mature AlgE. The sequence information is from online database: http://www.pseudomonas.com/.

Amino Acid			Number / percent % in AlgE sequence
Full name	Three Letter code	Single letter code	
Alanine	Ala	A	37 / 8.08
Arginine	Arg	R	33 / 7.21
Asparagine	Asn	N	20 / 4.37
Aspartic acid	Asp	D	46 / 10.04
Cysteine	Cys	C	0 / 0
Glutamine	Gln	Q	20 / 4.37
Glutamic acid	Glu	E	26 / 5.68
Glycine	Gly	G	51 / 11.14
Histidine	His	H	11 / 2.40
Isoleucine	Ile	I	16 / 3.49
Lysine	Lys	K	11 / 2.40
Leucine	Leu	L	36 / 7.86
Methionine	Met	M	4 / 0.87
Phenylalanine	Phe	F	21 / 4.59
Proline	Pro	P	14 / 3.06
Serine	Ser	S	28 / 6.11
Threonine	Thr	T	34 / 7.42
Tryptophan	Trp	W	15 / 3.28
Tyrosine	Tyr	Y	15 / 3.28
Valine	Val	V	18 / 3.93

To find out conserved regions or residues in AlgE protein, sequence alignment among 17 AlgE-like proteins, which are found by online tool BlastP (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi?PAGE=Proteins) to have highest sequence identity to AlgE from *P. aeruginosa*, is done using online tool: http://www.ebi.ac.uk/Tools/msa/muscle/. The result is shown in Figure 1.7, with conserva-

tion scores of each residue listed below the alignment result. Among *Pseudomonas* species, the highest identity comes from *P. putida*, which is 74 % and the lowest identity comes from *P. fulva*, which is 28 %. By looking at the conservation scores, it is clear that, in general, residues in the transmembrane segment are more conserved than those on the loops. In addition, L3, L4 and L7 are much more conserved than other loops and they all contain quite a few charged residues, such as Arg, Lys, Asp, Asn, Glu and Gln. Some of these residues are related to the function of AlgE (Whitney *et al.* 2011) and will be discussed in Chapter 4.


N-term signal sequence of AlgE
Conservation
Disorder
S1
L1
S2
S3
L2
S4
S5
L3
S6
S7
L4
S8
S9
L5
S10

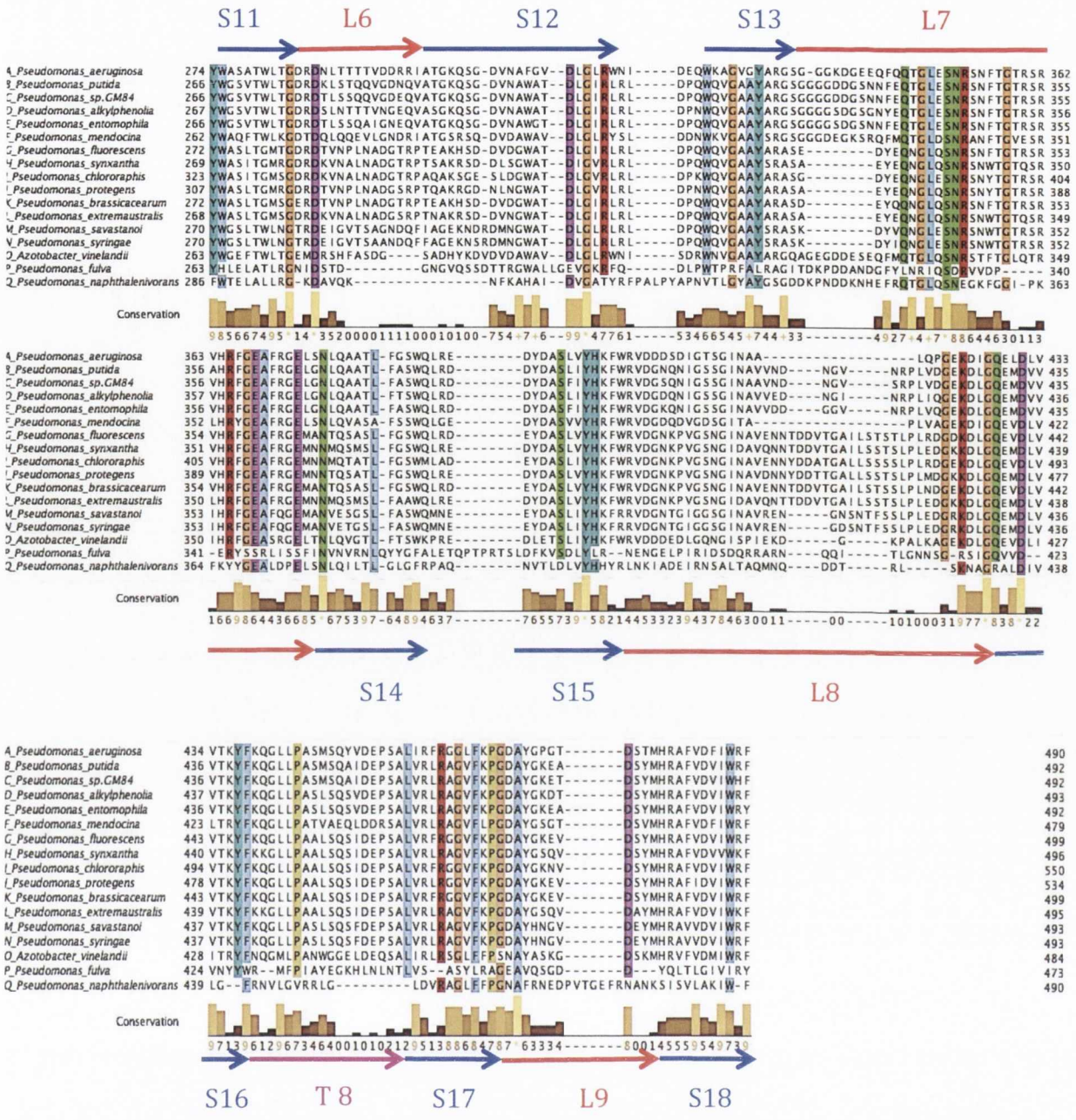


Figure 1.7 Sequence alignments among 17 AlgE or AlgE-like proteins. Conservation score is listed below each fraction of alignment. The bigger the number, the more conserved the residue. Residue numbers are listed on both sides of the alignment. N-term signal peptide of AlgE is labelled in green and indicated by arrow. S1 – S18 represent the transmembrane segments number 1 to number 18 and are labelled in blue with blue arrows. L1 – L9 represent loops number 1 to number 9 and are labelled in red with red arrow. Disordered is region is labelled in black with black arrow. The especially long turn T8 is labelled in purple with purple arrow. The dashed line represents gaps among different sequences.

1.3 Project overview

The main aim of this project is to gain more knowledge of *P. aeruginosa* alginate synthesis and transportation system. A structure of outer membrane transporter of alginate, AlgE, might contribute to a complete understanding of how alginate is secreted to the cell surface. By analysing a determined structure, it is possible that important residues or domains for the function of this protein can be identified and this information can possibly be used for drug design, which might eventually inhibit the transportation of alginate and help reduce the negative effects of *P. aeruginosa* biofilm on patients who suffer from diseases caused by this bacterium.

The plan was to clone AlgE gene into *E. coli* host cells and over-express the protein. Then AlgE protein was expected to be purified by immobilised metal affinity chromatography (IMAC) and size exclusion chromatography (SEC) before setting up crystallisation trials to find out optimal conditions for diffraction quality protein crystals. Heavy atom soaking was planned to generate phase information for structure determination.

However, after I cloned the gene, expressed and purified the protein, got native X-ray diffraction data set to 1.9 Å, which yielded some preliminary phasing information, a structure of AlgE was solved by Professor Lynne Howel's group in Hospital of Sick Children, Toronto and deposited into PDB as 3RBH (Whitney *et al.* 2011). This structure was obtained using a different method and will be described in details and compared with other structures in Chapter 4. As a result, the focus of this project was changed to solve the structure by molecular replacement and analyse the differences between structures or reveal more information about the transportation mechanism.

Another aim of this project is to use AlgE as a benchmark to test the effect of novel lipids on the *in meso* crystallisation process. The plan was to crystallise AlgE protein in different monoacylglycerol lipids (MAG) and then test the diffraction quality of those crystals, in order to assess which lipid would suit β-barrel protein *in meso* crystallisation better. Details of all the work and results will be explained in the following chapters.

Chapter 2

Cloning, expression, purification, identification and characterization of AlgE

2.1 Chapter overview

In this chapter, the method used to produce AlgE protein for this project will be described. Most of the methods are from published protocol but with changes, in order to make the protein purer and suitable for crystallisation work. Characterization work is also described at the end of the chapter in order to evaluate and control the quality of AlgE protein that will be used for crystallisation work in chapter 3.

2.2 Materials and instruments

All materials and instruments used for the work in this chapter are listed below.

Table 2.1 A list of materials and chemicals that have been used in this project.

Materials	Company	Catalogue number	Lot / Batch number
Acetone	Fisher scientific (MA, USA)	A/0606/17	959939
Agar	Merck (NJ, USA)	1.01613	VM135213
Agarose	Sigma (MO, USA)	A5304	108K1012
Ampicillin	Sigma (MO, USA)	A-9393	015K0521
Bis-Tris	Fluka (MO, USA)	14880	1240589
Buffer A (50 mM Tris-HCl, pH 7.2[a])	N/A	N/A	N/A
Buffer B (150 mM KCl, 50 mM Tris-HCl, pH 7.2[a])	N/A	N/A	N/A

Table 2.1 continued

Buffer C (50 mM Tris-HCl, pH 7.2[a], 0.1 %(w/v) LDAO)	N/A	N/A	N/A
Buffer D (50 mM Tris-HCl, pH 7.2[a], 0.1 %(w/v) LDAO, 300 mM NH_4Cl)	N/A	N/A	N/A
Buffer E (50 mM Tris-HCl, pH 7.2[a], 0.1 %(w/v)LDAO, 150 mM NaCl)	N/A	N/A	N/A
Centricon 50 kDa MWCO	Millipore (MA, USA)	UFC905024	R0KA67379
DNA marker (1 kbp): (0.5, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 6.0, 8.0 and 10 kbp)	NEB (MA, USA)	N3232L	710807
Dnase I	Sigma (MO, USA)	ND25	056K7680
dNTPs mixture	Greiner Bio One (Frickenhausen, Germany)	201900	N/A
E. coli BL21(DE3) Star competent cells	Invitrogen (Dun Laoghaire, Ireland)	44-0049	778271
E. coli DH5á competent cells	NEB (MA, USA)	C2977H	1211103
E. coli Top10 competent cells	Invitrogen (Dun Laoghaire, Ireland)	C404003	825023
1.5 mL Eppendorf tube	Sarstedt (Nümbrecht, Germany)	72.690.001	0000/0355001
Ethanol	Sigma (MO, USA)	E7023	07850LE
15 mL Falcon tubes	Sarstedt (Nümbrecht, Germany)	62.559.001	1040601
Gelcode Blue Safe (protein stain)	Thermo scientific (MA, USA)	24594	LG147135
Glycerol	Melford (Suffolk, UK)	G1345	56-81-5
HotStar Hifidelity DNA polymerase	Qiagen (Hamburg, Germany)	202602	1032734
IPTG	Melford (Suffolk, UK)	MB1008	A21345
IMAC Sepharose High Performance Matrix	GE Healthcare (Buckinghamshire , UK)	17-0920-06	10029650
Kanamycin	Melford (Suffolk, UK)	K0126	18022
KCl	Sigma-Aldrich (MO, USA)	P-4286	074K0024
LB High Salt Powder	Fluka (MO, USA)	51208	BCBC5118
LDAO	Sigma (MO, USA)	40236	BCBB8394
Lysozyme	Sigma (MO, USA)	L-6876	094K1454
MES	Sigma (MO, USA)	M3671	068K5421

Table 2.1 continued

MES 20X SDS-PAGE running buffer (1 M MES, 1 M Tris base, 2 %(w/v) SDS, 20 mM EDTA, pH 7.3)	Invitrogen (Dun Laoghaire, Ireland)	NP0002	871527
Methanol	Fisher scientific (MA, USA)	UN1230	1023754
MilliQ water (Resistance at 18.2 mΩ)	Produced by MilliQ Elix UV5 System	N/A	N/A
$MgCl_2$	Sigma (MO, USA)	M2670	129K00891
NaCl	Fisher scientific (MA, USA)	BP358-10	80786
NaOH	Merck (NJ, USA)	8627	B382182 411
NH_4Cl	Sigma (MO, USA)	3101	SZB82820
OG	Anatrace (CA, USA)	O311S	126780
PCR tubes	Sarstedt (Nümbrecht, Germany)	72.737.002	52501
35 mm Petri dish	Greiner Bio One (Frickenhausen, Germany)	627161	N/A
Primers and sequencing	MWG Biotech (Hamburg, Germany)	N/A	N/A
Protein marker for SDS-PAGE (ovotransferrin 76 -78 kD, bovine serum albumin 66.2 kD, ovalbumin 45 kD, carbonic anhydrase 30 kD, myoglobin 17.2 kD, Cytochrome c 12.3 kD)	BDH (MA, USA)	97063-716	104170
Pseudomonas aeruginosa PAO1 Holloway strain	Gift from Fergal O'Gara, University of College Cork	N/A	N/A
QiaPrep Spin mini prep kit	Qiagen (Hamburg, Germany)	27106	136231546
Q-Solution 10X	Qiagen (Hamburg, Germany)	1005485	127127178
SDS	Fluka (MO, USA)	71725	1344034
SDS-PAGE Bis-Tris gels	Thermo scientific (MA, USA)	84712	100407117
SOC medium	NEB (MA, USA)	B9020S	1211007

Table 2.1 continued

Tris-Glycine SDS sample buffer 2X (63 mM Tris-HCl, pH 6.8, 2 %(w/v) SDS, 20 %(w/v) glycerol, 0.002 %(w/v) bromophenol blue)	Invitrogen (Dun Laoghaire, Ireland)	LC2676	486397
Sybersafe	Invitrogen (Dun Laoghaire, Ireland)	S33102	643459
Trizma® base	Sigma (MO, USA)	T87602	088K0118
TOPO cloning kit	Invitrogen (Dun Laoghaire, Ireland)	450172	801495

a. All pHs only refer to the adjusted value of buffering stock solution before mixing with other chemicals in the solution

Table 2.2 A list of all equipments that have been used in the project.

Equipment	Company	Type
4 °C FPLC cold cabinet	Medical Supply (Ireland)	Ekofrigolab 1500
4 °C room	Aircold manufacturing (Ireland)	MasterCella - CAREI
42 °C water bath	Fisher scientific (MA, USA)	DMU5
AKTA Explorer	GE Healthcare (Buckinghamshire , UK)	AKTA Explorer
AKTA FPLC	GE Healthcare (Buckinghamshire , UK)	AKTA FPLC
Autoclave	Rodwell (Essex, UK)	Crystal MP 25
Balance (0.1 mg – 64 g)	Sartorius (Goettingen, Germany)	CP64
Balance (0 – 3,100 g)	Mettler Toledo (OH, USA)	B300T-S
Balance (0.01 mg – 220 g)	Mettler Toledo (OH, USA)	AX205
Benchtop minicentrifugre (0 – 16,300 g)	Spectrafuge (TX, USA)	24D
Biosafety cabinet	Labculture (Dorset, UK)	ESCO Class II Type A2
Bottle	Duran Group (Main, Germany)	50 mL, 100 mL, 200 mL, 1 L, 2 L
Cell Disruptor		
Centrifuge	Beckman Coulter (Clare, Ireland)	Allegra X-22R
Centrifuge rotor (0 – 10,000 rpm)	Beckman Coulter (Clare, Ireland)	C0650

Table 2.2 continued

Centrifuge	Beckman Coulter (Clare, Ireland)	Avanti J-E
Centrifuge rotor (0 – 9,000 rpm)	Beckman Coulter (Clare, Ireland)	JLA 9.1000
Chemical fume hood	Cameron (Longford, Ireland)	MACH-AIRE
Culture flasks	Schott Duran (Wertheim, Germany)	Nalgen baffled shake flask: 250 mL, 3.5 L, 5 L
Electrophoresis chamber	Bio-Rad (Hertfordshire, UK)	XCell SureLock
Electrophoresis power supply	Biorad (Wicklow, Ireland)	Powerpac basic
Freezer (- 20 °C)	Liebherr (PA, USA)	GS 1301
Freezer (- 80 °C)	Medical Supply (Ireland)	Platinum 500
French press	Fisher scientific (MA, USA)	FA-078A
Gel doc system	Alpha Innotech (CA, USA)	Fluorchem SP Multiimage II
Hiload 16/60 S200 SEC column	GE Healthcare (Buckinghamshire, UK)	17-1069-01 (lot. 433047)
Homogenizer (probe)	Griffin George (Leicestershire, UK)	LC9
Homogenizer (tube)	Wheaton (NJ, USA)	30 mL tube
HPLC fluorescence reader	Dionex (Dubin, Ireland)	RF-2000
Icemaker	Scotsman (Cumbria, UK)	AF 80
Incubator (Room temperature to 60 °C)	Labnet (New Jersey, US)	Mini incubator
Incubator (0 °C to 60 °C)	Rumed (Laatzen, Germany)	1001
Inoculation loop	Greiner Bio One (Frickenhausen, Germany)	1 µL
Magnetic stirrer	Stuart (Staffordshire, UK)	SB 162
Progad filter for MilliQ Elix UV5 system	Millipore (MA, USA)	PR0G00052
Simpack filter for MilliQ Elix 5 system	Millipore (MA, USA)	SIMPAK0RT
Nanodrop (spectrophotometer)	Mason Technology (Dublin, Ireland)	Nanodrop 1000 Spectrophotometer
PCR machine	Eppendorf (Cambridge, UK)	Mastercycler gradient
pH meter	Thermo scientific (MA, USA)	Orion 3 star
Pipettes	Gilson (WI, USA)	Pipetman kit (2 to 1000 µL pipettes)
Probe sonicator	Bandelin (CT, USA)	Sonoplus HD 200
Refrigerator (4 °C)	Microban (Staffordshire, UK)	Hotpoint RLA81

Table 2.2 continued

Nanodrop (spectrophotometer)	Mason Technology (Dublin, Ireland)	Nanodrop 1000 Spec trophotometer
PCR machine	Eppendorf (Cambridge, UK)	Mastercycler gradien
pH meter	Thermo scientific (MA, USA)	Orion 3 star
Pipettes	Gilson (WI, USA)	Pipetman kit (2 to 100 µL pipettes)
Probe sonicator	Bandelin (CT, USA)	Sonoplus HD 200
Refrigerator (4 °C)	Microban (Staffordshire, UK)	Hotpoint RLA81
SDS-PAGE apparatus	Invitrogen (Dun Laoghaire, Ireland)	Novex Mini-Cell
Shaking incubator (Room temperature to 80 °C)	AGB Scientific Ltd (Dublin, Ireland)	Sanyo Orbital Incubator
Spectrophotometer	Uvikon (Leeds, UK)	XL
Ultracentrifuge	Beckman Coulter (Clare, Ireland)	Optim L-100 XP
Ultracentrifuge rotor (0 – 70,000 rpm)	Beckman Coulter (Clare, Ireland)	70 Ti
Vortex mixer	Nichel Electro Ltd (North Somerset, England)	Clifton Cyclone CM-

2.3 Methods

All method sections in this thesis are written in protocol style for the convenience to the person who wants to repeat or continue working on this protein.

2.3.1 Primer design

- Design a pair of primers according to the AlgE sequence (Figure 2.1) in order to clone the gene into *E. coli* cells. The forward primer starts with a *CACC* overhang to facilitate amplified DNA fragment's ligation with pET200/D/TOPO vector from TOPO cloning kit and includes start codon (*ATG*): 5' ***CACCA TG***AAC AGCTC CCGTT CCG 3'. Reverse complementary nucleotide (*TCA*) to stop codon (*TGA*) is included in reverse primer: 5' ***TCA***GA AGCGC CAGAT GAAGT 3'. Thus the primers are determined to be suitable for use in the cloning of the full-length gene.

ATGAACAGCTCCCGTTCCGTCAACCCGCGGCCGTCCTTCGCGCCGCGCGCCCTGTCCC
TGGCCATCGCCCTGCTGCTCGGCGCGCCGGCGTTCGCCGCCAACAGCGGCGAGGCGC
CGAAGAACTTCGGCCTGGACGTGAAGATCACCGGCGAATCGGAAAACGATCGCGAC
CTCGGCACCGCTCCCGGCGGCACCCTCAACGACATCGGTATCGACCTGCGGCCCTGG
GCCTTCGGCCAGTGGGGCGACTGGAGCGCCTACTTCATGGGCCAGGCGGTGGCCGCC
ACCGACACCATCGAGACCGACACCCTGCAATCGGACACCGACGACGGCAACAACAGC
CGCAACGACGGTCGCGAGCCGGACAAGAGCTACCTCGCCGCACGCGAATTCTGGGTC
GACTACGCCGGCCTCACCGCCTACCCCGGCGAGCACCTGCGCTTCGGCCGCCAGCGC
CTGCGGGAAGACAGCGGCCAGTGGCAGGACACCAACATCGAGGCGCTGAACTGGAG
CTTCGAGACCACCCTGCTCAACGCCCACGCCGGGGTCGCCCAGCGTTTCAGCGAATA
CCGCACCGACCTCGACGAACTGGCTCCGGAGGACAAGGACCGCACCCATGTGTTCGG
CGACATTTCCACCCAGTGGGCGCCGCACCACCGCATAGGCGTGCGCATCCACCACGC
CGACGACAGCGGCCACCTGCGCCGCCCCGGCGAGGAAGTCGACAACCTCGACAAGA
CCTATACCGGCCAGCTCACCTGGCTCGGCATCGAGGCTACCGGCGATGCCTACAACT
ATCGTTCGAGCATGCCGCTGAACTACTGGGCCAGCGCCACCTGGCTGACCGGCGACC
GCGACAACCTGACCACCACCACGGTCGACGACCGGCGCATCGCCACCGGCAAGCAG
AGCGGCGACGTCAATGCCTTCGGCGTCGACCTCGGCCTGCGCTGGAACATCGACGAG
CAATGGAAGGCCGGCGTCGGCTACGCCCGCGGCAGCGGCGGTGGCAAGGACGGCGA
GGAGCAGTTCCAGCAGACCGGGCTGGAGAGCAATCGCTCCAACTTCACCGGCACCCG
CTCGCGCGTGCACCGCTTCGGCGAAGCCTTCCGCGGCGAACTGAGCAACCTCCAGGC
AGCCACCCTGTTCGGCTCCTGGCAACTGCGCGAGGACTACGACGCGAGCCTGGTCTA
CCACAAGTTCTGGCGCGTCGACGACGACTCCGACATCGGTACCAGCGGCATCAACGC
CGCCCTGCAACCGGGCGAGAAGGACATCGGCCAGGAACTCGACCTGGTGGTGACCAA
GTACTTCAAGCAAGGCCTGCTGCCGGCCTCGATGAGCCAGTACGTCGACGAGCCCTCG
GCGCTGATCCGCTTCCGCGGCGGCCTGTTCAAGCCGGGCGACGCCTACGGGCCGGGCA
CCGACTCGACCATGCACCGCGCCTTCGTCGACTTCATCTGGCGCTTCTGA

Figure 2.1 Full-length DNA sequence of AlgE gene with signal sequence from *P. aeruginosa* PAO1 Holloway strain (Winsor *et al.* 2009).

- The primers are then ordered and synthesized by MWG Biotech Ltd (Anzingerstr. 7a, 85560 Ebersberg, Germany). The cost is usually 12 Euro for 25 nucleotides.

2.3.2 Media preparation and cell culture

- Prepare 100 mL Luria Bertani (LB) agar media by adding 2.5 g LB High Salt powder and 1.5 g agar into 100 mL MilliQ water. Prepare 100 mL LB liquid media by adding 2.5 g LB High Salt powder into 100 mL MilliQ water. Autoclave the media at 121 °C for 15 min and pour the LB agar media into Petri dishes.

- Remove *P. aeruginosa* PAO1 cells from the – 80 °C freezer. Use an inoculation loop to pick up a loop of cells and spread onto the LB agar plate. Store the LB plate at 37 °C for 18 h and then transfer into the 4 °C refrigerator.

2.3.3 PCR amplification

- Autoclave to sterilize MilliQ water, pipette tips, Eppendorf tubes and PCR tubes at 121 °C for 15 min.

- Prepare 50 μL PCR mixture in 200 μL sterilized PCR tubes. The components in the mixture are listed in Table 2.3.

Table 2.3 A list of all components in 50 μL PCR mixture.

Contents	Quantity in 50 μL reaction system
Proofstart DNA polymerase	0.5 μL
10 X Proofstart PCR buffer	5 μL
dNTPs mixture (10 mM)	1.5 μL
Forward primer (10 μM)	5 μL
Reverse primer (10 μM)	5 μL
Sterilized MilliQ water	Fill up to 50 μL

- Use sterilized pipette tips to pick up *P. aeruginosa* PAO1 cells from Section 2.3.2 and transfer the cells into 50 μL PCR mixture. Stir the tips for 10 s to make sure cells are

dispensed into the mixture. Discard the tips. Prepare a negative control for PCR reaction by adding the PCR mixture into a sterilized PCR tube without transferring cells into it.

- Set up gradient PCR amplification. The running cycle is from standard protocol (HotStar HiFidelity PCR Handbook, Qiagen) but with a temperature gradient in the annealing step.

  Initial activation: 94 °C, 5 min

  Denaturing: 94 °C, 1 min

  Annealing: 1 min for each temperature: 58.1, 59.3, 60.0, 61.3, 62.1 and 63.1 °C

  Extension: 72 °C, 1.5 min

  Repeat from denaturing to extension for 30 cycles.

  End of PCR: 4 °C, hold.

2.3.4 PCR product purification

The following materials are provided by Qiagen Gel Extraction Kit (Qiagen):
50 mM Tris-HCl pH 8.0, 10 mM EDTA (P1 solution), 75 %(w/v) Ethanol, 25 mM NaCl, 5 mM Tris-HCl, pH 7.5 (PE solution), 10 mM Tris-HCl, pH 8.5 (EB buffer) and DNA-binding columns.

- Load PCR product onto 1.0 %(w/v) agarose gel, with NEB 1 kb DNA marker in a separate lane. Run electrophoresis at 100 V for 10 min. Stain agarose gel with SYBR Safe DNA Gel Stain and visualize the bands under UV light. The target band is expected to be close to 1.5 kb.

- Slice the 1.5 kb band from the gel using a clean blade. Measure and record the mass. Transfer the gel slice into 1.5 mL empty Eppendorf tubes and add P1 solution. The number of the volume for P1 solution (in microliters) is approximately 4 times the number of the mass for each sliced gel (in milligrams). Incubate the mixture at 55 °C until gel slice dissolved completely and transfer the mixture into DNA binding columns. Spin the column in benchtop centrifuge (Eppendorf) at 16,300 g and room temperature in the benchtop centrifuge for 1 min. Room temperature, if not specified, is

not measured and not known, with a possible range from 18.0 °C to 23.0 °C. Discard flowthrough and add 750 µL PE solution to the column. Spin the column again in benchtop centrifuge at 16,300 g and room temperature for 1 min. Discard flowthrough and add 20 µL EB buffer to the column. Elute DNA by spinning the column in bench-top centrifuge at 16,300 g and room temperature for 1 min.

- Store the purified product in 4 °C refrigerator. For long-term storage, leave the product in - 20 °C freezer.

2.3.5 TOPO cloning reaction

The following materials are provided in TOPO Cloning Kit (Invitrogen):
1.2 M NaCl, 60 mM $MgCl_2$ (Salt solution) and the pET200/D/TOPO vector.

- Prepare TOPO cloning reaction system as outlined in Table 2.4. Mix and then incubate the mixture in Eppendorf tube at room temperature for 5 min and then store it on ice.

Table 2.4 A list of all components in TOPO Cloning reaction system.

Contents	Volume in 6 µL system
Purified PCR product from Section 2.3.4	0.5 µL
Salt Solution	1 µL
pET200/D/TOPO vector	0.5 µL
Sterilized MilliQ water	Fill up to 6 µL

2.3.6 Transformation into cloning strain

- Remove one vial of *E. coli* Top10 chemically competent cells from the - 80 °C freezer and immediately leave it on ice. Prewarm Super Optimal broth with Catabolic suppressor (SOC) liquid media was pre-warmed in incubator (Labnet) at 37 °C. Set the waterbath to 42 °C.

- Add 1 μL of mixture from Section 2.3.5 to the vial of competent cells and mix by tapping the tube gently for 5 s. Incubate the mixture on ice for 15 min.

- Dip the tube containing the mixture from step 2 into the water bath for 30 s to heat shock the cells. Return tube onto ice immediately. Add 200 μL pre-warmed SOC liquid media into the tube and grow the cells in shaking incubator (Sanyo) at 37 °C and 180 rpm for 1 h.

- Apply the cultured cells onto fresh LB agar plates containing 50 μg / mL kanamycin using inoculation loop. Grow the cells in incubator (Labnet) at 37 °C for 18 h.

2.3.7 Plasmid isolation and sequencing

The following materials are provided in Qiaprep Mini Prep Kit 250 (Qiagen):
50 mM Tris-HCl, pH 8.0, 10 mM EDTA (P1 solution), 1 M NaOH, 10 %(w/v) SDS (P2 solution), 3 M KOAc pH 5.5 (N3 solution), 75 %(w/v) Ethanol, 25 mM NaCl, 5 mM Tris-HCl, pH 7.5 (PE solution), 10 mM Tris-HCl, pH 8.5 (EB buffer) and DNA-binding columns.

- Use sterilised pippete tips to pick single colonies from the plates from section 2.3.6 and then leave them in empty sterilized 15 mL Falcon tubes. Add 3 mL fresh LB liquid media with 50 μg / mL kanamycin to the tubes. Grow the cell cultures in a shaking incubator (Sanyo) at 37 °C and 180 rpm for 18 h.

- Prepare glycerol stocks for each culture by mixing 250 μL 80 %(w/v) glycerol with 750 μL cell culture. Flash freeze glycerol stocks in liquid nitrogen (- 196 °C) and store them in – 80 °C freezer.

- Transfer 1 mL of each cell culture into sterilised 1.5 mL Eppendorf tubes by pippete. Spin down the cells in benchtop centrifuge at 16,300 g and room temperature for 1 min. Discard the supernatant. Resuspend the cells in 250 μL P1 solution with pipette. Add 250 μL P2 solution to the suspension and invert the tube immediately 4 to 6 times. Add 350 μL N3 solution to the suspension and invert the tube immediately 4 to

6 times. Spin the tubes in benchtop centrifuge at 16,300 g and room temperature for 10 min. Transfer the supernatant to the DNA binding column and spin in benchtop centrifuge at 16,300 g and room temperature for 1 min. Discard the flowthrough and add 750 µL PE solution to the DNA binding column. Spin the DNA binding column in benchtop centrifuge at 16,300 g and room temperature. Discard the flowthrough and spin the DNA binding column again in benchtop centrifuge at 16,300 g and room temperature for 1 min to remove the residual PE solution. Put the DNA binding column into a 1.5 mL Eppendorf tube and add 30 µL EB buffer. Incubate the column and buffer for 1 min at room temperature and then spin in benchtop centrifuge at 16,300 g and room temperature for 1 min. Collect the flowthrough.

- Measure the A_{260} of the flowthrough in Nanodrop Spectrometer with MilliQ water as reference. The pathlength of Nanodrop spectrometer is 0.1 cm and the reading is an automatically normalized value to 1 cm given by the software. Because double-stranded DNA has extinction coefficient at 260 nm of 0.02 $(\mu g / mL)^{-1} cm^{-1}$ (Sambrook and Russell 2011), the plasmid concentration can be calculated using equation 1. If the plasmid concentration is lower than 50 ng / µL, concentrate the solution by spinning the tube in benchtop centrifuge at 16,300 g at rom temperature without closing the lid. If the plasmid concentration is higher than 100 ng / µL, dilute the solution with sterilised MilliQ water.
  Equation 1: $[DNA] = A_{260} \times 50$ mg / mL.

- Mail the plasmids to the following address for sequencing:
  Sequencing Department
  Anzingerstr. 7a
  85560 Ebersberg
  Germany
  The sequencing reaction is carried out by the company using both a T7 promoter primer and a T7 terminator primer, both of which are commercially available.

- Compare the sequencing results with the AlgE gene sequence, which is available on the website www.pseudomonas.com, using online alignment tool http://www.ebi.ac.uk/Tools/emboss/align/.

2.3.8 Transformation into the expression strain

- Examine the sequencing result to determine the correct clone. Remove the glycerol stock of the correct clone, from Section 2.3.7, from -80 °C freezer and grow the cells in 3 mL LB liquid media in a shaking incubator at 37 °C and 180 rpm for 18 h.

- Extract the plasmid as described in Section 2.3.7.

- Remove *E. coli* BL21 (DE3) Star chemically competent cells from the - 80 °C freezer and transform isolate plasmid into the cells as described in Section 2.3.6.

2.3.9 IPTG induction test

- Autoclave culture flasks at 121 °C for 15 min.

- Pick a single colony from plate in Section 2.3.8 and inoculate it into 10 mL fresh LB liquid media containing 50 μg / mL kanamycin. Grow the cell culture in shaking incubator at 37 °C and 180 rpm for 18 h.

- Prepare glycerol stocks as described in Section 2.3.7 and save 100 μL cells as negative control for later use. Transfer 0.1 mL of culture from step 2 into glass flasks containing 10 mL fresh LB liquid media. Add kanamycin to 50 at μg / mL. Grow the cells in a shaking incubator at 37 °C and 200 rpm. Measure and monitor the OD_{600} of cell culture by Nanodrop spectrometer. Unused LB medium serves as reference. When the OD_{600} reaches 0.6, which usually takes 2.5 h, add 10 μL 1 M IPTG solution into each of the 10 mL cultures. Divide the cultures equally into two sterilised Falcon tubes, 5 mL each. Grow the first tube of culture in a shaking incubator at 37 °C and 200 rpm for 3 h. Grow the second tube of culture in a shaking incubator at 18 °C and 200 rpm for 18 hours.

- Transfer 50 μL of each final culture and the negative control cells from step 3 into 1.5 mL empty Eppendorf tubes. Spin down the cells in a benchtop centrifuge at 16,300 g and room temperature for 1 min and discard the supernatant. Resuspend the cell pellet

with 20 µL 1 X Tris-Glycine SDS sample buffer (Invitrogen). Boil the resulting samples at 100 °C for 10 min before loading 14 µL of each sample onto 12 %(w/v) Bis-Tris gel. Leave the electrophoresis chamber in ice-water mixture and run electrophoresis at 100 V for 2 h.

- Remove the gel from electrophoresis chamber and rinse it 3 times with MilliQ water. Stain the gel for 1 h with Gelcode Bluesafe Protein Stain and destain it by soaking in MilliQ water for 1 h on shaker (Stuart), at 50 rpm and room temperature. Check the gel in Gel doc system and record picture.

2.3.10 Scale-up cell culture

- Prepare 6 X 500 mL LB liquid media in 3.5 L culture flasks and autoclave them at 121 °C for 15 min. Remove glycerol stock of *E. coli* BL21 (DE3) Star cells containing correct AlgE plasmid from the – 80 °C freezer. Scratch the frozen stock three to four times with sterilized pipette tip and use the tip to inoculate fresh 200 mL LB liquid media. Add kanamycin to 50 µg / mL. Grow the cell culture in shaking incubator at 37 °C and 180 rpm for 18 h.

- Transfer 10 mL of the culture into each 500 mL LB liquid media and add kanamycin 50 µg / mL. Grow the cell culture in shaking incubator at 37 °C and 200 rpm. Monitor OD_{600} with Nanodrop Spectrometer as described in section 2.3.9 until it reached 0.6. Usually this takes 3.5 h.

- Leave the cell culture in the 4 °C room for 30 min before adding 500 µL 1 M IPTG into each flask. Grow the cells in shaking incubator at 18 °C and 180 rpm for 18 h.

- Harvest the cells by spinning the cell culture in centrifuge (Beckman Coulter) at 3,000 g and 4 °C for 20 min. Discard the supernatant and place cell pellet on ice for later use.

2.3.11 AlgE purification flowchart

To purify AlgE protein from expression host *E. coli* cells, a protocol published before (Rehm *et al.* 1994) was modified. A size exclusion chromatography was added to the protocol.

Shown in Figure 2.2 is a flowchart briefly describing the protocol used for AlgE purification and details will be seen in later sections.

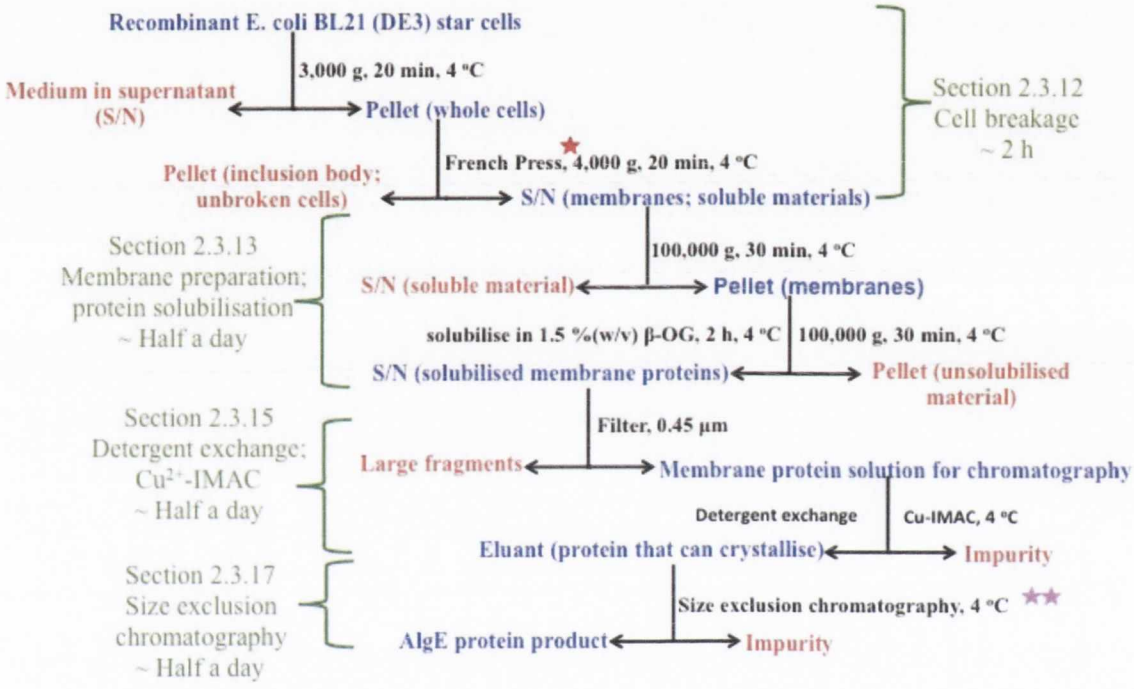


Figure 2.2 A flowchart for AlgE purification procedures. The flowchart covers all the procedures that will be described in details from section 2.3.12 to 2.3.17.

★ **In some preps, sonication or cell disruptor are used because French Press is broken and the preps show no difference in terms of yield and quality.**

★★ **This step is an additional procedure to the published protocol as it produces purer protein.**

2.3.12 Cell breakage

- Resuspend the recombinant cells from 4 L cell culture in section 2.3.10 in 100 mL of 50 mM Tris-HCl pH 7.2 (buffer A). Break the cells by French press or probe sonicator or cell disruptor: 1. Passing cells through the working cell of French Press, 3 times at working pressure of 1,000 psi, at 4 °C; 2. Sonication at 30 % power with 1 min pulse in ice-water mixture for 20 times; 3. Passing cells through the working cell of cell disruptor, 3 times at working pressure of 1,500 psi, at 4 °C. According to the manual of each instrument, 90 % of the cells should have been broken after the listed treatment. But I haven't done any experiment to test the efficiency.

2.3.13 Membrane preparation and protein solubilisation

- Spin the broken cells in centrifuge (Beckman Coulter) at 4,000 g and 4 °C for 20 min. Decant the supernatant into clean ultra-centrifuge tubes and spin the tubes in ultra-centrifuge at 100,000 g and 4 °C for 30 min. Discard the supernatant.

- Add KCl to buffer A to 150 mM and call it buffer B. Discard the supernatant. Vortex to resuspend the pellet in 100 mL buffer B. Add 1.5 g β-OG into the suspension. Stir the final suspension in a 200 mL beaker with magnetic stir bar on magnetic stir plate for 2 h at 4 °C.

2.3.14 Cu^{2+}-IMAC Column equilibration

- Pack 25 mL IMAC Sepharose High Performance Matrix into the column. Charge the matrix with Cu2+ by running 5 mL 0.1 M CuSO4 solution through the column. Control the flow rate at 0.1 mL / min by AKTA FPLC. Wash the matrix with 500 mL MilliQ water at the flowrate of 4.0 mL / min. Equilibrate the matrix with 100 mL buffer 50 mM Tris-HCl pH 7.2, 0.1 %(w/v) LDAO (buffer C) at the flowrate of 4.0 mL / min.
- Note: Cu^{2+}-IMAC is used instead of Ni^{2+}-IMAC because there is no His-tag in the mature form of AlgE protein and AlgE has the natural ability to bind Cu^{2+} ions.

2.3.15 Detergent exchange and Cu^{2+}-IMAC purification

- Spin the resulting solution from Section 2.3.13 in ultra-centrifuge machine at 100,000 g and 4 °C for 30 min. Decant the supernatant into a clean bottle and load onto the column prepared in Section 2.3.14 by AKTA FPLC. Control the flow rate at 4.0 mL / min.

- Wash the column with buffer C and monitor the A_{280}. When A_{280} becomes constant, run buffer C through the column at the flow rate of 0.1 mL / min for 18 h.

- Add NH_4Cl to buffer C to final concentration of 300 mM and call it buffer D. Elute the protein with a linear gradient formed by buffer C and buffer D. The concentration of NH_4Cl started at 0 mM and reached 300 mM within 1 h. Control the flow rate at 2.0 mL / min. Monitor the A_{280} and collect peak fractions in clean 15 mL Falcon tubes, 4 mL each. In total, 30 fractions are collected.

- Visualize fractions on SDS-PAGE as described in Section 2.3.9 and then pool fractions accordingly. The pooled solution usually is approximately 16 mL.

2.3.16 MALDI-TOF Mass Spectrometry

- Slice the main protein band (~ 51 kD) off from stained SDS-PAGE gel using a clean blade and then mail it out in Eppendorf tube to the following address:
  Astbury Centre for Structural Molecular Biology
  School of Chemistry
  University of Leeds, Leeds, LS2 9JT.
  The cost of this is around 150 Euro per sample.

- Trypsin digestion is then carried out before Matrix Assisted Laser Desorption Ionization Time-Of-Flight (MALDI-TOF) MS experiment. Analyse the result by inputting the sequence into on-line database http://prospector.ucsf.edu/ and look at the feedback as shown in result section 2.4.5.

2.3.17 Size exclusion chromatography

- Concentrate the pooled fractions from Section 2.3.15 to approximately 1 mL using an YM-50 centricon by centrifuging at 3,000 g in centrifuge (Beckman Coulter) at 4 °C. The resulting sample is injected onto the Hiload 16/60 Superdex 200 Column through the AKTA FPLC at 4 °C. Carry out gel filtration in 50 mM Tris-HCl pH 7.2, 0.1 %(w/v) LDAO and 150 mM NaCl (buffer E) at the flow rate of 1.2 mL / min at 4 °C. Monitor A_{280} and collect the peak fractions in clean Falcon tubes. Each fraction is 0.5 mL.

- Visualize the fractions on SDS-PAGE as described in Section 2.3.9 and then pool fractions accordingly. The final volume usually is approximately 4 mL.

2.3.18 Purity estimation

- Measure the A_{280} of the pooled protein solution in Section 2.3.17 with Nanodrop Spectrometer. Calculate the protein concentration using its extinction coefficient at 280 nm as in Section 2.1. The Equation 2 is shown below:
  Equation 2: [AlgE] = A_{280} / (105,150 $M^{-1}cm^{-1}$X 1 cm) = A_{280} / 105,150 M = A_{280} X (51,000 g / mol) / 105,150 (mol / L) = 0.51 x A_{280} (g / L) = 0.51 x A_{280} (mg / mL)

- According to the concentration of the solution, measure out 50 μg, 10 μg, 5 μg, 1 μg and 0.5 μg protein and transfer into 1.5 mL Eppendorf tubes. Make up to 100 μL with buffer E. Add 900 μL pre-cooled ethanol (– 20 °C) to each tube to a final concentration of 90 %(v/v). Incubate the tubes at -20 °C for 30 min before spinning the tubes in a benchtop centrifuge at 16,300 g and 4 °C for 10 min. Decant the supernatant and resuspend the protein pellet in 20 μL 1 X Tris-Glycine SDS sample buffer. Run electrophoresis as described in Section 2.3.9 and compare the band intensities to estimate the purity of the protein of interest.

2.3.19 Heat modification test

- Transfer 10 μL pure protein sample (about 10 mg / mL) from Section 2.3.17 into five 1.5 mL Eppendorf tubes with 10 μL 1 X Tris-Glycine SDS sample buffer (Invitrogen) added to each. Incubate the samples at 50 °C in PCR machine for different periods: 0, 5, 20 and 120 min.

- Visualize the samples by SDS-PAGE as described in Section 2.3.9 but don't boil the samples before loading them onto gels.

2.3.20 Protein storage

- Concentrate the combined peak fractions in Section 2.3.17 using YM-50 centricon at 3,000 g and 4 °C in benchtop centrifuge. Monitor A_{280} and calculate the protein concentration using Formula 2. The final concentration is approximately 10 mg / mL or 20 mg / mL, depending on the usage of the final product. The concentrated protein solution is aliquoted into 500 μL Eppendorf tubes at 10 μl solution in each tube. The final protein solution is always in buffer E, which contains 50 mM Tris-HCl, pH 7.2, 150 mM NaCl, 0.1 %(w/v) LDAO. The yield is listed in result section 2.4.6.

- Flash freeze the protein in liquid nitrogen (- 196 °C) and store at - 80 °C for future use. The AlgE protein is considered stable for up to at least one month as they can produce crystals with same quality after one month, although no experiments have been done to prove the stability.

2.4 Results and discussion

2.4.1 PCR amplification

After PCR reaction with temperature gradient (from 58.1 °C to 63.1 °C with about 1 degree difference in each step), the product was visualized on 1.0 %(w/v) agarose gel. Multiple bands appeared on the gel as in Figure 2.3.

The negative control sample shows a band smaller than 100 bp at the front indicated by the lower arrow. This band is also present in lane 2 to lane 7 but with much lower intensity. This indicates that forward primer and reverse primer formed primer aggregate to a certain degree during the reaction.

As the upper arrow points out, there are bands of approximately 1.5 kb in lane 2 to 7, with the strongest intensity in lane 7 and the weakest one in lane 4. The size of bands matches the theoretical size of AlgE gene and the band is absent in the negative control sample. This indicates that this amplified DNA fragment present did not arise from contamination and so can be considered possible to be the target gene of this study.

In lane 2, 3 and 6, besides the possible target band, bands of about 800 bp or 1000 bp are also seen. These bands are the product of PCR reaction due to non-specificity between primers and template DNA. Compared to these lanes, there are no such bands shown in lane 7, indicating that the PCR reaction yields pure product under these two conditions.

Adding up all the results, it is determined that 63.1 °C (lane 7) is the optimum of those tested annealing temperatures for AlgE colony PCR reaction.

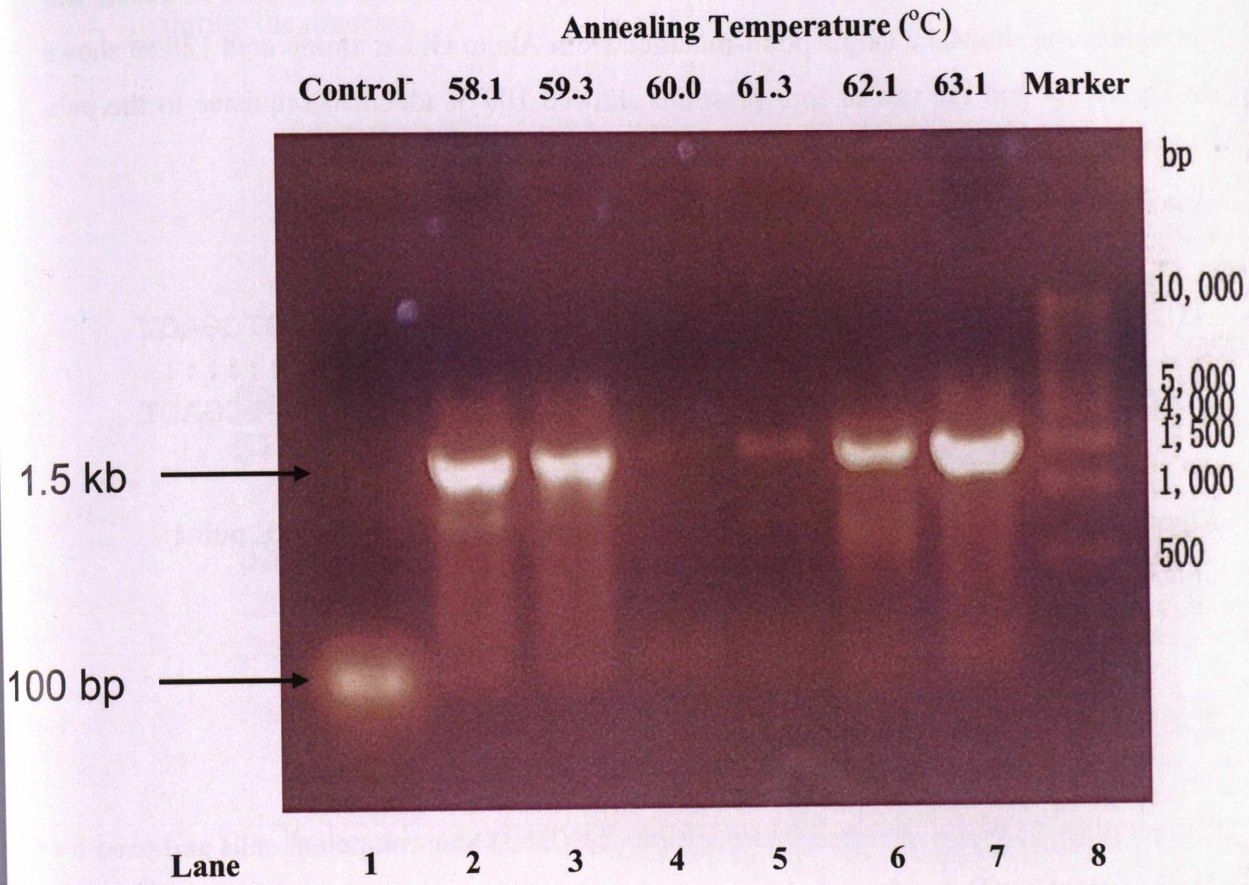


Figure 2.3 Result for annealing temperature gradient of PCR to find out the optimal condition for AlgE gene amplification. The theoretical molecular weight of AlgE gene is 1.5 kb, as indicated by the upper arrow on the left, and the lower arrow indicates the primer aggregate. Control⁻: Colony PCR system without *P. aeruginosa* cells as a negative control. Lanes 2 to 7 shows products generated by differing annealing temperatures. Lane 8: NEB 1 kb DNA ladder and the molecular weight of each band is listed on the right-hand side of agarose gel.

2.4.2 Sequencing result

Following successful PCR reaction, the PCR product was transformed into *E. coli* Top10 chemically competent cells using TOPO cloning kit. Six isolated plasmids for AlgE cloning work were sent out for sequencing.

Of all the plasmids, one failed to pass the quality control of company and no result was provided, one showed a single point-mutation from Ala to Glu at amino acid 128 as shown in Figure 2.4 and the rest of four plasmids showed 100 % identical sequence to the published AlgE gene sequence in *P. aeruginosa* PAO1 strain.

A correct plasmid was then chosen and used for all the downstream work.

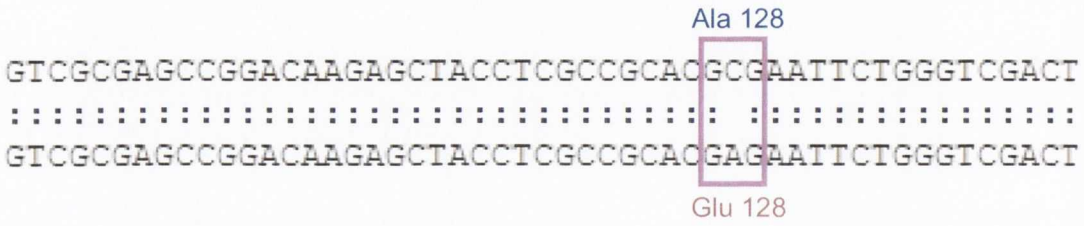


Figure 2.4 One of the six plasmids sent for sequencing, showing A128E point mutation

2.4.3 IPTG induction test

Correct plasmid was transferred into *E. coli* BL21 (DE3) Star competent cells and used for AlgE production. But before starting to produce AlgE protein in *E. coli* cells, the optimal expression condition must be determined and IPTG induction test was designed for this purpose.

One major factor that affects protein production is the temperature so IPTG induction was carried out at two temperatures and the result is shown in Figure 2.5.

Because AlgE protein has a molecular weight of 54 kD according to the full-length gene sequence, bands in the range between 40 kD to 70 kD are analysed. Comparing the intensity of those bands in Figure 2.6, two bands showed obvious difference in differing lanes. These two bands are identified by the arrows on the left-hand side of SDS-PAGE. The top band, which is close to 59 kD, is strongest in Lane 3 but middle band, which is close to 51 kD, is strongest in Lane 2. Both bands are observed to be of much lower intensity in lane 1. This demonstrates that, expression of the proteins represented by these two bands was induced by the addition of 1 mM IPTG. There is another band at about 35 kD position as indicated by the bottom arrow. The identity of this protein is unknown but it is also induced by IPTG as it is not present in the negative control sample. This is possibly due to the natu-

ral response of cells to IPTG and some proteins are co-expressed with the target protein during the process.

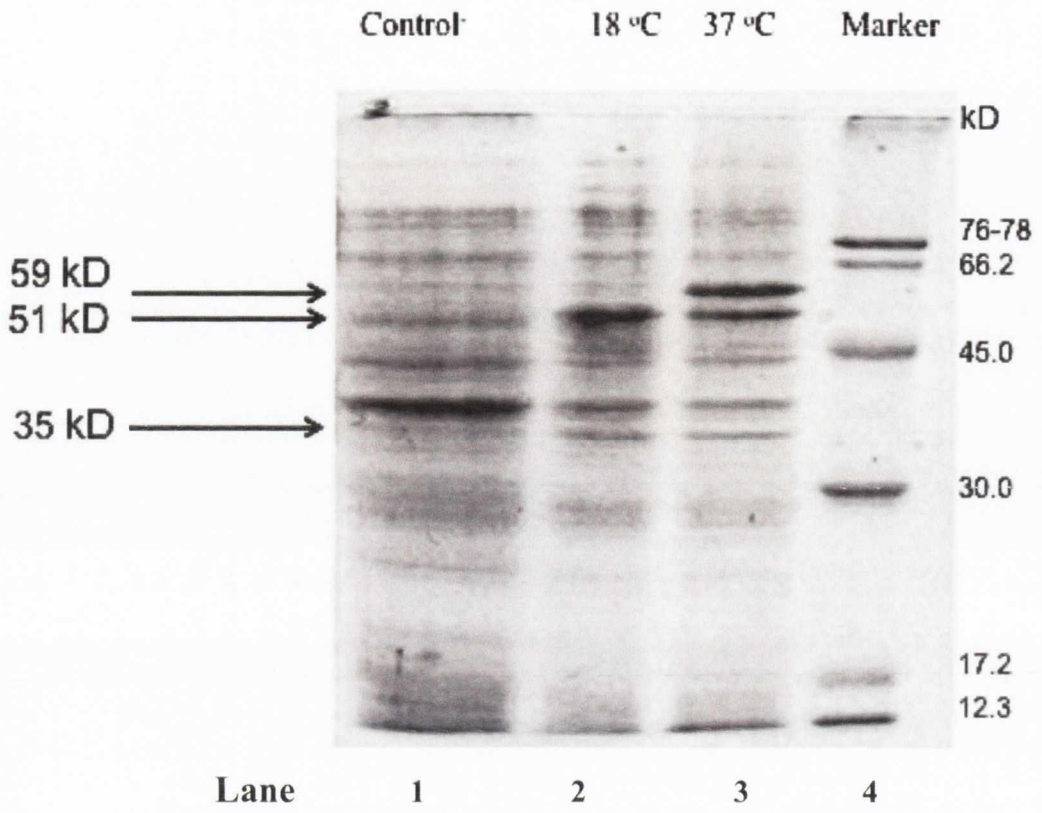


Figure 2.5 The result of IPTG induction test. The samples of whole cells were boiled to 100 °C before loading and the gel was stained by Gelcode Blue Safe stain. Lane 1: *E. coli* BL21 (DE3) cells containing AlgE plasmid, not induced by 1 mM IPTG, as negative control. Lane 2: *E. coli* BL21 (DE3) Star cells containing AlgE plasmid, induced by 1 mM IPTG at 18 °C. Lane 3: *E. coli* BL21 (DE3) Star cells containing AlgE plasmid, induced by 1 mM IPTG at 37 °C. Lane 4: BDH protein marker and the molecular weight in kD of each band is listed on the right-hand side of SDS-PAGE gel. Three arrows on the left indicate three bands that show different intensity between negative control and induced samples. The top arrow shows a protein with molecular weight at 59 kD, the middle arrow shows a protein with molecular weight of 51 kD and the bottom arrow 35 kD.

The theoretical molecular weight calculated using full length DNA sequence of AlgE gene from *P. aeruginosa* PAO1 strain is 54 kD. However, as a 32-residue signal sequence is

predicted on the N-terminal, the mature AlgE protein should be 51 kD. So it is possible that the lower 51 kD band corresponds to the mature AlgE protein.

To understand what the 59 kD band represents, one must look at Figure 2.6. As shown in panel A of Figure 2.6, AlgE gene is cloned into pET200/D-TOPO vector between the two sites indicated by the two blue "TOPO" labels. As a result, the product of this construct is a recombinant protein, which includes the sequence of mature AlgE protein, signal sequence and extra residues from the vector.

One can see that the extra residues from the vector include the codon for a 6His-tag, the codon for an Xpress^{TM} Epitope site, the codon for an enterokinase recognition site and Pro-Phe-Thr (from the sequence CCC TTC ACC before the start codon of AlgE protein). In total, there are 102 nucleotides. As a result, 34 extra amino acids are existent before the signal sequence of AlgE. The addition of the 34 residues from the vector, the 32 residues from signal sequence and 458 residues from the mature AlgE protein is 524 residues, which is approximately 59 kD, as shown in panel B of Figure 2.6. So the upper band in Figure 2.5 correlates to the full-length recombinant AlgE protein perfectly. In summary, IPTG induction at 18 ^{o}C only produces the mature AlgE protein and IPTG at 37 ^{o}C produces both full-length recombinant protein and the mature form. The cloning strategy was flawed in a way that signal peptidase may not cleave off signal peptide with extended residues, but mature AlgE protein was obtained so the strategy was valid for this project.

As two forms of AlgE proteins were produced by *E. coli* cells, choice has to be made between them. The full-length AlgE form is analysed by online disorder prediction programme (Protein Disorder Prediction System) at http://www.prdos.hgc.jp/cgi-bin/top.cgi. The first 66-residue N-terminal fragment is predicted to be disordered and so not preferred for later crystallogenesis work, it is abandoned. So, in this study, 18 ^{o}C is chosen for IPTG induction in order to produce the mature AlgE protein only.

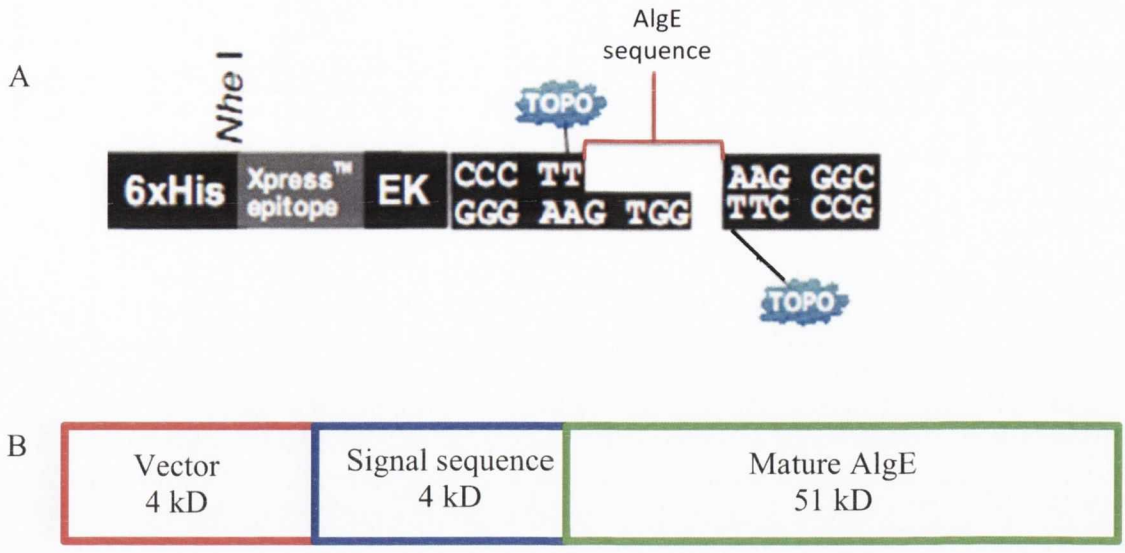


Figure 2.6 A cartoon representation of major fragments in full-length recombinant AlgE protein. In panel A, the representation of DNA sequence is shown. In panel B, the representation of protein sequence is shown. If AlgE protein is expressed in the full-length form, then the molecular weight will be (4 + 4 + 51) = 59 kD. If AlgE protein is expressed in the mature form then the molecular weight will be 51 kD.

2.4.4 Cu^{2+}-IMAC chromatography for AlgE

In an attempt to purify AlgE without His-tag in its mature form, a published protocol (Rehm *et al.* 1994) was followed and modified as in the flowchart in section 2.3.11. Because it is shown that AlgE can bind to Cu^{2+} naturally in the paper, Cu^{2+}-IMAC was used as the first step for purification.

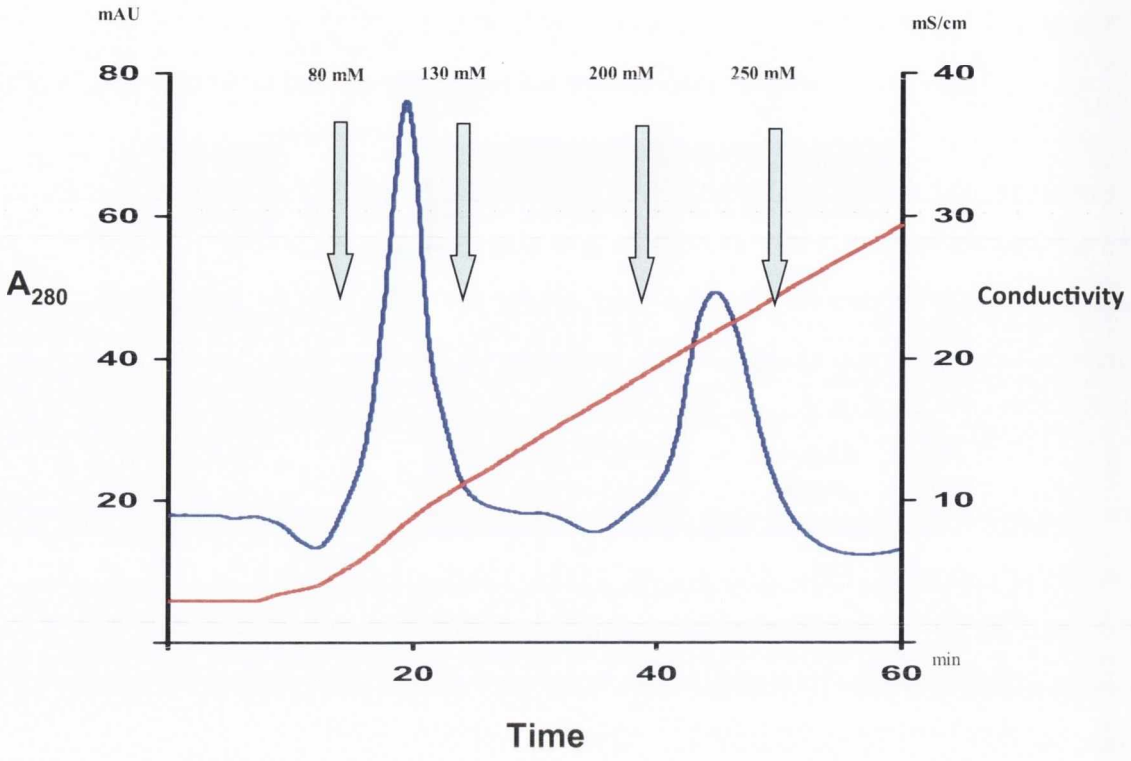


Figure 2.7 Cu^{2+}-IMAC chromatography for AlgE purification. The protein sample shows two peaks at A_{280}. As shown in the figure, the blue curve represents A_{280} and two peaks appear between the arrow-pointing range of 80 – 130 mM and 200 – 250 mM NH_4Cl. The red curve represents the conductivity of solution. X-axis is elution time, Y-axis on the left is the absorbance at 280 nm and Y-axis on the right is the conductivity of elution buffer.

As in Figure 2.7, there are 2 major peaks in the chromatogram, with the first elution between 80 mM and 130 mM of NH_4Cl while the second elution between 200 mM and 250 mM of NH_4Cl. AlgE protein was expected to be within one of these ranges and so the range of interest needed to be decided using SDS-PAGE (Rehm *et al.* 1994). After electrophoresis, the fractions of the second elution showed nothing on the gel, possibly because the protein in these fractions were too small and ran out of the gel at early stage. The fractions of the first elution were instead shown below in Figure 2.8.

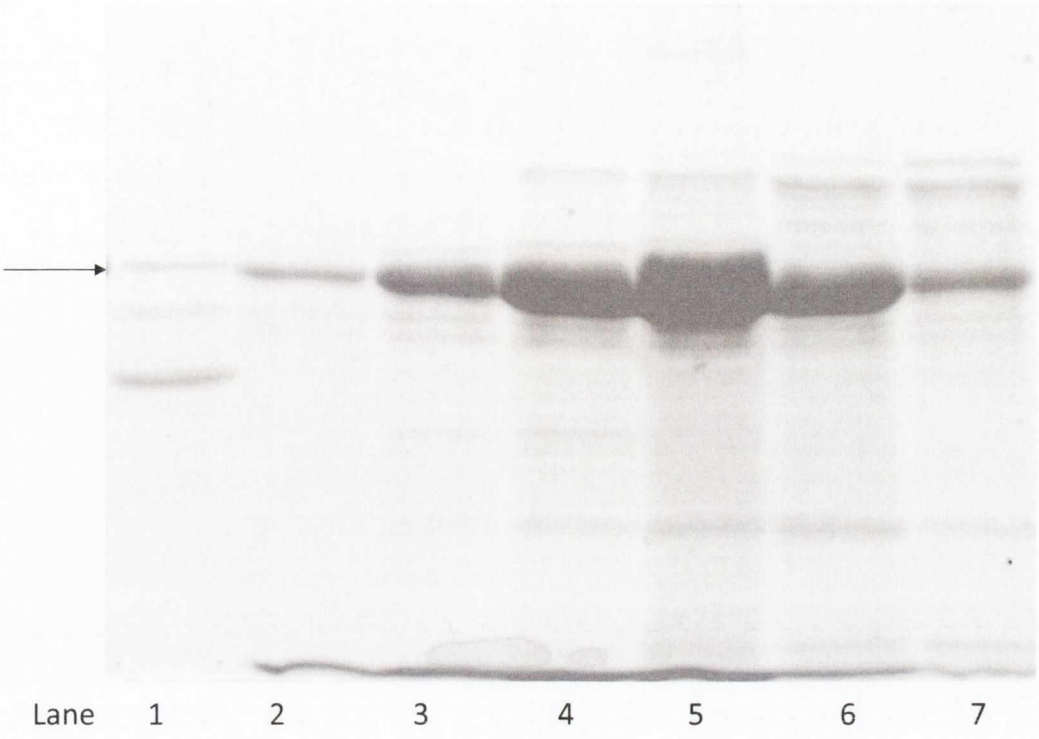


Figure 2.8 The purity of fractions collected in the 80 mM - 130 mM range from Cu^{2+}-IMAC checked by SDS-PAGE. The samples were boiled at 100 °C for 5 min before loading and the gel was stained by Gelcode Blue Safe stain. From lane 1 to 7, the fractions eluted between 80 mM and 130 mM NH_4Cl from Cu^{2+}-IMAC were loaded. The arrow indicates the position of one major band, with a molecular weight of 51 kD.

As in Figure 2.8, multiple bands were present in each lane but one major band was overwhelmingly dominating in lane 5 and present across all the lanes, as pointed out by the arrow on the left. Although this major band was eluted mainly between 100 mM to 120 mM of NH_4Cl, which differs from the 0 to 40 mM NH_4Cl reported in the paper (Rehm *et al.* 1994), the molecular weight of it showed good match with mature AlgE protein so this band was sliced off from SDS-PAGE gel with a clean blade and sent out for MALDI-TOF identification.

In the meantime, judging from the gel, further purification was decided to be necessary. So the fractions were pooled and were loaded for size exclusion chromatography.

2.4.5 MALDI-TOF MS identification

The experiment was performed in the way that protein sample was first digested into fragments by trypsin and then analysed by MALDI-TOF. The result will give the masses of each fragment that are found by machine. The result is shown in Figure 2.9.

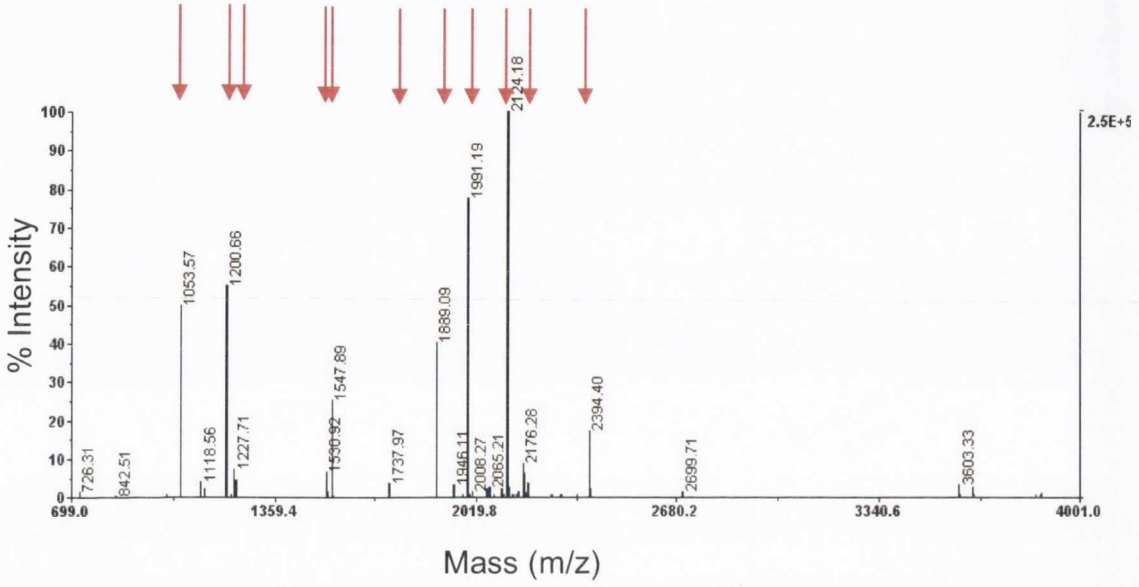


Figure 2.9 MALDI-TOF result for the protein purified by Cu2+IMAC in section 2.3.14. X-axis is the molecular mass of each peak and the Y-axis shows the intensity. The red arrows indicate the peaks picked out manually that show much higher intensity than background contaminants and chosen for later analysis.

In the result, there are 25 peaks altogether, among which 11 peaks showed much higher intensity than the rest. These 11 major peaks should be attributed to the major band in Figure 2.8 and the rest of minor peaks could be due to minor contaminantion bands in the same figure.

These 11 peaks are picked out manually and a search of their masses is done against online database: http://prospector.ucsf.edu/. In this database deposited the masses of theoretical fragments of all proteins with known sequences after trypsin digestion. By matching the masses of these 11 peaks with the ones in database, the identity of the unknown protein sample can be confirmed. The results are shown in Figure 2.10 and Figure 2.11.

MS-Fit search selects 78 entries (results displayed for top 5 matches).

Results Summary

	MOWSE Score	#/11(%) Masses Matched	% Cov	% TIC	Mean Err Da	Data Tol Da	MS-Digest Index #	Protein MW (Da)/pI	Accession #	Species	Protein Name
1	8.93e+7	11 (100)	29.4	100.0	0.166	0.223	6194	54422/4.9	P18895	PSEAE	Alginate production protein algE precursor
2	922	4 (36)	23.2	36.4	0.120	0.154	64597	18854/9.3	Q89AZ6	BUCBP	Flagellar fliJ protein
3	599	4 (36)	17.4	36.4	0.167	0.245	240341	47598/6.4	P07776	ACIAD	Hypothetical protein ACIAD1443 (ORF1)
4	269	4 (36)	7.1	36.4	0.0960	0.183	166498	87167/9.1	P57276	BUCAI	Ribonucleoside-diphosphate reductase alpha subunit (EC 1.17.4.1) (Ribonucleotide reductase)
5	230	4 (36)	9.5	36.4	0.212	0.207	118393	66169/8.4	Q4KJB2	PSEF5	Lipid A export ATP-binding/permease protein msbA (EC 3.6.3.-)

6194	x	x	x	x	x	x	x	x	x	x
64597	.	.	x	x	x	x	.	.	.	.
240341	.	.	.	.	x	.	x	x	x	.
166498	v			v					v	v

Figure 2.10 Database search result showing the tested protein sample is AlgE. The best match in the database has an Index number as 6194 and Accession number as P18895. The protein is called Alginate production protein algE from PSEAE (PSE-AE: short for *Pseudomonas aeruginosa* in database) species, namely the target protein of this project. The eleven fragments match one hundred percent with the ones in database for AlgE. The eleven fragments cover 29.4 % of the total mass of AlgE protein and the mean error is 0.166 Dalton.

Index Number: **6194**
Acc. #: P18895 **Species:** PSEAE **Name:** Alginate production protein algE precursor
pI of Protein: **4.9**
Protein MW: **54422**
Amino Acid Composition: **A44 D46 E26 F23 G52 H11 I17 K13 L41 M5 N22 P18 Q20 R36 S33 T34 V19 W15 Y15**

1 MNSSRSVNPR PSFAPRALSL AIALLLGAPA FAANSGEAPK NFGLDVKITG ESENDRDLGT APGGTLNDIG IDLRPWAFGQ
81 WGDWSAYFMG QAVAATDTIE TDTLQSDTDD GNNSRNDGRE PDKSYLAARE FWVDYAGLIA YPGEHLRFGR QRLREDSGQW
161 QDTNIEALNW SFETTLLNAH AGVAQRFSEY RTDLDELAPE DKDRTHVFGD ISTQWAPHHR IGVRIHHADD SGHLRRPGEE
241 VDNLDKTYTG QLTWLGIEAT GDAYNYRSSM PLNYWASATW LTGDRDNLTT TTVDDRRIAT GKQSGDVNAF GVDLGLRWNI
321 DEQWKAGVGY ARGSGGGKDG EEQFQQTGLE SNRSNFTGTR SRVHRFGEAF RGELSNLQAA TLFGSWQLRE DYDASLVYHK
401 FWRVDDDSDI GTSGINAALQ PGEKDIGQEL DLVVTKYFKQ GLLPASMSQY VDEPSALIRF RGGLFKPGDA YGPGTDSTMH
481 RAFVDFIWRF

The matched peptides cover **29%** (144/490AA's) of the protein.

Figure 2.11 Database search result showing which regions in protein sequence the identified fragments cover. The regions labelled in red show where the fragments are in the AlgE sequence. In total, there are only 8 labelled regions because some of the fragments overlap with each other.

As shown in Figure 2.10 and 2.11, after the comparison of the mass value of the 11 major fragments against the on-line database, protein number 6194 is found to show 100% match with the 11 fragments. The name of protein number 6194 is Alginate production protein

AlgE precursor, namely the target of this study, AlgE. The second best match is protein number 64597 in the database, which showed only 4 matches out of 11 fragments. Furthermore, these 11 fragments from covered 29 % of AlgE full length.

The result is convincing and we can say that the purified protein after Cu-IMAC chromatogram is AlgE.

2.4.6 Size exclusion chromatography

After confirming the purified sample from Cu^{2+}-IMAC was AlgE, the sample was concentrated and loaded for size exclusion chromatography. The purpose of this experiment is not only to purify the protein further but also to see in what oligomeric form AlgE is in the detergent solutions. The chromatogram is shown in Figure 2.12. The values of V_o and V_t are provided by GE Healthcare.

One major peak is indicated by the two green arrows in the Figure 2.13. The peak position is 67.5 mL, which suggests an apparent molecular weight of about 150 kD, as calculated from the calibration cureve of Hiload 16/60, Superdex 200 column, from GE Healthcare. The calibration curve is shown in Appendix 1.

As the micelle size of LDAO is about 17 kD (Moller and M. 1993) and mature AlgE protein is 51 kD, the theoretical sizes of AlgE dimer and trimer are (17 + 51 X 2) = 119 kD and (17 + 51 X 3) = 170 kD, respectively. Because protein can bind detergents to a certain amount (Moller and M. 1993) and some natural lipids might remain bound with protein after purification, the final sizes of AlgE dimer and AlgE trimer should be slightly bigger than the above two values. As a result, the apparent molecular weight of AlgE (150 kD) is between dimer and trimer. No clear conclusion can be drawn without further analysis but AlgE protein in detergent solution should be either dimer or trimer.

There are minor peaks in the chromatogram at 42 mL and 78 mL. The peak at 42 mL is likely to be aggregate of protein as this is immediately after the void volume. The peak at 78 mL indicates some protein with an apparent molecular weight of 45 kD and would be too small to be AlgE monomer (17 + 51) = 68 kD. It is most probably the contaminant pro-

teins that are present in the sample before gel filtration and now have been removed by size exclusion chromatography.

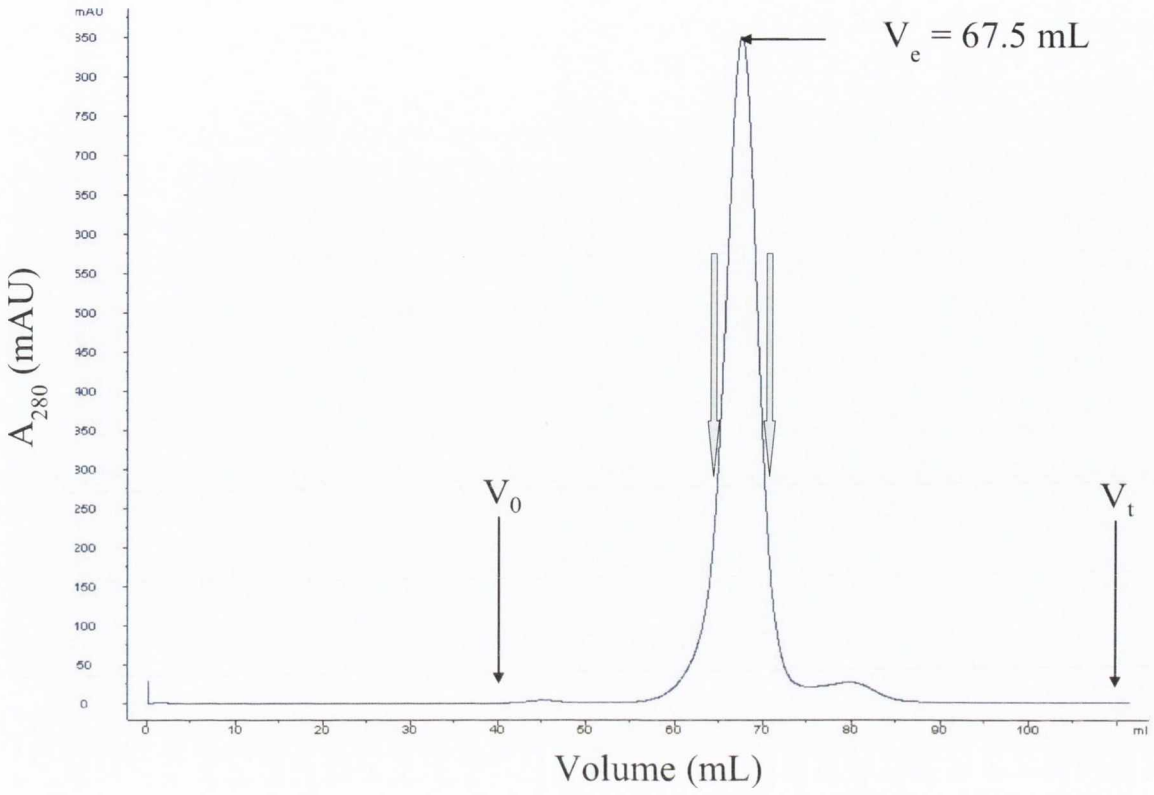


Figure 2.12 SEC chromatogram for AlgE purification. The elution volume is shown on x-axis and the value of A_{280} is shown on Y-axis. The void volume and total volume are 45 mL and 128 mL respectively, indicated by V_0 and V_t arrows. There is one major peak at 67.5 mL as black arrow on the top points out. The two green arrows show where fractions are taken for SDS-PAGE analysis.

The fractions collected across the peak were then analysed by SDS-PAGE gel. And the result is shown in Figure 2.13. As shown in the figure, there are fewer contamination bands when compared to Figure 2.8, where obvious bands can be found above or below the main band in the samples before gel filtration, although no experiment was done to characterize exactly how purer the sample is.

The typical yield of AlgE preps ranges from 1.0 mg per liter cell culture
to 2.0 mg per liter cell culture. But the lowest yield was 0.2 mg per liter cell culture and the highest yield was 3.5 mg per liter cell culture. Details are shown in The Table 2.5.

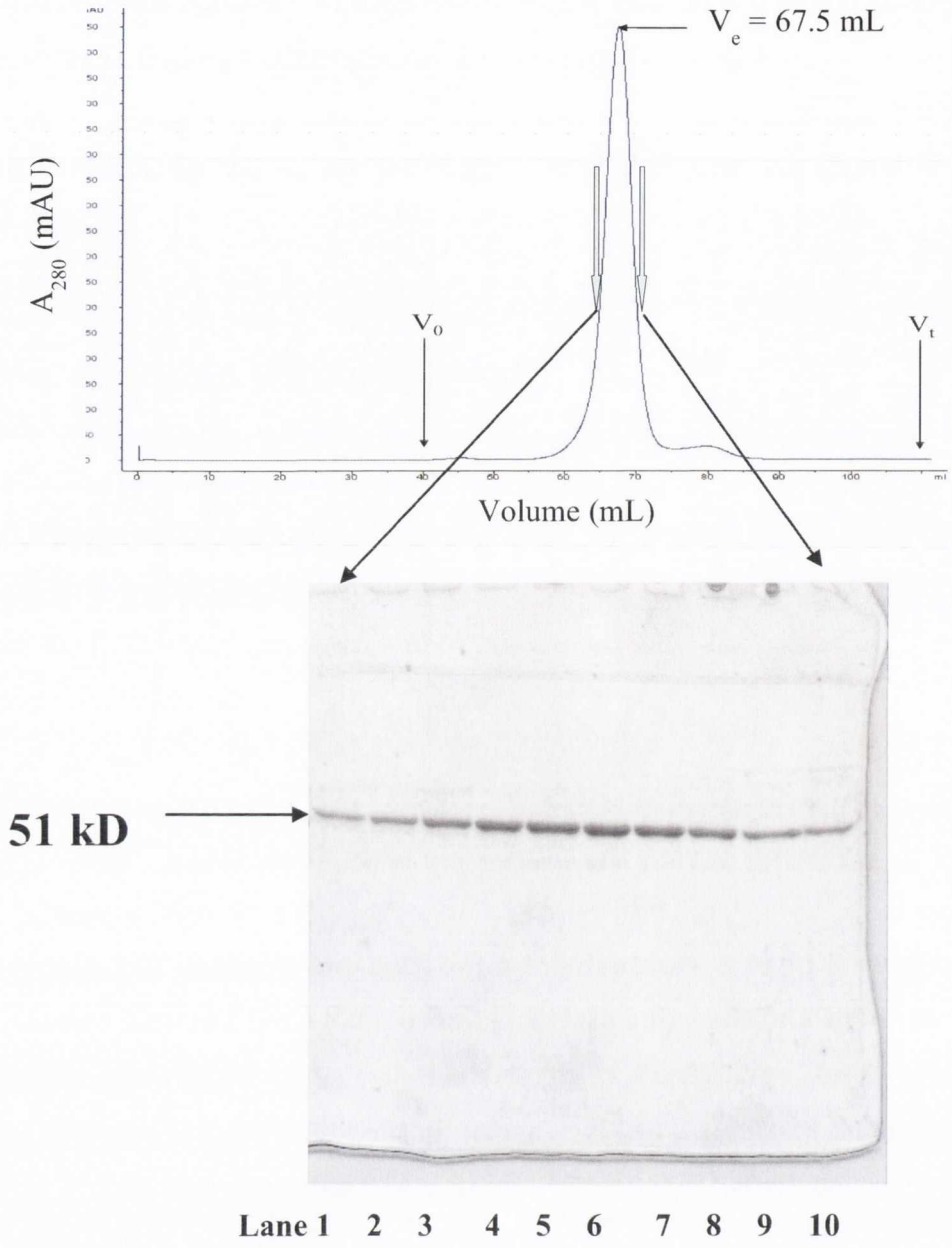


Figure 2.13 The purity of fractions from size exclusion chromatography checked on SDS-PAGE. The samples were boiled to 100 °C before loading and the gel was stained by Gelcode Blue Safe stain. Lane 1 to 10 represents the fractions collected from the peak in the size exclusion chromatography shown above (the same chromatogram as Figure 2.12). The arrow indicates the position of protein bands with molecular weight of 51 kD.

Table 2.5 Summary for the yield of AlgE preps.

Prep name	Yield (mg per Litre cell culture)
AlgE1	0.5
AlgE2-2	1.4
AlgE3	1.2
AlgE4	2.7
AlgE5	3.5
AlgE6	2.3
AlgE8-1	2.0
AlgE8-2	1.3
AlgE9	1.2
AlgE10-1	2.1
AlgE10-2	2.4
AlgE11-2	1.4
AlgE12	0.2
AlgE14	1.8
AlgE16	0.7
AlgE17	1.2
AlgE20	1.0
AlgE22	1.4

2.4.7 Purity estimation for final AlgE sample

The purity of AlgE was determined on SDS-PAGE based on an assumption that equal amount of any protein would show as bands of equal intensity on the SDS-PAGE gel. This is not always true because binding of the Coomassie-type dyes to basic amino acids is much more efficient then to acidic amino acids and this effect (Fazekas *et al.* 1963, Meyer and Lamberts 1965) may cause slight differences in staining of proteins in gels. However, this could still be used as a rough indication of how pure the protein sample is. The result is shown below in Figure 2.14.

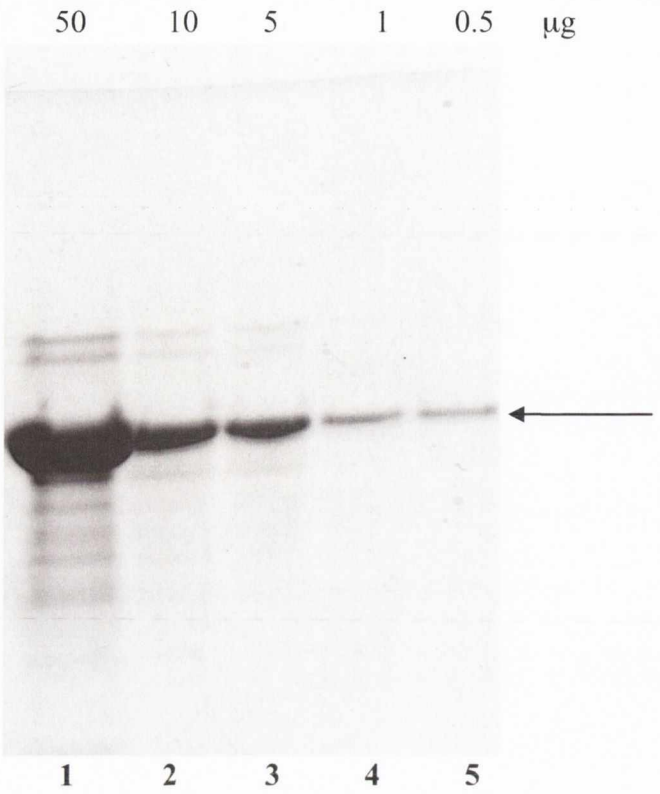


Figure 2.14 AlgE purity estimation. It was determined by SDS-PAGE by loading a series of amount of protein sample into each lane and a rough estimation by eye. The samples were boiled to 100 °C before loading and the gel was stained by Gelcode Blue Safe stain. The arrow points out where AlgE band should be. From lane 1 to 5, the amount of AlgE that's loaded is listed on top the gel.

As shown in gel, from left to right, 50 μg, 10 μg, 5 μg, 1 μg and 0.5 μg of final AlgE sample was loaded onto the gel.

In the first lane, there are about 7 major contamination bands along with AlgE protein band. Each contamination bands' intensity is close to the band's intensity in Lane 5, which should be 0.5 μg loading.

So in total, the contamination bands sum up to about 3.5 μg in Lane 1, which means the purity of AlgE protein should be about (50 – 3.5) / 50 = 93 %. The purity was good for membrane protein purification considering how difficult it is to get a pure membrane pro-

tein (Kloppmann *et al.* 2012) and the protein sample was determined to be suitable be used for crystallisation trials, should it show basic heat modifiability as natural β-barrel proteins.

2.4.8 Heat modifiability of AlgE

β-barrel outer membrane proteins have characteristic heat modifiability (Nakamura and Mizushima 1976). Before and after heat treatment at a temperature higher than a certain degree, the proteins appear in both folded and unfolded form on SDS-PAGE, with the ratio between the two depending on the duration of heat treatment. The two forms will have distinct molecular weight and can easily be differentiated.

As AlgE starts to unfold when the temperature reaches 50 °C (Grabert *et al.* 1990), the experiment was carried out at 50 °C. AlgE in Tris-Glycine SDS sample buffer is treated at 50 °C for different duration, from 5 min to 120 min. The result is shown below in Figure 2.15.

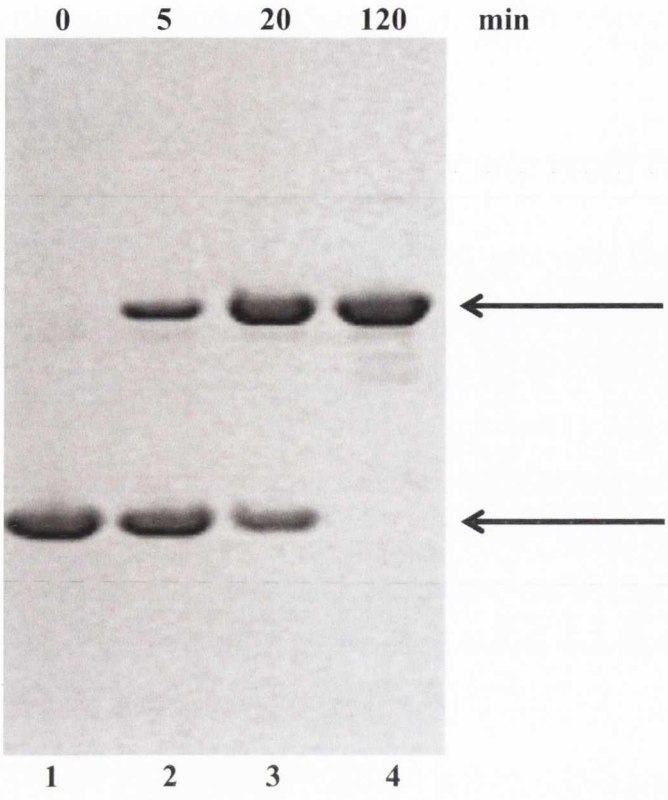


Figure 2.15 The heat-modification behaviour of AlgE checked by SDS-PAGE. After 50 °C incubation and before loading, the samples were not heated. The gel was stained by Gelcode Blue Safe stain. From lane 1 to lane 4, treatment time increases as shown on top of gel. AlgE shows two bands after treatment as indicated by the two arrows: upper band is unfolded AlgE, which has an apparent molecular weight of 51 kD; lower band is folded AlgE, which has an apparent molecular weight of 34 kD.

As shown in Figure 2.15, the lower arrow at 34 kD indicates the position of the unheated and folded AlgE protein on SDS-PAGE approximately. The heated and unfolded AlgE protein was expected to run at a higher position as indicated by upper arrow with a molecular weight of 51 kD.

With no treatment, AlgE was folded with apparent molecular weight as 34 kD, as expected and shown in lane 1 of Figure 2.15. With the increase in time for heat treatment, the intensity of upper band increased while that of lower band decreased. This indicates that AlgE was gradually being unfolded with heat treatment. After 120 min treatment, only the upper band was left indicating that AlgE had reached a stable unfolded state. This behaviour is

identical to that of most β-barrel outer membrane proteins, according to previous studies described earlier (Nakamura and Mizushima 1976).

2.5 Conclusions

2.5.1 According to PCR and sequencing results, AlgE gene has been successfully cloned into *E. coli* BL21 (DE3) Star chemically competent cells and is suitable for use in protein expression experiments.

2.5.2 The IPTG expression result showed that, expression of two proteins was induced by the addition of IPTG. These two proteins have molecular masses of 59 kD and 51 kD, respectively, and represent the two different forms of expressed AlgE protein: full length recombinant AlgE (with signal peptide and 37 residues from expression vector) and mature AlgE (with signal peptide and residues from expression vector removed by signal peptidase in *E. coli* cells).

2.5.3 When the cells were induced under 37 °C for 3 h, almost equal amounts of full-length recombinant form and mature form of AlgE protein was observed. The full-length recombinant form has a 66 amino-acid-long disorder region on the N-terminal and was not preferred in structural study, as the disordered region might cause problems in the crystallisation stage. When the cells were induced under 18 °C for 18 h, the full-length recombinant form was nearly 100 % absent and only the mature form of AlgE remained. So, expression under 18 °C for 18 h is established as better method in future production protocol.

2.5.4 The expressed protein was partially purified by Cu^{2+}-IMAC chromatography and then was sent for trypsin digestion followed by MALDI-TOF MS experiment. The MALDI-TOF MS experiment revealed eleven major fragments after trypsin digestion and, by comparing the masses of these fragments with online database, it is convincingly concluded that the protein sample is AlgE protein. After size exclusion chromatography, AlgE reached a purity of approximately 93 %. Pure AlgE runs with the apparent molecular weight of between a dimer and a trimer on size exclusion chromatogram.

The typical yield is 1.0 to 2.0 mg per litre cell culture.

2.5.5 Pure AlgE shows characteristic consistent to that of common β-barrel proteins.

50 °C incubation can unfold AlgE gradually and unfolded AlgE protein shows higher molecular weight than the folded AlgE protein on SDS-PAGE. Although further experimentation is required to verify that the purified AlgE protein remains active, the protein can be used for crystallisation trials. And because AlgE can crystallise after storage in – 80 °C freezer as observed in later experience, it is considered to be relatively stable compared to some other sensitive membrane proteins, although no experiments have been done to verify this.

Chapter 3

AlgE crystallisation and experimental approaches for structure determination

3.1 Introduction

3.1.1 Methods for protein structure determination

Many methods have been used for protein 3D conformational structure determination. They include X-ray crystallography, nuclear magnetic resonance spectroscopy (NMR spectroscopy), electron microscopy, fiber diffraction, electron diffraction, neutron diffraction and solution scattering.

Among those methods, X-ray crystallography and NMR are the most popular ones so far. According to the statistics from Protein Data Bank (PDB) and Membrane Protein Data Bank (MPDB), there have been 81,538 deposited proteins or protein complex structures (PDB statistics, 22th August 2012) and among them 1,348 structures are membrane proteins or have at least one membrane protein in the complex (MPDB, 25th June 2012). Of these 81,538 structures, 72,369 are solved by X-ray crystallography and 8,551 are solved by NMR. As a comparison, only 426 structures were solved by electron microtroscopy (EM) and 146 were solved by other methods (PDB Statistics, 22th August 2012).

In the MS&FB group, X-ray crystallography is used as a main method to solve membrane protein structures.

This method was introduced into protein structure field in the late 1950s by Max Perutz and Sir John Cowdery Kendrew, who were awarded Nobel Prize in 1962 for this great contribution (Kendrew 1958). For this method to work, diffraction-quality crystals of protein must be available. Then the crystals are exposed to X-ray beam and the scattering pattern is recorded. Analysis of these data can reveal the information on protein structure.

3.1.2 General protein crystallisation methods

As stated above, for macromolecular crystallography to be successful, diffraction-quality crystals of the target protein must be obtained. However, it has been widely accepted that to obtain diffraction-quality crystals for membrane proteins, especially integral membrane proteins who have at least one segment embedded inside membrane lipid bilayer, is one of the major obstacles in structure determination work (Cecchini 2003, Newstead *et al.* 2008, Newstead *et al.* 2008, Bill *et al.* 2011).

Several techniques are used to crystallise membrane proteins, and most of these techniques directly use the protein-detergent solution by the surfactant-based methods (also known as *in surfo* methods) (Caffrey 2009). Various set-ups are applied to grow protein crystals by *in surfo* method, such as hanging-drop vapour-diffusion, sitting-drop vapour-diffusion, micro-batch, counter diffusion, dialysis, gel crystallisation and various seeding techniques (Sutton and Sohi 1994, Fromme 2003, Loll *et al.* 2003, Ng *et al.* 2003, Ng *et al.* 2008). *In surfo* methods have been generally successful in crystal growth of many membrane proteins. However, these methods have their own limits because many integral membrane proteins have both hydrophobic and hydrophilic regions and, in most studies, are extracted from membranes by use of detergents (Helenius and Simons 1975). Detergents, at certain concentration, can form micelles and membrane proteins are buried into these micelles so that they can be soluble in solution (Jumpertz *et al.* 2011). Although there have been many successful cases that have detergent in crystallisation systems, it still happens to some membrane proteins that the existence of micelles hamper the crystallogenesis process, resulting in the failure to produce structure-quality crystals, and in some incidences, the failure to produce any crystals at all (Féthière 2007).

To supplement *in surfo* crystallisation method, two lipid bilayer based methods have been developed in order to successfully crystallise membrane proteins in the absence of detergent: bicelle method (Faham and Bowie 2002) and *in meso* method (also known as LCP method) (Landau and Rosenbusch 1996, Rummel *et al.* 1998). In this project, *in meso* method has again proved to be successful and details will be described later in this chapter.

3.1.3 *In meso* crystallisation method

After studying the roles that detergents and phase science play in producing membrane protein crystals using *in surfo* method, Rosenbusch's group invented a novel method. This method starts with mixing membrane proteins with lipids, monoolein, to form a transparent gel-like cubic phase (Landau and Rosenbusch 1996, Rummel *et al.* 1998) so the method is now called Lipidic Cubic Phase (LCP) method or *in meso* method. Under cubic phase, a highly curved lipid bilayer partitions into two continuous but non-penetrating water channels so cubic phase is viewed as a bicontinuous system (Luzzati *et al.* 1968). A cartoon representation of cubic phase is shown in Figure 3.1. Cubic phase also has some other interesting properties, such as its ability as a liquid crystal to diffract X-rays at small angles (Luzzati 1997), its proposed involvement as an intermediate in fat digestion (Patton and Carey 1979) and membrane fusion (Siegel *et al.* 2006).

Because of the high viscosity of cubic phase, in the beginning, the method was not easy to be applied widely, which involved multiple steps of centrifugation tubes with increasing speed. To make it more efficient to use, the MS&FB group has developed a syringe-type mixing kit, a manual dispensing kit and a robot to employ high-throughput automatic setup for *in meso* crystallisation (Cheng *et al.* 1998, Cherezov *et al.* 2004, Cherezov and Caffrey 2005). The group also spent years in researching the mechanisms of *in meso* method by studying the phase transition (Caffrey 1989), designing and testing novel monoacylglycerol lipids (MAGs) (Li *et al.* 2011) and producing membrane protein structures using this method (Caffrey *et al.* 2012). Although exactly how the *in meso* system produces protein crystals is not completely understood, a hypothesis is brought out and tested.

As in Figure 3.2, the process begins with the mixing of lipid and protein-detergent solution and integral membrane proteins preferably reconstituting into curved lipid bilayer of cubic phase. After precipitant solution is applied to cubic phase, local lamellar phase domain is induced. Lamellar phase is different from cubic phase in that it has a layered structure and it is thought this phase separation is essential to *in meso* crystallisation. At this stage, protein molecules start to diffuse from cubic phase domain into lamellar phase domain. By contact between protein molecules, nuclei can form. And by feeding protein from cubic phase reservoir to the nuclei, crystals begin to appear. (Caffrey 2009)

Although detergents have profound effect on the cubic phase system due to its partitioning into lipid bilayer and flattening the curvature, it is inappropriate to conclude its role in *in meso* method as the mechanism is not completely understood (Ai and Caffrey 2000, Misquitta and Caffrey 2003). However, it has been shown that, because detergent flattens the lipid bilayer, it destabilizes cubic phase and facilitates the phase transition from cubic to lamellar. Thus, at low concentration, the existence of detergent may be helpful to *in meso* crystallisation because the proposed mechanism for the method requires a cubic phase reservoir to feed protein into local lamellar conduit (Ai and Caffrey 2000, Misquitta and Caffrey 2003).

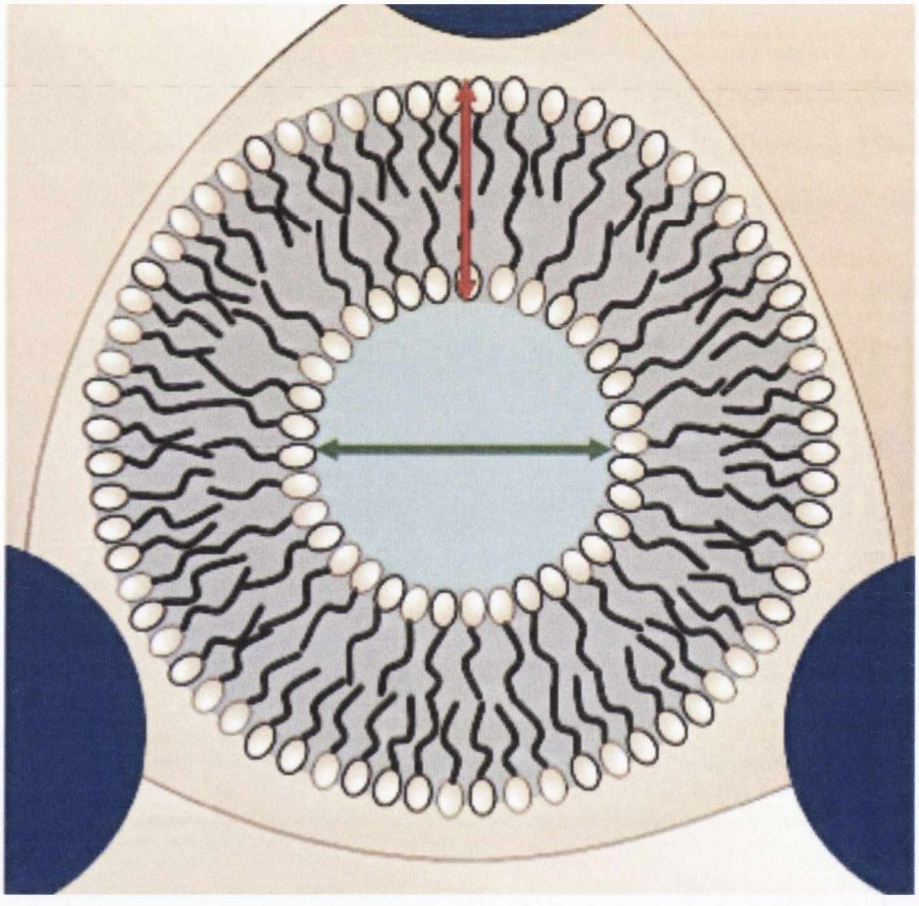


Figure 3.1 A cartoon representation of cubic phase system formed by water and monoolein. The bilayer is coloured in brown and the thickness of bilayer is indicated by red double-headed arrow. The two non-penetrating water channels are coloured in light blue and dark blue. The green double-headed arrow indicates the thickness of water channel. In fully swollen cubic phase of monoolein the bilayer thickness is on the order of 34-36 Å while the water channel diameter is approximately 33 Å (Briggs *et al.* 1996). Copied from (Lyons 2012).

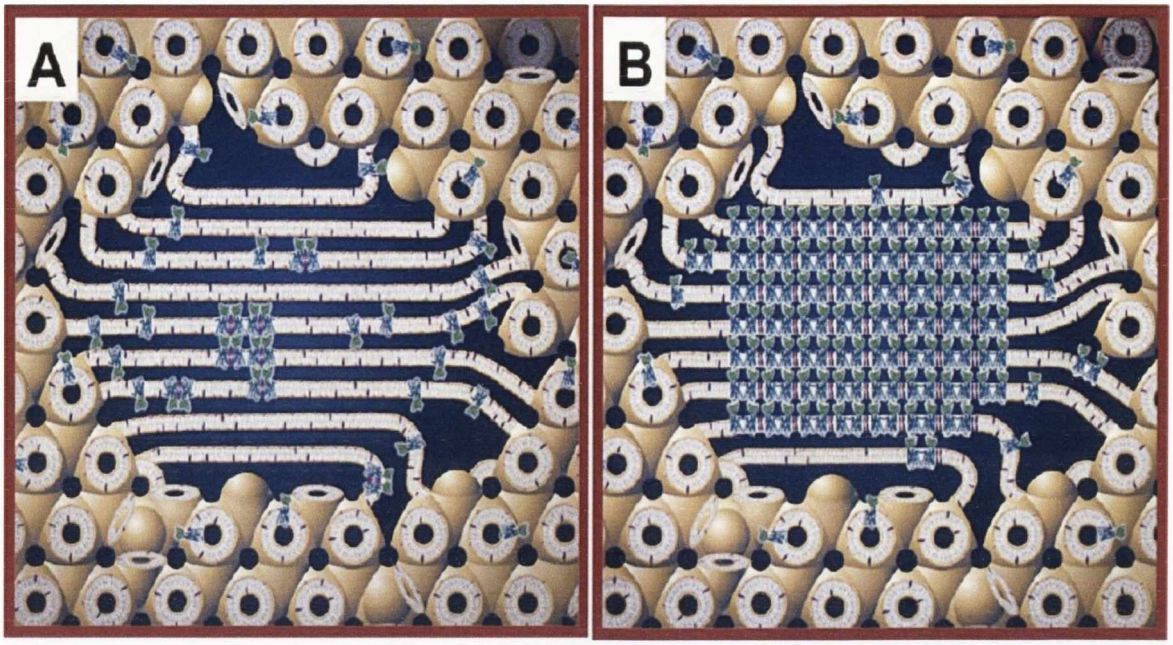


Figure 3.2 Cartoon representation of the proposed mechanism for *in meso* crystallisation process. The brown objects represent curved lipid bilayer in cubic phase and the dark blue background represents the water channel in cubic phase system. White tails extending from the curved bilayer and forming sheetlike structure in the center represent local lamellar phase induced by the addition of crystallisation component. Small blue and green objects represent integral membrane proteins that have reconstituted into lipid bilayer. In panel A, proteins start to diffuse from curved bilayer of cubic phase into the layered bilayer of lamellar phase. In panel B, by contact between protein molecules, nuclei forms and more protein molecules keep feeding it by diffusion from cubic phase to lamellar phase. Eventually, crystals may appear. Copied from (Caffrey *et al.* 2012).

Up to the date when this thesis is submitted, 106 entries in PDB, including the three unreleased entries for AlgE *in meso* structure models from this thesis work, are generated by *in meso* crystallisation method (Caffrey *et al.* 2012). This corresponds to about 10 % of all membrane protein entries that have been deposited in PDB and represents at least eight distinct classes (Caffrey *et al.* 2012). Some information for the 103 entries, not including the ones for AlgE, is listed in Table 3.1. In addition to the entries in this table, AlgE has generated three *in meso* entries in PDB recently: 4AFK, 4AZL and 4B61. These three entries will be described in full details in Chapter 4.

Table 3.1 Structures in Membrane Protein Data Bank (MPDB) that are solved by *in meso* method. Rearranged from (Caffrey *et al.* 2012).

Protein (Organism) [PDB Record Count]	Hosting Lipids	PDB ID (Resolution, Å)
β-Helix		
Gramicidin D (*Brevibacillus brevis*) [3]	7.7 MAG; 8.8 MAG; 9.9 MAG	2Y5M (1.08); 2Y6N (1.26); 2XDC (1.70)
β-Barrel		
BtuB (*Escherichia coli*) [1]	9.9 MAG	2GUF (1.95)
OpcA (*Neisseria meningitidis*)[1]	9.9 MAG	2VDF (1.95)
OmpF (*Escherichia coli*) [3]	9.9 MAG	3POQ (1.90); 3POX (2.00); 3POU (2.80)
Intimin (*Escherichia coli*) [1]	9.9 MAG	4E1S (1.86)
Invasin (*Yersinia pseudotuberculosis*) [1]	9.9 MAG	4E1T (2.26)
a-helical (non-GPCR)		
Halorhodopsin (*Halobacterium salinarum*) [3]	9.9 MAG	2JAF (1.70); 1E12 (1.80); 2JAG (1.93)
Sensory Rhodopsin II (*Natronomonas pharaonis*) [6]	9.9 MAG	3QAP (1.90); 1H68 (2.10); 1GU8 (2.27); 1GUE (2.27); 1JGJ (2.40); 3QDC (2.50)
Sensory Rhodopsin II/Transducer Complex (*Natronomonas pharaonis*) [3]	11.7 MAG	1H2S (1.93); 2F93 (2.00); 2F95 (2.20)
Sensory Rhodopsin (*Nostoc sp. pcc 7120*) [1]	9.9 MAG	1XIO (2.00)
Channelrhodopsin (*Chlamydomonas reinhardtii*) [1]	9.9 MAG	3UG9 (2.30)
Acetabularia Rhodopsin II (*Acetabularia acetabulum*) [1]	9.9 MAG + 10 %(w/w) Cholesterol	3AM6 (3.20)

Table 3.1 continued

Light Harvesting Complex II (*Rhodoblastus acidophilus*) [1]	9.9 MAG	2FKW (2.45)
Bacterorhodopsin (*Halobacterium salinarum*) [38]	9.9 MAG; β-XylOC$_{16+4}$: 1-O-(3,7,11,15-tetramethylhexadecyl)-β–d-xyloside	1M0K (1.43); 1M0M (1.43); 1M0L (1.47); 1P8H (1.52); 2NTU (1.53); 2NTW (1.53) ; 1C3W (1.55); 1O0A (1.62); 1P8U (1.62); 1KGB (1.65); 1F50 (1.70); 3NS0 (1.78); 3NSB (1.78); 1C8R (1.80); 1F4Z (1.80); 1KG9 (1.81); 2I21 (1.84); 1P8I (1.86); 1QHJ (1.90); 1C8S (2.00); 1KG8 (2.00); 2I1X (2.00); 1JV6 (2.00); 1MGY (2.00); 3MBV (2.00); 2I20 (2.08); 1E0P (2.10); 1QKO (2.10); 1QKP (2.10); 2WJL (2.15); 1CWQ (2.25); 1JV7 (2.25); 1BRX (2.30); 2WJK (2.30); 1S8J (2.30); 1S8L (2.30); 1VJM (2.30); 1AP9 (2.35)
Photosynthetic Reaction Center (*Blastochloris viridis*) [5]	9.9 MAG	2WJN (1.86); 2WJM (1.95); 2X5U (3.00); 2X5V (3.00); 4AC5 (8.20)
Photosynthetic Reaction Center (*Rhodobacter sphaeroides*) [4]	9.9 MAG	2GNU (2.20); 1OGV (2.35); 2BNS (2.50); 2BNP (2.70);
ba_3 Cytochrome c Oxidase (*Thermus thermophilus*) [2]	9.9 MAG	3S8F (1.80); 3S8G (1.80)
caa_3 Cytochrome c Oxidase (*Thermus thermophilus*) [1]	7.7 MAG	2YEV (2.36)

Table 3.1 continued

Na^{+}-Ca^{2+} Exchanger (*Methanocaldococcus jannaschii*) [1]	9.9 MAG	3V5U (1.90)
a-helical (GPCRs)		
μ-Opioid (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	4DKL (2.80)
κ-Opioid (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	4DJH (2.90)
Nociceptin (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	4EA3 (3.01)
δ-Opioid (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	4EJ4 (3.40)
Sphingosine 1-Phosphate Subtype 1 (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	3V2Y (2.80) 3V2W (3.35)
Muscarinic Acetylcholine M2 (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	3UON (3.00)
Muscarinic Acetylcholine M3 (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	4DAJ (3.40)
Histamine H1 (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	3RZE (3.10)
Dopamine D3 (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	3PBL (3.15)
CXCR4 Chemokine (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	3ODU (2.50); 3OE0 (2.90); 3OE8 (3.10); 3OE9 (3.10); 3OE6 (3.20)
A_{2A} Adenosine (Inactive)	9.9 MAG + 10 %(w/w) Cholesterol	4EIY (1.80); 3EML (2.60)
A_{2A} Adenosine (Active)	9.9 MAG + 10 %(w/w) Cholesterol	3QAK (2.70)

Table 3.1 continued

β_2 Adrenergic - Gs Protein (Active)	9.9 MAG + 10 %(w/w) Cholesterol	3SN6 (3.20)
β_2 Adrenergic (Active)	9.9 MAG + 10 %(w/w) Cholesterol	3P0G (3.50)

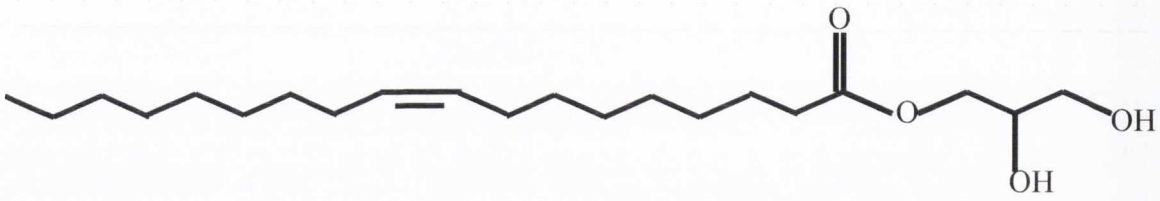
3.1.4 Novel monoacylglycerol lipids (MAGs)

In the lab of MS&FB group, 9.9 MAG (monoolein) is always used as the default hosting lipid for commercial screen trials at the start of any new projects. But novel MAGs are synthesized in the group and they have been widely used in various projects. As will be described in the later part of this chapter, some novel lipids were also used to obtain better quality of AlgE crystals.

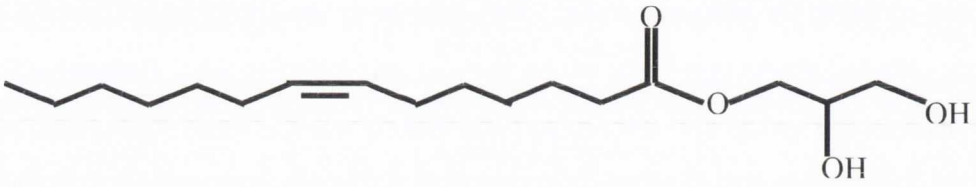
Figure 3.3 shows the chemical structures for all the MAGs that are used in this chapter. As in the figure, the MAGs are named by the length of hydrocarbon chains on both sides of the double bond. Take 9.9 MAG for an example, as indicated by the name, there are 9 carbons in both neck and tail regions. These lipids have different phase diagram and so form cubic phase at 20 °C at different ratio of protein solution to lipid (Qiu and Caffrey 2000, Misquitta *et al.* 2004). Except 9.9 MAG, the other MAGs need a ratio at around 1 : 1 (w : w) to form cubic phase.

In a cubic phase system, these novel MAGs behave different values for bilayer thickness and water channel radius. These are thought to be critical values for *in meso* crystallization method. By varying the chain length of MAGs, these two values can be tailored to facilitate the crystallization process of certain type of proteins, in order to yield diffraction-quality protein crystals.

9.9 MAG Tail Neck Head



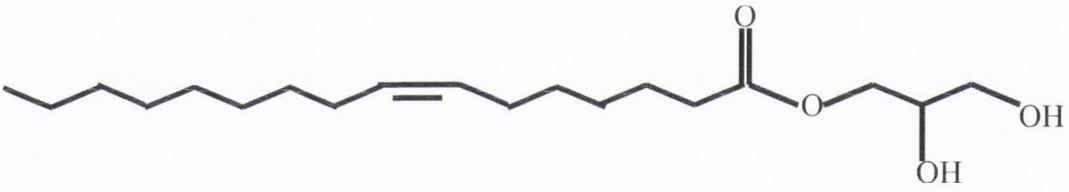
7.7 MAG



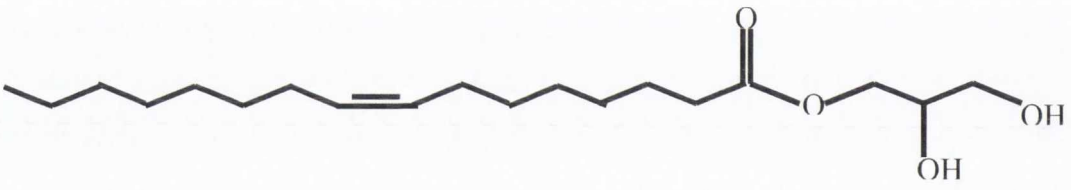
7.8 MAG



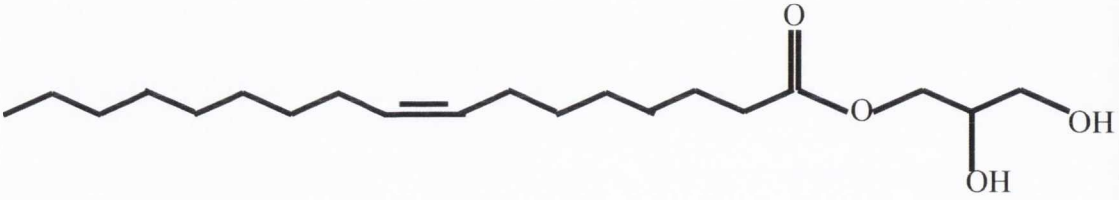
7.9 MAG



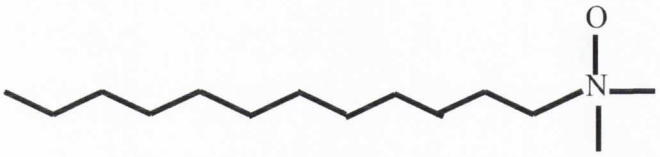
8.8 MAG



8.9 MAG



LDAO



β-OG

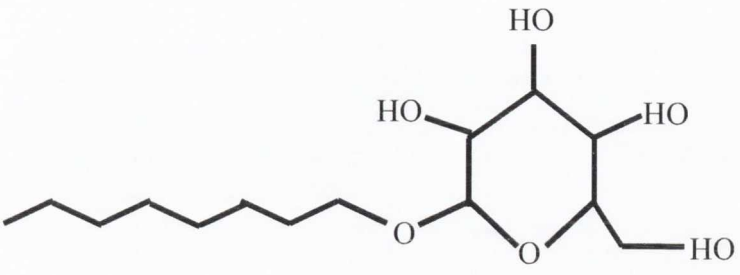


Figure 3.3 The chemical structures of all monoacylglycerol lipids and detergents used in the project. The headgroups of MAG lipids are indicated by red bracket. Both neck and tail are in blue brackets. In addition, LDAO and β-OG are shown for comparison and discussion in the following chapters.

3.1.5 Phasing strategies

After protein crystals have been obtained and optimised to a suitable size, usually at least 20 μm in the longest dimension, they can be harvested and mounted onto an X-ray beam. Then the crystals are irradiated with a beam of monochromatic X-ray and a diffraction pattern recorded as the crystals is gradually rotated.

The data collected from an X-ray diffraction experiment, diffraction pattern of regularly spaced spots usually known as reflections, is a reciprocal space representation of the protein crystal lattice. By rotating the crystal, a set of images corresponding to different crystal orientation is recorded. To transform these images into final structure, the crystallographers must index, integrate, merge and scale the diffraction data and obtain the phase information. Indexing, merging and scaling can be done by mathematical methods but phasing usually is more complicated (Drenth 2007).

Initial phasing can be obtained from this result by various methods, such as *ab initio* phasing or direct method (Usón and Sheldrick 1999), molecular replacement (Taylor 2003), single-wavelength dispersion (SAD) or multi-wavelength dispersion (MAD) (Cowie and Cohen 1957, Hendrickson *et al.* 1990, Ealick 2000) and multiple isomorphous replacement (MIR) (Smyth and Martin 2000).

At the beginning of this project, direct method and molecular replacement were not considered. This was because: first, it was almost impossible to use direct method for membrane proteins due to the resolution of data obtained by X-ray diffraction experiment; second, there was no known structure for any homologue of *P. aeruginosa* AlgE while molecular replacement has to start with a known model.

One of the most successful SAD or MAD methods to determine protein structure is to replace methionine with selenium-methionine (SeMet) by expressing proteins in a methionine auxotroph bacterial strain in SeMet rich medium. Due to the toxicity of selenium to cells, a much lower than normal yield of proteins is usually seen (Cowie and Cohen 1957). And because a milligram scale yield of AlgE prep was observed in Chapter 2, very low yield of SeMet-AlgE prep was anticipated. Although, in principle, only one crystal is needed, usually multiple crystals are used to collect the high-redundancy data set for SAD or MAD method. This suggests this method might not be a good idea, because low yield of proteins might create more uncontrolled batch-to-batch differences and damage the reproducibility of crystal production (Gonzalez 2002).

As a result, MIR was chosen as the main focus of phasing strategy. This method is done by soaking crystal of target protein with heavy atom or co-crystallise the target protein with heavy atom. Data sets of native crystals and heavy-atom derivative crystals are collected and used in the calculation of Patterson difference map, which can reveal phase information and solve the target structure (Harker 1956). In this method, the quality of derivative is called phasing power. If phasing power is above 1.5, it is excellent. If above 1.0, it is good. If above 0.5, it is useful but not able to solve the structure. If lower than 0.5, it is useless (Tickle 1996). Dr. David Aragão and Dr. Valerie Pye did most of data collection work and all of data processing work.

Whilst working on the structure of AlgE, preliminary X-ray crystallography work was also being carried out by Professor Lynne Howell's group at Hospital for Sick Children, Toronto (http://biochemistry.utoronto.ca/howell/bch.html), using a strategy that was completely different from this work. In that work, the AlgE gene was cloned into *E. coli* without signal peptide and the protein expressed as inclusion body. The unfolded protein was then solubilised in guanidine-HCl and refolded into detergent solution by rapid dilution and overnight incubation before being purified by Ni-IMAC and size exclusion chromatography. The purified protein was then used for *in surfo* crystallisation trials. Initial hits were found and crystals were optimised. X-ray preliminary experiment showed diffraction pattern, at best, to 3.0 Å (Whitney *et al.* 2009). And in May 2011, the *in surfo* structure was solved at 2.3 Å resolution using SeMet labeled crystals and deposited in PDB as 3RHB (Whitney *et al.* 2011). Project focus was then changed and details can be found in the next chapter.

3.2 Materials and instruments

Table 3.2 A list of materials that have been used in this chapter.

Item	Company	Catalogue number	Lot number
7.7 MAG	N/A	N/A	JPL-4-92
7.8 MAG	N/A	N/A	JPL-3-77, JPL-3-89, RT-1-10, FM2-21, SS-49, SS-72, TasA12, TasA21
7.9 MAG	N/A	N/A	RT-01-17S(B)
8.8 MAG	N/A	N/A	TASA25
8.9 MAG	N/A	N/A	ASN17
9.9 MAG (Monoolein)	Nuchek (MN, USA)	M239	M239-F15-L; M239-A21-Q
AlgE Screen Kit 1 (see Appendix 2)	N / A	N / A	N / A
AlgE Screen Kit 2 (see Appendix 3)	N / A	N / A	N / A
AlgE Screen Kit 3 (see Appendix 4)	N / A	N / A	N / A
AlgE Screen Kit 5 (see Appendix 5)	N / A	N / A	N / A
$AuCl_3$	Hampton Research (CA, USA)	HR2-444-04	244205
Bis-Tris	Fluka (MO, USA)	14880	1240589

Table 3.2 continued

$CdCl_2$	Hampton Research (CA, USA)	HR2-448-08	245005
Centricon 50 kDa MWCO	Millipore (MA, USA)	UFC905024	R0KA67379
Crystal Screen 1 crystallisation screen	Hampton Research (CA, USA)	HR2-110	211080
Crystal Screen 2 crystallisation screen	Hampton Research (CA, USA)	HR2-112	211251
Eppendorf tubes	Sarstedt (Nümbrecht, Germany)	72.690.001	0000/0355001
Ethanol	Sigma (MO, USA)	E7023	07850LE
Falcon tubes	Sarstedt (Nümbrecht, Germany)	62.559.001	1040601
$HAuCl_3$	Hampton Research (CA, USA)	HR2-444-05	244205
$HoCl_3$	Hampton Research (CA, USA)	HR2-450-16	245005
INDEX crystallisation screen	Hampton Research (CA, USA)	HR2-144	214424
$IrCl_3$	Hampton Research (CA, USA)	HR2-448-11	245005
KCl	Sigma-Aldrich (MO, USA)	P-4286	074K0024
K_2OsCl_6	Hampton Research (CA, USA)	HR2-448-17	244805

Table 3.2 continued

K_2PtCl_4	Hampton Research (CA, USA)	HR2-442-02	244206
K_2PtBr_4	Hampton Research (CA, USA)	HR2-442-08	244206
K_2PtBr_6	Hampton Research (CA, USA)	HR2-442-09	244206
$K_2(NH_4)_2IrCl_6$	Hampton Research (CA, USA)	HR2-448-14	245005
LDAO	Sigma (MO, USA)	40236	BCBB8394
MembFac crystallisation screen	Hampton Research (CA, USA)	HR2-114	211423
MemGold crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-39	004-1-39
MemPlus crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-45	BN002-1-45
MemStart crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-21	011-1-21
MemSys crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-25	010-1-25
MES	Sigma (MO, USA)	M3671	068K5421
Methanol	Fisher scientific (MA, USA)	UN1230	1023754
$MgCl_2$	Sigma-Aldrich (MO, USA)	M2670	129K00891
MilliQ water (resistance at 18.2 mΩ)	Produced by UV5 Elix system	N/A	N/A

Table 3.2 continued

NaCl	Fisher scientific (MA, USA)	BP358-10	80786
NaOH	Merck (NJ, USA)	8627	B382182 411
NdCl3	Hampton Research (CA, USA)	HR2-450-15	245005
NH_4Cl	Sigma-Aldrich (MO, USA)	3101	SZB82820
$(NH_4)_2IrCl_6$	Hampton Research (CA, USA)	HR2-448-13	245005
$(NH_4)_2OsCl_6$	Hampton Research (CA, USA)	HR2-448-16	244805
$(NH_4)_2PtCl_2$	Hampton Research (CA, USA)	HR2-442-02	244206
$(NH_4)WS_4$	Hampton Research (CA, USA)	HR2-450-02	245005
$OsCl_3$	Hampton Research (CA, USA)	HR2-448-18	244805
PACT premier crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-36	BN002
$PrCl_3$	Hampton Research (CA, USA)	HR2-450-17	245005
$Pt(NH_3)_2(NO_2)_2$	Hampton Research (CA, USA)	HR2-442-07	244206
Spacer	Saunders (CA, USA)	3M TAPE 9500PC 77MM	N/A
Structure screen I crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-30	BN016

Table 3.2 continued

Structure screen II crystallisation screen	Molecular Dimensions (Suffolk, UK)	MD1-30	BN016
Sybersafe	Invitrogen (Dun Laoghaire, Ireland)	S33102	643459
Ta_6Br_{12}	Hampton Research (CA, USA)		JBS 00022605
Trizma® base	Sigma (MO, USA)	T87602	088K0118
Wizard I crystallisation screen	Emerald Biosystems (WA, USA)	EBS-WIZ-1	EBS000322200499
Wizard II crystallisation screen	Emerald Biosystems (WA, USA)	EBS-WIZ-2	EBS000322200499
Wizard III crystallisation screen	Emerald Biosystems (WA, USA)	EBS-WIZ-3	EBS00006152009299

Table 3.3 A list of instruments that have been used in this chapter.

Equipment	Company	Type
-20 °C freezer	Liebherr (PA, USA)	GS 1301
20 °C plate incubator	Rumed (Leipzig, Germany)	Rubarth apparate
4 ° C refrigerator	Microban (Staffordshire, UK)	Hotpoint RLA81
4 °C room	Aircold manufacturing (Ireland)	MasterCella - CAREL
-80 °C freezer	Medical Supply (Ireland)	Platinum 500

Table 3.3 continued

Balance (for below 1 g)	Sartorius (Goettingen, Germany)	CP64
Balance (for over 1 g)	Mettler Toledo (OH, USA)	B300T-S
Chemical fume hood	Cameron (Longford, Ireland)	MACH-AIRE
Crystallisation plate imager	Formulatrix (MA, USA)	Rock Imager 1500
Icemaker	Scotsman (Cumbria, UK)	AF 80
Incubator	INFORS HT (Bottmingen, Switzerland)	Multitron
Magnetic stirrer	Stuart (Staffordshire, UK)	SB 162
Microscope	Nikon (Surrey, UK)	SMZ1500
Progad filter for MilliQ Elix 5 system	Millipore (MA, USA)	PR0G00052
Simpack filter for MilliQ Elix 5 system	Millipore (MA, USA)	SIMPAK0RT (lot. F9HN06031)
Nanodrop (spectrophotometer)	Mason Technology (Dublin, Ireland)	Nanodrop 1000 Spectrophotometer
pH meter	Thermo scientific (MA, USA)	Orion 3 star

Table 3.3 continued

Pipettes	Gilson (WI, USA)	Pipetman kit (2 to 1000 µL pipettes)
Robot for *in meso* crystallisation	Sias AG (Hombrechtikon, Switzerland)	Xantus *in meso* 8-tip
Robot for *in surfo* crystallisation	TTP Labtech (Hertfordshire, UK)	Mosquito
Syringe for *in meso* crystallisation	Hamilton	Hamilton 100 mL syringe
Tabletop minicentrifuge	Spectrafuge (TX, USA)	24D

3.3 Methods

3.3.1 AlgE *In meso* crystallisation trials in 9.9 MAG

The method described below can be found in published work (Caffrey and Cherezov 2009, Caffrey and Porter 2010, Li *et al.* 2012, Li *et al.* 2012).

3.3.1.1 Cubic phase preparation

- Remove AlgE protein solution in section 2.3.20 from - 80 °C freezer and thaw it quickly in hands. Spin the sample in benchtop centrifuge at 16,300 g and room temperature for 1 min, in order to remove any denatured protein due to the freezing and thawing procedure. Room temperature, if not specified, is not measured and not known, with the possible range from 18 °C to 23 °C. There is usually no obvious pellet at the bottom of the tube and based on this observation, the concentration of AlgE protein solution is considered to be 20 mg / mL as when it was stored. Transfer the supernatant into a new Eppendorf tube and leave it on ice immediately.
- Note: For commercial screen trials in 9.9 MAG, the AlgE concentration was 20 mg / mL. But for the commercial trials in other novel MAGs, the AlgE concentration was 10 mg / mL. And AlgE is always in a solution in buffer E, which has 50 mM Tris-HCl, pH 7.2, 150 mM NaCl, 0.1 %(w/v) LDAO.

- Remove 9.9 MAG from - 80 °C freezer and incubate it in 37 °C incubator.

- Clean two Hamilton 100 μL syringes (example shown in Figure 3.4) with methanol and then dry it under nitrogen gas. Measure the mass of one empty syringe with plunger inside by Mettler Toledo analytical balance and record the number. The normal weight of an empty syringe is about 13.3 g. Use pippete to transfer AlgE protein solution into the syringe from the front as shown in Figure 3.4. Measure the mass of the syringe with plunger again in analytical balance (Mettler Toledo) and record the number. Calculate how much protein solution is added into the syringe. The normal weight of 10 μL AlgE protein is approximately 9.4 mg.

- Measure the mass of another empty syringe with plunger in the same way and record the number. Transfer melted 9.9 MAG from 37 °C incubator to room temperature. Then transfer lipid into the empty syringe by pippete the same way as AlgE solution.

Measure how much lipid is added as above and calculate the ratio of protein solution to lipid. Ideally, the ratio should be kept at about 2 : 3 (w : w) as at this ratio and 20 °C 9.9 MAG forms cubic phase (Luzzati *et al.* 1968). If the ratio is below 2 : 3 (w : w), suck out lipid from the front by pippete from the front end of syringe. If the ratio is above 2 : 3 (w : w), add more lipid the same way as described. Because of the accuracy limit of human and pippete, 5 % error is accepted.

- Note: For other novel MAGs that are used in this chapter, 1 : 1 (w : w) ratio is used instead of 2 : 3 (w : w) as their phase diagrams are different from 9.9 MAG (Qiu and Caffrey 2000, Misquitta *et al.* 2004).

- Remove air gap by pushing plungers in both syringes up to the point where solution or lipid is about to come out from the front end of syringe. Connect ferrule and coupler to the syringe that contains lipid. Remove air inside ferrule and coupler by pushing the plunger up until lipid is about to come out from the front end of coupler. Then connect two syringes with coupler and ferrules as shown in Figure 3.4. Mix protein solution and lipid by gently pushing the plungers back and forth until a transparent cubic phase is formed. Usually, at room temperature, the mixing process takes about 1 min and one can see through the gel-like cubic phase in syringe and read the scales on the other side. This observation is true to all the lipids that I have used in this chapter except for 7.7 MAG. Under room temperature, at 1 : 1 (w : w) ratio, it is never transparent although cubic phase is considered to have been achieved according to its phase diagram (Misquitta *et al.* 2004).

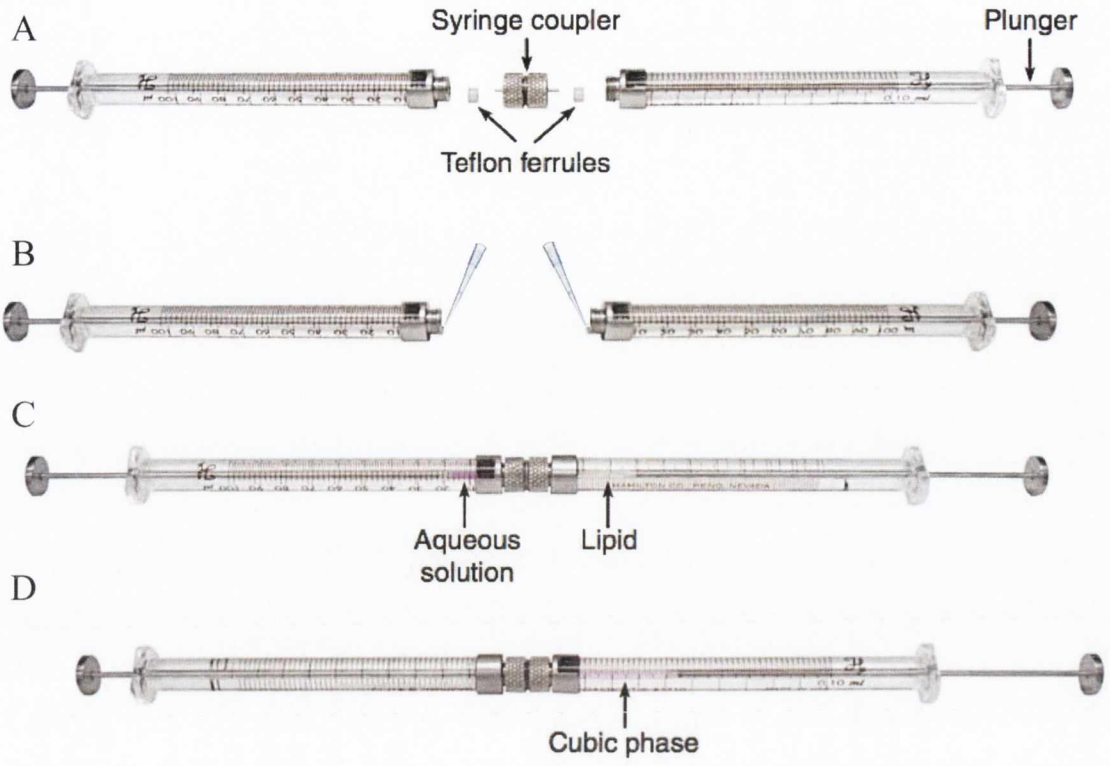


Figure 3.4 The lipid mixing device. The coupled syringe mixer is shown (A) disassembled and assembled (B, C). In (B) the mixer is shown loaded and ready for mixing with lipid (white) in one syringe and bacteriorhodopsin solution (pink) in the other. The two pippete tips show the front of syringes where samples are loaded or sucked out. The homogenized, protein-loaded mesophase (light pink) in the syringe on the right is shown in (C). Copied from (Caffrey and Cherezov 2009).

3.3.1.2 Commercial screen trial

- Flush the base glass plates with MilliQ water carefully on both sides and dry with Kimwipe tissue. Clean one side of plates with RainX to make the plate hydrophobic so that crystallisation solution can stay in the center of the well instead of spreading into the spacer. Air-dry the plate and clean the surface again with MilliQ water before drying with Kimwipe tissue.

- Remove the bottom layer of double-stick 96-well spacer and paste the spacer onto the clean plates. The base plate with spacer can be seen in Figure 3.5. Remove the top layer of the spacer before crystallisation trial and transfer the plate to the rack of *in meso* 8 tip robot in 20 °C room, which is shown in Figure 3.6.
- Note 1: The temperature of 20 °C room controlled by air-conditioning. It is measured by thermometer beside the robot to be within the range of 18 °C to 21 °C.
- Note 2: The spacers used in this project are usually 5 mm or 6 mm in diameter and 140 μm the thickness. The full size of spacer is 112 mm X 75 mm and it is made from 3M TAPE 9500PC roll.

- Set up the crystallisation plates using *in meso* 8 tip robot in 20 °C room. The humidity of the room is not measured and not known. In each well of the plates, Arm 2 in Figure 3.4 dispenses 50 nL mesophase onto the crystallisation plate and Arm 1 lays 800 nL precipitant solution from commercial screens onto the top of mesophase. Crystallisation conditions start with A1 well and end with H12 well for each commercial screen. Then a reverse plate is set up in the same way but crystallisation conditions start with H12 well and end with A1 well for each commercial screen. Eleven commercial screen kits used in this experiment are listed below: Crystal Screen HT, Grid Screen, Index HT, MemFac HT, SaltRX, MemStar MemSys HT96, MemGold™ HT96, JBScreen PACT, Wizard I & II and JBScreen Membrane.
- Note 1: The usual volume for mesophase is 50 nL but it can vary depending the purpose of experiment and projects. In this project, the following volumes have been tried as one of the optimisation strategies: 10 nL, 25 nL, 50 nL and 100 nL. Crystals that grow from different volume of cubic phase show no difference in terms of size or diffraction quality.
- Note 2: The robot was calibrated by Dr. Bo Sun regularly from 2006 to 2009 and then by Dr. Joseph Lyons from 2009 till now. It usually takes 7 min for one 96-well run and the reproducibility is considered to be good as diffraction-quality crystals are reproducible.

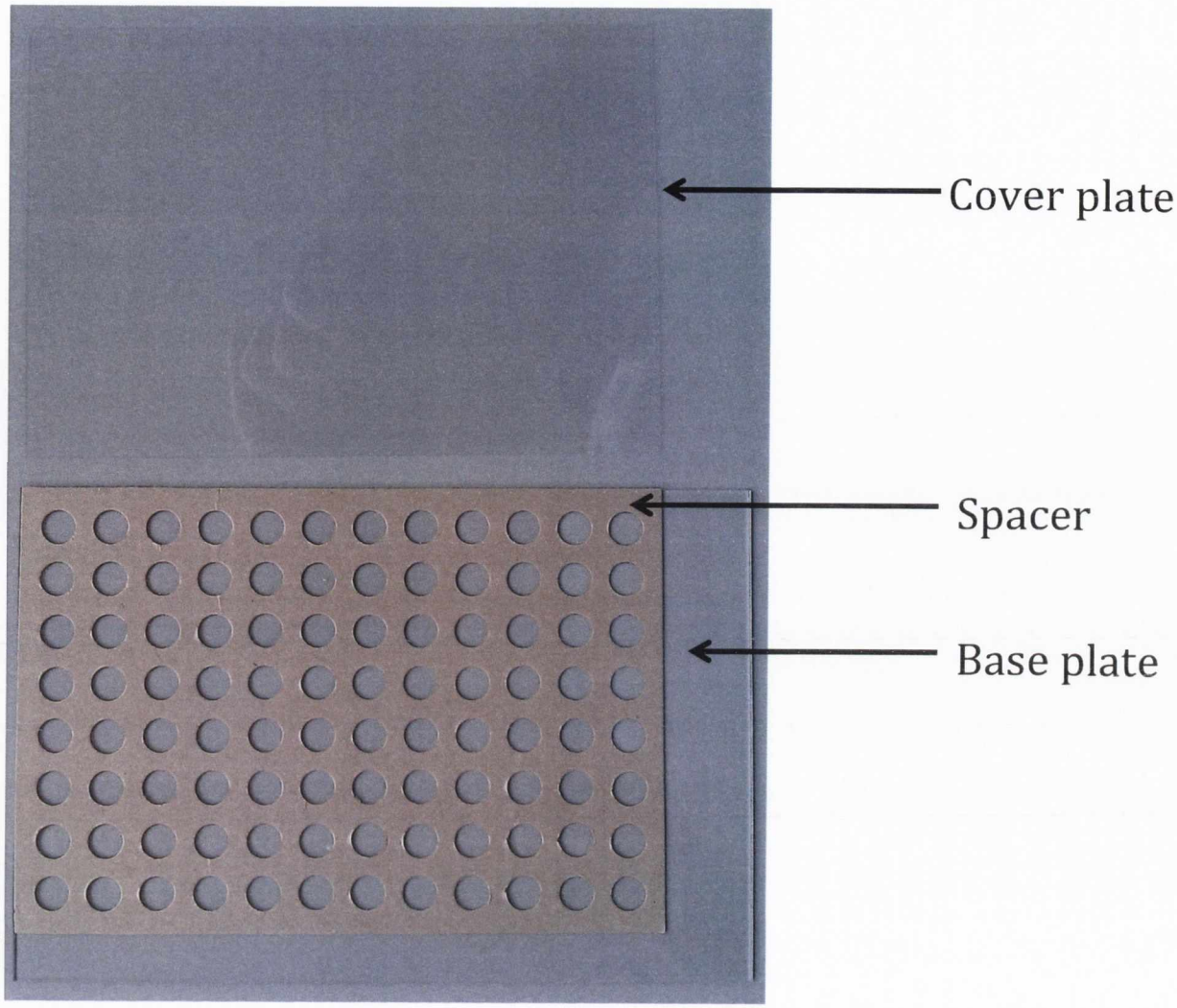


Figure 3.5 The set up of *in meso* crystallisation plate. In panel A, the top arrow points to an example of cover plate, the middle arrow points to spacer and round wells are crystallisation well. The brown surface of spacer will be removed during the experiment and then cover plate can be laid over the spacer to seal crystallisation wells. The bottom arrow points to the base plate. In panel B, the roller used to press and seal the sandwich system is shown.

- When all 96 crystallisation wells are filled with 800 nL precipitant solution on top of 50 nL mesophase, carefully lay a cover plate on top of the crystallisation plate to seal all the wells. Use a roller to press every corner of the cover plate a few times and make sure the sealing is airtight.
- Note: The cover plate is not washed by MilliQ water or RainX in this project.

- Transfer all plates into Formulatrix imager that is fixed at 20 °C. Set the imaging schedule as: day 0, 1, 3, 5, 7, 14, 21 and 30. Both normal light and cross polarized lights are used when the imager take pictures of each well. In order to capture the images for different depths in each well, the imager takes 7 slices of images with 0.04 motor-step between neighbouring slices. Then imager generates a composite picture by combining all slices together.
- Note: The internal temperature of imager is controlled by its separate air-conditioning and the range of temperature is 19.8 °C to 20.2°C

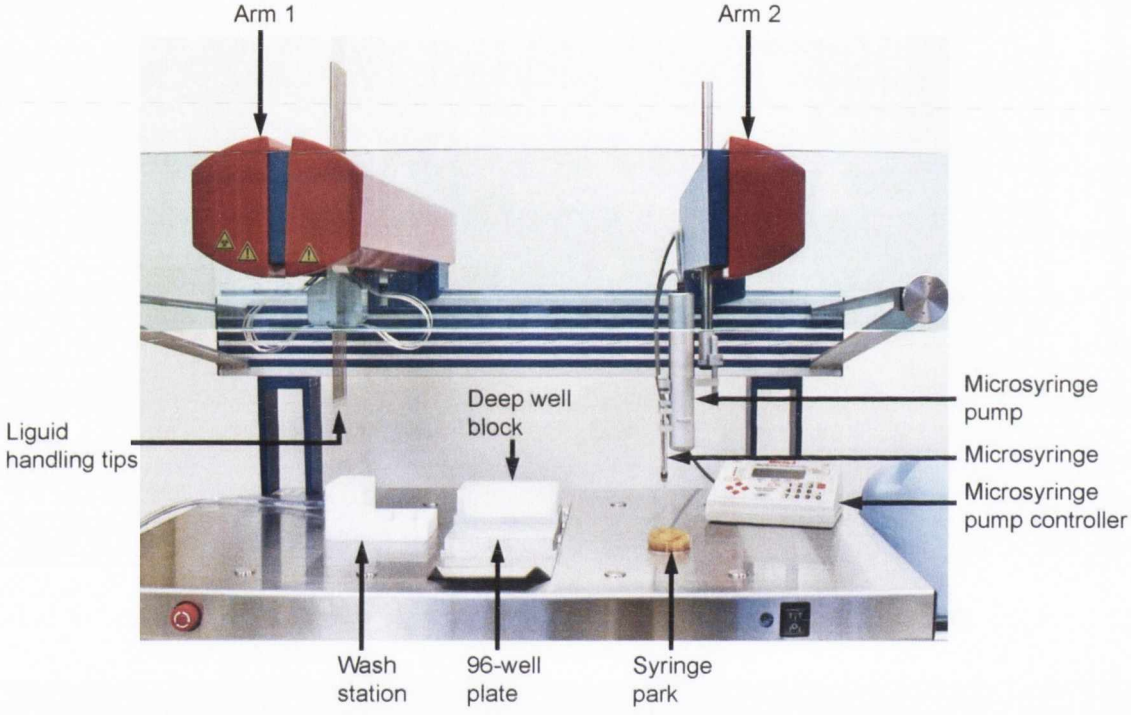


Figure 3.6 The *in meso* crystallisation robot. The robot has two dispensing arms. Arm 1 includes eight tips and is used for handling liquid precipitant solutions. Arm 2 supports a microsyringe for dispensing the protein/lipid dispersion. The microsyringe dispenses mesophase by the action of a microsyringe pump, which in turn is driven by a controller. When not in use, arm 2 goes to a park position where the microsyringe needle tip is placed in a moist sponge. Also shown in the figure are a 96-well glass plate, a 96-deep-well block of precipitant solutions and a station where the tips on arm 1 are washed. (Caffrey and Cherezov 2009)

3.3.1.3 Optimisation trial

- Prepare optimisation screen kits on the basis of the initial hits found in Section 3.3.1.2. The pHs of solution, salt concentration and precipitant concentration are varied in each of the 96 wells in optimisation screen kits in order to get well-diffracting crystals. In total, three kits are prepared and named as AlgE screen kit 1, 2 and 3. The contents are listed in Appendix 2 – 4.

- Prepare the cubic phase and set up the crystallisation plates as described in Section 3.3.1.1 and 3.3.1.2. Look through the wells on crystallisation plates and find out the best condition for AlgE / 9.9 MAG *in meso* crystallisation. In the case of AlgE, best conditions yield big crystals (usually at least 50 μm) reproducibly.

3.3.2 AlgE *In meso* crystallisation trials in novel MAGs

3.3.2.1 Commercial screen trials

As mentioned in section 3.1.4, novel MAGs are used in this chapter: 7.7 MAG, 7.8 MAG, 7.9 MAG, 8.8 MAG and 8.9 MAG.

- All the procedures in this section are the same as described in Section 3.3.1.1 and 3.3.1.2 except that: 1. The cubic phase is prepared with AlgE and individual MAGs that are listed above and the protein to lipid ratio is kept at 1 : 1 (w : w); 2. The AlgE protein concentration is 10 mg / mL. Usually transparent cubic phase is observed but in the case of 7.7 MAG transparent mesophase is never achieved.
- Note: The experiment will still be continued even when there is never transparent mesophase for 7.7 MAG.

3.3.2.2 Optimisation trials

- Prepare optimisation screen kit on the basis of the initial hits found in Section 3.3.2.1. The pHs of solution, salt concentration and precipitant concentration are varied in each of the 96 wells in optimisation screen kit in order to get well-diffracting crystals. One kit is prepared and named as AlgE screen kit 5. The contents are listed in appendix 5.

The optimization screen made for 9.9 MAG, AlgE screen kit 1, is also used in this section.

- Prepare cubic phase and set up the crystallisation plates as described in section 3.3.1.1 and 3.3.12. Look through the wells on crystallisation plates and find out the best condition for AlgE / 9.9 MAG *in meso* crystallisation. Again, in the case of AlgE, best conditions yield big crystals (usually at least 50 μm) reproducibly.

3.3.3 AlgE *in surfo* sitting-drop crystallisation trials

- Remove AlgE protein solution in section 2.3.20 from - 80 °C freezer and thaw it quickly in hands. Spin the sample in benchtop centrifuge at 16,300 g and room temperature for 1 min, in order to remove any denatured protein due to the freezing and thawing procedure. There is usually no obvious pellet at the bottom of the tube and based on this observation, the concentration of AlgE protein solution is considered to be 20 mg / mL as when it was stored. Transfer the supernatant into a new Eppendorf tube and leave it on ice immediately.

- Set up *in surfo* sitting-drop crystallisation trials on MRC plate (Figure 3.7) using Douglas Oryx 8 Robot in 20 °C room.
- Note: The temperature of 20 °C room controlled by air-conditioning. It is measured by thermometer on the wall to be within the range of 18 °C to 21 °C.

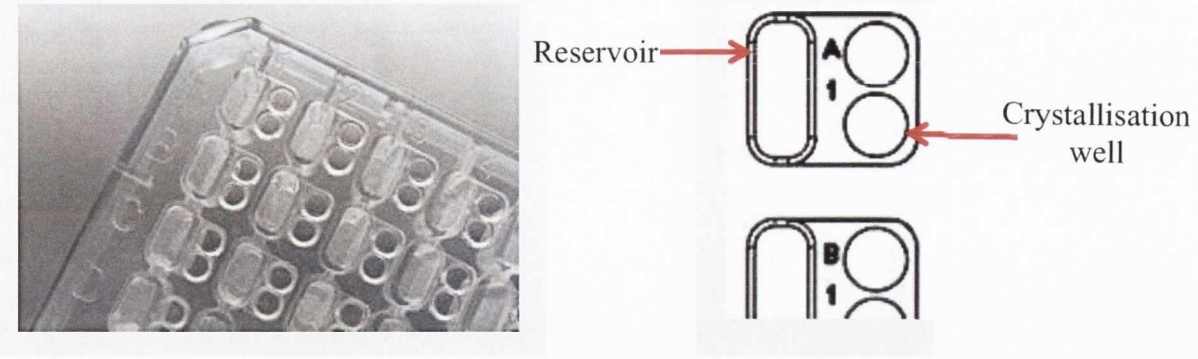


Figure 3.7 Layout of the SwiisC1 MRC 2-well plate used for *in surfo* sitting-drop crystallisation trials. The picture on the left is an example of real object and the picture on the right is cartoon representation of the real object. Reservoir and crystallisation wells are labelled by red arrow. The reservoir can hold up to 80 μL of precipitant solution. Redrawn from website of Douglas Instrument Ltd.: http://www.douglas.co.uk/products.htm#DICONSUMABLES.

- In each condition, 80 μL precipitant solution is dispensed in the reservoir, 100 nL AlgE protein plus 100 nL precipitant solution dispensed in the well, as shown in Figure 3.4. Precipitant solutions come from the four commercial screens listed: MemStar MemSys HT96, MemGold™ HT96, Wizard I & II and JBScreen Membrane.

- Cover the plates with Crystal Clear Tape. Keep the plates in the imager at 20 °C and all settings for imager are the same as described in Section 3.3.1.2.

3.3.4 Crystal harvesting and data collection at synchrotrons

- Take out plates from imager after the crystals stop growing, which usually take 7 – 14 days. Leave them on the bench of 20 °C room. Label the crystallisation wells with crystals of at least 50 μm with black marker. The arrangement of a single well is represented by cartoon in panel A in Figure 3.8.
- Note: The temperature of 20 °C room controlled by air-conditioning. It is measured by thermometer on the wall to be within the range of 18 °C to 21 °C.

- Score circle 1 carefully on the cover plate over the spacer close to the edge of chosen well with a glass cutter and the glass plate not being broken at this stage, as shown by

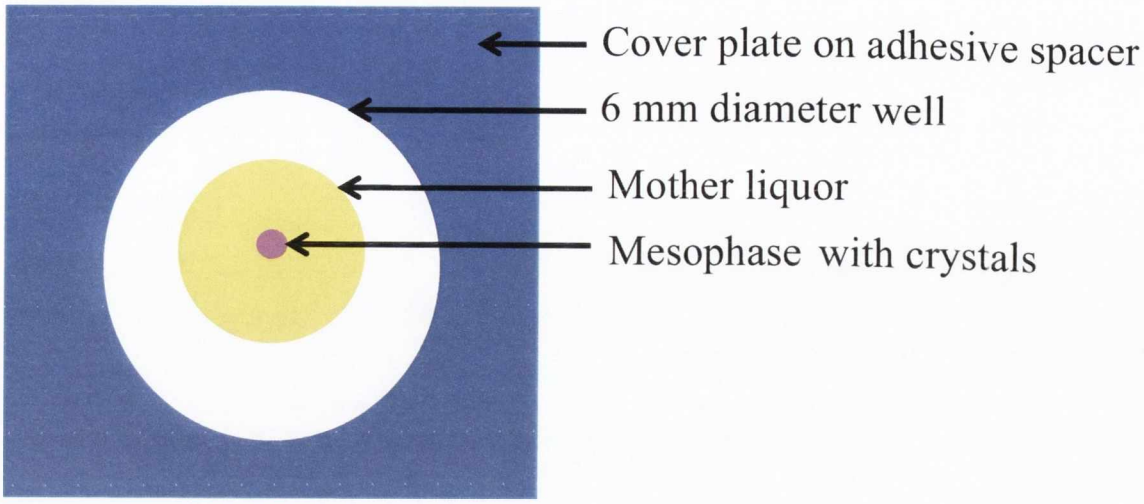
the red dotted line in panel B in Figure 3.8. Score circle 2 just outside circle 1 as shown by the purple dotted line in panel C in Figure 3.8. Break the glass plate along circle 2 as in panel D in Figure 3.8 by applying force on glass with glass cutter. The existence of circle 1 is going to stop the force applied on the circle 2 from impacting the cover glass within circle1 so that this area will remain intact during the cutting. Remove both spacers and glass between circle 1 and circle 2 immediately to prevent them from entering the crystallisation drop. At this stage, the cover plate within the region of circle 1 should not fall and crash the mesophase drop. Then lift the cover plate within the region of circle 1 with tweezers and expose mesophase drop as in panel E in Figure 3.8. Mother liquor is usually left in the well at this stage. But if mother liquor is sucked away by spacers or the lifted cover plate, use pippete to add 1 μL of the crystallisation solution onto the top of mesophase drop immediately to keep moist.

- At this stage, crystals are visualized under harvest microscope with both normal light and cross polarized light. Use mitegen harvesting loops to fish crystals that were wrapped in viscous cubic phase. Snap freeze the loop with liquid nitrogen and put into storage puck. Leave storage pucks in liquid nitrogen dewars and then ship out to synchrotron sources for X-ray diffraction experiment (by FEDEX).

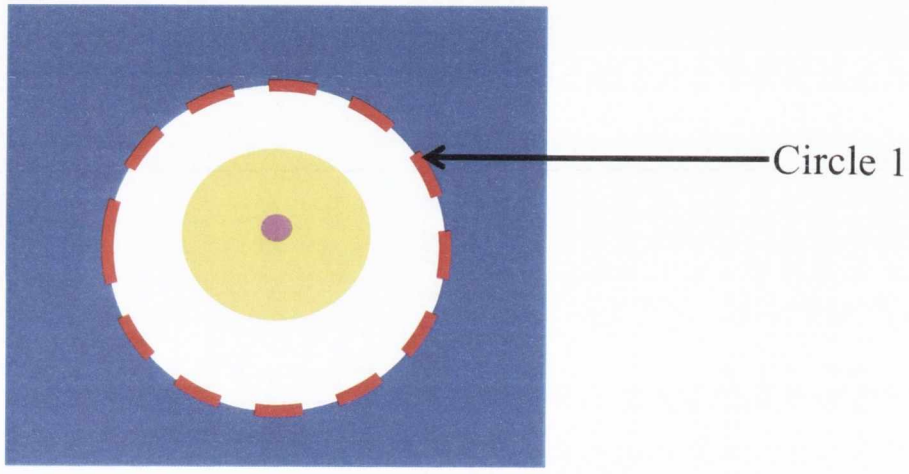
- X-ray diffraction experiments are carried out either on the 23-ID-D beamline of the General Medicine and Cancer Institutes Collaborative Access Team (GM/CA-CAT) at the Advanced Photon Source (APS), Argonne, Illinois, USA or the I24 beamline at the Diamond Light Source (DLS), Didcot, Oxford, UK. Data are acquired using a 10 μm or 20 μm minibeam at GM/CA-CAT, while a 10 μm and a 20 μm microfocus beam is used at DLS, respectively. Attenuated (10X) images are used to locate, in many cases, invisible crystals buried in mesophase and to centre on highly ordered regions of the crystals (Cherezov *et al.* 2009, Hilgart *et al.* 2011). Then diffraction data are collected at 12.0 kev energy.

- Dr. David Aragão and Dr. Valerie Pye did most of the data collection from 2008 to 2012 and data processing work throughout the project. I did data collection from 2011 to 2012 but no data processing work.

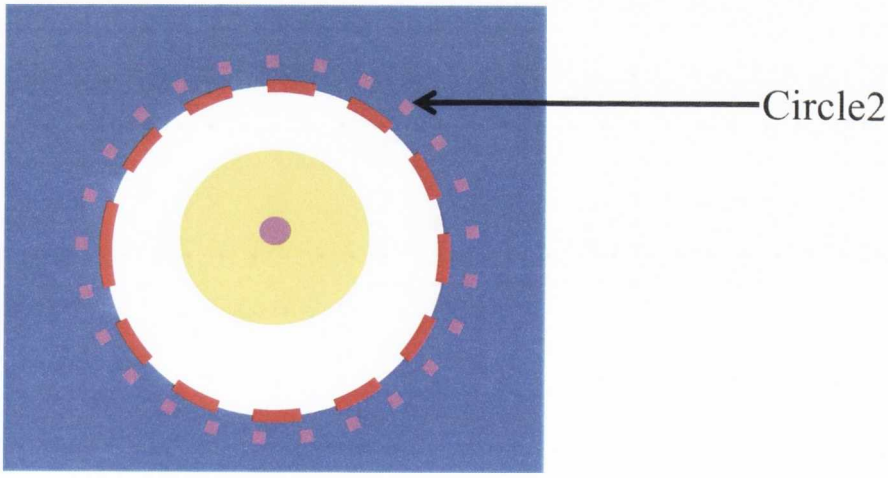
A


Cover plate on adhesive spacer
6 mm diameter well
Mother liquor
Mesophase with crystals

B


Circle 1

C


Circle2

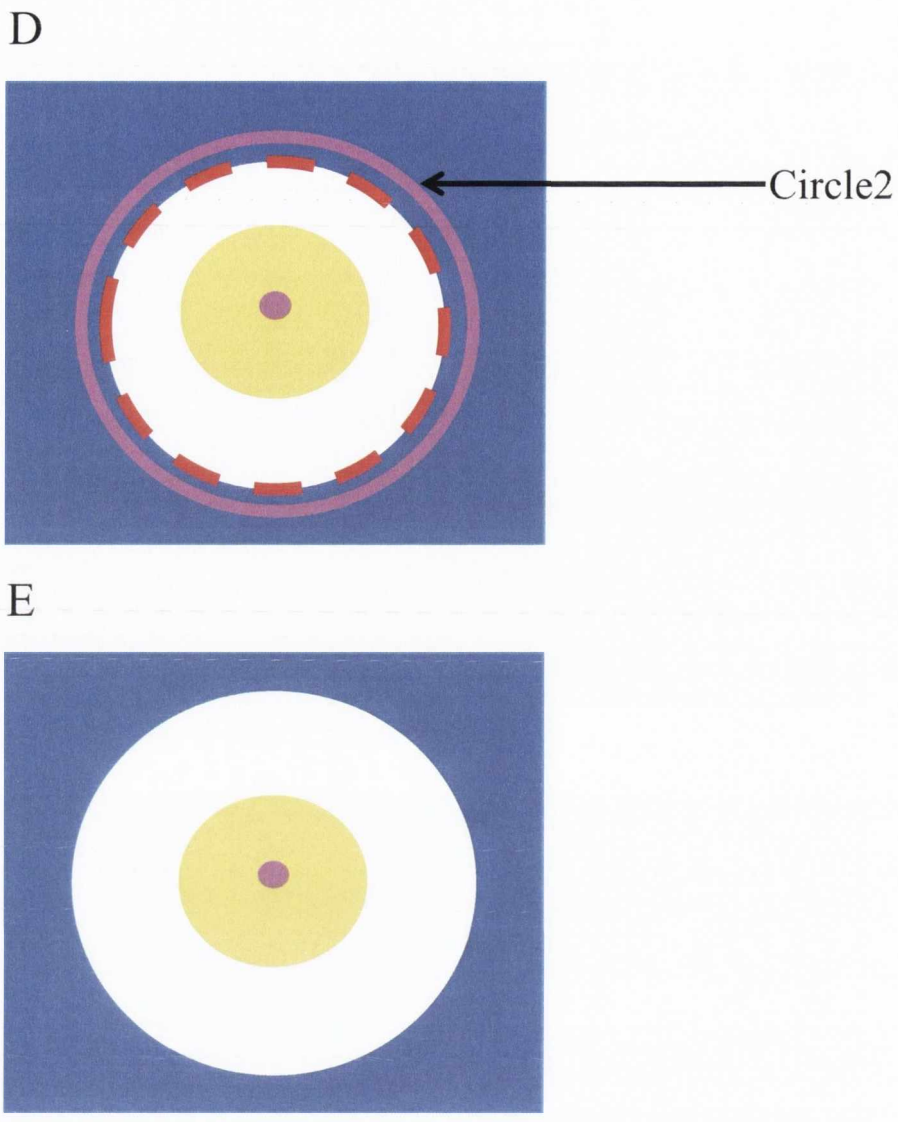


Figure 3.8 Cartoon representation of harvesting procedure from *in meso* crystallisation plate. A: representation of single well of *in meso* crystallisation plate. The diameter of a cubic phase bolus is usually 550 μm, that of precipitant solution usually 1.8 mm and that of a well usually 5-6 mm. Cover plate, well, mother liquor and mesophase are colour coded and used in the rest of panels. B: Circle 1 is scored as indicated by the arrow on the right-hand side. C: Circle 2 is scored as indicated by the arrow on the right-hand side. D: Cover plate is broken along circle 2 as indicated by the arrow on the right-hand side. E: Mesophase with mother liquor is exposed after cover plate is removed.

3.3.5 MIR phasing work for AlgE structure determination

3.3.5.1 Native AlgE crystals production

- Produce crystal production the same as described in Section 3.3.1.1 and 3.3.1.2. For this purpose, AlgE screen kit 5 from Section 3.3.2.2 is always used because it yields AlgE crystals of at least 50 μm most reproducibly and the crystals have the second best diffraction quality of all crystal forms, which is 2.8 Å as shown in result section 3.4.2.2.

- Crystals usually grow to maximum size after 7 to 14 days at 20 °C and are then used for heavy atom soaking experiment.

3.3.5.2 Heavy atom solution preparation

To find out which heavy atoms could be used to solve AlgE phase and eventually yield structure information, 17 coloured heavy atoms are chosen in initial screening experiment in order for easy visualisation for the effect of each heavy atom on crystals. The idea is to soak AlgE crystals in mesophase with these coloured chemicals for 1 h and then back-soak them with heavy-atom-free solutions for 2 min at room temperature and back-soaking is to remove non-specific binding between chemicals and protein crystals, if there was any. Then observe the colour changes of crystals under microscope. If there is colour change in AlgE crystals and the colour is not the same as in mesophase background, the heavy atom is considered a possible choice to determine AlgE phase information by MIR method.

- A list of chemicals and their concentrations used in the initial screening is shown below in Table 3.4.

- Measure he weight of heavy atom chemicals was measured on Mettler Toledo analytical balance in a dedicated working fume hood and then transfer it into labelled Eppendorf tubes.

- Add the precipitant solution used for AlgE *in meso* crystallisation for 7.8 MAG (0.1 M sodium cacodylate, pH 6.5, 1.0 M sodium acetate) into each tube to dissolve the chemicals and the final concentration of each chemical is listed in Table 3.4.
- Note: Usually, cacodylate is considered to be a bad choice for crystallisation work because the arsenic in it can cause extra radiation damage and make crystals die faster under X-ray beam. It is recommended to remove cacodylate by back-soaking after obtaining crystals (Garman 2010). However, in the case of AlgE, I haven't observed large radiation damage done to crystals by the existence of cacodylate as the AlgE crystals survive well under X-ray beam and the best diffraction data set to 1.9 Å is collected from a single crystal.

- Keep heavy atom solutions in a box sealed with tape in fume hood for future use.

Table 3.4 List of heavy atoms with working concentrations in colour screening experiment.

Name of chemical	Concentration (mM)	Colour
$AuCl_3$	20	Yellow
$HAuCl_3$	20	Yellow
Ta_6Br_{12}	10	Green
K_2PtCl_4	10	Brown
K_2PtBr_6	10	Brown
K_2PtBr_4	10	Brown
$Pt(NH_3)_2(NO_2)_2$	10	Brown
$(NH_4)_2PtCl_2$	10	Brown
$OsCl_3$	20	Black
K_2OsCl_6	20	Black
$(NH_4)_2IrCl6$	20	Brown
$IrCl_3$	20	Brown
$(NH_4)WS_4$	20	Light yellow
$HoCl_3$	20	Yellow
$NdCl_3$	20	Red purple
$PrCl_3$	20	Black
$CdCl_2$	10	Brown

3.3.5.3 Colour screening experiment

- Take out plates from imager and leave them on the bench of 20 °C room. Label the crystallisation wells with crystals of at least 50 μm with black marker.
- Note: The temperature of 20 °C room controlled by air-conditioning. It is measured by thermometer on the wall to be within the range of 18 °C to 21 °C.

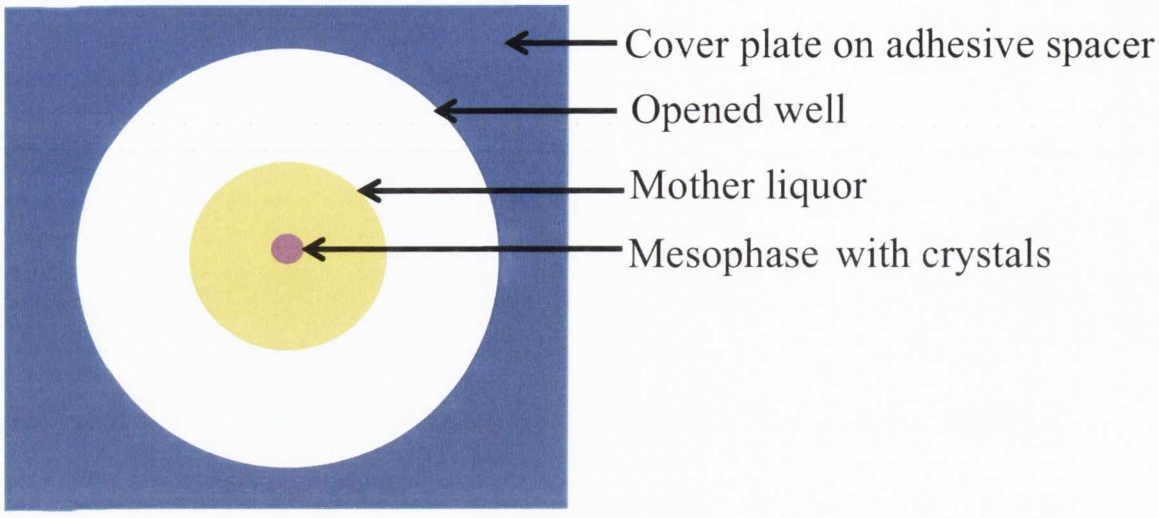
- Take out heavy atom solutions from the dedicated fume hood and place them onto clean tissue on the bench in the 20 °C room. This is to avoid contamination from heavy atom solutions.

- Open the crystallisation wells as described in Section 3.3.4. The opened well is represented by cartoon in panel A of Figure 3.9. If the mother liquor is left in the well, use pippete to remove it, as in panel B of Figure 3.9. Use pippete and add 1 μL heavy atom solution that is prepared in section 3.3.5.2 onto the mesophase drop as in panel C in Figure 3.9. Then cover the well with a new glass slide and incubate the system at 20 °C for 1 h, as shown in panel D in Figure 3.9. Heavy atom is coloured in red. The system is not sealed by tape or any other way but it can keep hydration and stable cubic phase for around 4 h according to my experience.

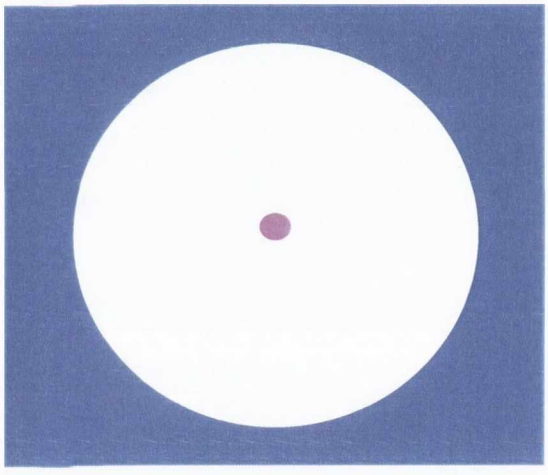
- After 1 h incubation at room temperature, remove glass slide by tweezers and take out heavy atom solution by pippete. Add 1 μL native precipitant solution without heavy atom to the exposed mesophase immediately. Again cover the well by glass slide and incubate for 2 min, as shown in panel E in Figure 3.9. The final precipitant solution after back-soaking is coloured in brown.

- Observe the crystals under microscope and the colour change is judged by eye.

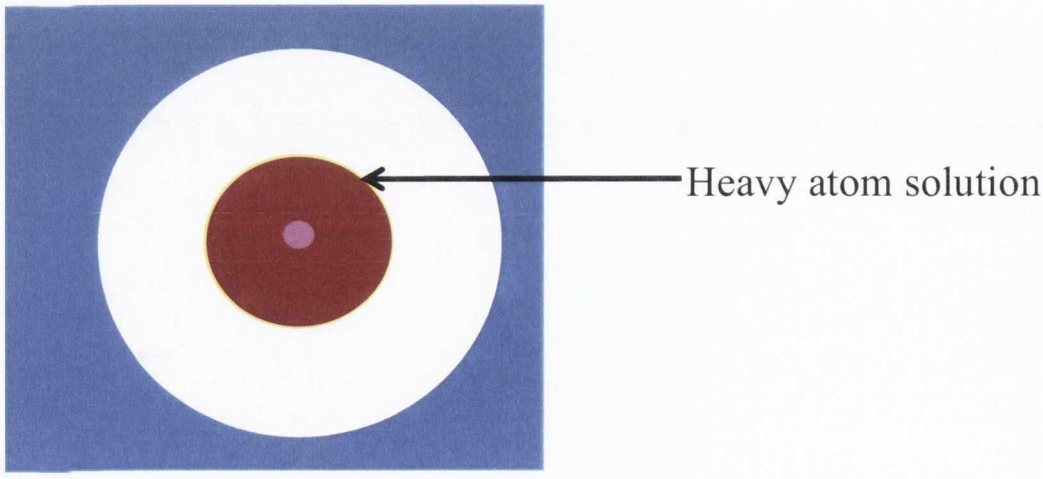
A


Cover plate on adhesive spacer
Opened well
Mother liquor
Mesophase with crystals

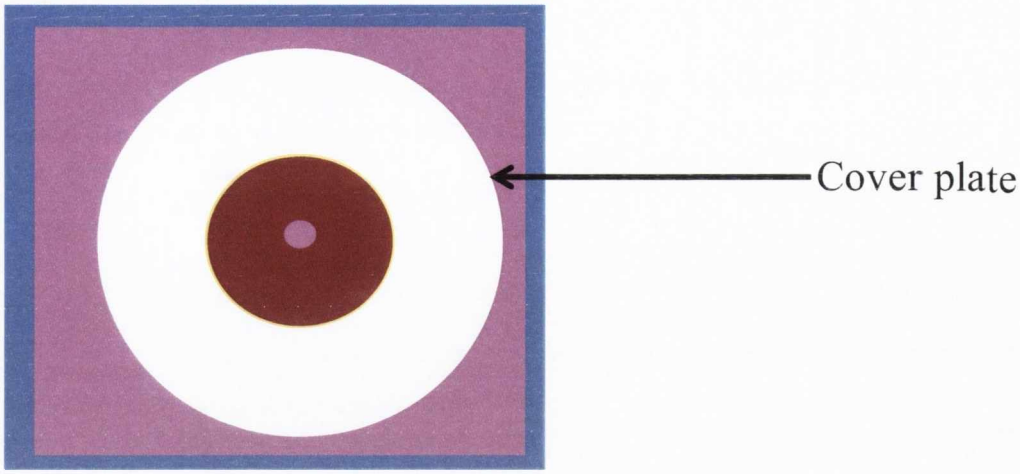
B



C


Heavy atom solution

D



E

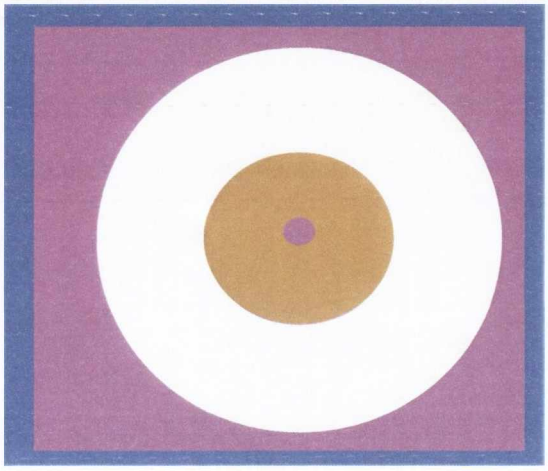


Figure 3.9 Cartoon representation of heavy atom soaking procedure from *in meso* crystallisation plate. A: representation of an opened well of *in meso* crystallisation plate for heavy atom soaking. Cover plate, well, mother liquor and mesophase are colour coded and used in the rest of panels. B: Mother liquor is removed and mesophase is exposed to air. C: Heavy atom solution is added to the top of mesophase immediately. D: The mesophase with heavy atom solution is covered by cover plate, which is indicated by the purple colour and arrow on the right-hand side. E: Replace heavy atom solution with fresh precipitant solution for back-soaking and cover again with cover plate.

3.3.5.4 Preliminary study on phasing power of K_2OsCl_6 and $(NH_4)_2IrCl_6$

According to the result of colour screening experiment as in result part, K_2OsCl_6 and $(NH_4)_2IrCl_6$ are identified as promising choices. Experimental phasing power of each chemical in AlgE structure work needs to be tested.

- Soak AlgE native crystals with K_2OsCl_6 and $(NH_4)_2IrCl_6$ separately as described above. The concentration of heavy atoms and soaking time are listed in Table 3.5. In this experiment, back-soaking is not performed to ensure the maximum binding between heavy atom and protein for maximum phasing power determination under these conditions.

- After soaking, harvest the crystals as described in section 3.3.4 and ship out to synchrotrons for X-ray diffraction test. These crystals were tested at GM/CA-CAT 23 ID-B beamline at APS in 10th March and 26th Nov of 2010 at 12.0 keV energy to find out the maximum phasing power of both heavy atoms. Dr. Valerie Pye did all of data collection and processing work.

Table 3.5 Soaking conditions used for shorter soak approach to determine the phasing power of K_2OsCl_6 and $(NH_4)_2IrCl_6$.

Chemical	Soaking condition	
	Concentration (mM)	Time (min)
K_2OsCl_6	10	60
K_2OsCl_6	10	45
K_2OsCl_6	5	30
$(NH_4)_2IrCl_6$	20	60
$(NH_4)_2IrCl_6$	20	45
$(NH_4)_2IrCl_6$	10	30

3.3.5.5 Systematic study on K_2OsCl_6's phasing power

Based on the result of preliminary study on the phasing power of K_2OsCl_6 and $(NH_4)_2IrCl_6$, $(NH_4)_2IrCl_6$ is abandoned because of its low phasing power and high damage to crystal quality. In addition, a systematic study is planned to study the phasing power of K_2OsCl_6.

- Soak and back-soak AlgE crystals in the different conditions listed in Table 3.6 and harvest the crystals. The crystals are stored in liquid nitrogen dewar.
- Note: Because *in surfo* AlgE structure came out at the stage, these crystals were never tested at any synchrotron source and the phasing strategy was changed to molecular replacement.

Table 3.6 Conditions used for systematic study on the phasing power of K_2OsCl_6.

K_2OsCl_6 concentration (mM)	Time (min)	
	Soaking	Back-soaking
5	30	0
5	30	2
5	60	0
5	60	2
5	90	0
5	90	2
5	120	0
5	120	2
10	30	0
10	30	2
10	60	0
10	60	2
10	90	0
10	90	2
10	12	0
10	120	2
15	10	0

Table 3.6 continued

15	10	2
15	20	0
15	20	2
15	30	0
15	30	2
15	60	0
15	60	2
15	90	0
15	90	2
15	120	0
15	120	2
20	10	0
20	10	2
20	20	0
20	20	2
20	30	0
20	30	2
20	60	0
20	60	2
20	90	0
20	90	2
20	120	0
20	120	2

3.4 Results and discussion

3.4.1 *In meso* crystals in 9.9 MAG

3.4.1.1 Initial hits from commercial screen trials

In the commercial screening trial in 9.9 MAG, 1,056 conditions were screened. 11 conditions were found to yield possible crystal hits and 6 of them were obvious crystals. Other conditions are microcrystals that are less than 2 μm and can barely be seen. Crystals range from 2 μm microcrystals to 10 μm small needles to 30 μm small plates. The images for the obvious crystals are shown in Figure 3.9 to Figure 3.14. No better images or images under normal light are available because of a data loss on the imager server that affected the whole lab. A summary of all initial hits is listed in Table 3.7.

Table 3.7 Summary of all initial hits in AlgE / 9.9 MAG in meso commercial trials. The condition that yielded 2 D crystal is labelled in red.

Hosting lipid	Conditions that yield initial hits	Maximum crystal size (μm) / shape
9.9 MAG	0.2 M magnesium chloride, 0.1 M HEPES, pH 7.5, 30 %(w/v) iso-propanol	10 / needle
9.9 MAG	0.2 M sodium citrate, 0.1 M Tris-HCl, pH 7.5, 30 %(w/v) PEG400	10 / needle
9.9 MAG	0.1 M sodium acetate, pH 4.6, 0.4 M magnesium formate	10 / needle
9.9 MAG	0.1 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 18 %(w/v) PEG400	30 / plate
9.9 MAG	0.2 M sodium citrate, 0.1 M Tris-HCl, pH 8.5, 30 %(w/v) PEG400	10 / needle
9.9 MAG	0.02 magnesium chloride, 0.1 M sodium citrate, pH 5.6, 20 %(w/v) PEG400	20 / needle
9.9 MAG	0.5 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 2 %(w/v) PEG400	2 / microcrystal

Table 3.7 continued

9.9 MAG	0.1 M sodium citrate, 0.1 M HEPES, pH 7.5, 0.1 M magnesium chloride, 30 %(w/v) PEG400	2 / spot
9.9 MAG	0.1 M sodium citrate, pH 4.6, 10 %(w/v) iso-propanol	2 / spot
9.9 MAG	0.1 M Tris-HCl, pH 8.5, 0.1 M calcium chloride, 25 %(w/v) tert-butanol	2 / spot
9.9 MAG	0.1 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 0.7 M sodium acetate	2 / spot

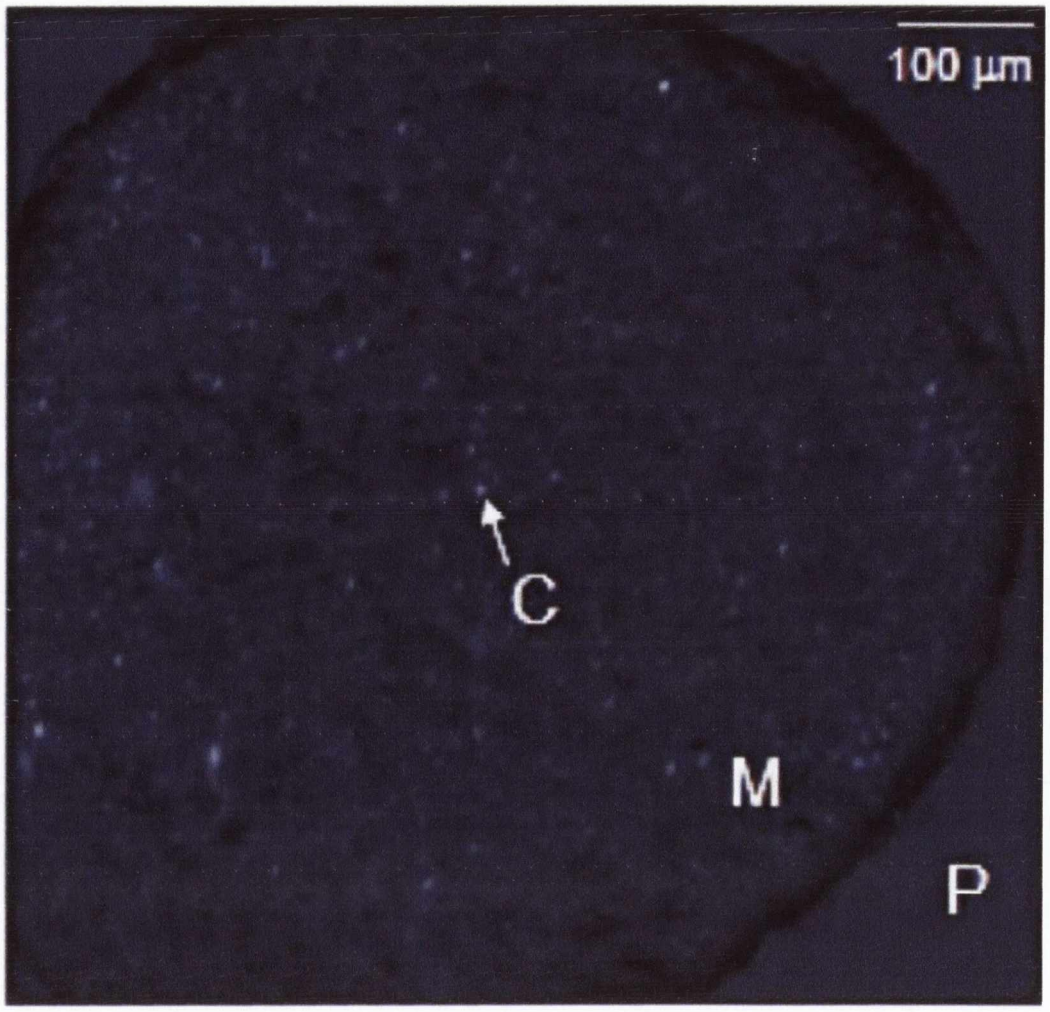


Figure 3.9 Example 1 of initial hits found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 20.0 mg / mL and the crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 21. Crystallisation condition is: 0.2 M magnesium chloride, 0.1 M HEPES, pH 7.5, 30 %(w/v) iso-propanol. Cross polarized light was used to identify possible protein crystals more easily as shining objects in dark background. M: mesophase drop, C: crystals, P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

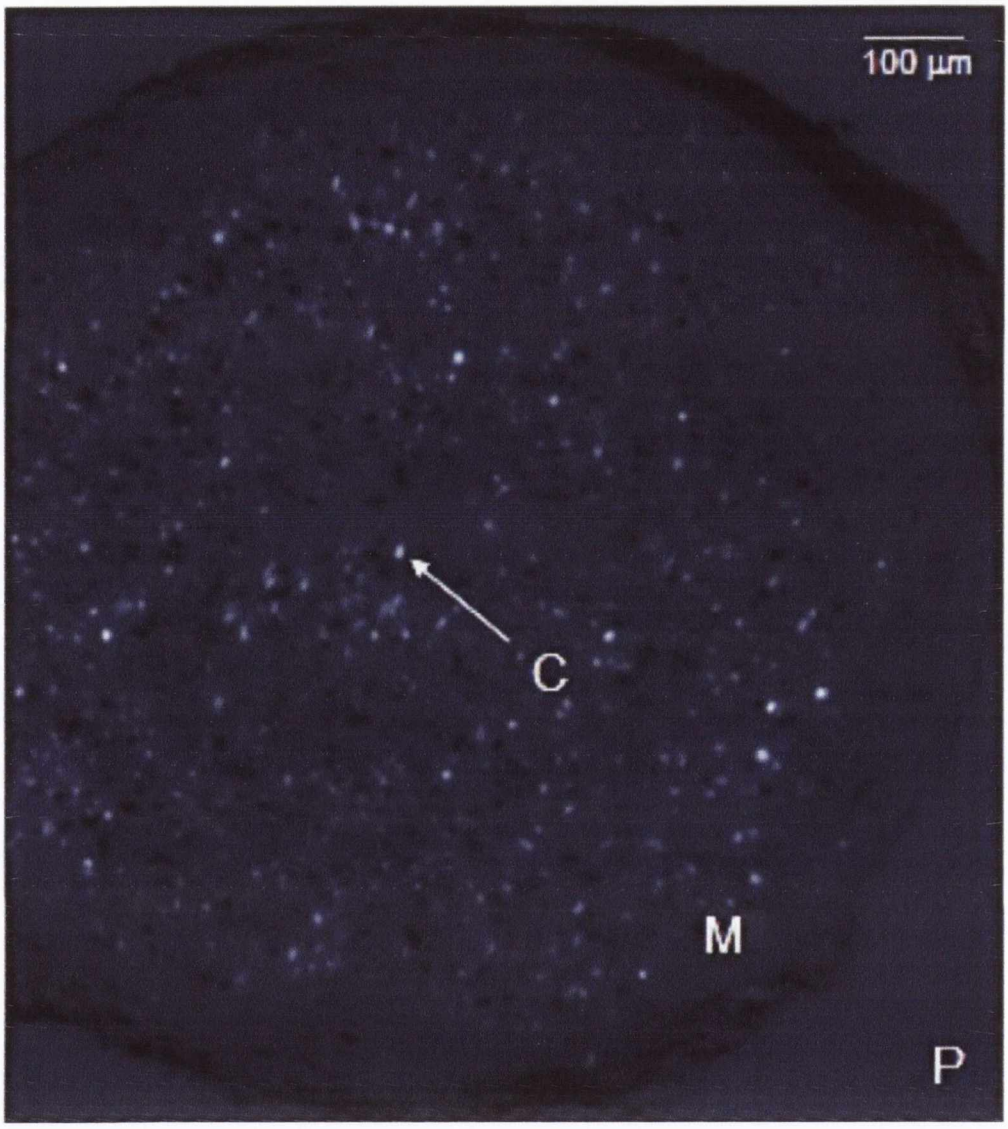


Figure 3.10 Example 2 of initial hits found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 20.0 mg / mL and the crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 21. Crystallisation condition is: 0.2 M sodium citrate, 0.1 M Tris-HCl, pH 7.5, 30 %(w/v) PEG400. Cross polarized light was used to identify possible protein crystals more easily as shining objects in dark background. M: mesophase drop, C: crystals, P: precipitant. In this case, the mesophase remains cubic, as the boundary of it is still clear.

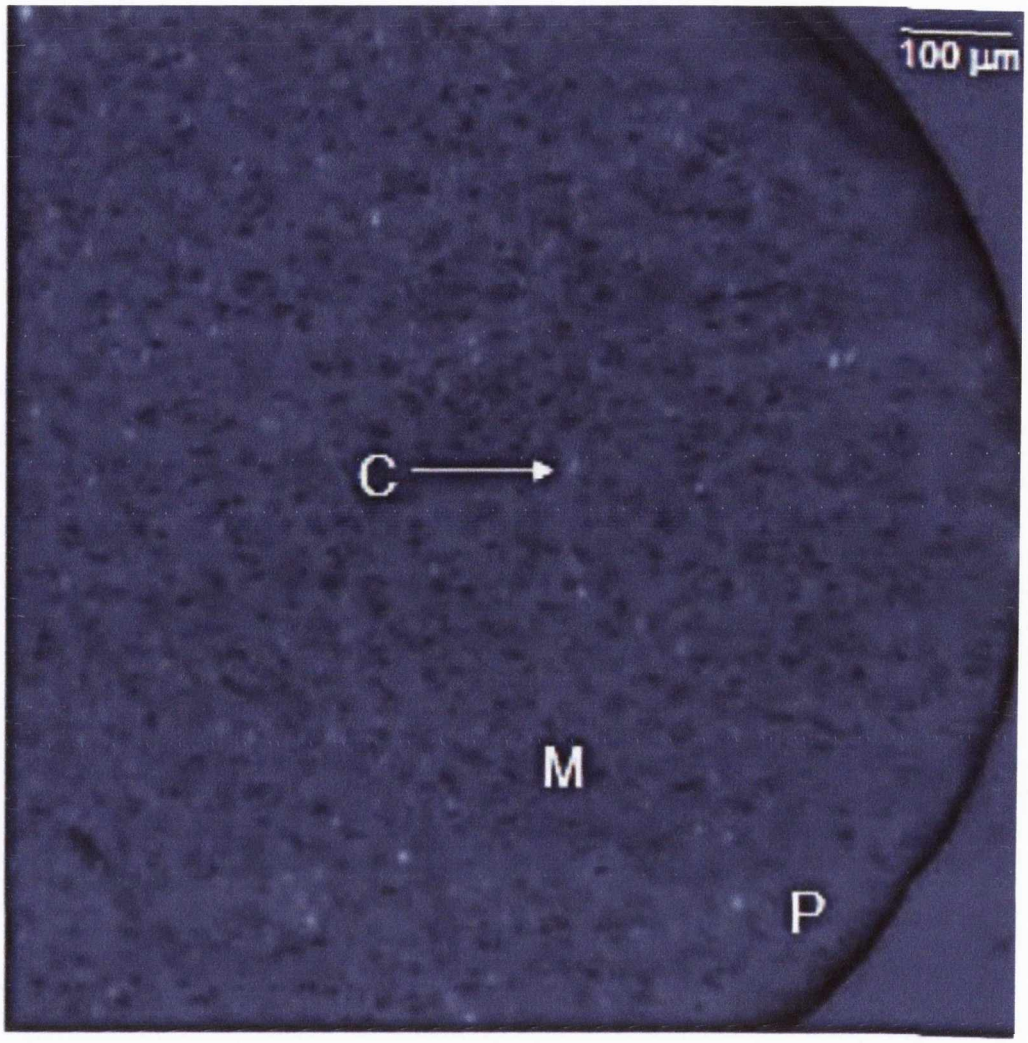


Figure 3.11 Example 3 of initial hits found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 20.0 mg / mL and the crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 21. Crystallisation condition is: 0.1 M sodium acetate, pH 4.6, 0.4 M magnesium formate. Cross polarized light was used to identify possible protein crystals more easily as shining objects in dark background. M: mesophase drop, C: crystals, P: precipitant. In this case, the mesophase remains cubic, as the boundary of it is still clear.

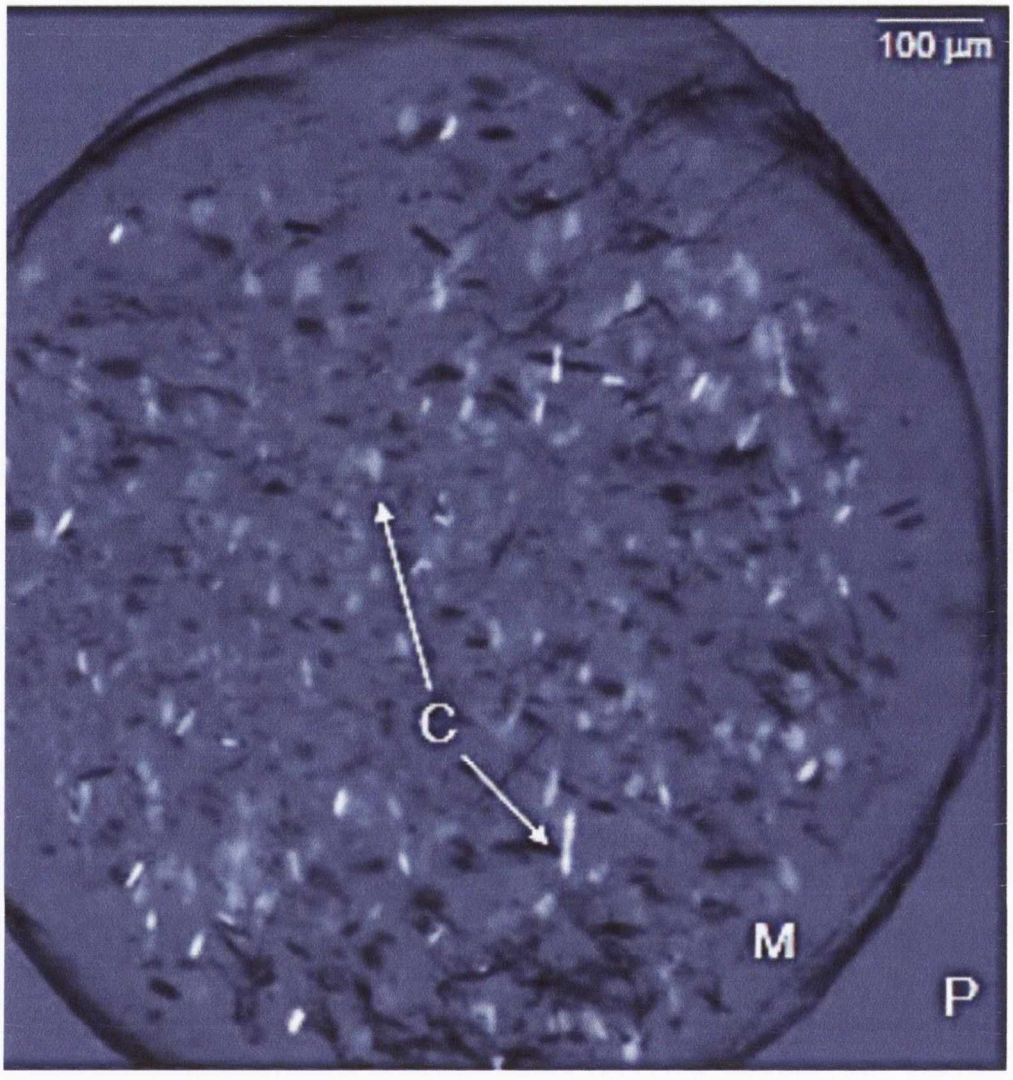


Figure 3.12 Example 4 of initial hits found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 20.0 mg / mL and the crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 21. Crystallisation condition is: 0.1 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 18 %(w/v) PEG400. Cross polarized light was used to identify possible protein crystals more easily as shining objects in dark background. M: mesophase drop, C: crystals, P: precipitant. In this case, the mesophase remains cubic, as the boundary of it is still clear. Under this condition, some crystals showed two dimensions as small plates but most of them are needles.

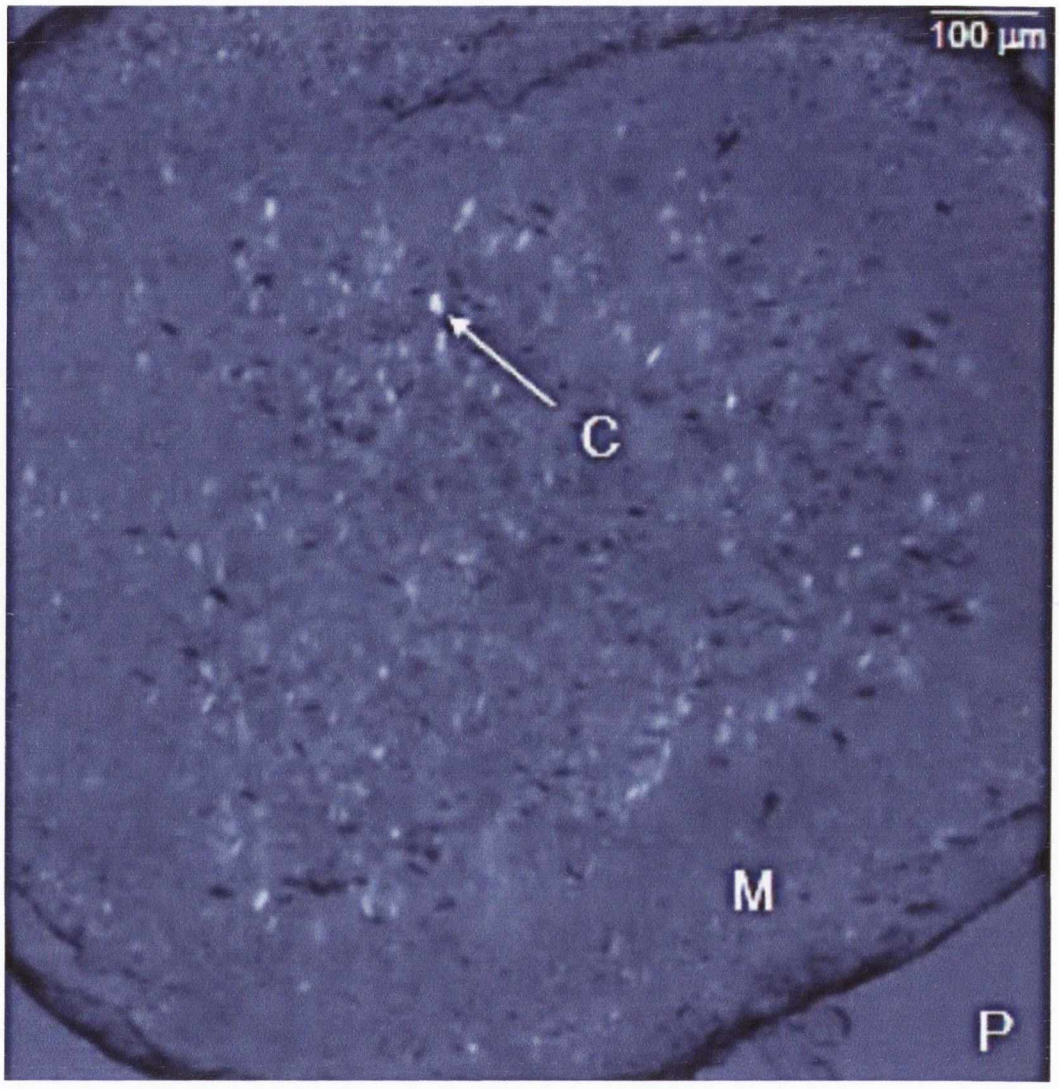


Figure 3.13 Example 5 of initial hits found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 20.0 mg / mL and the crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 21. Crystallisation condition is: 0.2 M sodium citrate, 0.1 M Tris-HCl, pH 8.5, 30 %(w/v) PEG400. Cross polarized light was used to identify possible protein crystals more easily as shining objects in dark background. M: mesophase drop, C: crystals, P: precipitant. In this case, the mesophase remains cubic, as the boundary of it is still clear.

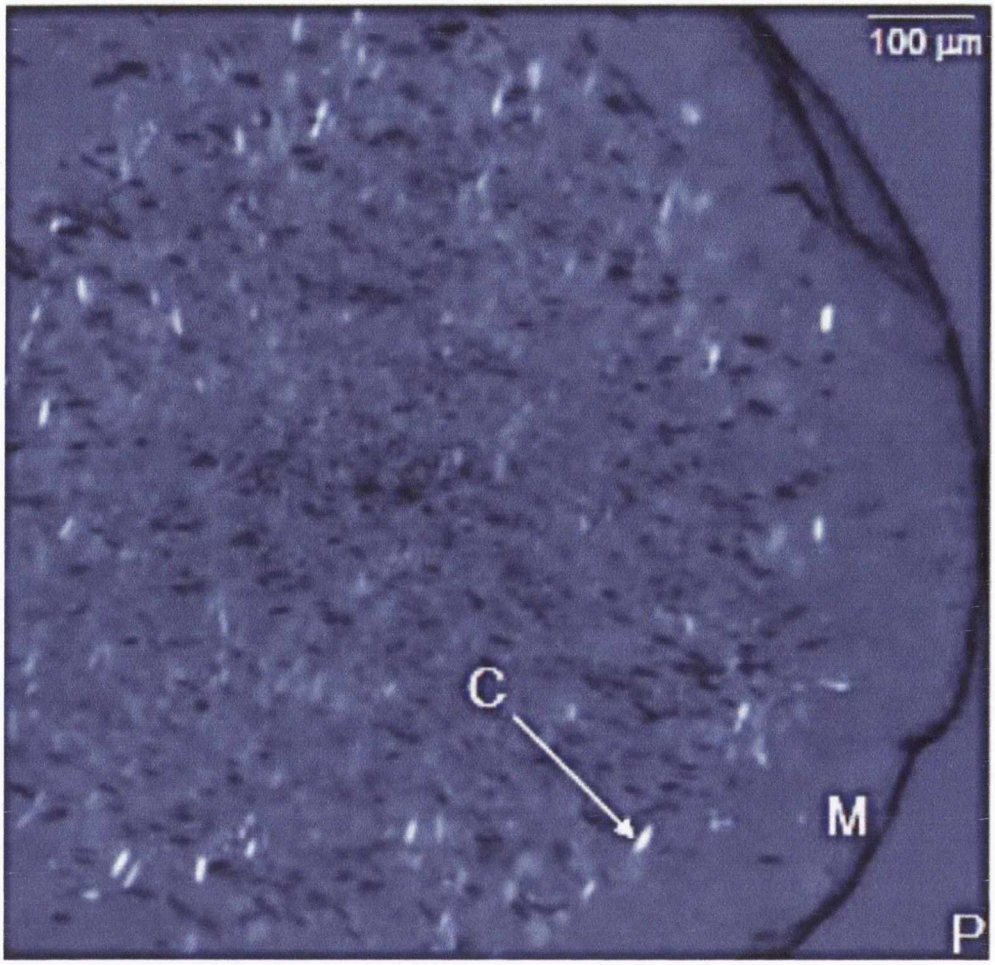


Figure 3.14 Example 6 of initial hits found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 20.0 mg / mL and the crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 21. Crystallisation condition is: 0.02 magnesium chloride, 0.1 M sodium citrate, pH 5.6, 20 %(w/v) PEG400. Cross polarized light was used to identify possible protein crystals more easily as shining objects in dark background. M: mesophase drop, C: crystals, P: precipitant. In this case, the mesophase remains cubic, as the boundary of it is still clear.

In the above figures, the biggest crystals were about 30 μm plates as in Figure 3.12. Crystals were found in a wide range of pHs from 4.6 to 8.5. The presence of PEG400 was found in most of the conditions, with only two exceptions as in Figure 3.9 and Figure 3.11. Under all of these conditions, mesophase remain cubic because the boundary is still clear.

The crystals were harvested and tested by Dr. David Aragão at APS on 5th April of 2008 and the diffraction pattern best to 10 Å proved them to be protein crystals. Based on these results, optimisation was carried out to look for optimal crystallisation conditions.

3.4.1.2 Crystals after optimisation

After optimisation (as described in section 3.3.1.3) around protein, salt concentrations, PEG400 or isopropanol concentrations and pHs, using AlgE Screen Kit 1, 2 and 3, three types of crystals were obtained. The best screen kit was AlgE Screen Kit 1, which had PEG400 in all conditions. Two types of two-dimensional crystals were yielded in some conditions and other conditions only generated needle crystals. Some examples of each crystal form are shown below in Figure 3.15 to Figure 3.17.

In Figure 3.18, distribution of the crystal forms is shown. Among conditions that yielded crystals, when PEG400 concentration was below 27 %(w/v), only tiny needle crystals were obtained as indicated by blue area, which were quite similar to initial hits. When PEG400 concentration was higher than 27 %(w/v), two types of 2D crystals were obtained as indicated by red area: Parallelogram and hexagonal crystals. Parallelogram crystals grew to max of 75 μm x 30 μm and the hexagonal ones were about 20 μm in each dimension.

Surprisingly, same crystallisation conditions could yield two different forms of crystals, although the two forms could only be found in separate wells. Examples are shown in Figure 3.15 and 3.16.

These crystals were harvested and used for X-ray diffraction study.

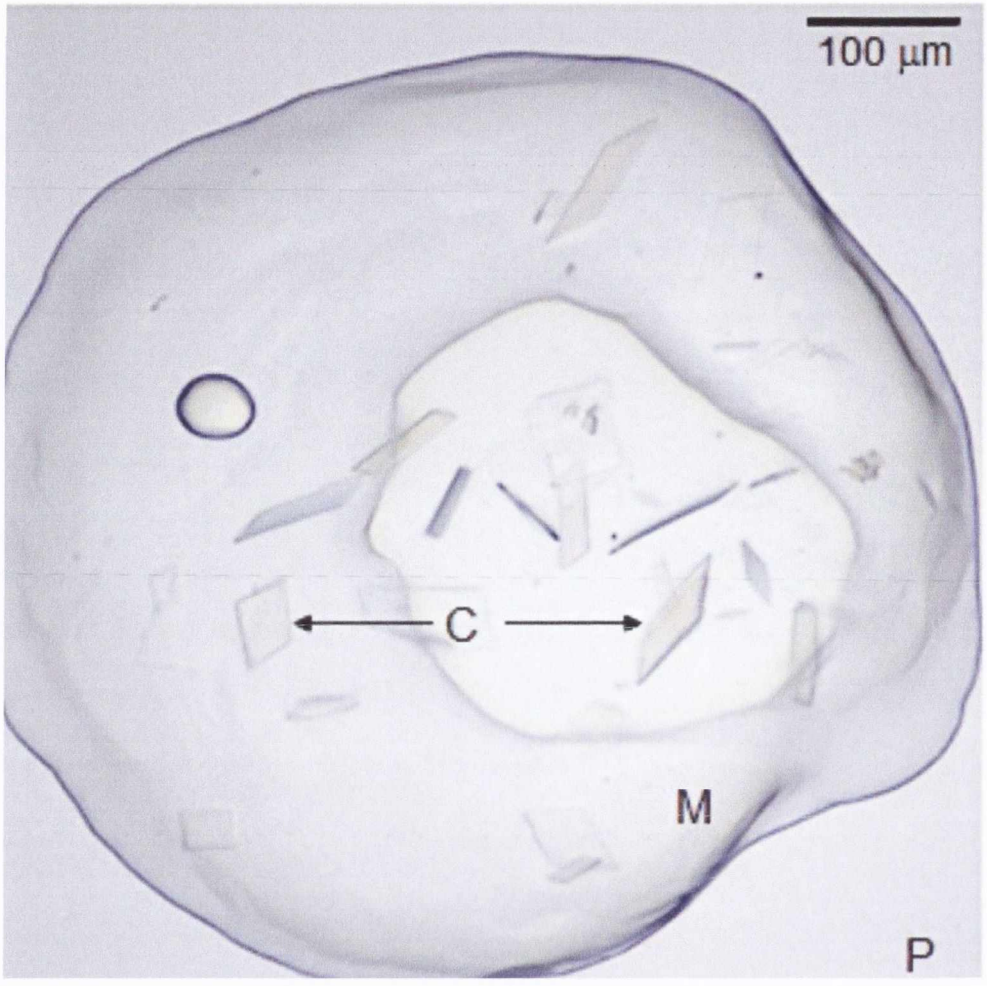


Figure 3.15 Typical parallelogram type crystals found for AlgE / 9.9 MAG *in meso* trials. The crystal pointed by arrow is about 75 µm X 20 µm and the thickness is unknown. The protein concentration was 10 mg / mL and crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 14. Crystallisation condition is: 0.1 M Bis-Tris, pH 6.0, 0.1 M sodium citrate, 42 %(v/v) PEG 400. M: mesophase drop. C: crystals. P: precipitant solution. The thickness of crystals is not known. In this case, the mesophase remains cubic, as the boundary of it is still clear.

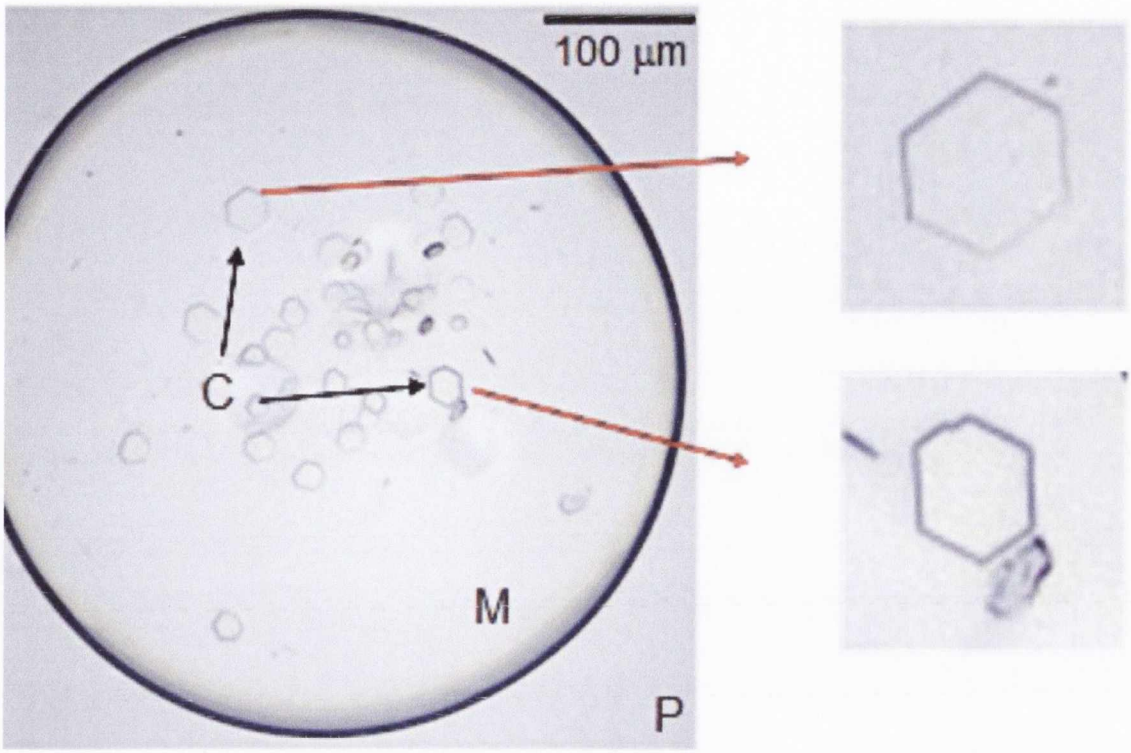


Figure 3.16 Typical hexagonal type crystals are found for AlgE / 9.9 MAG *in meso* trials. The upper crystal pointed by arrow is about 15 μm in all dimensions and the thickness is unknown. The protein concentration was 10 mg / mL and crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 14. Crystallisation condition is: 0.1 M Bis-Tris, pH 6.0, 0.1 M sodium citrate, 42 %(v/v) PEG 400. M: mesophase drop. C: crystals. P: precipitant solution. An expanded view of the two biggest hexagonal crystals is shown on the right-hand side. The thickness of crystals is not known. In this case, the mesophase remains cubic, as the boundary of it is still clear.

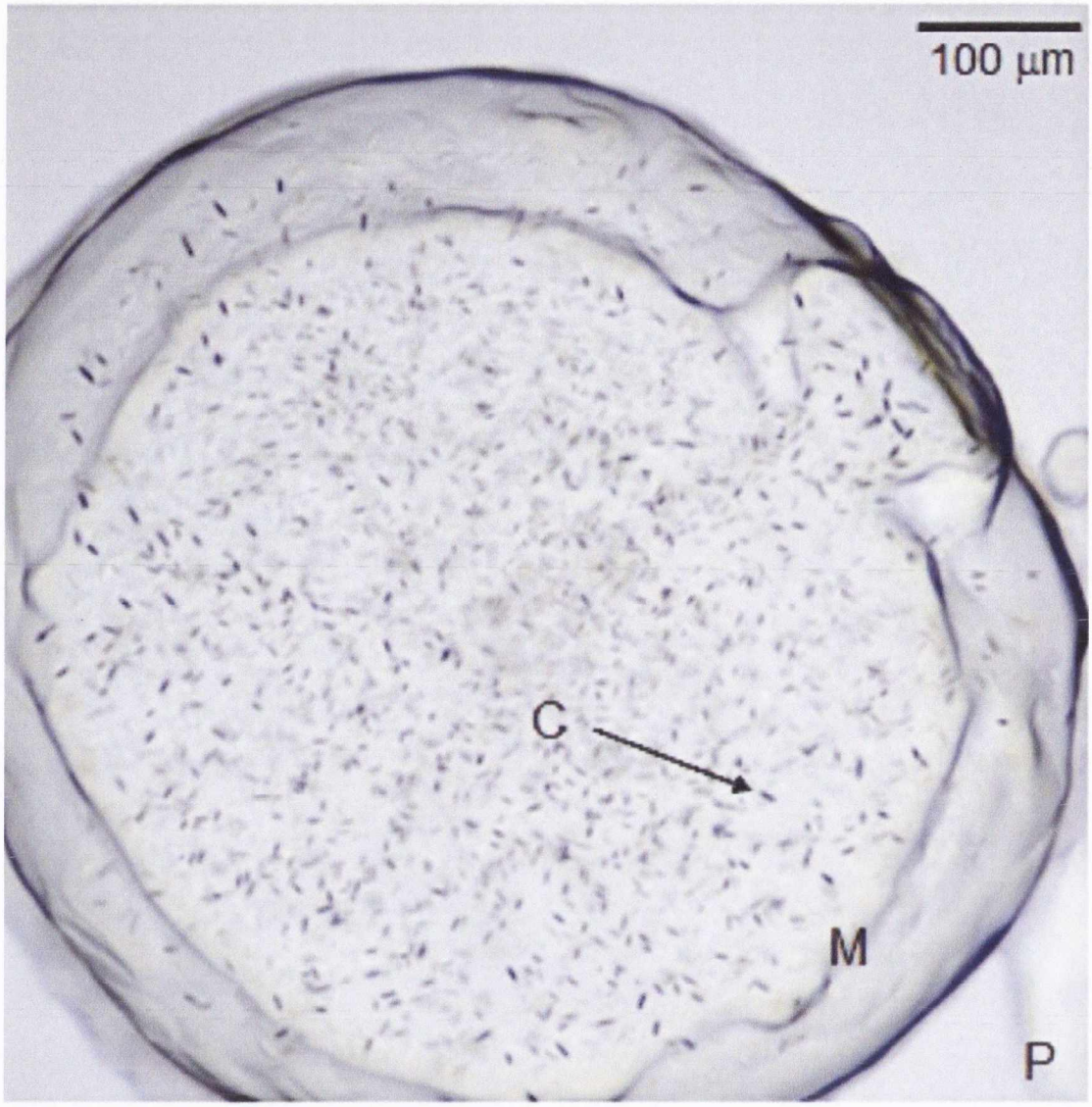


Figure 3.17 Typical needle type crystals are found for AlgE / 9.9 MAG *in meso* trials. The protein concentration was 10 mg / mL and crystals grew at 20 °C in lipid 9.9 MAG. The image was taken on day 14. Crystallisation condition is: 0.1 M Bis-Tris, pH 6.0, 0.1 M sodium citrate, 21 %(v/v) PEG 400. M: mesophase drop. C: crystals. P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

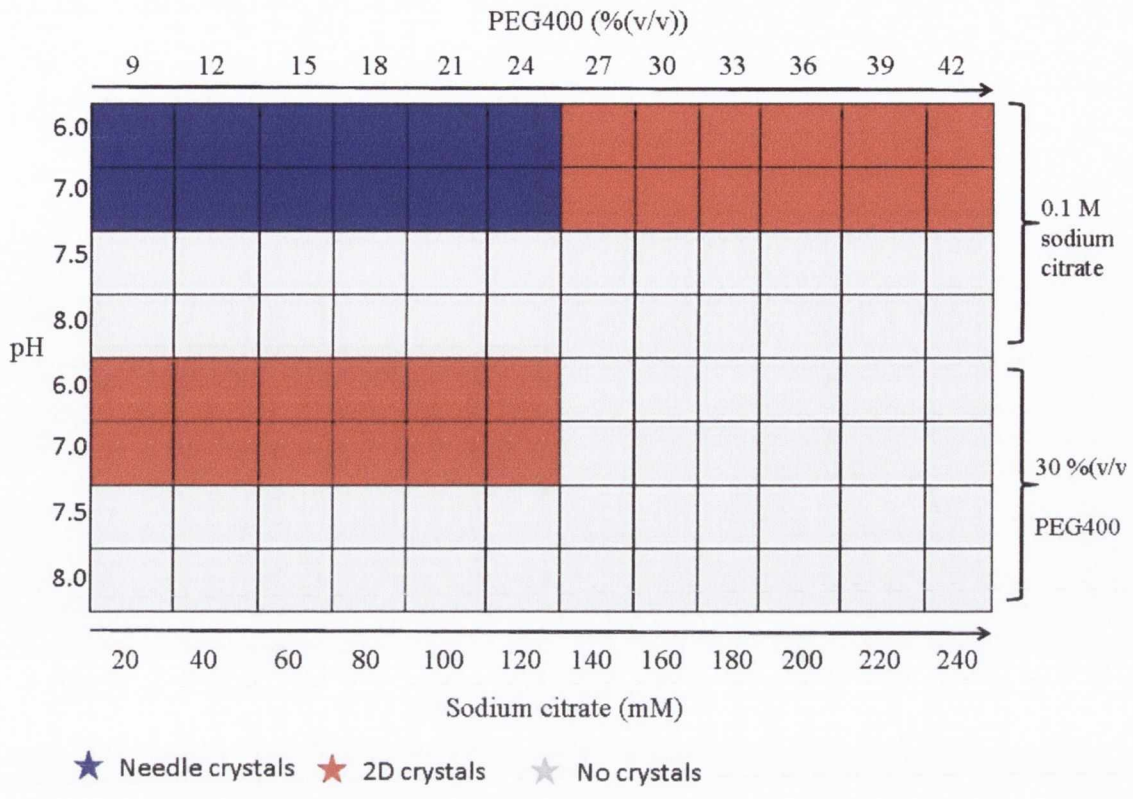


Figure 3.18 Conditions of crystals found on crystallisation plates that are set up using AlgE Screen Kit 1. In grey region, no crystals are found. In blue region, only needle crystals are found and the PEG400 concentrations are below 27 %(v/v). In the red region, two types of 2D crystals are found and the PEG400 concentrations are above 27 %(v/v).

3.4.1.3 X-ray diffraction

The crystals were harvested from cubic phase drop and shipped to synchrotrons for X-ray diffraction tests. Dr. David Aragão and Dr. Valerie Pye did all data collection and processing work.

For crystals that grow in regular *in surfo* drops, cryo-protection is needed in many cases. For AlgE crystals that grow by *in meso* method, the mesophases can function as cryo-protectant so there was no need to test cryo-protectant. However, in some cases, because of the high viscosity of cubic phase system, big loops of mesophase were harvested together

with crystals. This might prevent the crystals from being cryo-cooled because it took longer time for bigger mesophase to cool down than the small piece of mesophase. How this was going to affect the diffraction quality of crystals was not explored but should be kept in mind.

An image of the X-ray diffraction pattern is shown in Figure 3.19. After analysis, the space group was indexed in C222 with unit cell: a = 58.2 Å b = 75.0 Å c = 250.1 Å. The best resolution was up to 2.8 Å. There were two molecules in each asymmetric unit and the solvent content was 48.7 %(v/v). It was also observed by Dr. David Aragão that, in general, the bigger the crystal of AlgE is, the better they diffract under X-ray beam.

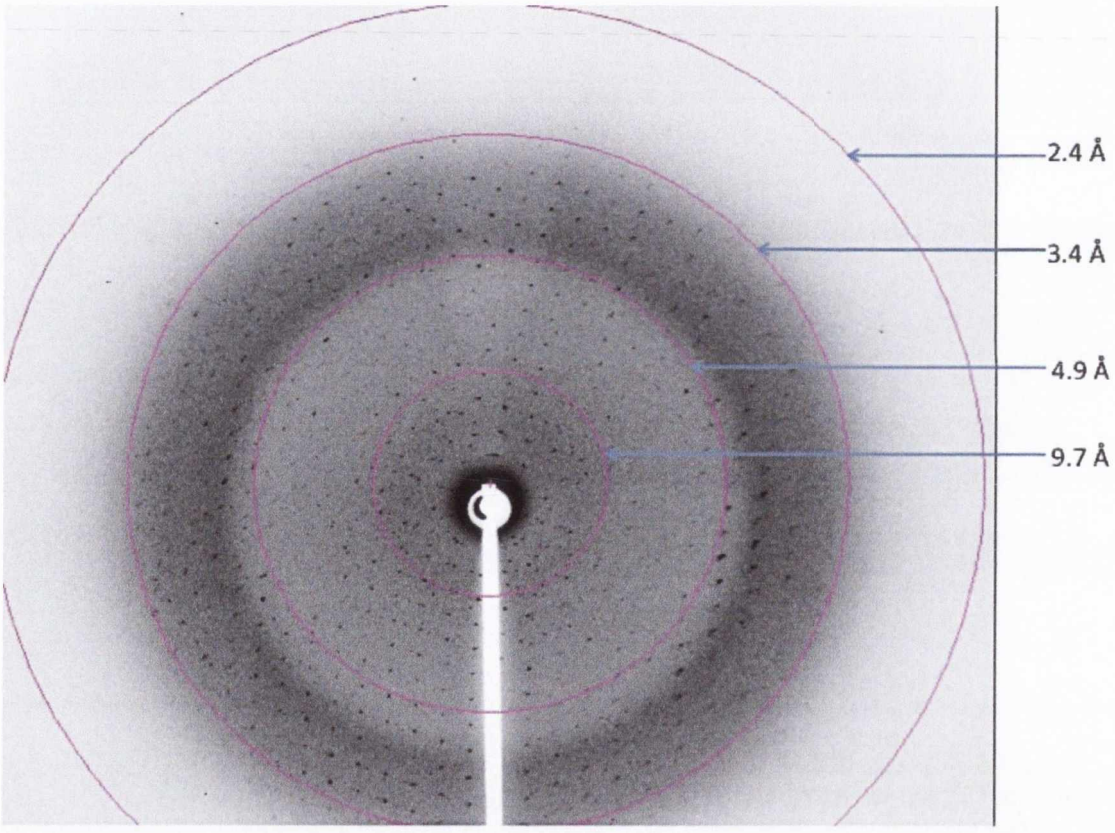


Figure 3.19 The best diffraction data of crystals from AlgE / 9.9 MAG *in meso* trials. The four purple rings represent different resolution shell as indicated by arrows on the right. Diffraction pattern consists of the black dots in the figure. The best resolution in this case was up to about 2.8 Å. The data were collected at GM/CA-CAT 23 ID-B beamline at APS. The energy of data collection was 12.0 keV. The beam size was 20 µm X 20 µm and flux density was about 5 X 10^{14} photon $s^{-1}mm^{-2}$ (Fischetti *et al.* 2009).

3.4.2 *In meso* crystals in Novel MAGs

3.4.2.1 Initial hits from commercial screen trials

For the initial commercial screen trials in novel MAGs, 1,056 conditions were screened in each lipid. In 7.7 MAG, 3 hits were found. In 7.8 MAG, 24 hits were found. In 7.9 MAG, 8 hits were found. In 8.8 MAG, 16 hits were found. And in 8.9 MAG, 15 hits were found. A summary of all the initial hits is shown Table 3.8. Most of the conditions in these initial hits contained different PEGs and ions just as the result for 9.9 MAG. Except for 7.7 MAGs, the mesophase remain cubic under most of these conditions and the boundary of it is clear. But in the case of 7.7 MAG, all three conditions spongify mesophase and the boundary of it is no longer clear.

Table 3.8 Summary of all initial hits found in different novel MAGs. The conditions that yielded two-dimensional crystals are labelled in red text.

Hosting lipid	Conditions that yield initial hits	Maximum crystals size (μm) / shape
7.7 MAG	2.0 M ammonium phosphate	10 / needle
7.7 MAG	0.1 M Tris-HCl, pH 8.5, 1.5 M lithium sulphate	10 / small needle cluster
7.7 MAG	0.1 M Tris-HCl, pH 8.5, 0.1 M sodium chloride, 30 % (w/v) PEG400	20 / plate
7.8 MAG	0.05 M magnesium acetate, 0.05 M sodium acetate, 25 %(w/v) PEG400	10 / small needle
7.8 MAG	0.1 M sodium acetate, 0.2 M calcium acetate, 30 %(w/v) PEG400	20 / rod
7.8 MAG	0.1 M HEPES, pH 7.5, 1 M potassium sodium tartrate	10 / small needle
7.8 MAG	0.1 M magnesium chloride, 0.1 M CAPSO, pH 9.5, 30 %(w/v) PEG400, 0.1 M sodium chloride	10 / small needle
7.8 MAG	2 M sodium formate, pH 4.6	10 / small needle
7.8 MAG	1 M $(NH_4)_2HPO_4$, 0.1 M $NH_4H_2PO_4$, pH 6.5	10 / small needle

Table 3.8 continued

7.8 MAG	0.1 M magnesium chloride, 0.1 M HEPES, pH 7.0, 18 %(w/v) PEG400	10 / small needle
7.8 MAG	0.1 M sodium cacodylate, pH 6.5, 0.7 M sodium acetate	20 / plate
7.8 MAG	0.2 M calcium chloride, 0.1 HEPES, pH 7.5, 14 %(w/v) PEG400,	10 / small needle
7.8 MAG	0.05 M magnesium acetate, 0.05 M sodium acetate, pH 5.0, 28 %(w/v) PEG400	10 / small needle
7.8 MAG	0.05 M magnesium acetate, 0.05 M sodium acetate, pH 5.4, 24 %(PEG400)	10 / small needle
7.8 MAG	0.2 M sodium chloride, 0.1 M Tris-HCl, 44 %(w/v) PEG400	10 / small needle
7.8 MAG	0.2 M lithium sulphate, 0.1 M sodium acetate, pH 3.5, 28 %(w/v) PEG400	10 / small needle
7.8 MAG	0.2 M calcium chloride, 0.1 HEPES, pH 7.0, 33 %(w/v) PEG400	10 / small needle
7.8 MAG	0.2 M calcium chloride, 30 %(w/v) PEG400	10 / small needle
7.8 MAG	0.35 M sodium chloride, 0.01 M MOPS, pH 7.0, 28 %(w/v) PEG1000	10 / small needle
7.8 MAG	0.23 M MES/Bis-Tris, pH 6.0, 6.6 %(w/v) PEG6000, 0.2 M sodium chloride, 30 %(w/v) PEG400	10 / small needle
7.8 MAG	0.2 M magnesium acetate, 0.1 M MES, pH 6.0, 10 %(w/v) PEG3500	10 / small needle
7.8 MAG	0.2 M magnesium acetate, 0.05 M sodium cacodylate, pH 6.5, 11 %(w/v) PEG3350	20 / plate
7.8 MAG	0.1 M imidazole, pH 6.5, 1 M sodium acetate	20 / rod
7.8 MAG	0.1 M Tris-HCl, Bis-Tris propane, pH 7.0, 3.5 M NH_4Cl	10 / small needle
7.8 MAG	0.1 M sodium acetate, pH 4.6, 2 M sodium chloride	10 / small needle

Table 3.8 continued

7.8 MAG	0.1 M sodium acetate, pH 4.6, 1.5 M ammonium chloride	10 / small needle
7.8 MAG	0.1 M sodium cacodylate, pH 6.5, 1.4 M sodium acetate	10 / small needle
7.9 MAG	0.2 M lithium sulphate, 0.1 M MES, pH 6.0, 20 %(w/v) 1,4-butanediol	20 / needle cluster
7.9 MAG	0.1 M sodium chloride, 0.005 M magnesium chloride, 0.1 M Tris-HCl, pH 8.0, 30 %(w/v) PEG 2000 MME	10 / needle
7.9 MAG	2.8 M sodium acetate, 0.1 M Bis-Tris propane, pH 7.0	10 / needle
7.9 MAG	0.1 M sodium acetate, pH 4.5, 3.0 M sodium chloride	10 / needle
7.9 MAG	1.0 M ammonium sulphate, 0.1 M Bis-Tris, pH 5.5, 1 %(w/v) PEG3350	10 / needle
7.9 MAG	20% (w/v) PEG 3350, 0.2 M ammonium citrate dibasic	10 / needle
7.9 MAG	20 %(w/v) PEG6000, 0.1M citric acid, 0.1 M sodium hydroxide, 1M lithium chloride	10 / needle
7.9 MAG	20 %(w/v) PEG400, 0.1 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 0.02 M magnesium chloride	10 / needle
8.8 MAG	30 %(w/v) PEG2000, 0.5 M sodium phosphate, 0.2 M sodium chloride	10 / needle
8.8 MAG	0.1M MES, pH 6.5, 1.4 M sodium acetate	10 / needle
8.8 MAG	0.1M MES pH 6.5, 1.5 M sodium phosphate	10 / needle
8.8 MAG	0.1 M HEPES, pH 7.5, 0.8 M sodium dihydrogen phosphate, 0.8 M potassium dihydrogen phosphate	10 / needle

Table 3.8 continued

8.8 MAG	0.2 M lithium sulphate, 0.1 MES, pH 6.0, 20 %(w/v) 1,4-butanediol	10 / needle
8.8 MAG	0.36 M sodium chloride, 0.1 %(w/v) sodium azide, 0.015 M sodium phosphate, pH 7.0, MAG 9.9 %(w/v) PEG4000	10 / needle
8.8 MAG	0.05 M calcium chloride, 0.05 M barium chloride, 0.1 M Tris-HCl, 32 %(v/v) PEG 400	10 / needle
8.8 MAG	0.1 M magnesium chloride, 0.1 M Tris-HCl, pH 7.5, 22 %(v/v) PEG 400	10 / needle
8.8 MAG	0.1M MES pH 6.5, 1.0 M sodium acetate	10 / needle
8.8 MAG	8% (w/v) PEG 4000, 0.1 M sodium acetate, pH 4.6	10 / needle
8.8 MAG	0.35 M sodium chloride, 0.01 M MOPS, pH 7.0, 28 %(w/v) PEG1000	10 / needle
8.8 MAG	0.1 M sodium cacodylate, pH 6.5, 0.7 M sodium acetate	10 / needle
8.8 MAG	0.2 M calcium chloride, 20 %(w/v) PEG3350, pH 5.1	10 / needle
8.8 MAG	20 %(w/v) PEG1000, 0.1 M imidazole, pH 8.0, 0.2 M calcium acetate	10 / needle
8.8 MAG	2.8 M sodium acetate, 0.1 M Bis-Tris propane, pH 7.0	10 / needle
8.8 MAG	0.2 M potassium thiocynate, 20% w/v PEG 3350	40 / plate
8.9 MAG	0.01 M zinc sulphate, 0.05 M HEPES, pH 7.5, 21 % (w/v) PEG600	10 / needle
8.9 MAG	0.05 M sodium chloride, 20 %(w/v) PEG400, 0.1 M sodium acetate	10 / needle
8.9 MAG	0.1 M sodium chloride, 25 %(w/v) PEG300, 0.05 M glycine	10 / needle

Table 3.8 continued

8.9 MAG	0.02 M sodium chloride, 18 % (w/v) PEG400, 0.02 M magnesium chloride	20 / needle
8.9 MAG	0.1 M imidazole, 15 %(v/v) ethanol, 0.2 magnesium chloride	10 / needle
8.9 MAG	0.2 M calcium acetate, 0.1 M imidazole, 20 %(w/v) PEG1000	10 / needle
8.9 MAG	0.2 M Potassium thiocynate, 20% w/v PEG 3350	10 / needle
8.9 MAG	24 %(w/v) Peg1500, 20 % (w/v) glycerol	10 / needle
8.9 MAG	0.2 M ammonium phosphate monobasic, 50 %(w/v) 2-methyl-2,4-pentanediol, 0.1 M Tris-HCl, pH 8.5	10 / needle
8.9 MAG	0.02 M magnesium formate, 20 %(w/v) PEG1500	10 / needle
8.9 MAG	0.2 M calcium chloride, 30 %(w/v) PEG400	10 / needle
8.9 MAG	0.2 M potassium nitrate, 20 %(v/v) PEG1500	10 / needle
8.9 MAG	0.1 M HEPES, pH 7.5, 20 %(v/v) Jeffamine M-600	10 / needle
8.9 MAG	4.4% w/v PEG 8000, 0.08 M sodium cacodylate, pH 6.5, 0.16 M calcium acetate, 20% v/v glycerol	30 / plate
8.9 MAG	0.1 M magnesium chloride, 0.1 M Tris-HCl, pH 7.5, 22% v/v PEG 400	30 / rod

Of all these 66 initial hits, some crystals have the shape of two-dimensional plates and they are much bigger than other needle crystals or spot crystals. Typical examples of these two-dimensional plate crystals are shown below in Figure 3.20 – 3.24.

These crystals were harvested and tested at APS and Diamond as described. One condition from 7.8 MAG yielded crystals of better diffraction quality than others as the crystals diffracted to 2.8 Å: 0.1 M sodium cacodylate, pH 6.5, 1.0 M sodium acetate.

A summary of the diffraction quality of crystals grown in different MAGs can be found in Table 3.9. In general, in the case of AlgE *in meso* crystallization trials, needle crystals diffracted very poorly. Two-dimensional plate crystals diffracted much better. And as observed for the crystals that grow in 9.9 MAG, the bigger the size of crystals, the better they diffracted. As such, a direction for later optimisation experiments in novel MAGs was to look for bigger two-dimensional or even three-dimensional crystals.

Based on this result, AlgE Screen Kit 5 was prepared for optimisation around concentration of sodium acetate and pH as described in Section 3.3.2.2. In addition, AlgE Screen kit 1 was used as well because it yielded the best diffraction quality for crystals that were grown in 9.9 MAG.

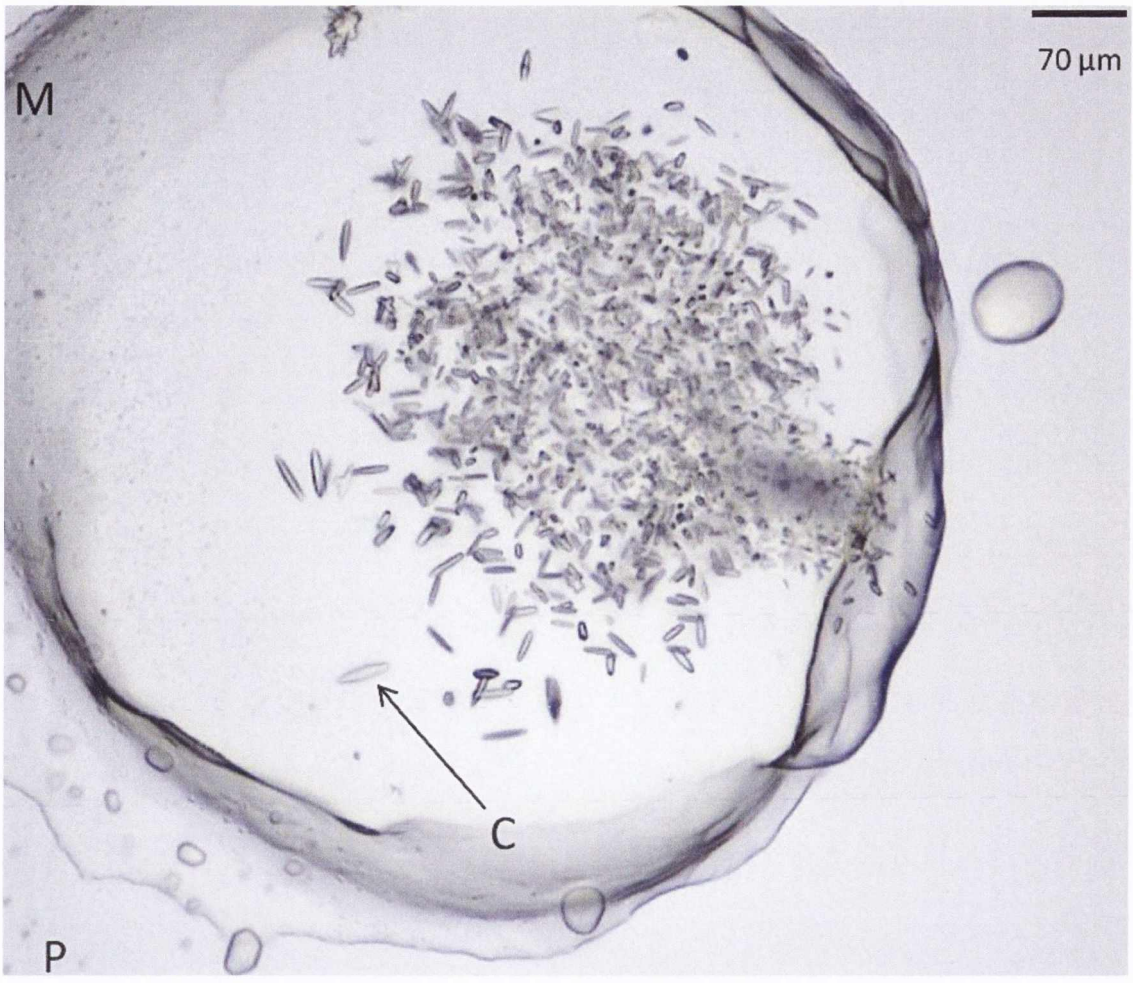


Figure 3.20 Example 1 of 2D crystals found for AlgE / 8.9 MAG *in meso* trials. The protein concentration was 10.0 mg / mL and the crystals grew at 20 °C in lipid 8.9 MAG. The image was taken on day 7. Crystallisation condition is: 0.1 M magnesium chloride, 0.1 M Tris, pH 7.5, 22 %(v/v) PEG 400. M: mesophase drop, C: crystals, P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

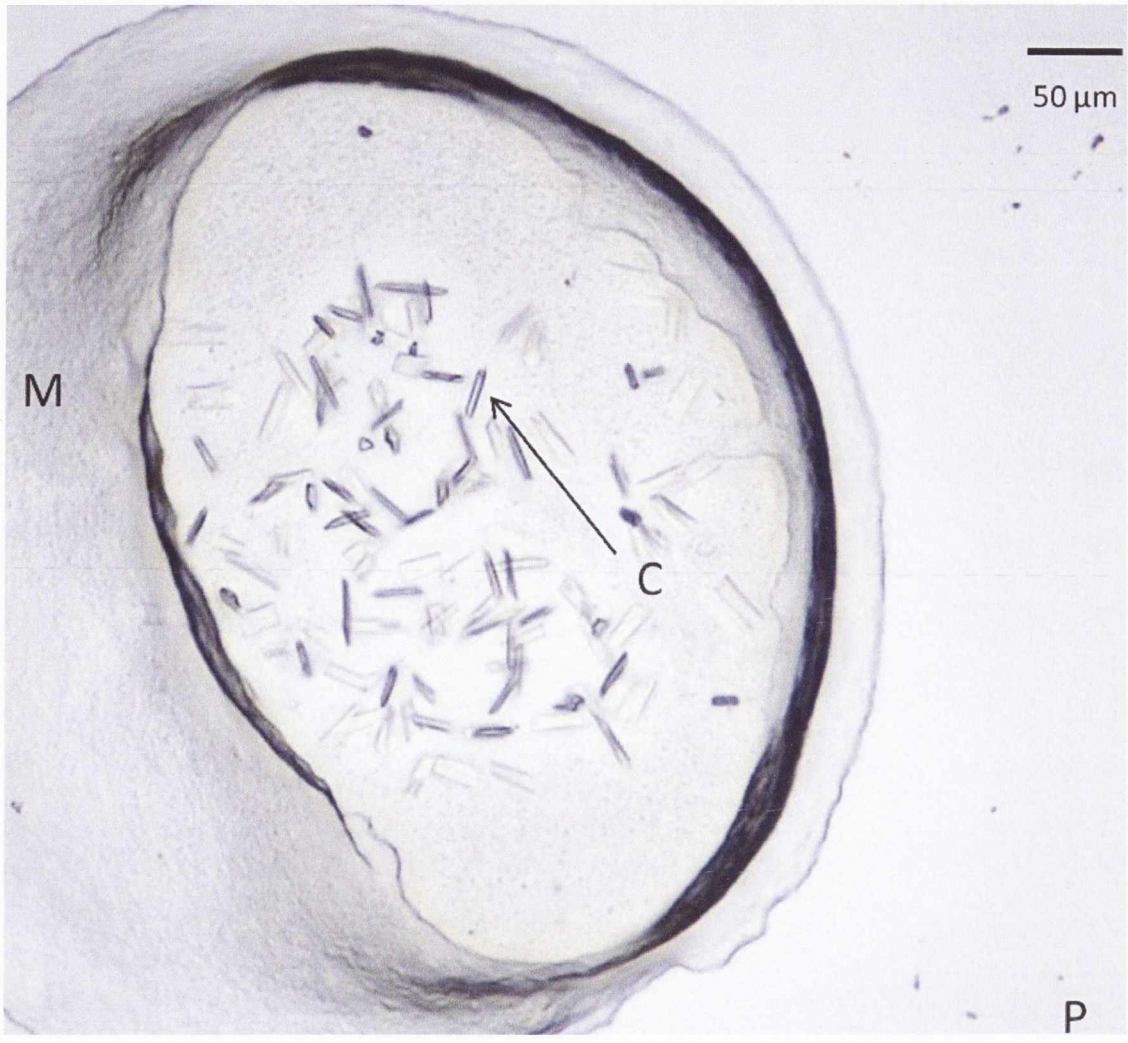


Figure 3.21 Example 2 of 2D crystals found for AlgE / 8.9 MAG *in meso* trials. The protein concentration was 10.0 mg / mL and the crystals grew at 20 °C in lipid 8.9 MAG. The image was taken on day 7. Crystallisation condition is: 4.4 %(w/v) PEG 8000, 0.08 M sodium cacodylate, pH 6.5, 0.16 M calcium acetate, 20 %(v/v) glycerol. M: mesophase drop, C: crystals, P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

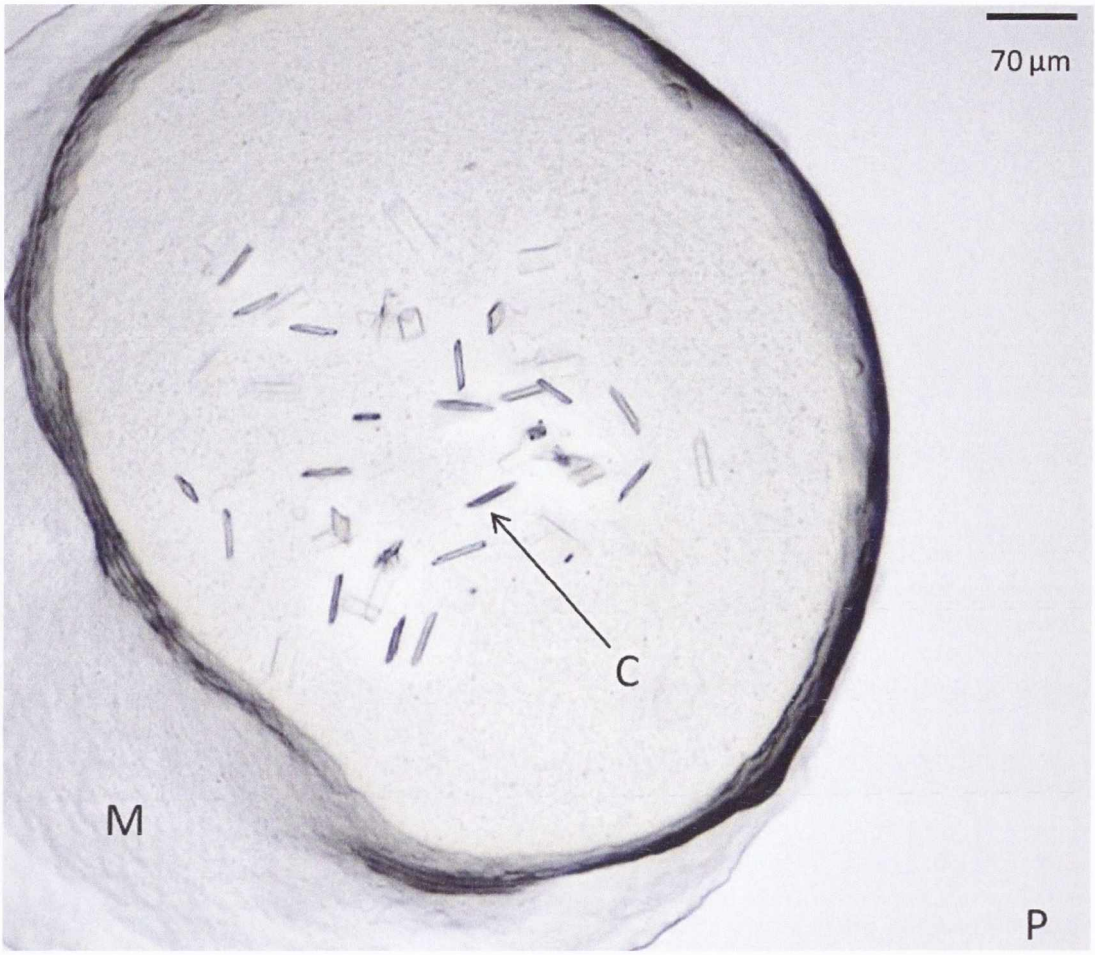


Figure 3.22 Example 1 of 2D crystals found for AlgE / 8.8 MAG *in meso* trials. The protein concentration was 10.0 mg / mL and the crystals grew at 20 °C in lipid 8.8 MAG. The image was taken on day 14. Crystallisation condition: 0.2 M potassium thiocynate, 20 %(w/v) PEG 3350. M: mesophase drop, C: crystals, P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

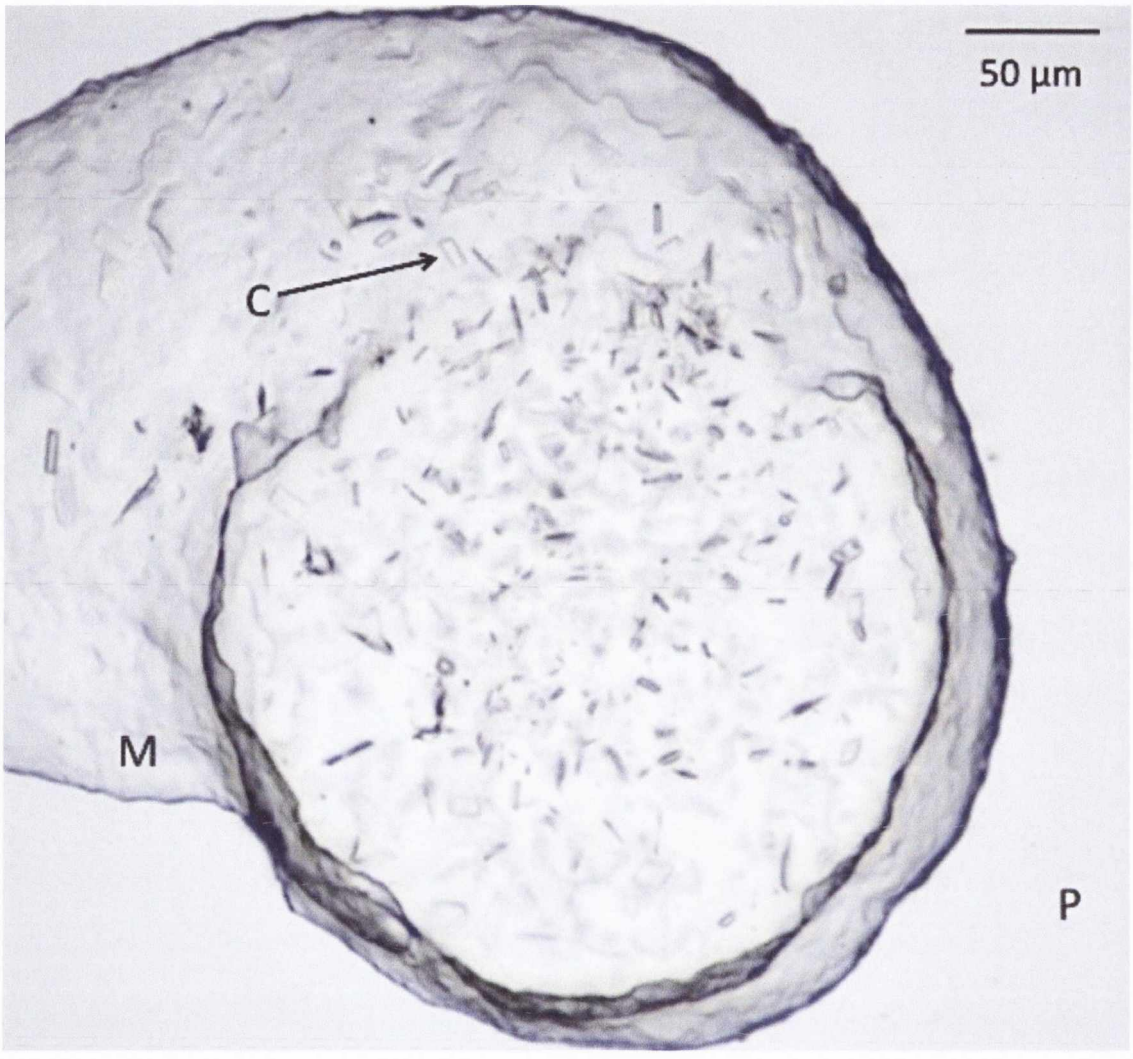


Figure 3.23 Example 1 of 2D crystals found for AlgE / 7.8 MAG *in meso* trials. The protein concentration was 10.0 mg / mL and the crystals grew at 20 °C in lipid 7.8 MAG. The image was taken on day 14. Crystallisation condition is: 0.1 M sodium cacodylate, pH 6.5, 1.4 M sodium acetate. M: mesophase drop, C: crystals, P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

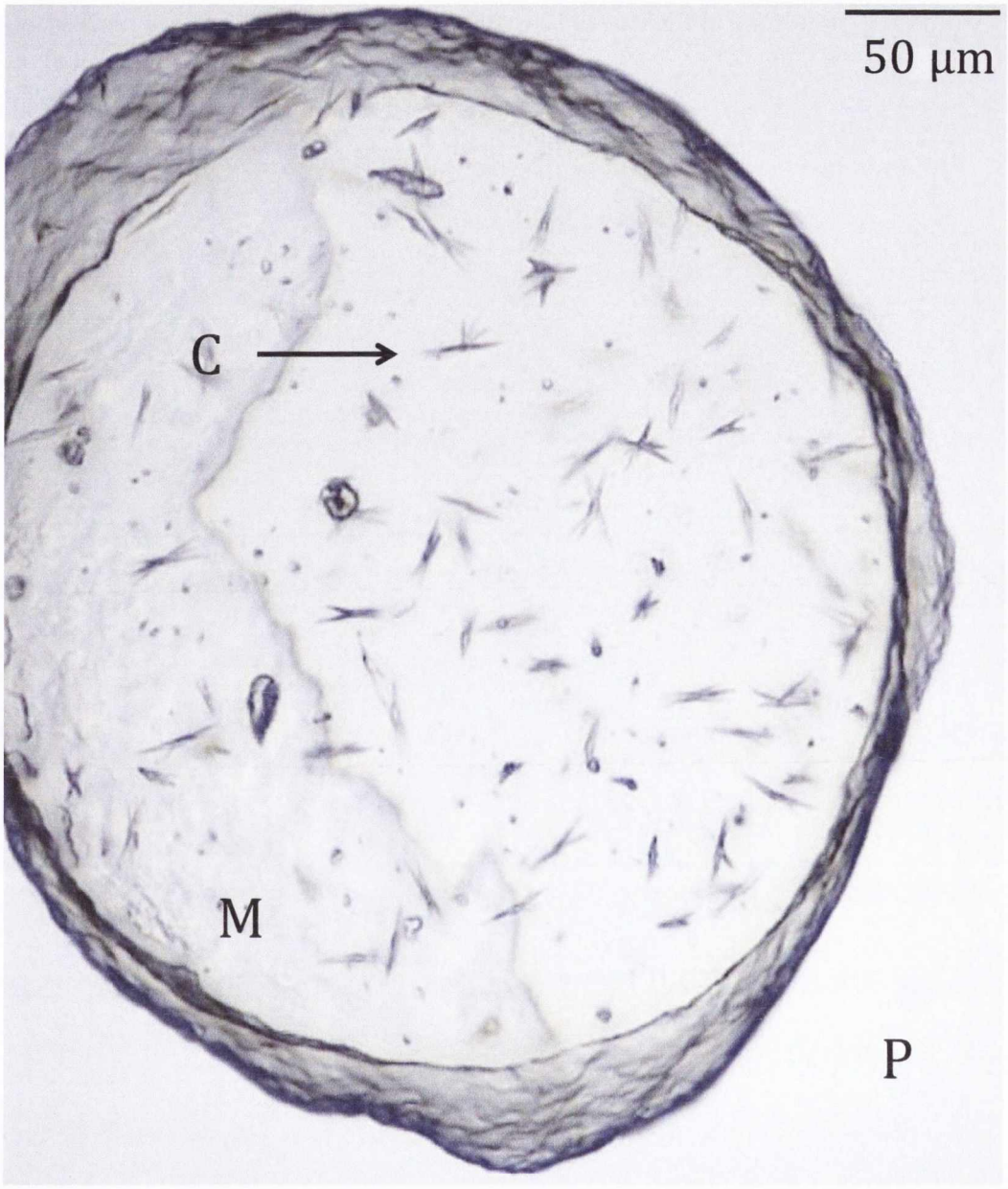


Figure 3.24 Example 1 of crystal cluster found for AlgE / 7.9 MAG *in meso* trials. The protein concentration was 10.0 mg / mL and the crystals grew at 20 °C in lipid 7.8 MAG. The image was taken on day 14. Crystallisation condition is: 0.2 M lithium sulphate, 0.1 MES, pH 6.0, 20 %(w/v) 1,4-butanediol. M: mesophase drop, C: crystals, P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

Table 3.9 Summary of diffraction quality for the initial hits found in all MAGs. The highest resolution comes from 7.8 MAG, which is labelled in red.

Lipid name	Number of initial hits out of 1,056 conditions (Harvested on day ~21)	Best diffraction of initial hits (Å)
7.7 MAG	3	3.4
7.8 MAG	24	2.8
8.8 MAG	16	3.3
7.9 MAG	8	3.5
8.9 MAG	15	3.8
9.9 MAG	11	10.0

3.4.2.2 Crystals after optimisation

As a result of optimisation, crystals of at least 50 μm in the longest dimension were found and examples were shown in Figure 3.25 – 3.25.

As in figures, two very similar crystal forms were seen under completely different conditions. The crystals in Figure 3.25 and Figure 3.26 normally had slightly longer second dimension, which is usually 20 – 50 μm. And in some cases the crystals had two equal dimensions as a "square" shape. The crystals in Figure 3.26 were more like rectangular plates. The crystals in Figure 3.27 usually has shorter second dimension, which is around 20 μm.

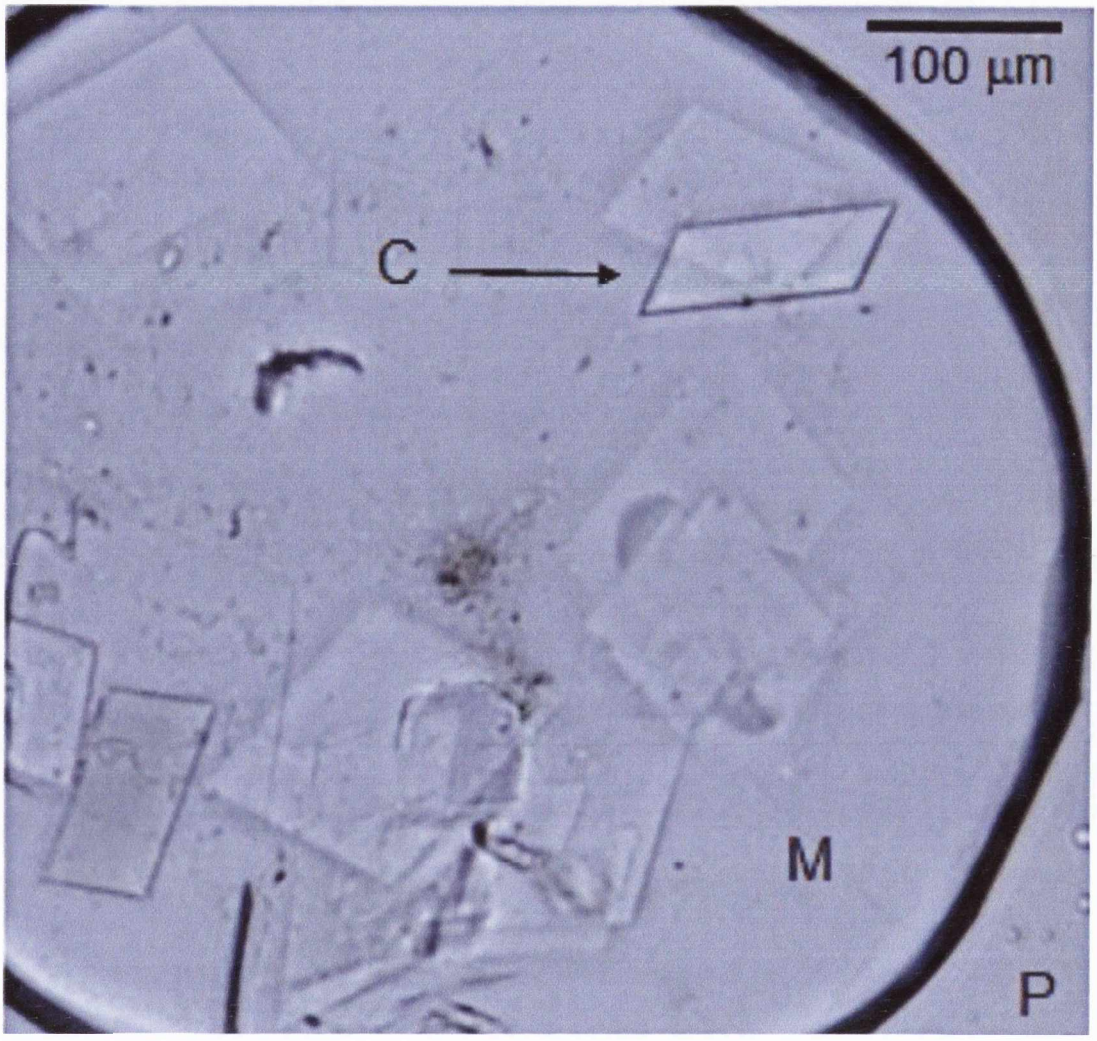


Figure 3.25 Typical parallelogram type crystals for AlgE / 7.8 MAG *in meso* trials using AlgE Screen Kit 1. The crystal pointed by arrow is about 100 µm X 30 µm and the thickness is unknown. The protein concentration was 10 mg / mL and crystals grew at 20 °C in lipid 7.8 MAG. The image was taken on day 21. The crystallisation condition was 0.1 M Bis-Tris, pH 7.0, 0.1 M sodium citrate, 21 %(v/v) PEG 400. Cross polarized light was not used in this case as the crystals can be easily found. M: mesophase drop. C: crystals. P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

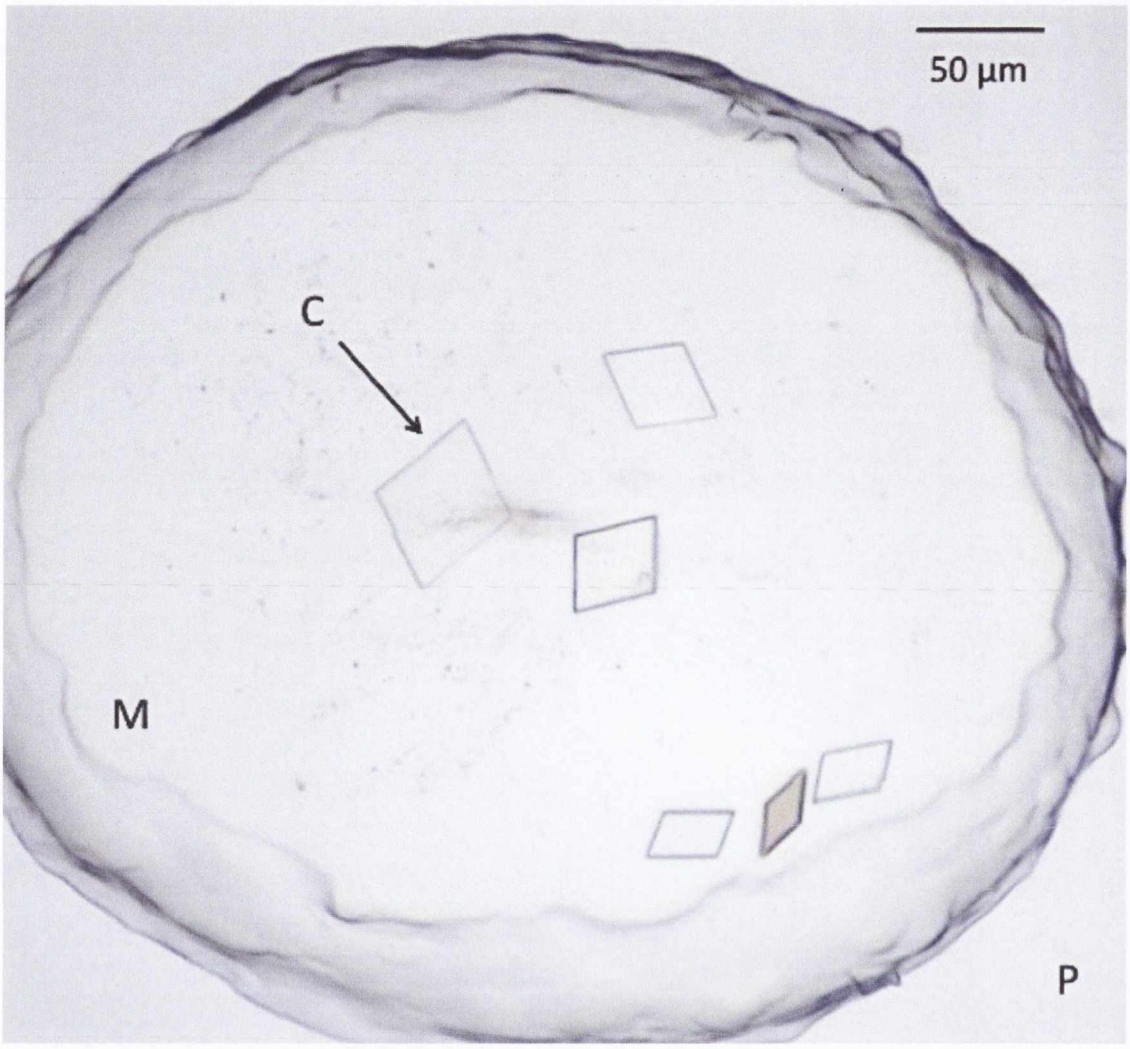


Figure 3.26 Typical "square" type crystals for AlgE / 7.8 MAG *in meso* trials using AlgE Screen Kit 1. The crystal pointed by arrow is about 50 µm X 50 µm and the thickness is unknown. The protein concentration was 10 mg / mL and crystals grew at 20 °C in lipid 7.8 MAG. The image was taken on day 21. The crystallisation condition was 0.1 M Bis-Tris, pH 7.0, 0.1 M sodium citrate, 21 %(v/v) PEG 400. Cross polarized light was not used in this case as the crystals can be easily found. M: mesophase drop. C: crystals. P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

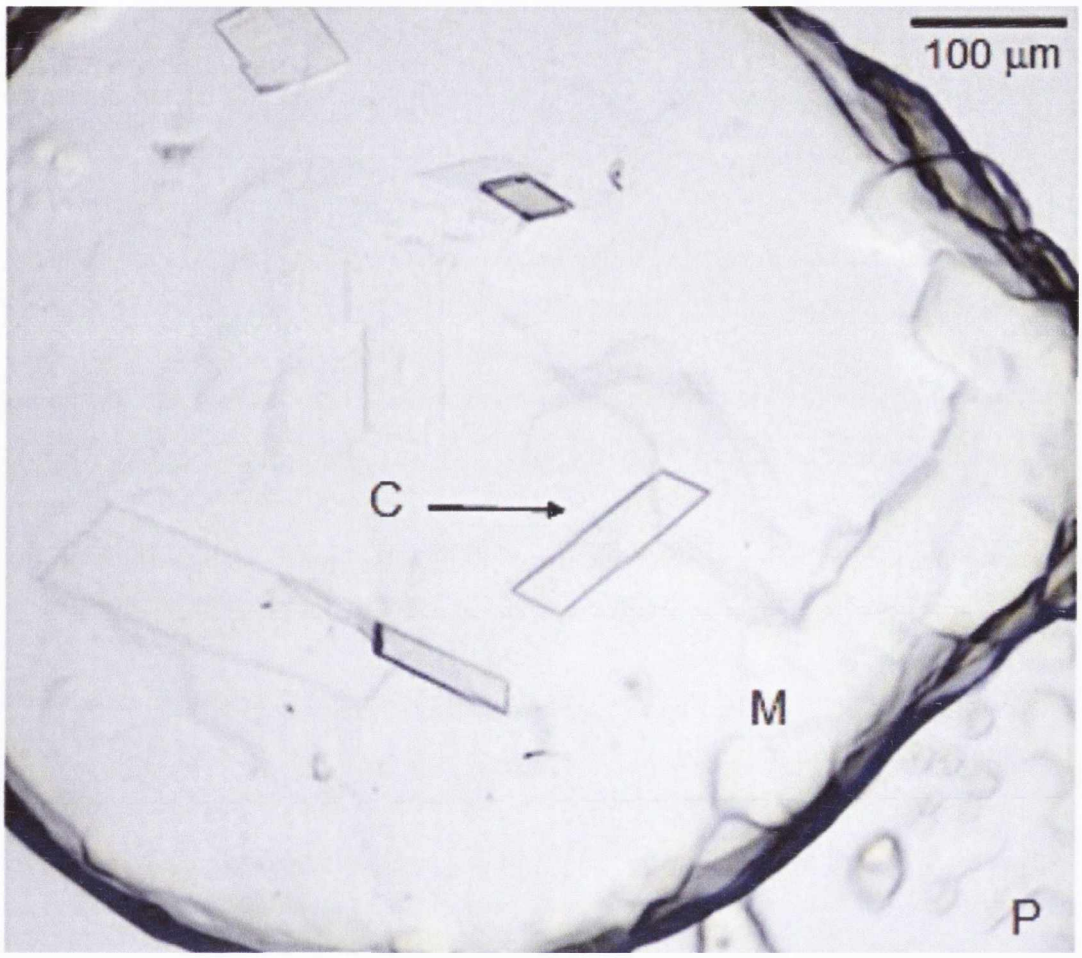


Figure 3.27 Typical rectangular type crystals for AlgE / 7.8 MAG *in meso* trials using AlgE Screen Kit 5. The crystal pointed by arrow is about 100 µm X 20 µm and the thickness is unknown. The protein concentration was 10 mg / mL and crystals grew at 20 °C in lipid 7.8 MAG. The image was taken on day 21. The crystallisation condition was 0.1 M sodium cacodylate buffer, pH 6.5, 0.9 M sodium acetate. Cross polarized light was not used in this case as the crystals can be easily found. M: mesophase drop. C: crystals. P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

3.4.2.3 X-ray diffraction

The crystals were harvested from cubic phase bolus, snap cooled in liquid nitrogen and shipped to synchrotrons for X-ray diffraction tests. Dr. David Aragão and Dr. Valerie Pye did all data collection and processing work. In total, three complete diffraction data sets were collected and indexed in three different space groups. Statistics for each crystal form are listed in Table 3.10.

After analysis, data set 1, which came from crystal form in the Figure 3.25 and Figure 3.26, was indexed in space group C2 with unit cell: a = 57.2 Å b = 74.3 Å c = 115.5 Å, $\alpha = \Upsilon =$ 90, $\beta = 101.2^{\circ}$. The native data set diffracted to max 1.9 Å and 94.51 % complete. The completeness is slighty low for high-resolution data but acceptable. There was one molecule in each asymmetric unit and the solvent content was 43.8 %(v/v).

Data set 2, which came from crystal form in the Figure 3.27, was indexed in the space group $P2_12_12_1$ with unit cell: a = 61.6 Å b = 71.4 Å c = 240.3 Å, $\alpha = \beta = \Upsilon = 90$. The native data set diffracted to max 2.40 Å and 89.45 % complete. There were two molecules in each asymmetric unit and the solvent content was 53.5 %(v/v).

Data set 3, which also came from crystal from in the Figure 3.27, was indexed in the space group $P2_1$, with the unit cell: a = 47.1 Å b = 245.8 Å c = 47.1 Å, $\alpha = \Upsilon = 90$, $\beta = 104.4^{\circ}$. The native data set diffracted to max 2.8 Å and 94.44 % complete. There were two molecules in each asymmetric unit and the solvent content was 48.7 %(v/v).

An X-ray diffraction pattern of data set 1 (1.9 Å with C2 space group) is shown in Figure 3.30. This image comes from the best data set of all that have been collected throughout this project. This complete data set comes from a single crystal, which is shown in Figure 3.29. The image of the crystal was taken on day 4 and the size of the crystal was around 50 μm X 50 μm. However, according to the observation of Dr. David Aragão at APS, the crystal was measured as about 150 μm X 150 μm before being harvested out from plate and mounted onto the beam on day 10. That indicated the crystal in plate continued to grow during shipment from Ireland to United States, from day 4 to day 10.

It is possible that this single crystal went through temperature change, pressure change and even vibration during the flight from Dublin to Chicago. One or multiple of these factors may be beneficial to the diffraction quality of this crystal. This fact itself is interesting to

crystallogenesis work. However, this crystal form could be obtained only in one of the 7.8 MAG batches (batch number JPL-3-77), which were synthesised by Dr. Jean Lee in the chemistry lab of the group. After changing to any other batches that were synthesized later by a modified purification protocol, the crystal form could never be obtained again. And after that, no AlgE crystals could diffract to the same resolution ever again, the best diffraction of which was only up to 2.4 Å as shown by the second data set.

Table 3.10 Summary of data reduction statistics for three crystal forms of AlgE.

		Data set 1	Data set 2	Data set 3
Space group		C2	$P2_12_12_1$	$P2_1$
Resolution (Å)		19.77 – 1.90	45.88 – 2.40	44.90 – 2.80
Cell dimensions	a, b, c (Å)	57.2, 74.3, 115.5	61.6, 71.4, 240.3	47.1, 245.8, 47.1
	α, β, Υ (°)	α = Υ = 90, β = 101.2	α = β = Υ = 90	α = Υ = 90, β = 104.4
Completeness (%)		94.51 (84.9)[a]	98.6 (99.9)[a]	93.25 (95.3)[a]
Redundancy		6.4 (2.5)[a]	9.3 (9.4)[a]	2.7 (2.7)[a]
Solvent content (%)		43.8	53.5	48.7
R_{merge}[b]		0.140 (0.348)[a]	0.093 (0.546)[a]	0.217 (0.908)[a]
I / σ(I)		8.0 (2.0)[a]	6.9 (1.8)[a]	5.8 (2)[a]

a. The values in parentheses are for the highest resolution shell

b. R_{merge} = ΣhΣi |I(h)i - <I(h)>| / Σi I(h)i, where <I(h)> is the mean intensity of equivalent reflections (Rupp 2010).

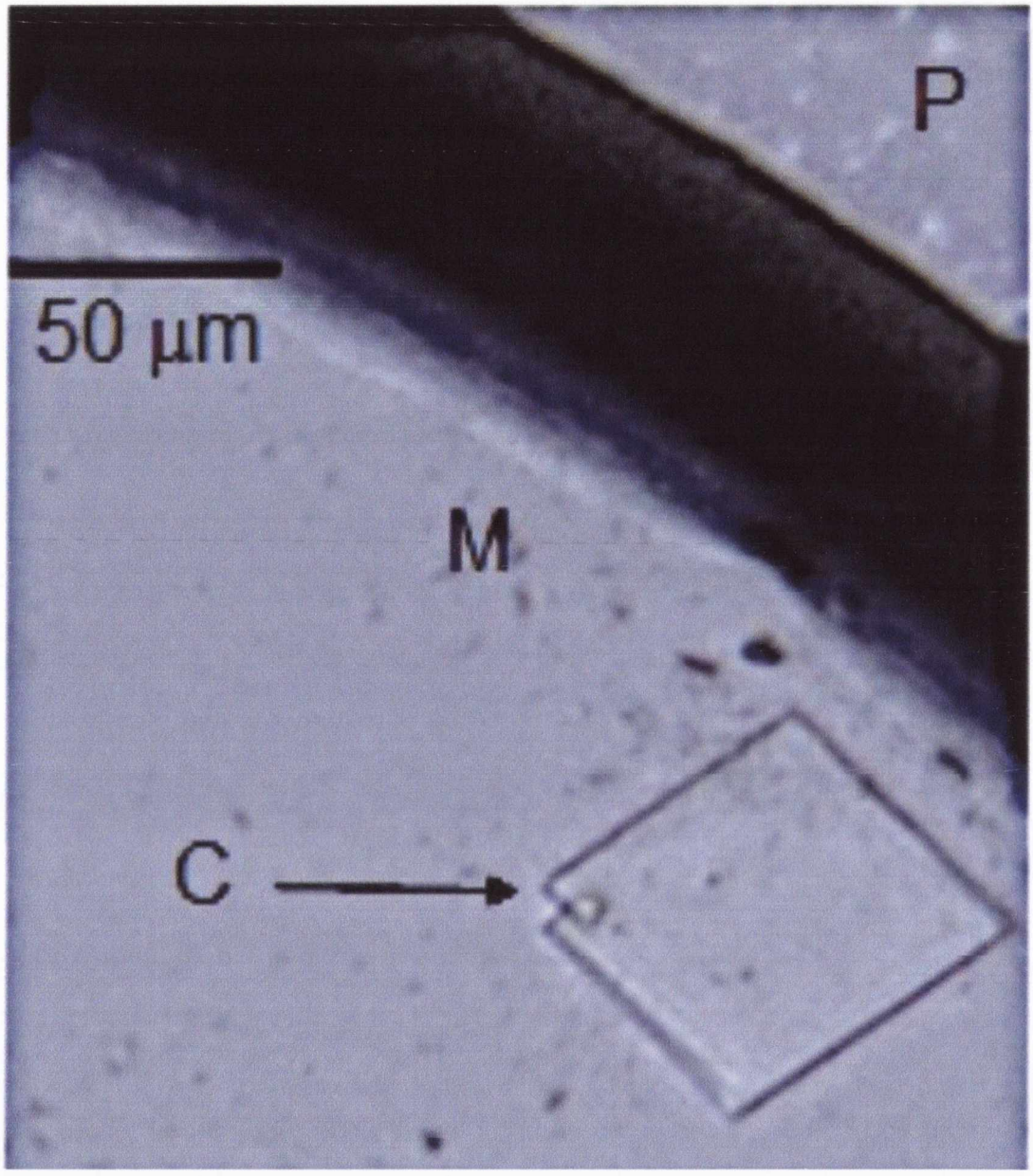


Figure 3.28 The crystal that yielded the best and complete data set to 1.9 Å. The crystal pointed by arrow is about 50 µm X 50 µm and the thickness is unknown. The protein concentration was 10 mg / mL and Crystals grew at 20 °C in lipid 7.8 MAG. The image was taken on day 4. The crystallisation condition was: 0.1 M Tris-HCl, pH 7.5, 0.1 M Sodium citrate, 18 %(v/v) PEG 400. M: mesophase drop. C: crystals. P: precipitant solution. In this case, the mesophase remains cubic, as the boundary of it is still clear.

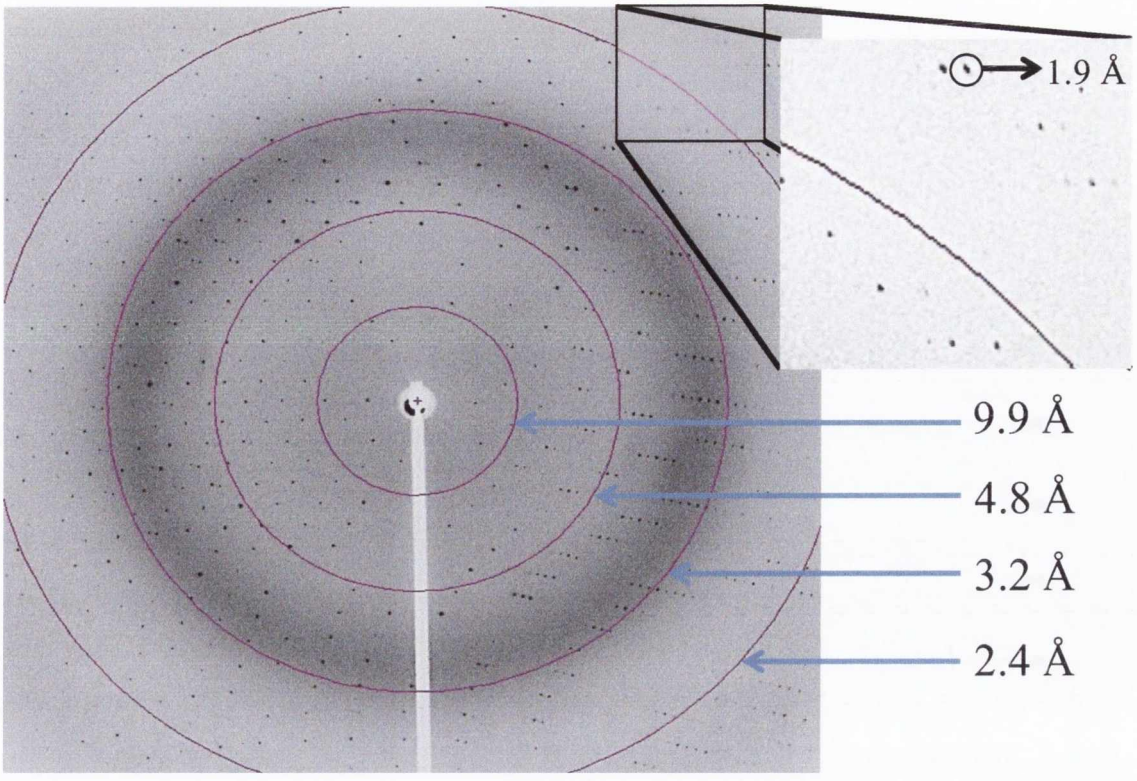


Figure 3.29 The best diffraction data of the crystal in Figure 3.28 from AlgE / 7.8 MAG *in meso* trials. The four purple rings represent different resolution shell as indicated by arrows on the left. Diffraction pattern consists of the black dots in the figure. The best resolution was up to about 1.9 Å, as indicated in the expanded region. The data were collected at GM/CA-CAT 23 ID-B beamline at APS. The energy for data collection was at 12.0 keV. The beam size was 20 µm X 20 µm and flux density was about 5 X 10^{14} photon $s^{-1}mm^{-2}$ (Fischetti *et al.* 2009).

Two batches of 7.8 MAG were tested after JPL-3-77 ran out, using the same AlgE batch that yielded the best diffraction data set, and no crystals were obtained. Because AlgE protein is usually stable after storage in – 80 °C freezer as mentioned in section 2.5.5, it might then be concluded that protein was the reason for irreproducibility.

Chemists in the group suspect that this ireproducibility is related to the impurities in synthesized lipids. One of the most common impurities is the unsaturated ester that is used in the coupling reaction for MAG synthesis, as pointed out by red arrow in Figure 3.30. This impurity is called 5.1 MAG for simplicity reason (Lyons 2012). As the impurity has distinct peaks in the proton NMR profile and some of these peaks are not masked by any oth-

er signals, a comparison of NMR profiles between two lipid batches (JPL-3-77 and RT-1-10) is done and shown in Figure 3.31. RT-1-10 is used as a representative of later batches of 7.8 MAG as these batches show very little difference in terms of NMR profile. The boxed region shows where the unique signals for 5.1 MAG will be in the NMR profiles of final product. An expanded view of the boxed region is shown in Figure 3.32.

As shown in both figures, there was around 6.0 %(w/w) 5.1 MAG in the batch JPL-3-77 and there was none in the batch RT-1-10. That was the major noticeable difference between the two batches. Because 5.1 MAG has much shorter chain length than 7.8 MAG, it was then proposed that this small unsaturated MAG partitions into the cubic phase system and flattens the membrane, which, in an unknown way, changed the crystal packing or the interaction between lipid and protein, which eventually improved the quality of AlgE crystals. By mixing purified 5.1 MAG into 7.8 MAG in different ratio and using the different mixture for AlgE *in meso* crystallisation trials, it may be possible to find out whether the proposed idea is right or wrong. But this is not done, as it isn't the main objective of this project.

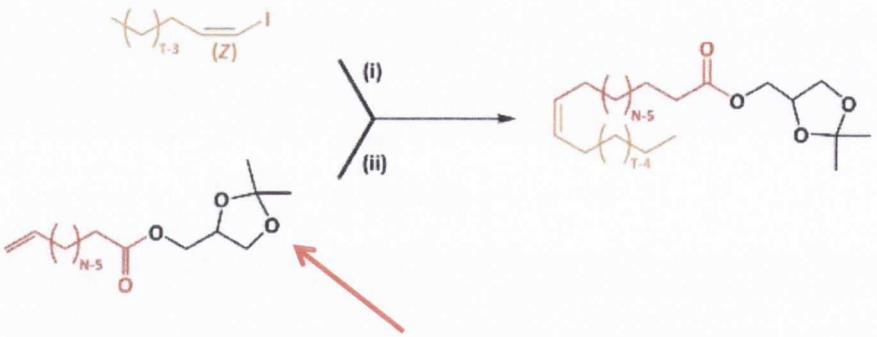


Figure 3.30 Suzuki-Miyaura coupling of an iodoalkene (orange) and the terminally unsaturated ester (red and black) to prepare MAG acetonide with different length of necks and tails. The red arrow points out one of the most common impurities in synthesized lipids. Copied from (Lyons 2012).

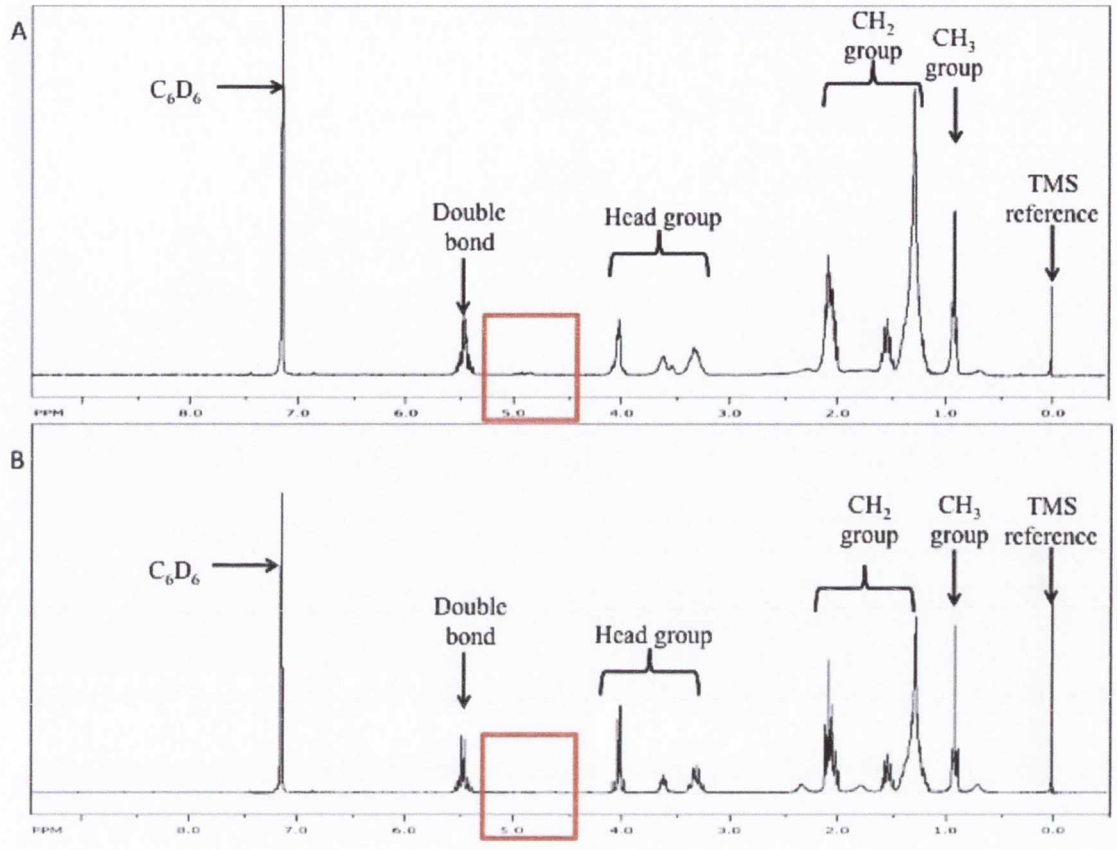


Figure 3.31 Proton NMR profiles for two different batches of 7.8 MAG. Panel A is the NMR profile for JPL-3-77 that yielded the best diffraction data set and panel B is for RT-1-10, a representative of all later batches. The boxed region is where impurity, 5.1 MAG, can be found. An expanded version of the boxed region can be seen in Figure 3.32. NMR experiment was done in 0.8 mL to 1.0 mL of C_6D_6 (D_6-benzene) and the experiment was done at room temperature. About 10 mg of each sample was used. The NMR profiles are provided by Dr. Joseph Lyons.

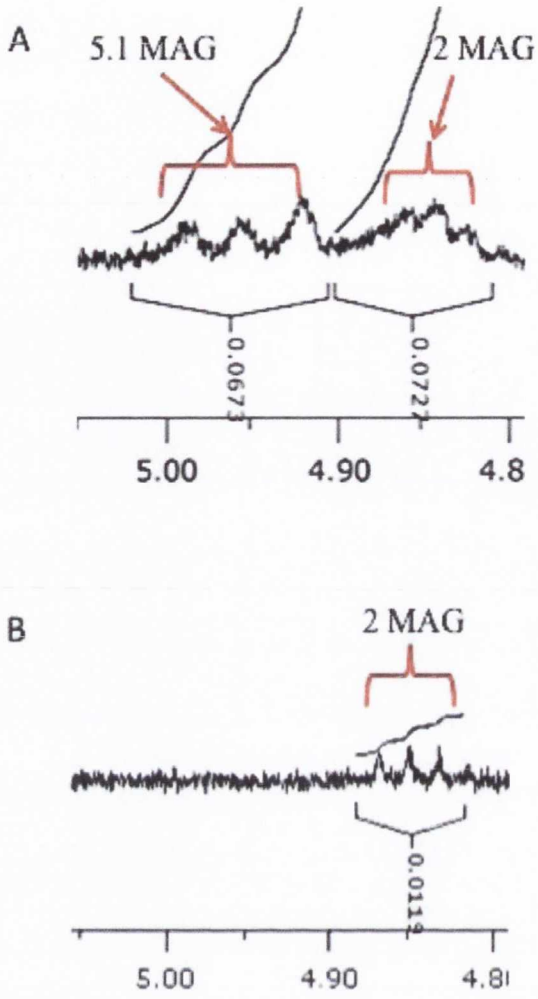


Figure 3.32 Expanded view of the boxed region in Figure 3.31. As stated in Figure 3.31, by integration, both 5.1 MAG and 2 MAG show obvious different amount in the two profiles. There is around 6.7 %(w/w) of 5.1 MAG and 7.3 %(w/w) of 2 MAG in panel A, which is 7.8 MAG (JPL-3-77). There is none 5.1 MAG and 1.1 %(w/w) of 2 MAG in panel B, which is 7.8 MAG (RT-1-10). But because 5.1 MAG also contains 2 MAG, the major difference between the two NMR profiles is the concentration of 5.1 MAG. The NMR profiles are provided by Dr. Joseph Lyons.

Back to AlgE work, a summary of the results that were obtained for crystals that grew in 9.9 MAG and 7.8 MAG are shown in Table 3.11. Crystals from 7.8 MAG *in meso* trials showed consistently better diffraction quality than 9.9 MAG and yielded the best native data set. As a result, crystals growing from 7.8 MAG were used for later phasing work.

Table 3.11 Summary of X-ray diffraction qualities of all crystal forms

Hosting lipid	Crystal size (μm)	Space group	Best diffraction (Å)
9.9 MAG	75 X 30	C222	2.80
7.8 MAG	100 X 30	$P2_1$	2.80
7.8 MAG	100 X 30	$P2_12_12_1$	2.40
7.8 MAG	150 X 150	C2	1.90

3.4.3 AlgE *in surfo* crystallisation trials

Out of 384 conditions, no obvious crystal hits except some very thin strings were found for AlgE *in surfo* crystallisation. The screened conditions included the conditions that were used for AlgE *in surfo* structure work (Whitney *et al.* 2009, Whitney *et al.* 2011), but no crystals were obtained in them. As a result, *in surfo* methods were abandoned for AlgE structure determination.

3.4.4 Phasing work

3.4.4.1 Colour screening experiment

When the colour screening experiment was planned, photos of each condition were planned to be taken. However, because the mesophase got interrupted during soaking process and had a rough surface, it showed very poor transparency and crystals were all buried inside it, it was not possible to take clear photos. As a result, no photos will be seen in this part. Of course, later phasing power study indicated that the judgement made by eye was valid.

After soaking AlgE *in meso* crystals with sixteen coloured heavy atom solutions, colour changes were observed in some cases. Details are listed in the Table 3.12.

Table 3.12 Results for colour screening experiment. Pictures of crystals after soaking are not available.

Name of chemical	Concen-tration (mM)	Colour of chemicals	Colour of cubic phase after soaking / back-soaking	Colour of crystals after soaking / back-soaking
$AuCl_3$	20	Yellow	Colourless	Colourless
$HAuCl_3$	20	Yellow	Colourless	Colourless
Ta_6Br_{12}	10	Green	Colourless	Colourless
K_2PtCl_4	10	Brown	Colourless	Colourless
K_2PtBr_6	10	Brown	Colourless	Colourless
K_2PtBr_4	10	Brown	Colourless	Colourless
$Pt(NH_3)_2(NO_2)_2$	10	Brown	Colourless	Colourless
$(NH_4)_2PtCl_2$	10	Brown	Colourless	Colourless
$OsCl_3$	20	Black	Black	Same as background
K_2OsCl_6	20	Black	Black	Darker than background
$(NH_4)_2IrCl6$	20	Brown	Colourless	Brown
$IrCl_3$	20	Brown	Colourless	Colourless
$(NH_4)WS_4$	20	Light yel-low	Colourless	Colourless
$HoCl_3$	20	Yellow	Colourless	Colourless
$NdCl_3$	20	Red purple	Red	Same as background
$PrCl_3$	20	Black	Colourless	Colourless
$CdCl_2$	10	Brown	Colourless	Colourless

According to colour screening experiment, AlgE crystals changed colour after being soaked with 20 mM $OsCl_3$, 20 mM K_2OsCl_6, 20 mM $(NH_4)_2IrCl_6$ and 20 mM $NdCl_3$. With 20 mM $OsCl_3$, the colour of both cubic phase and crystals was the same after soaking /

back-soaking, indicating that this heavy metal stained the lipid cubic phase and crystals at the same time.

With 20 mM K_2OsCl_6, crystals turned black and darker than the background, indicating that this heavy metal preferred to bind protein more than other components in the lipid cubic phase system.

With 20 mM $(NH_4)_2IrCl_6$, crystals turned slightly brown and the background was colourless, indicating this heavy atom also preferred to bind protein more than other components in the lipid cubic phase system.

With 20 mM $NdCl_3$, the colour of both cubic phase and crystals was the same red, indicating that he heavy metal stained lipid cubic phase and crystals at the same time. Overall, K_2OsCl_6 and $(NH_4)_2IrCl_6$ may have bound to AlgE crystals and were explored for phasing power.

3.4.4.2 Phasing power for K_2OsCl_6 and $(NH_4)_2IrCl_6$

Dr. David Aragão and Dr. Valerie Pye did all data collection and Dr. Valerie Pye did all the data processing work. The result is shown in Table 3.13.

Table 3.13 Phasing power of K_2OsCl_6 and $(NH_4)_2IrCl_6$ in AlgE structure work.

Chemical	Soaking Condition	Resolution (Å)	Phasing power Isomorphous
K_2OsCl_6	10 mM, 45 min	2.9	0.997
K_2OsCl_6	10 mM, 1 h	3.0	0.437
K_2OsCl_6	10 mM, 1 h	3.5	0.615
K_2OsCl_6	5 mM, 30 min	2.9	0.267
K_2OsCl_6	5 mM, 30 min	2.8	0.229
K_2OsCl_6	5 mM, 30 min	3.6	0.243
$(NH_4)_2IrCl_6$	20 mM, 1 h	3.7	0.293

6 data sets were collected for K_2OsCl_6 and only 1 data set was collected for $(NH_4)_2IrCl_6$.

From the diffraction quality of the only data set collected for $(NH_4)_2IrCl_6$, it looked like that after 1 h soaking with 20 mM $(NH_4)_2IrCl_6$, the AlgE crystals were damaged by the heavy atom and this might indicate that we need to lower the concentration or shorten the soaking time when using $(NH_4)_2IrCl_6$ for AlgE phasing work.

The result for K_2OsCl_6 looked particularly interesting. First, the phasing power was reasonably reproducible and the crystals after soaking experiment still diffracted best up to 2.8 Å. Second, the isomorphous phasing power increased with the increasing soaking concentration, although none of them was higher than 1.0, which is usually the minimum level required to solve the structure. Third, based on very limited data, Dr. Valerie Pye was able to build a crude model for AlgE structure.

An overview of the crude model is shown in Figure 3.33 and Figure 3.34. We could see the typical Type I packing of the crystal grown by *in meso* method and a very crude shape of protein plus solvent as a whole. The shape of the barrel, when looked at from extracellular side, is almost a square and this result agrees with the solved structure that will be presented in the next chapter. But without better phase, there is no way to trace the AlgE sequence and no correct unit cell dimension information can be derived from the model.

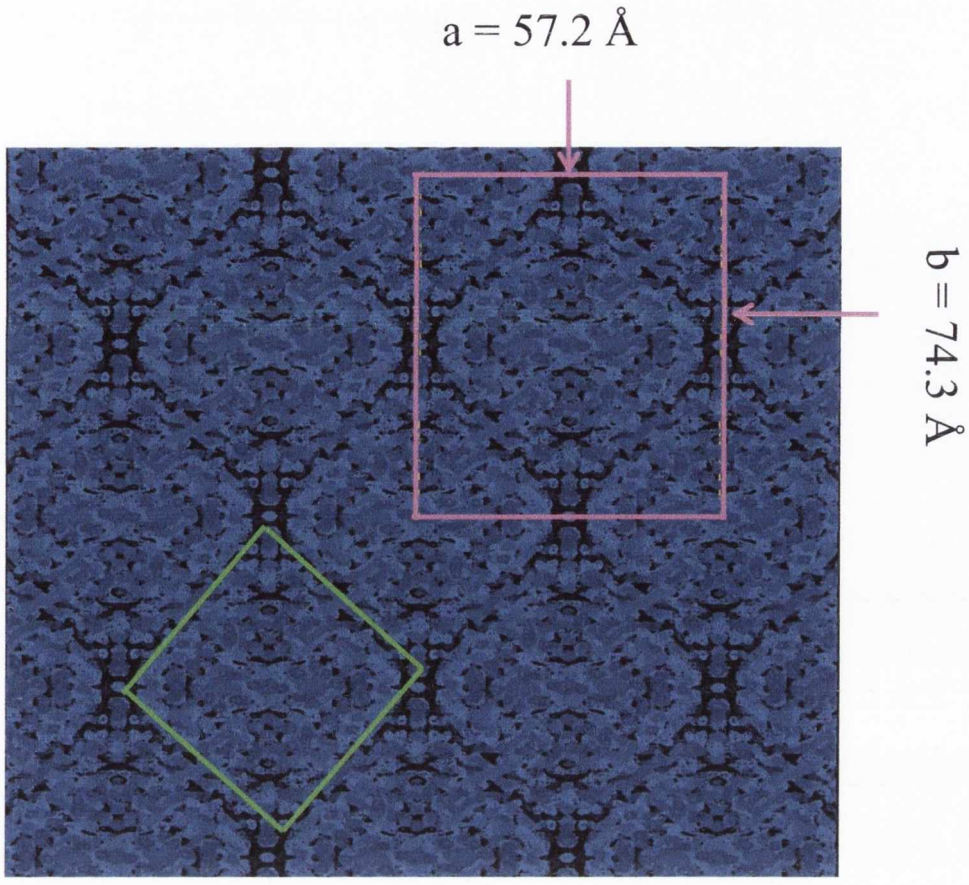


Figure 3.33 View of probable AlgE β-barrel in the unit cell, along c;ab face. Unit cell is defined by purple frame and cell dimensions of a and b are listed. The dark blue pattern in the green frame represents the view of a single AlgE protein molecule plus the solvent molecules around it from extracellular side. The overall shape of one protein barrel is almost a square.

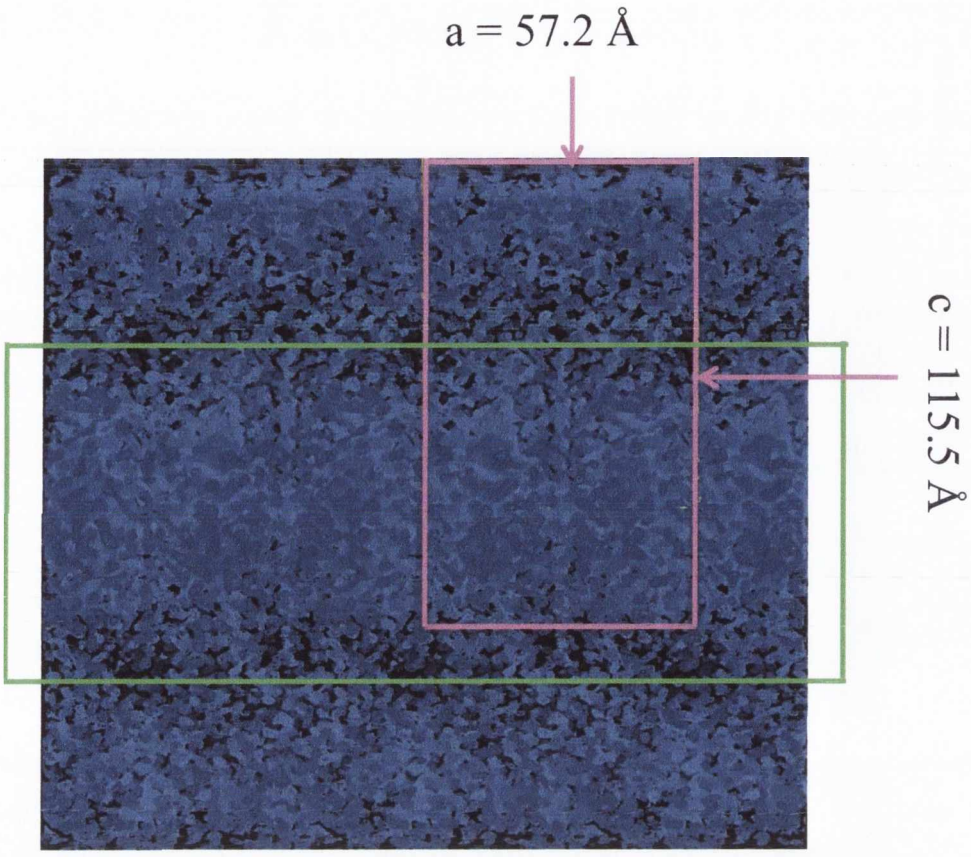


Figure 3.34 View of layer of density, along b;ac face. Unit cell is defined by purple frame and cell dimensions of a and c are listed. The dark blue pattern in the green frame represents a single layer of AlgE proteins plus solvent molecules in the Type I crystal packing style. The view is taken from membrane side.

3.5 Conclusions

3.5.1 Two types of crystallisation methods were explored for AlgE protein crystallisation at 20 °C.

After initial screenings using commercial kits, *in surfo* method didn't yield obvious crystals but *in meso* method generated small crystals in many conditions in different hosting lipids. 11 initial hits were found for 9.9 MAG *in meso* trials out of 1,056 conditions, 3 initial hits were found for 7.7 MAG out of 1,056 conditions, 24 initial hits were found for 7.8 MAG out of 1,056 conditions, 8 hits were found for 7.9 MAG out of 1,056 conditions, 16 initial hits were found for 8.8 MAG out of 1,056 conditions and 15 initial hits were found for 8.9 MAG out of 1,056 conditions. The conditions found for all hosting lipids were mainly PEG400 based and with a wide range of pHs, salts and other additives. Some conditions with only buffer and salt also worked and yielded crystals. For MAGs other than 7.7 MAG, under most of the conditions, mesophase remains cubic and the boundary can be clearly seen. But in the case of 7.7 MAG, under all three conditions, mesophase become spongified and the boundary of it is no longer clear.

3.5.2 After optimisation, three different crystal forms were found in 9.9 MAG and two different crystal forms were found in 7.8 MAG. In 9.9 MAG, parallelogram crystals and hexagonal crystals were obtained from absolutely same conditions on same plates due to some unknown reason, when PEG400 concentration is higher than 27 %(v/v). Needle crystals were only obtained from conditions with PEG400 concentration lower than 27 %(v/v). The two crystal forms in 7.8 MAG were all 2D-plates and they looked very similar to each other with only slight difference in the shorter dimension. Crystals in 7.8 MAG generally grow to 100 μm X 30 μm while the ones in 9.9 MAG generally grow to 75 μm X 30 μm. The mesophase remains cubic during the crystallisation process in almost all conditions except that it becomes spongy when PEG400 concentration is above 36 %(v/v). This stable cubic phase allows an efficient harvesting procedure as described Section 3.3.4.

3.5.3 Three complete X-ray diffraction data sets were collected for the crystals from 7.8 MAG. The first data set came from a single crystal obtained in PEG400 based conditions

and the best resolution was 1.9 Å. But this crystal form was not reproducible anymore. Second best data set came from crystals that grew in sodium acetate based conditions and the resolution was 2.4 Å. The third data set was collected from the crystals that also grew in the sodium acetate base condition has a different unit cell and space group, with the best resolution of 2.8 Å. In summary, the native data collected for AlgE had very reasonable resolution for a membrane protein (Kloppmann *et al.* 2012). At that stage, efforts were then focused on solving the phase and determining the structure.

3.5.4 The colour screening experiment identified K_2OsCl_6 and $(NH_4)_2IrCl_6$ as choices for AlgE phasing work. The phasing power tests showed that $(NH_4)_2IrCl_6$ likely damages the crystals during soaking experiment as the crystals diffracted very poorly. But the use of K_2OsCl_6 looked promising. The isomorphous phasing power increased with increasing concentration of K_2OsCl_6 but the phasing power was still lower than 1.0 and couldn't be used to solve the structure. A preliminary model had been built based on the limited phasing information. The model showed that the typical Type I packing which is usually found in most *in meso* structures (Caffrey *et al.* 2012). The crude shape of protein molecules plus solvent molecules around them was also seen. But there was no way to trace the protein sequence due to the lack of enough phase information.

3.5.5 A systematic study on the phasing power from K_2OsCl_6 had been started. However, before those crystals could be tested at any synchrotrons, Professor Lynne Howell's group at Hospital for Sick children, Toronto, solved the structure of AlgE from *P. aeruginosa.* The structure was deposited into PDB as 3RBH. As a result of that, experimental phasing work has been changed to molecular replacement and an *in meso* AlgE structure is discussed in chapter 4.

Chapter 4 *In meso* Structure of AlgE

4.1 Introduction

Although the alginate synthesis and secretion system (the 12 proteins in algD operon shown in Figure 1.3 and a separate protein named AlgC), is important to the life cycle of *P. aeruginosa*, little structure information was available. The first structure of this system came from AlgC, phosphomannomutase, in 2002 (Regni *et al.* 2002). The second structure came from AlgD, GDP-mannose dehydrogenase, in 2003 (Snook *et al.* 2003). These two proteins are both soluble proteins. After them, there had been a eight-year gap before the third structure was obtained from AlgK, a membrane associated protein with tetratricopeptide repeat (TPR)-like domain and unclear function (Keiski *et al.* 2010). Then in 2011, whilst my project was going on, the structure of AlgE, alginate transporter, was solved (Whitney *et al.* 2011).

The published structure of AlgE was solved using different construct and different crystallisation method from my work. The protein was cloned from *P. aeruginosa* PAO1 into *E. coli* host cells without the 32-residue N-terminal signal peptide. Then it was over-expressed as inclusion body (IB) form and solubilised in solution containing 8 M urea. The solubilised AlgE-IB was purified by Ni-NTA chromatography and refolded by rapid dilution into solutions containing decyl β-D-maltopyranoside (DM) before going onto the second size exclusion chromatography. On the second size exclusion chromatography, the detergent was exchanged to n-octyl tetraoxyethylene (C_8E_4) (Whitney *et al.* 2009).

The purified and refolded protein was used for *in surfo* hanging-drop vapour-diffusion crystallisation trials and yielded initial hits, which diffracted best to 3.0 Å (Whitney *et al.* 2009). Mr. Toufic El Arnaout, a fellow PhD student in MSFB group, had tried to repeat this published work in order to use *in surfo* crystals as an alternative method for phasing AlgE structure. Although the crystals diffracted best to 3.0 Å, the diffraction quality of crystals was not reproducible (unpublished thesis work: El Arnaout 2012) so this was not pursued further. The *in surfo* structure was eventually solved by MAD method using SeMet-labelling to a resolution at 2.3 Å and can be found in PDB with code 3RBH. The final statistics for the structure are shown in Table 4.1 (Whitney *et al.* 2009, Whitney *et al.* 2011).

An overview of the *in surfo* structure is shown in Figure 4.1. Overall, this structure model has eighteen transmembrane β-strands (S1 – S18) as predicted by previous bioinformatics and experimental work (Rehm *et al.* 1994, Hay *et al.* 2010). It has nine loops (L1 – L9) and L7 is exceptionally long, which folds into the pore region. The periplasmic turns (T1 – T9) are very short except for the T8. A calcium binding site is present in L1 and proposed to stabilize the protein. The transmembrane β-strands have a tilt angle of about 52° and a shear number of 22. This structure leads to the finding of a proposed secretion portal by bioinformatics work and reveals a closed state, in which two highly charged loops, L2 and T8, block the proposed secretion portal on both extracellular and periplasmic sides (Whitney *et al.* 2011). The details of this structure will be discussed together with my *in meso* structure model in Result section 4.4.3.

Using 3RBH as model for molecular replacement, molecular replacement was carried out and three *in meso* structure models were obtained from the X-ray diffraction data sets that are described in full details in Chapter 3. The full details of all *in meso* structures and a comparison with *in surfo* structure will be described in the following sections.

Table 4.1 Data collection and refinement statistics for AlgE in surfo structure model, 3RBH. Copied from (Whitney *et al.* 2011).

Space group		C2
Unit cell dimension (Å, °)		a = 162.0, b = 90.7, c = 160.3, β = 107.6
Resolution (Å)		50.00 – 2.30
No. of unique reflections		98,204
Redundancy		3.8 (3.6)[a]
Completeness (%)		99.8 (99.5)[a]
Average I/σI		20.2 (4.0)[a]
R_{merge} (%)[b]		9.3 (45.8)[a]
R_{work} / R_{free} [c]		0.221 / 0.259
No. of atoms	Protein	12,963
	C_8E_4	363
	Calcium	4
	Ethylene glycol	96
	Water	622
Average B factor (Å^2)	Protein	40.75
	C_8E_4	46.6
	Calcium	64.5
	Ethylene glycol	47.0
	Water	41.1
R.m.s. deviations	Bond length (Å)	0.008
	Bond angle (°)	1.086
Ramachandran plot (%)[d]	Total favoured	96.8
	Total allowed	95.5
Coordinate error (Å)[e]		0.34

a. Values in parentheses correspond to the highest resolution shell.

b. R_{merge} = ΣhΣi |I(h)i - <I(h)>| / Σi I(h)i. <I(h)> is the mean intensity of equivalent reflections (Rupp 2010).

c. R_{work} = Σ |*Fo-Fc*| / Σ|*Fo*|, where *Fo* and *Fc* are the observed and calculated structure factor amplitudes, respectively (Rupp 2010). R_{free} is calculated as of R_{work} but using a test-set of randomly sampled reflections (5 % of total observed reflections) that is omitted in the refinements (Brunger 1997).

d. As calculated using PHENIX (Adams *et al.* 2010).

e. Maximum-Likelihood Based Coordinate Error, as determined by PHENIX (Adams *et al.* 2010).

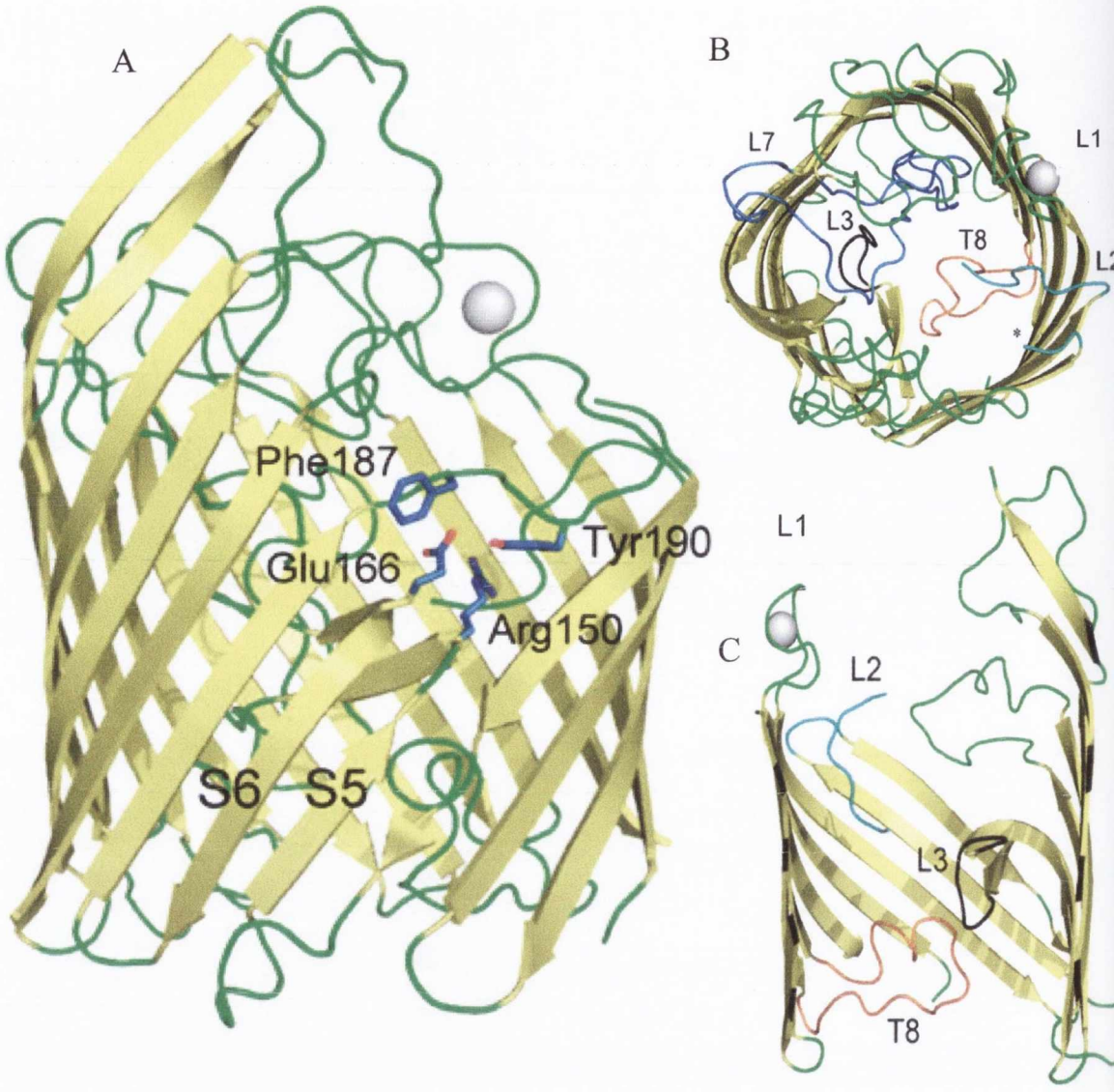


Figure 4.1 Overall *in surfo* structure of AlgE. (A) View of AlgE from membrane. Conserved residues Phe187 and Tyr190 that extend the hydrophobic surface area above Glu166 and Arg150 of strands S6 and S5, respectively, are shown in blue. The calcium ion bound to loop L1 is shown in gray. (B) View of AlgE from the extracellular space. Loops L2, L3, L7, and T8 are shown in cyan, black, blue, and orange, respectively. The asterisk indicates a region of L2 where no interpretable electron density is observed in the structure. (C) Slab view of AlgE showing occlusion of the pore constriction by loops L2, L3 and T8. Copied from (Whitney *et al.* 2011).

4.2 Method

Thanks very much to Mr. John Whitney in Professor Lynne Howell's group. The starting model (PDB ID 3RBH) used for molecular replacement was kindly provided by him on April 25th of 2011, ahead of publication in August 2011. Based on the *in surfo* structure model, Dr. David Aragão and I carried out molecular replacement and generated three *in meso* AlgE structure models. These models are then refined both manually in COOT and computationally in PHENIX. The final models are verified by online programme MOLPROBITY before submitting to PDB. The resulting models are analysed in order to reveal more information as to the mechanism for alginate secretion. The details of how molecular replacement and molecular refinement is done will be described in the following protocol.

4.2.1 Molecular replacement and structure refinement

- Prepare molecular replacement search model (3RBH) and remove all non-protein atoms using CHAINSAW in CCP4i (version: 6.1.2). Obtain the initial phases by molecular replacement using the program PHASER (version: 3.0) (McCoy *et al.* 2007). Three structure models are generated from the three X-ray diffraction data sets described in the Table 3.10 in section 3.4.2.

- Simulated annealing and initial stages of structure refinement are performed on resulted structure models using CNS (Brunger *et al.* 1998) with subsequent rounds of refinement carried out in PHENIX (version: 1.8-1069) (Adams *et al.* 2010). A criterion from one round to the next is that R_{free} should decrease with R_{work} proportionally and the structure model fits the resulted electron density map.
- Note 1: The refinement cycle for the structure in C2 space group (1.90 Å) includes: XYZ coordinates refinement, individual B-factors refinement, occupancies refinement and TLS refinement. The TLS groups are calculated using online server: http://skuld.bmsc.washington.edu/~tlsmd/. 4 TLS groups are used for Molecule A: 37 – 118; 119 – 232; 233 – 313; 314 – 490. The final structure model is then named as AlgE_1.9.
- Note 2: The refinement cycle for the structure in P212121 space group (2.40 Å) in-

cludes: XYZ coordinates refinement, individual B-factors refinement, occupancies refinement, TLS refinement and reference model refinement. The TLS groups are calculated using online server: http://skuld.bmsc.washington.edu/~tlsmd/. 5 TLS groups are used for Molecule A: 39 – 99; 100 – 118; 119 – 236; 237 – 358; 359 - 490; 4 TLS groups are used for Molecule B: 38 – 117; 118 – 215; 216 – 352; 353 - 490. The completely refined structure in C2 structure is used as a reference model in PHENIX to define secondary structure restraints due to the lack of enough reflections of this data set (Adams *et al.* 2010). The final structure model is then named as AlgE_2.4.

- Note 3: The refinement cycle for the structure in $P2_1$ space group (2.80 Å) includes: XYZ coordinates refinement, group B-factors refinement and reference model refinement with NSC restraints. The completely refined structure in C2 structure is used as a reference model in PHENIX to define secondary structure restraints due to the lack of enough reflections of this data set. The final structure model is then named as AlgE_2.8.

- Revise the structural model after each refinement cycle in real space with the program COOT (version: 0.7-pre-1) using sigma-A weighted $2F_o$-F_c and F_o-F_c electron density maps (Emsley and Cowtan 2004). The geometric quality of the model is continuously assessed with MolProbity (Chen *et al.* 2010). Both surface associated and integral water molecules were assigned based on sigma-A weighted $2F_o$-F_c electron density maps contoured at 1σ using standard geometrical and chemical restraints like coordination number and length or angle of hydrogen bond. In some cases, after refinement by PHENIX, the modelled water molecules show more than 4 coordinations while real water molecules should only have a maximum of 4. Then Na^+ or Mg^{2+} ions, which can be picked up during various stages from cell culture to crystallisation trials, are modelled to replace these water molecules and refined by PHENIX. Detergent and lipid (7.8 MAG) molecules are modelled and refined in the same way by PHENIX. The pdb files used for metal ions, detergents and water molecules come from online COOT library. The pdb file for lipid is generated by Dr. Joseph Lyons. In some cases where long lobes of electron density are observed, it is not possible to differentiate between lipids and detergents. For those densities, hydrocarbon chains are modelled instead. Overall refinement statistics are presented in Table 4.2 and Table 4.3 in section 4.3.1. The PDB headers for all three structures can be found in Appendix 6 – 8.

4.2.2 Analysis

Analysis is done in COOT, Chem3D Ultra (Version: 10.0) or CCP4i toolset. Then the results are displayed in and all figures in this chapter are prepared by and CCP4mg (version: 2.6.2).

4.3 Results and discussion

4.3.1 Molecular replacement and refinement

By molecular replacement, the three X-ray diffraction data sets from section 3.4.2.3 yielded three structure models that are very similar to each other. The r.m.s.d between all AlgE structure models is shown in Table 4.2 and confirms that these structures look very alike. In the meantime, these structures do vary in unit cell dimensions, crystal packing, the conformation of loops, metal ions and other solvent molecules. Structure in C2 space group has the best resolution as 1.90 Å (AlgE_1.9). Structure in $P2_12_12_1$ space group has resolution as 2.40 Å (AlgE_2.4). And structure in $P2_1$ space group has resolution as 2.80 Å (AlgE_2.8). And from now on, the *in surfo* AlgE structure, 3RBH, will be referred to as AlgE_in_surfo.

Before deposition into PDB, the quality of all three models has been assessed by both MolProbity (Chen *et al.* 2010). A summary of the structure refinement and geometry statistics for all three models is listed in the Table 4.3 and Table 4.4. The Ramachandran plots are shown in Figure 4.3 to Figure 4.5. In all *in meso* and even the *in surfo* AlgE structures, Gly261 is found as Ramachandran outlier. Gly residues are sometimes found as outliers because they have only one hydrogen atom in the place for side chain and is the least restrained of all amino acids, in terms of ϕ and ψ angles that define the ramachandran plot as shown in Figure 4.2. But there is no obvious reason that this particular Gly261 is an outlier in all AlgE structures. These structure models have been deposited in PDB as HPUB status: 4AFK (AlgE_1.9) on 19th Jan 2012, 4B61 (AlgE_2.4) on 8th Aug 2012 and 4AZL (AlgE_2.8) on 26th Jun 2012. The entries will be released when publication is ready or after being deposited in PDB for over a year.

Table 4.2 A summary of root-mean-square deviation (r.m.s.d)[a] of C_α between each of four AlgE structure models.

Between models of	R.m.s.d (Å)
AlgE_1.9 / AlgE_2.4	0.47
AlgE_1.9 / AlgE_2.8	0.98
AlgE_1.9 / AlgE_in_surfo	0.37 (very similar)
AlgE_2.4 / AlgE_2.8	0.6
AlgE_2.4 / AlgE_in_surfo	1.27
AlgE_2.8 / AlgE_in_surfo	0.45

a. The r.m.s.d value is calculated in COOT by superposing structures together.

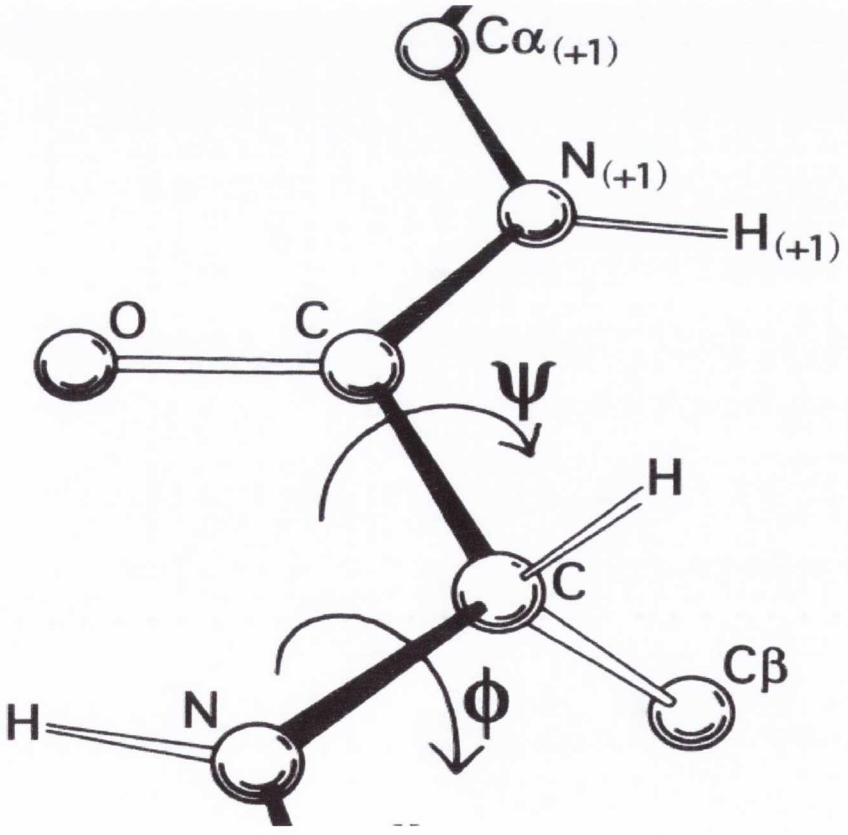


Figure 4.2 Illustration of ϕ and ψ angles of peptide bond that define Ramachandran plot for L- amino acids: http://en.wikipedia.org/wiki/Ramachandran_plot.

Table 4.3 Refinement and geometry statistics for AlgE *in meso* structure models.

		AlgE_1.9	AlgE_2.4	AlgE_2.8
PDB ID		4AFK	4B61	4AZL
Number of crystals used for data set		1	1	2
Number of reflections		35,580	45,596	23,738
Completeness (%)		94.51 (84.9)	98.60 (99.9)	93.25 (95.3)
R_{work} / R_{free}[b]		0.1656 / 0.2086	0.2217 / 0.2436	0.2393 / 0.2803
Average B-factors (Å^2)	Overall	23.51	40.72	44.24
	Protein	20.93	A: 38.64 B: 41.69	A: 43.49 B: 45.21
	Metal ions	38.87	60.37	65.98
	Detergents / lipids / hydrocarbon chain	47.75 / 45.61 / 44.54	66.84 / 58.42 / 49.51	50.46 / N/A / 36.23
	Waters	31.073	38.192	43.944
	Others	Citrate: 37.818 PEG fragment: 28.307	Acetate: 57.735	N/A
R.m.s d[c]	Bond lengths (Å)	0.010	0.019	0.006
	Bond angles (°)	1.344	1.109	1.218
Ramachandran plot statistics (%)[c]	Favoured region	97.41	96.85	97.45
	Allowed region	2.35	2.92	2.31
	Outlier region	0.24	0.23	0.24
Rotamer outlier (%)[c]		0	0.57	0.15
Clash score (all atoms)[c]		7.31	10.08	7.69
Cβ deviations > 0.25 Å[c]		0	0	0

a. Values in parentheses correspond to the highest resolution shell.

b. $R_{work} = \Sigma |Fo-Fc| / \Sigma|Fo|$, where *Fo* and *Fc* are the observed and calculated structure factor amplitudes, respectively (Rupp 2010). $R_{free} = \Sigma |Fo-Fc| / \Sigma|Fo|$, calculated using a test-set of randomly sampled reflections (5 % of total observed reflections) omitted in the refinements (Brunger 1997).

c. These values are calculated by MolProbity (Chen *et al.* 2010).

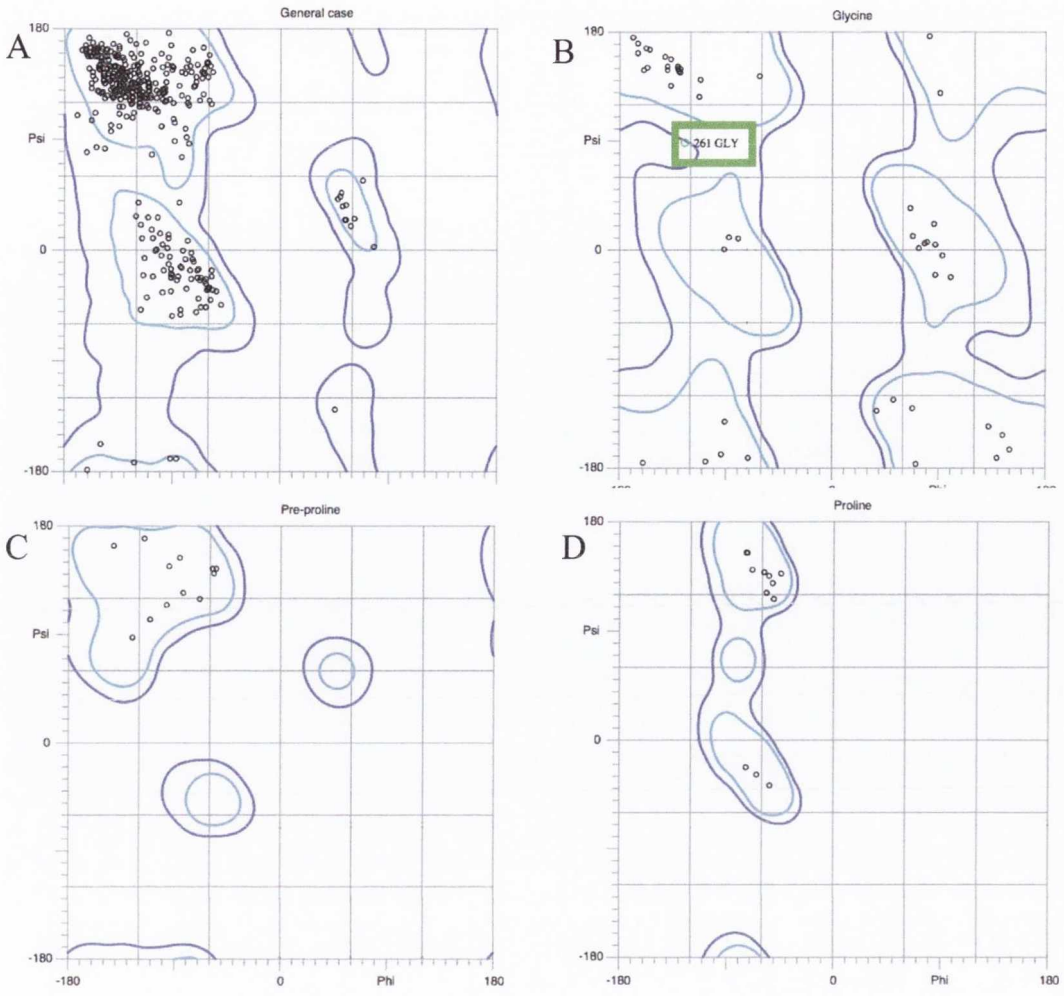


Figure 4.3 Ramachandran plots for AlgE_1.9. In panel A, a plot for the general case, which includes all amino acids but Gly, Pro and pre-Pro (non-Gly and non-Pro residues preceding Pro in protein sequence), is shown. In panel B, C and D, plots for Gly, pre-Pro and Pro are shown respectively. The dark blue outer line represents the allowed region and the inner light blue line represents the favoured region. Region outside the dark blue line is disallowed region. Only one outlier is found, Gly261, and labelled in green frame in Panel B.

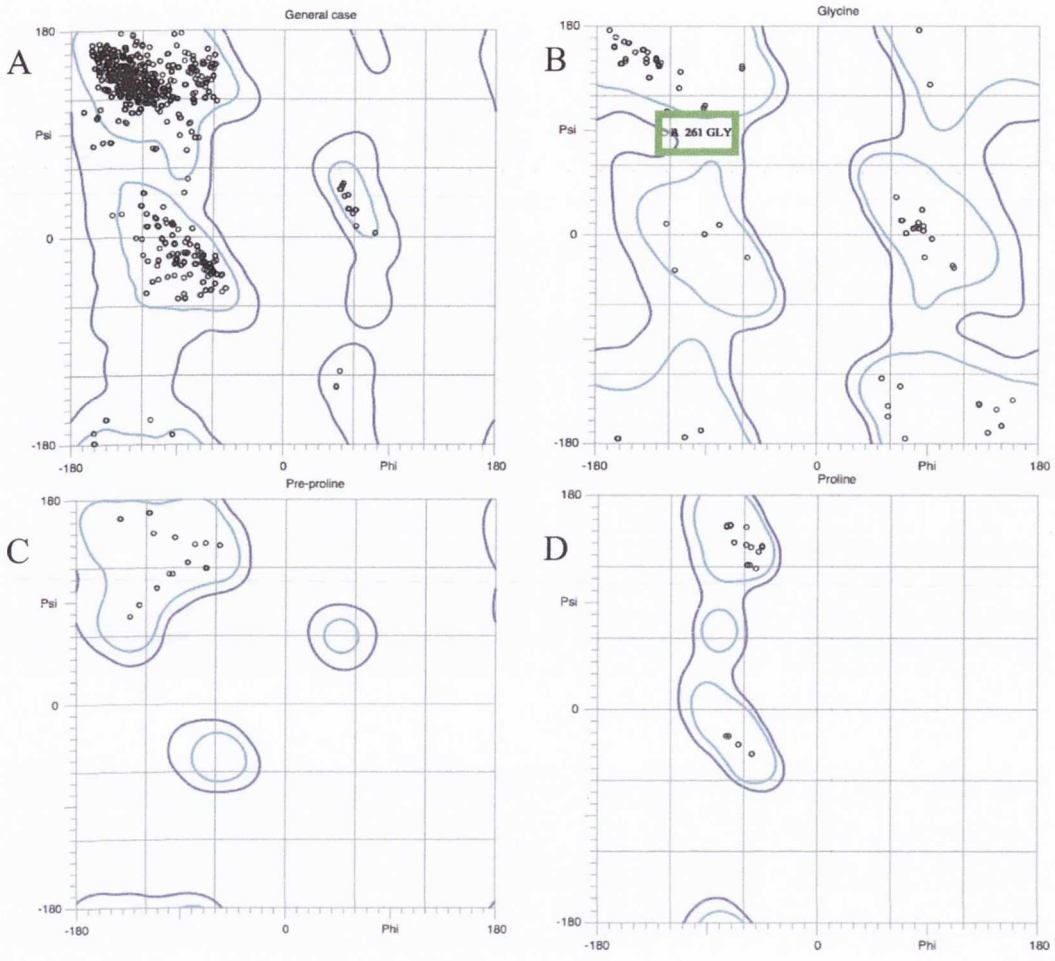


Figure 4.4 Ramachandran plots for AlgE_2.4. In panel A, plot for the general cases is shown not including Gly, Pro or pre-Pro. In panel B, C and D, plots for Gly, pre-Pro and Pro are shown respectively. The dark blue outer line represents the allowed region and the inner light blue line represents the favoured region. Region outside the dark blue line is disallowed region. Only one outlier is found, Gly261, and labelled in green frame in Panel B.

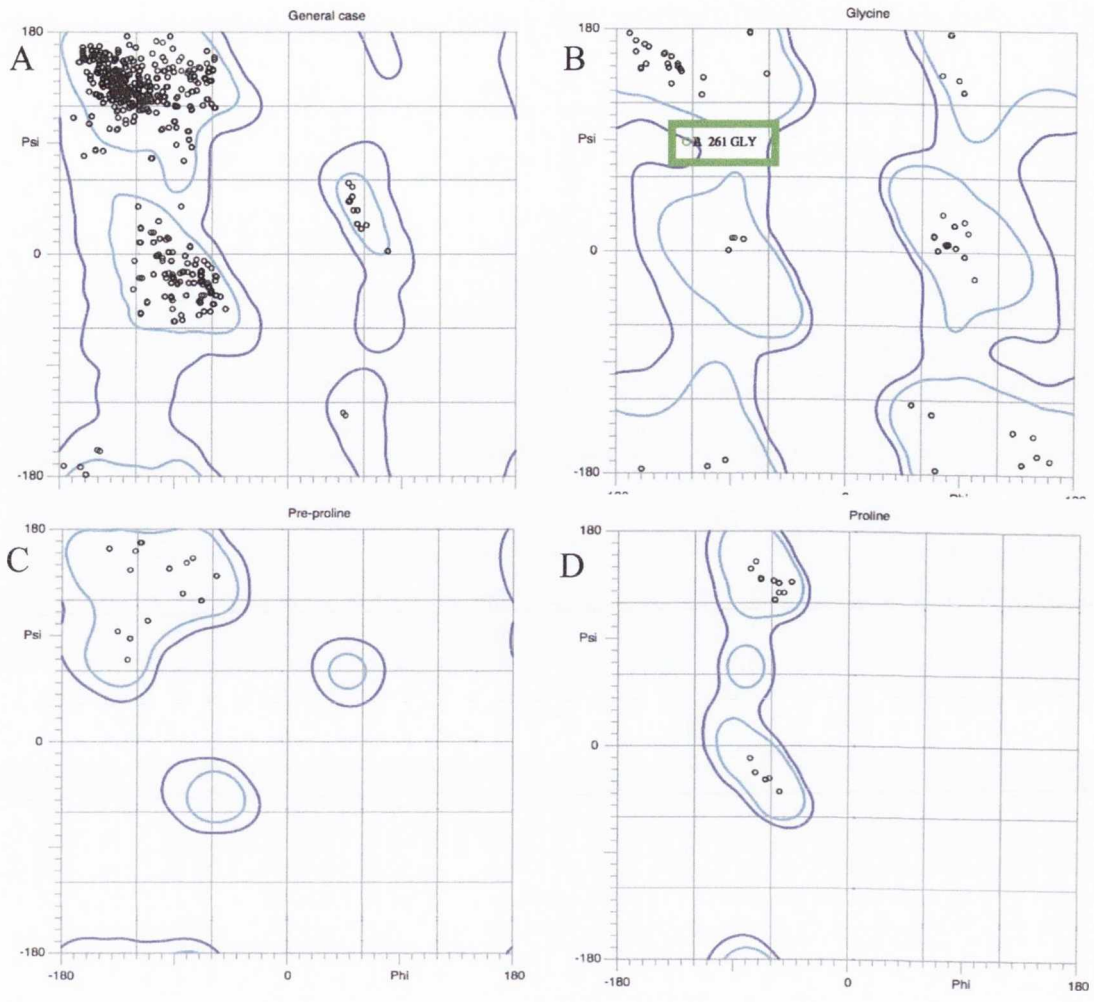


Figure 4.5 Ramachandran plots for AlgE_2.8. In panel A, plot for the general cases is shown not including Gly, Pro or pre-Pro. In panel B, C and D, plots for Gly, pre-Pro and Pro are shown respectively. The dark blue outer line represents the allowed region and the inner light blue line represents the favoured region. Region outside the dark blue line is disallowed region. Only one outlier is found, Gly261, and labelled in green frame in Panel B.

Table 4.4 Summary of molecules in one asymmetric unit for all four AlgE models.

	AlgE_1.9	AlgE_2.4	AlgE_2.8	AlgE_in_surfo
PDB ID	4AFK	4AZL	4B61	3RBH
Number of total atoms	4,073	7,510	6,942	14,048
Protein molecule per asymmetric unit: disordered residues that can't be mod-elled	A: 33 - 36, 109 - 116, 439 - 454	A: 33 - 38, 106 -115, 441 - 453	A: 33 - 39, 104 -120, 240 - 246	A: 33 – 39, 105 - 120
		B: 33 - 37, 106 -116, 472 - 474	B: 33 - 38, 62 - 64, 105 - 119, 242 - 244	B: 33 – 39, 104 – 120
				C: 33 – 39, 105 – 120
				D: 33 – 39, 105 - 121
Number of lipids per asymmetric unit	7.8 MAG: 6	7.8 MAG: 5	7.8 MAG: 5	None
Number of detergents per asymmetric unit	LDAO: 6	LDAO: 1	0	C8E4: 29
Number of hydrocar-bon chains per asymmetric unit	9[a]	19[b]	8[c]	None
Number of metal ions per asymmetric unit	Ca^{2+}: 2 Na^{+} : 1	Cu^{2+} : 4 Mg^{2+} : 1	Ca^{2+}: 2	Ca^{2+}: 4
Number of waters per asymmetric unit	242	205	50	622
Other molecules per asymmetric unit	Citrate: 1 PEG frag-ment: 2	Acetate: 1	None	EDO: 26

a. The length of hydrocarbon chains in AlgE_1.9 range from 8 to 12 carbons.
b. The length of hydrocarbon chains in AlgE_2.4 range from 7 to 15 carbons.
c. The length of hydrocarbon chains in AlgE_2.8 range from 7 to 13 carbons.

4.3.2 Overall structure

The fully refined *in meso* models are shown individually in Figure 4.6 to Figure 4.8. In all of them, one protein molecule consists of 18 transmembrane β-barrel strands (S1 to S18), 9 extracellular loops (L1 to L9) and 9 periplasmic turns (T1 to T9). L7 is the longest loop and folds into the pore region. T8 is unusually long compared to other periplasmic turns of AlgE protein and consists of 19 residues, which is disordered and so can't be modelled in AlgE_1.9 and the molecule A of AlgE_2.4.

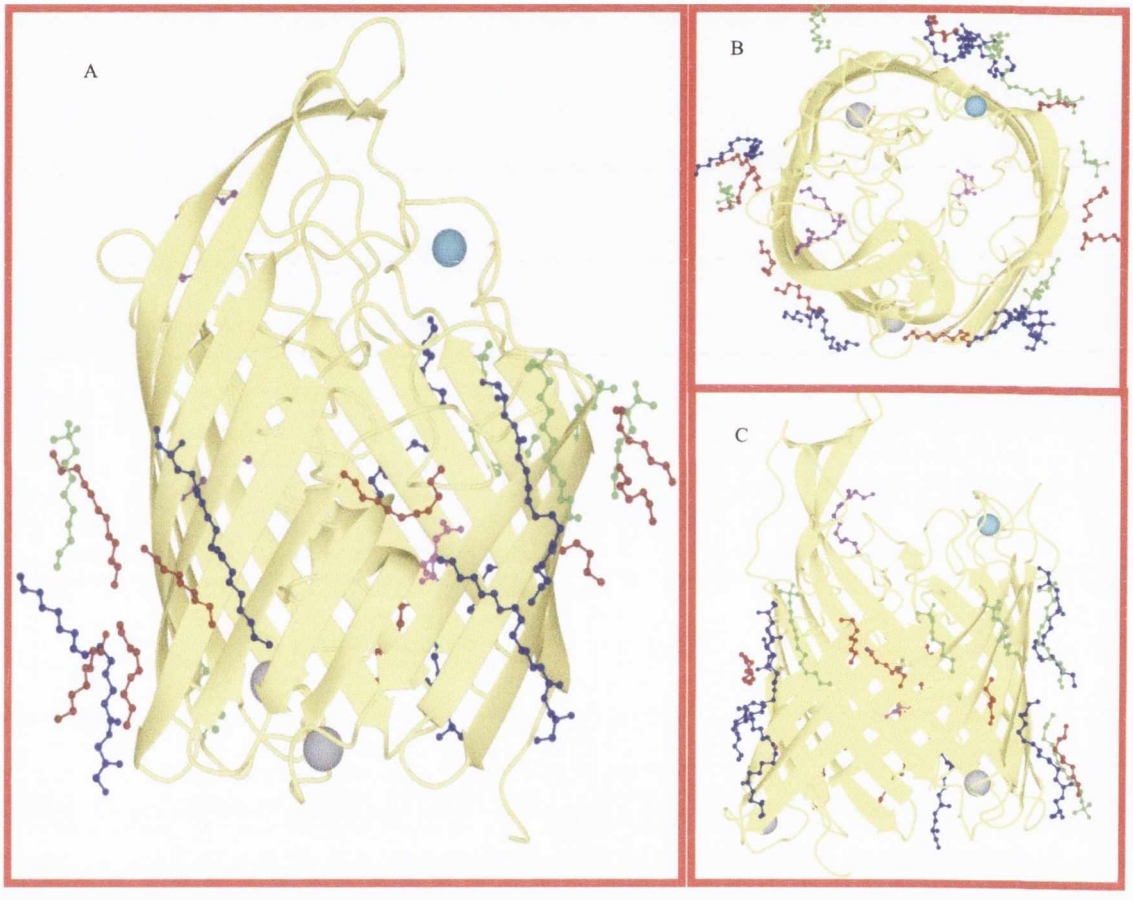


Figure 4.6 Overview of AlgE_1.9 (code 4AFK in PDB). In panel A, view of AlgE_1.9 from membrane is shown. The protein molecule is drawn in ribbon style and coloured in yellow. The LDAO molecules are drawn in ball-and-stick style and coloured in green, the 7.8 MAG molecules in blue and the hydrocarbon chains in brown. The two sodium ions are drawn in sphere style and coloured in grey, the calcium ion in cyan. A citrate is inside the barrel region and coloured in magenta. Two PEG fragments are coloured in purple. In panel B, view of AlgE_1.9 from extracellular side is shown. In panel C, the view is taken by rotating panel A 90° horizontally.

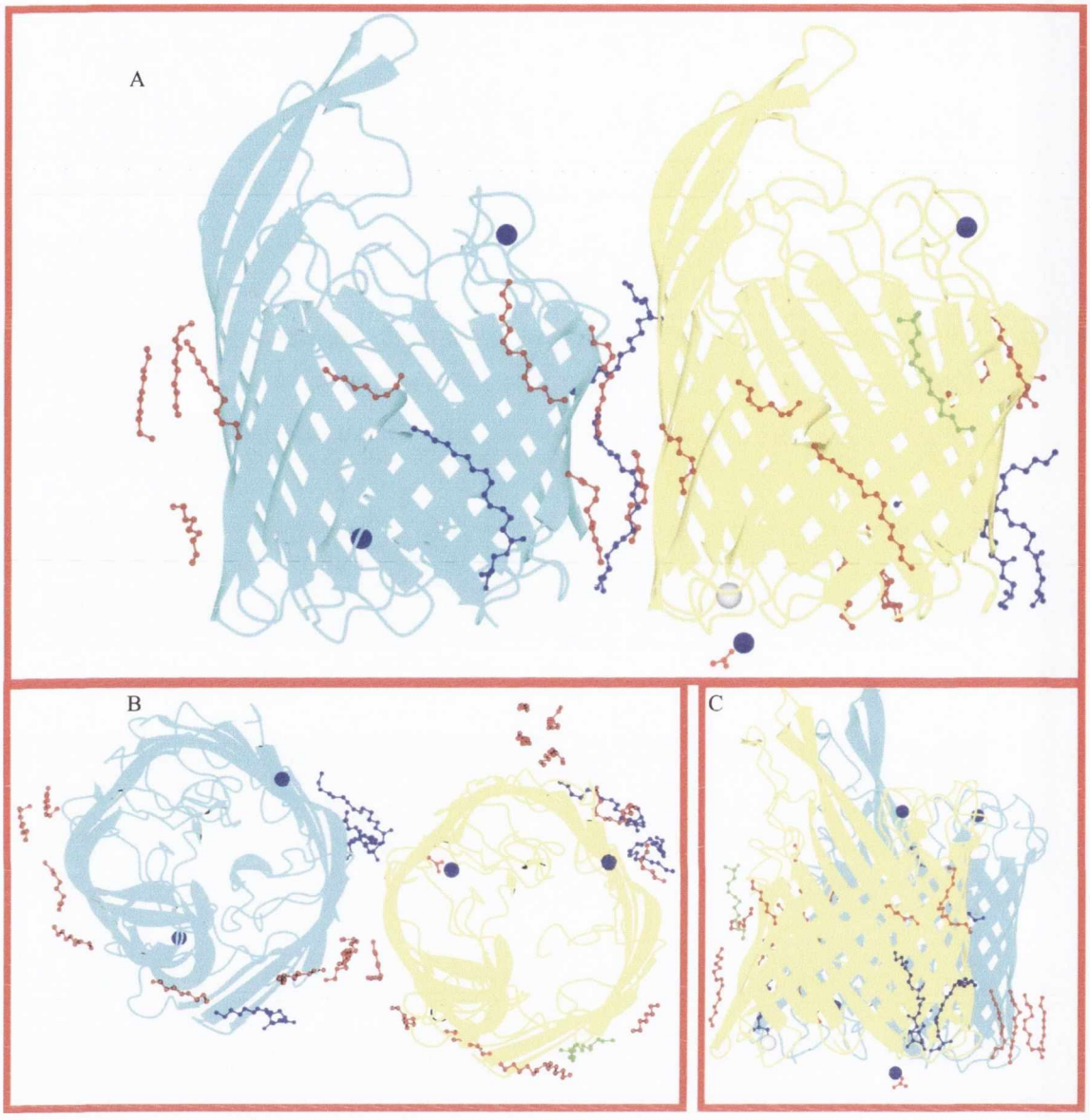


Figure 4.7 Overview of AlgE_2.4 (code 4B61 in PDB). In panel A, view of AlgE_2.4 from membrane is shown. The protein molecules are drawn in ribbon style. Molecule A is coloured in cyan and molecule B in yellow. The LDAO molecules are drawn in ball-and-stick style and coloured in green, the 7.8 MAG molecules in blue and the hydrocarbon chains in brown. The sodium ion is drawn in sphere style and coloured in grey, the copper ions in blue. In panel B, view of AlgE_2.4 from extracellular side is shown. In panel C, the view is taken by rotating panel A 90° horizontally.

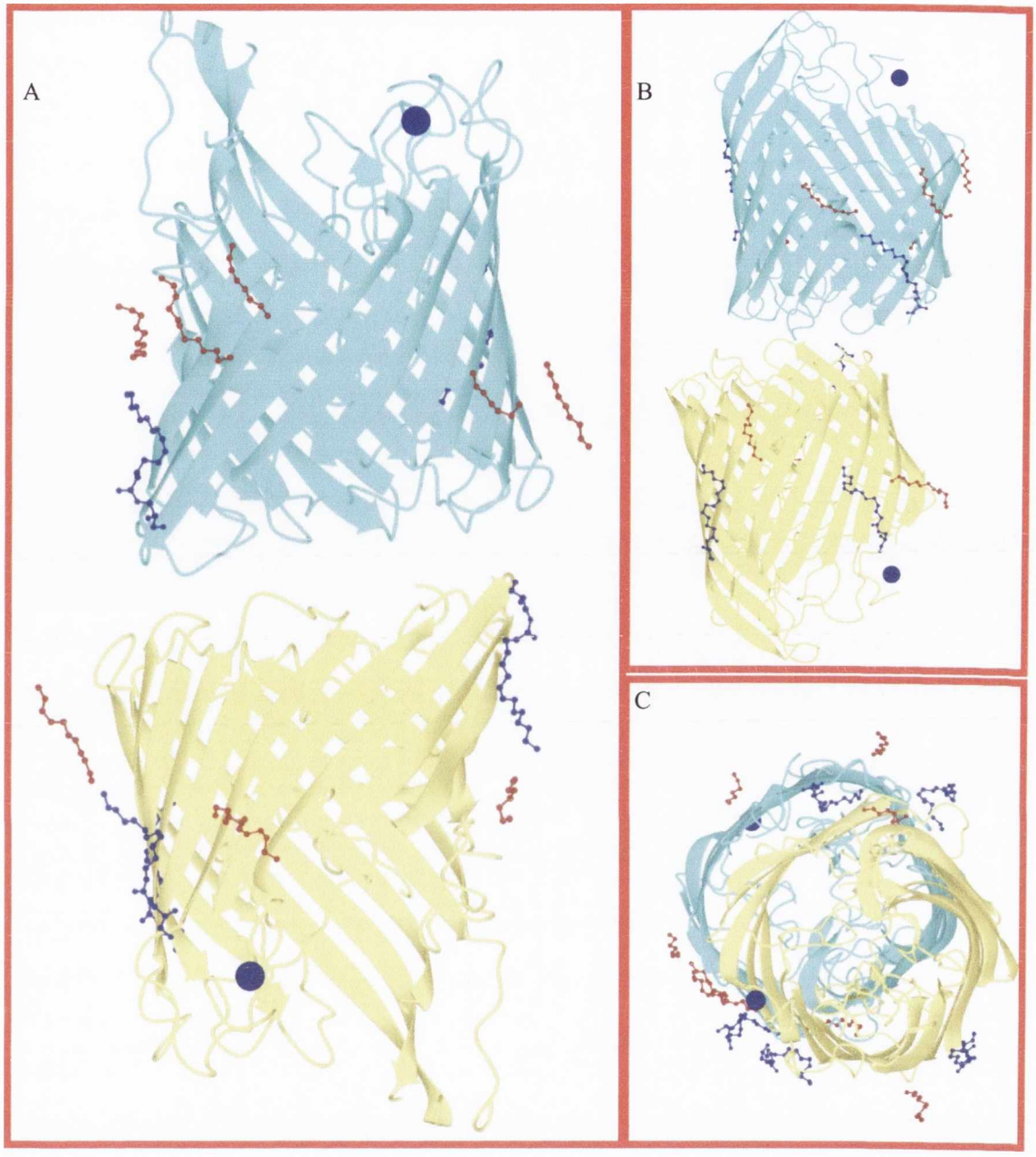


Figure 4.8 Overview of AlgE_2.8 (code 4AZL in PDB). In panel A, view of AlgE_2.8 from membrane is shown. The protein molecules are drawn in ribbon style. Molecule A is coloured in cyan and molecule B in yellow. The 7.8 MAG molecules are coloured in blue and the hydrocarbon molecules in brown. The calcium ions are drawn in sphere style and coloured in blue. In panel B, view of AlgE_2.8 from extracellular side is shown. In panel C, the view is taken by rotating panel A 90° horizontally.

When superposed altogether, with the *in surfo* structure 3RBH, it is clear that all structure models are very similar to each other as already suggested by the r.m.s.d in the Table 4.2. The superposed structure models are shown in Figure 4.9 and it shows that the transmembrane segments are almost identical but the loop regions have some difference. As a result, AlgE_1.9 is used as a representative for calculation of the hydrophobic thickness for AlgE molecule. The boundaries are shown in Figure 4.10 and the hydrophobic thickness is calculated to be 24.8 Å using online OPM server: http://opm.phar.umich.edu/server.php. This value is considered normal when compared to other known β-barrel proteins in OPM database: from the minimum 20.6 Å of OprH (PDB ID: 2LHF; 8 transmembrane strands) from *P. aeruginosa,* to the maximum 28.6 Å of OMP (PDB ID: 3DZM; 8 transmembrane strands) from *Thermus thermophilus*, with all 18-stranded β-barrel proteins ranging from 21.4 Å to 24.2 Å.

It is known that the bilayer thickness of 9.9 MAG and 7.7 MAG is 32.3 Å and 25.8 Å, respectively (Misquitta *et al.* 2004). Based on the assumption that the bilayer thickness added by one carbon number is the same for all lipids, the bilayer thickness of other novel MAGs used in this project is calculated to be as follows: ~27.4 Å for 7.8 MAG, ~29.0 Å for 7.9 MAG and 8.8 MAG and ~30.6 Å for 8.9 MAG. It is clear that the value of bilayer thickness for 7.7 MAG or 7.8 MAG is closer, than other MAGs, to the value of hydrophobic thickness of AlgE protein. In addition, as shown in Table 3.9 in section 3.4.2.1 of the last chapter, it is found out that the initial crystal hits from 7.8 MAG yielded the best diffraction resolution, which is 2.8 Å and made 7.8 MAG my choice for *in meso* AlgE structure work. Thus, it is reasonable to come to the conclusion that 7.8 MAG yielded best initial crystal hits because its bilayer thickness matches AlgE hydrophobic thickness better than all the other MAGs, except for 7.7 MAG. The logic behind this conclusion is that the more matching between the two values, the more likely the lipid bilayer resembles the native environment for reconstituted protein and the more likely the protein will maintain a natural conformation, which eventually results in crystals of better quality (Li *et al.* 2011). However, it has to be noted that 7.7 MAG, in fact, provides the best hydrophobic matching but the best initial crystal hits grown from it only diffracted to 3.4 Å, which is even slightly worse than 3.3 Å from 8.8 MAG. The discrepancy between this observation and the conclusion is probably because of the following reasons:

A. The initial crystal hits are not optimised so the result in Table 3.9 doesn't necessarily reflect the best diffraction quality that can possibly be obtained from these lipids after extensive optimisation. However, it has to be kept in mind that the best diffraction resolution obtained from 7.8 MAG and 9.9 MAG after extensive optimisation is 1.90 Å and 2.80 Å, respectively, which does support that the better hydrophobic matching is, the better diffraction quality of crystals will be obtained by using the hosting lipid.

B. The behaviour of 7.7 MAG is very different from the other MAGs (Misquitta *et al.* 2004) in that it never forms transparent cubic phase during my experiment set up and gets easily spongyfied or turns lamellar with the existence of high concentration of precipitant solutions. As a result, many conditions that are included in the commercial screen trials may not be compatible with this lipid and that lowers the chance to find proper initial crystal hits in this hosting lipid.

C. As stated in the previous literature, the hydrophobic matching may be too perfect in 7.7 MAG so AlgE protein would like to stay inside this stable environment instead of diffusing into the lamellar conduit as required by the proposed mechanism of *in meso* crystallisation method (Li *et al.* 2011).

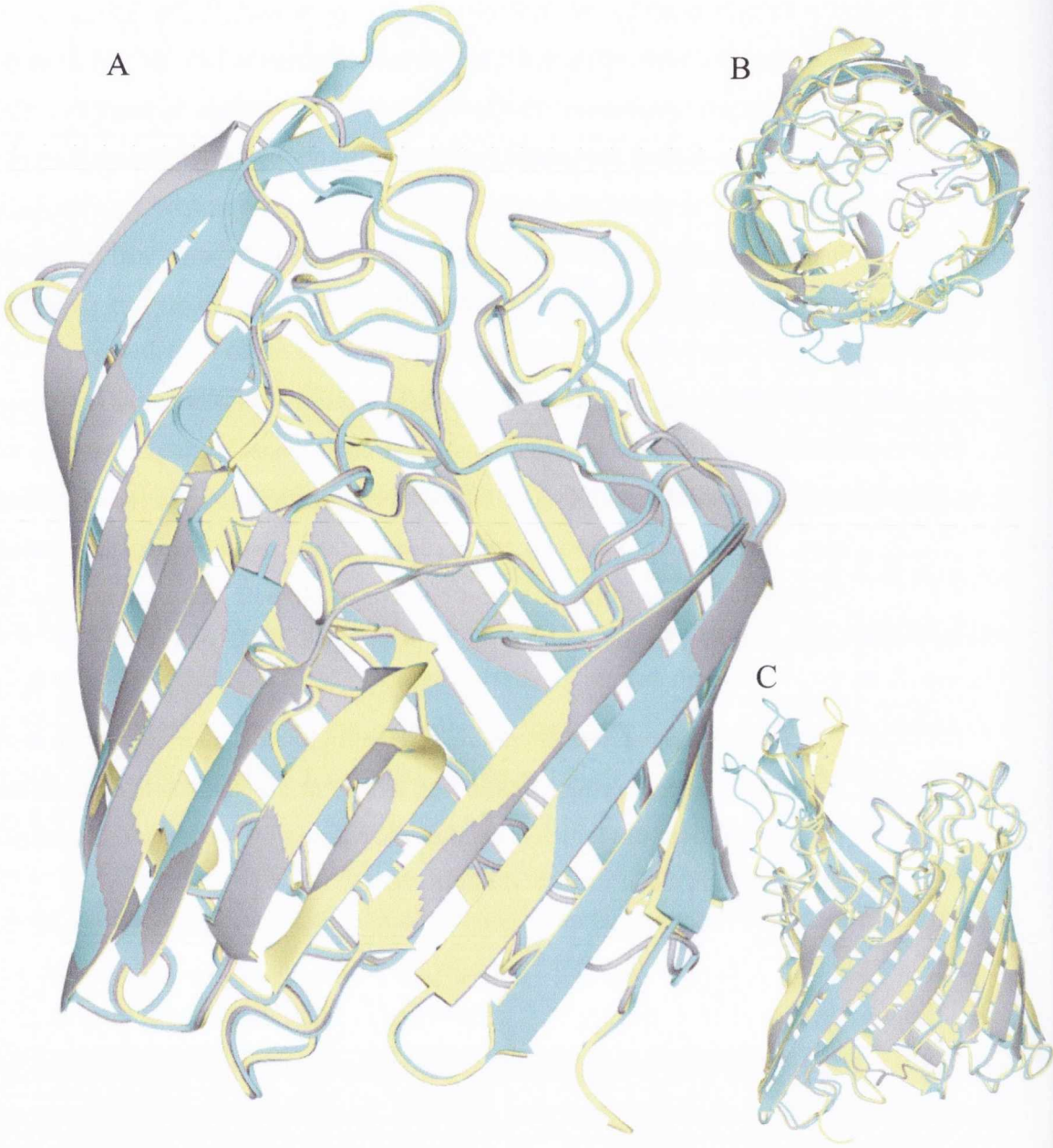


Figure 4.9 A superposed view of all three *in meso* structures. All structure models are shown in ribbon style. AlgE_1.9 is coloured in yellow, AlgE_2.4 in cyan, AlgE_2.8 in grey. In panel A, view from membrane is shown. In panel B, view of AlgE molecules from extracellular side is shown. In panel C, the view is taken by rotating the molecules of panel A 90° horizontally.

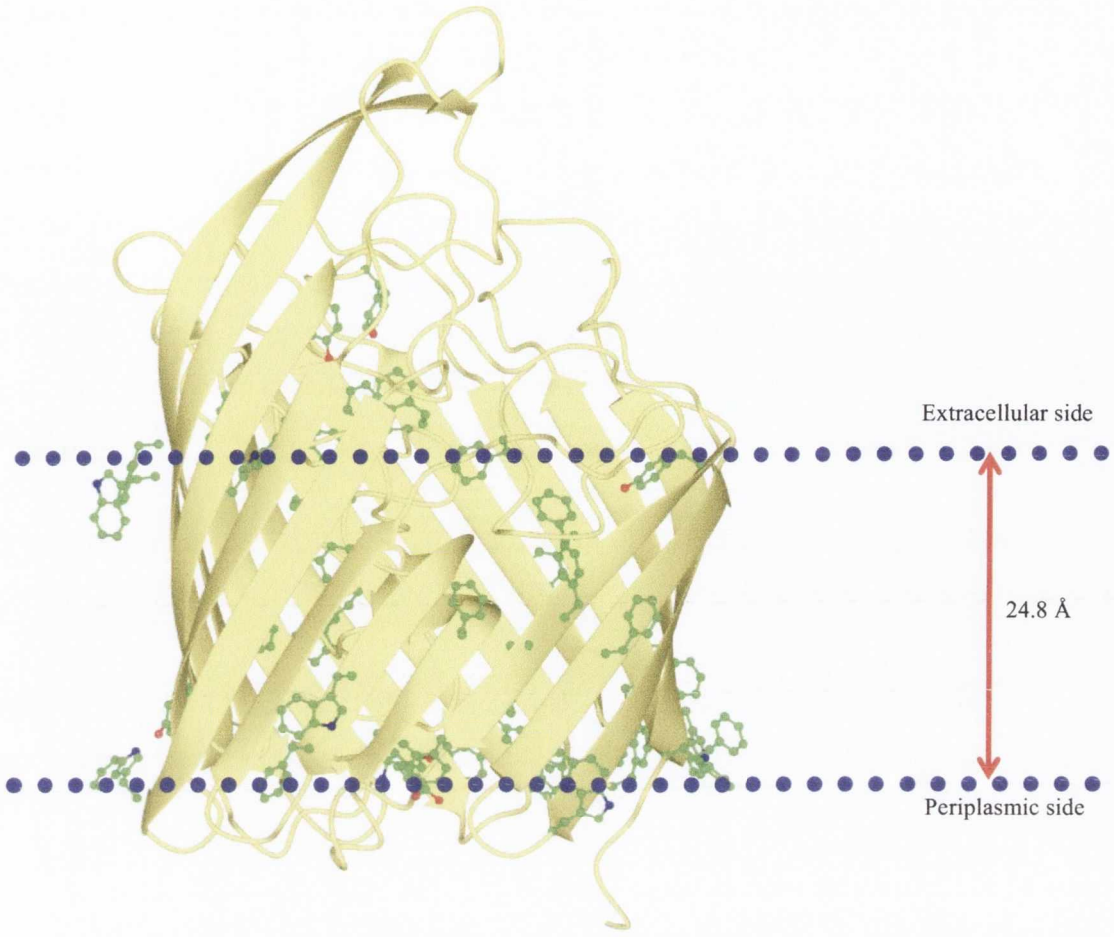


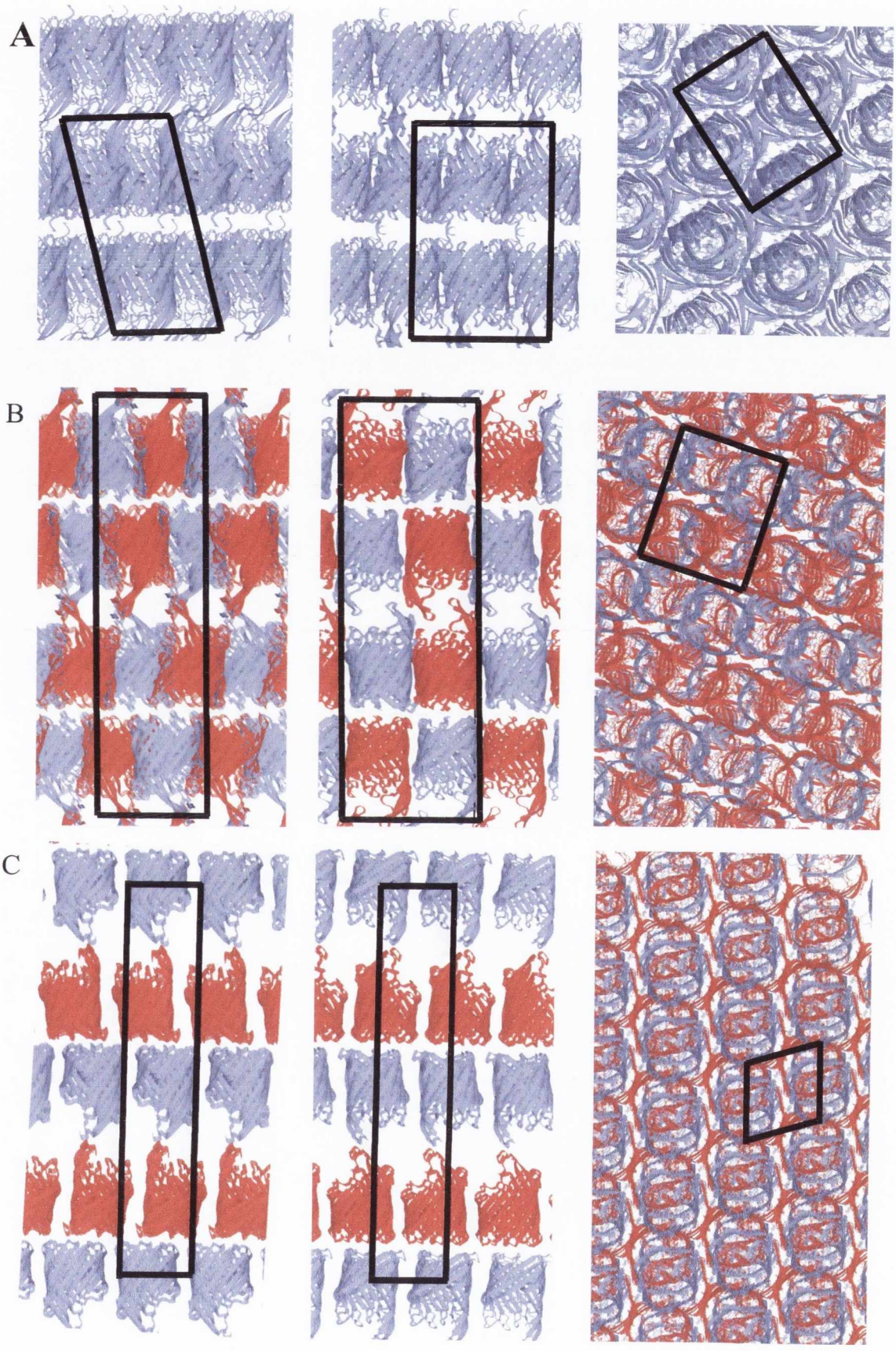
Figure 4.10 The hydrophobic thickness calculated by PPM server. The yellow molecule is the structure model for AlgE_1.9, which is the representative of available AlgE structure models and used for calculation because all models share nearly the same structure in transmembrane segment. The blue dots forming two planes showing where the hydrophobic boundaries are, as predicted by online PPM sever: http://opm.phar.umich.edu/server.php. The hydrophobic thickness is 24.8 Å as indicated in the figure. All Trp, Tyr and Phe residues are highlighted in ball and stick style. It is obvious that a lot of them are located on or very close to the interface between hydrophobic region and hydrophilic region.

Although very similar to each other, the *in meso* structures do differ both slightly among themselves and much from the *in surfo* structure, in terms of crystal packing.

First, the *in meso* structures all adopt typical Type I packing, namely layered packing (Ostermeier and Michel 1997), and the molecules in same layers all have the same orienta-

tions so the contact between layers is either head-to-head or tail-to-tail, as shown in panel A, B and C of Figure 4.11. The Type I packing can be viewed as one of the common characteristics of *in meso* grown crystals (Caffrey *et al.* 2012) as all of them adopt this packing. The observation could be related to the proposed mechanism of *in meso* crystallisation method that the protein molecules need to diffuse into local lamellar conduit, which is layered, and there they start the packing to form crystals. Slight differences in the crystal contact regions are found. These differences mainly come from the two extracellular loops, L5 and L6. L5 and L6 are stretching out completely in all AlgE models and form the crystal contact regions. Because of subtle differences in the conformations they adopt, the *in meso* models are packed in three different ways. An expanded view of the two loops is shown in Figure 4.12 but it is not clear whether these differences have any physiological meanings or are just responsible for crystal contact.

Second, when compared to the *in meso* models, AlgE_in_surfo shows quite different packing style as panel D in Figure 4.10, although it also adopts Type I packing. The molecules in same layers all have the alternating orientations so that the contact between layers is head-to-tail.


A
B
C

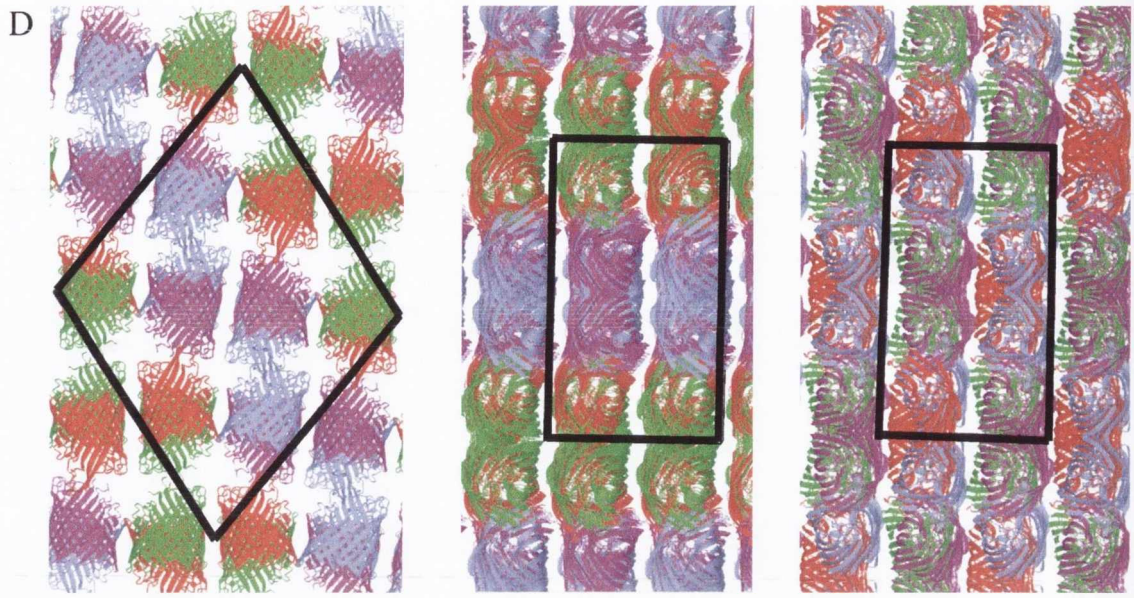
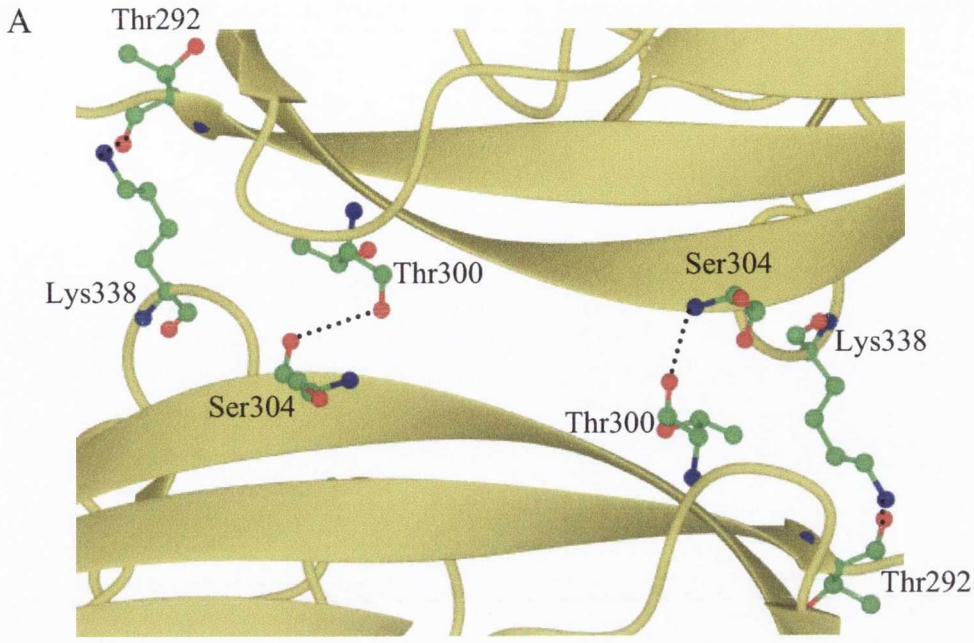
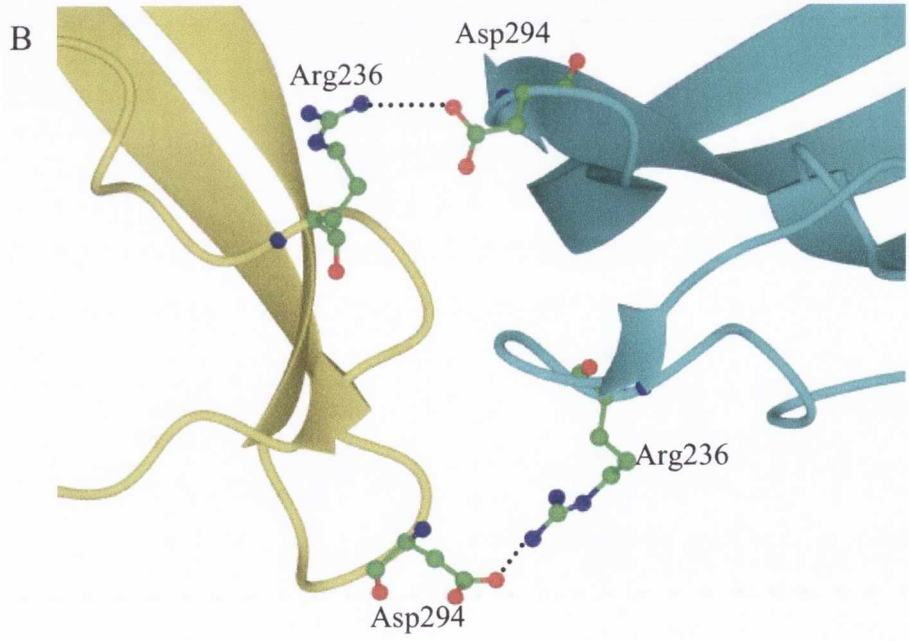


Figure 4.11 Crystal packing styles for all AlgE structure models. In panel A to C, the *in meso* structure models, AlgE_1.9 to AlgE_2.8, are shown. Protein molecules A and B in one asymmetric unit of AlgE_2.4 and AlgE_2.8 are coloured in darkblue and red. In panel D, AlgE_in_surfo structure model is shown with protein molecules A to D in one asymmetric unit coloured in darkblue, red, green and magenta. Views are taken alongside the unit cell edges for each of AlgE model.


A
Thr292
Lys338
Thr300
Ser304
Ser304
Thr300
Lys338
Thr292


B
Asp294
Arg236
Arg236
Asp294

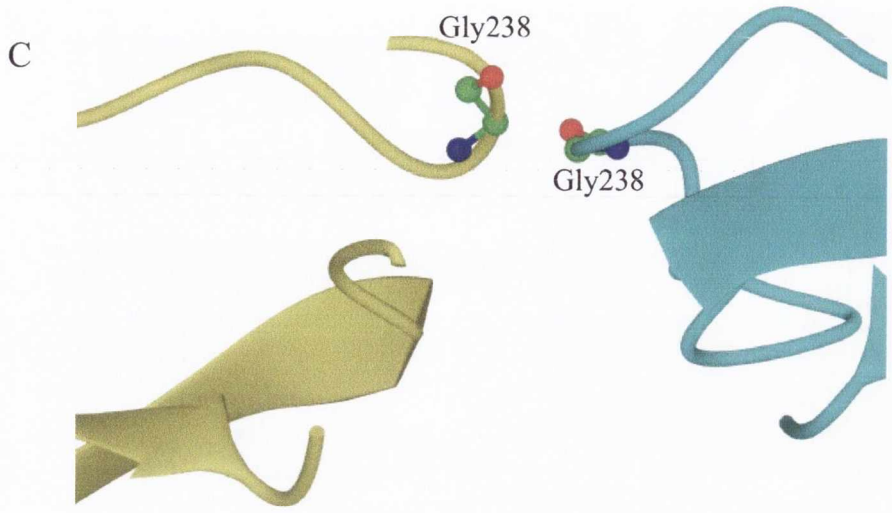


Figure 4.12 Zoomed view in the crystal contact regions on loops L5 and L6 of AlgE *in meso* structures. In panel A, AlgE_1.9 is shown. Hydrogen bonds are in the crystal contact region, between Ser304 (L6) and Thr300 (L6) and between Thr292 (L6) and Lys338 (L7). In panel B, AlgE_2.4 is shown. Hydrogen bond is in the crystal contact region, between Arg236 (L5) and Asp294 (L6). In panel C, AlgE_2.8 is shown. Two Gly238 (L5) from two different molecules show the closest contact of this crystal form but there are no obvious hydrogen bonds in the region of L5 and L6.

Besides the differences in crystal contact regions, various organic and inorganic molecules are modelled into each *in meso* structure model and they show significant difference, as in Table 4.3. Some of these differences may be important and related to the protein's biological properties. They will be discussed in full detail in the following sections.

4.4.3 Citrate binding site

Among all the different organic and inorganic molecules that are modelled into these AlgE structure models, one of the most important one is found in AlgE_1.9. In this structure model, positive electron density is found in the pore region. Based on the crystallisation condition (0.1 M Tris-HCl, pH 7.5, 0.1 M sodium citrate, 18 %(v/v) PEG400) and the shape of the positive electron density, a citrate is proposed by Dr. Joseph Lyons and Dr. Valerie Pye. The molecule is then modelled into the position and validated by electron

density, B-factor and previous experience, as will be described below. The citrate binding site is shown in Figure 4.13 and Figure 4.14.

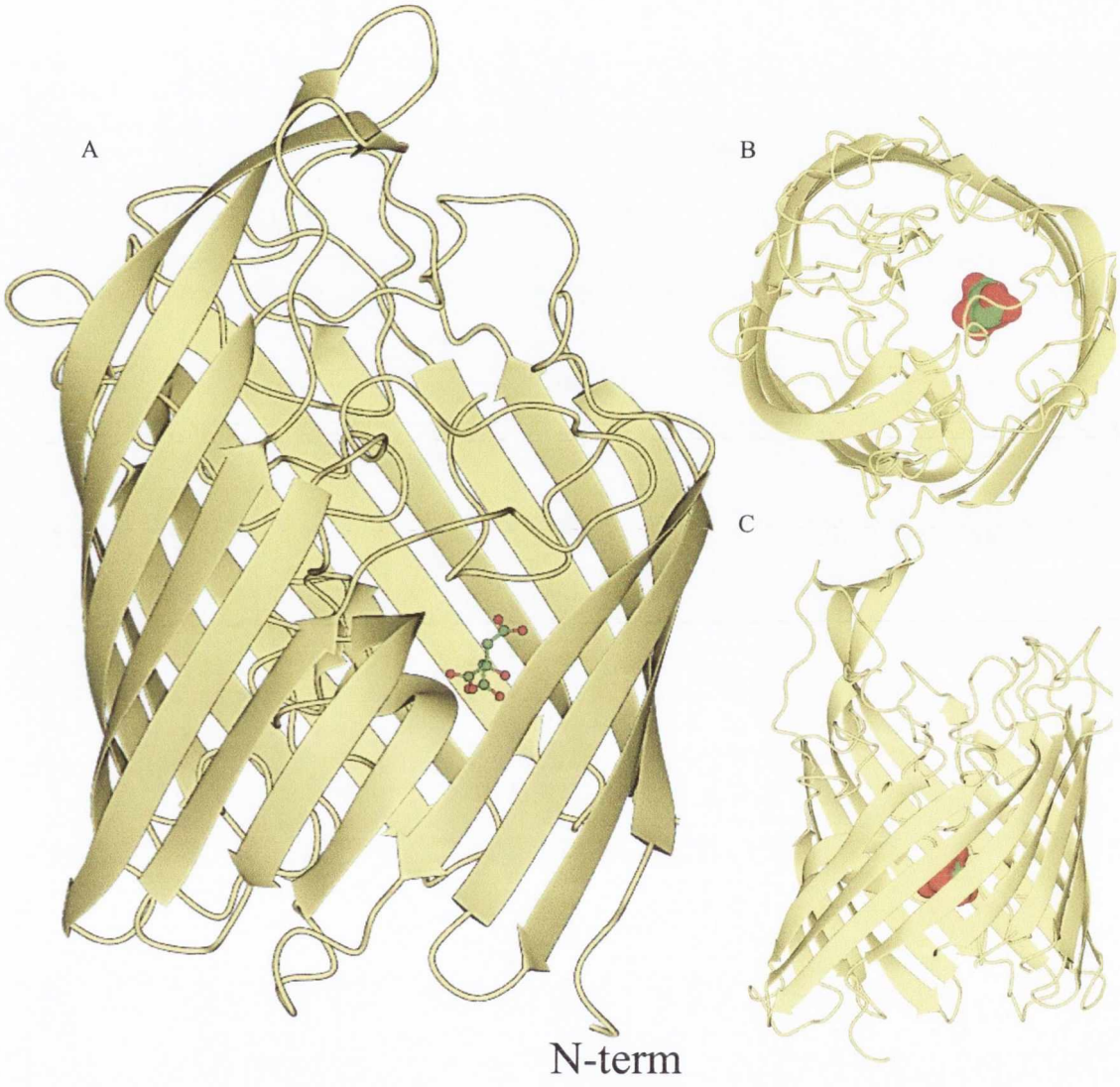


Figure 4.13 Position of citrate molecule found in AlgE_1.9. The protein molecule is drawn in ribbon style and coloured in yellow. The citrate is drawn in ball-and-stick style with colours by atom types: oxygen is red and carbon is green. View of AlgE from membrane is shown in A and the view from extracellular side is shown in B. In panel C, another view from membrane is shown by rotating the protein molecule in panel A 90° horizontally. The citrate molecules in B and C are drawn in space-filling style for clarity.

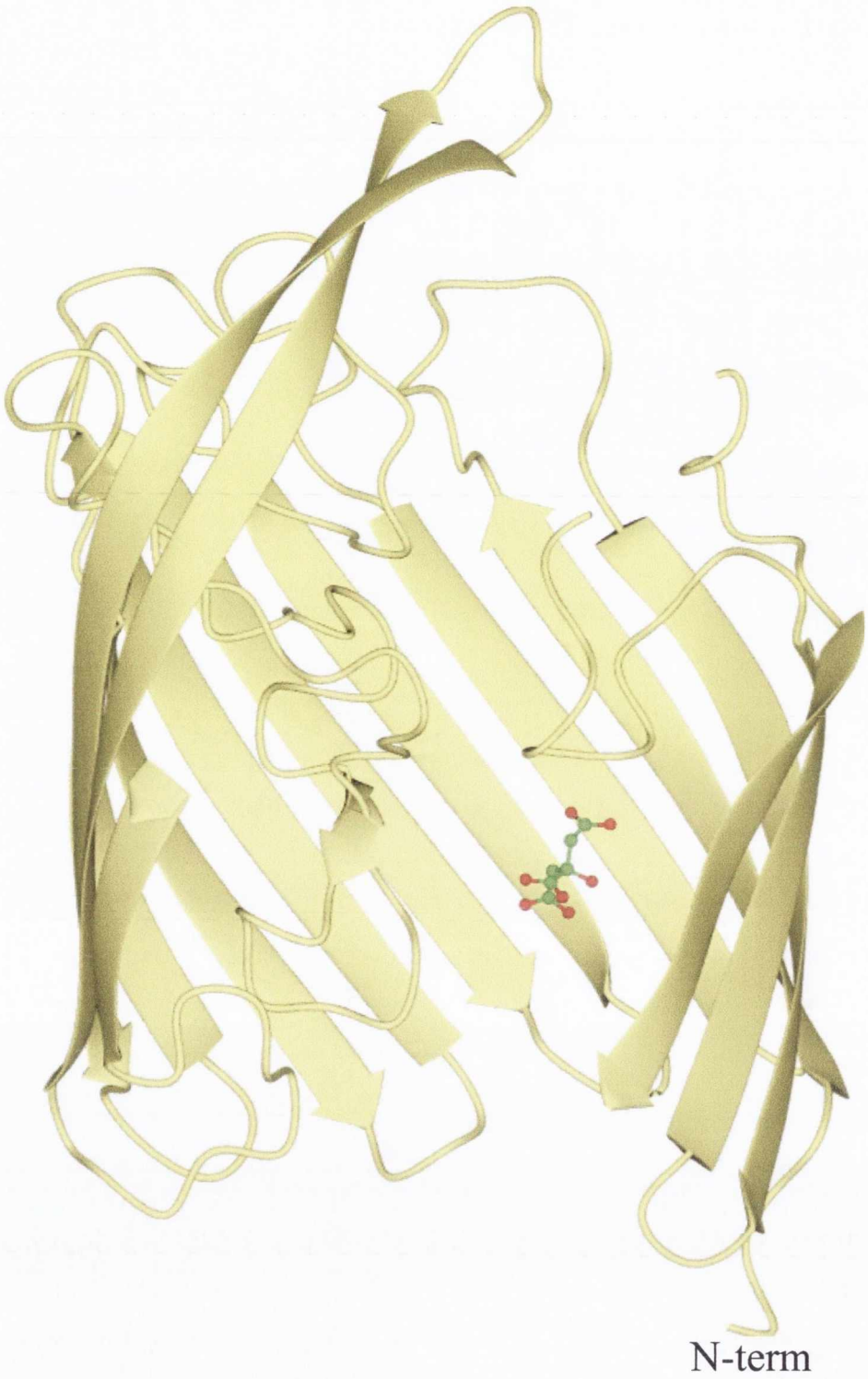
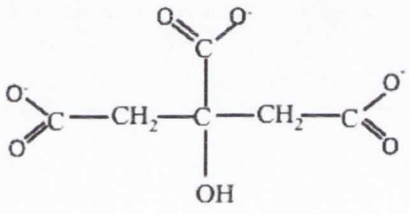


Figure 4.14 Slab view of AlgE_1.9 reveals the position of citrate molecule. The protein molecule is in the same orientation as Figure 4.13 and drawn in ribbon style and coloured in yellow. The citrate is drawn in ball-and-stick style and coloured in atom types: oxygen is red and carbon is green.

The completely refined electron density map for citrate site is shown in Figure 4.15 in three different angles. The electron density map is considered to fit the citrate model because all atoms of the modelled citrate are inside electron density when the density map is contoured at 1σ. The B-factor for citrate site is also reasonable: the overall structure is 23.5 and the average B-factor for citrate molecule is 37.8. Citrate molecule coordinates with Thr103, Arg74 and Arg152, whose average B-factors are 18.7, 18.0 and 18.5, respectively. The B-factor of citrate is higher than that of the protein molecule but not to a great extent.

Compared to AlgE_1.9, no positive density is found in this region in all the other structure models and nothing could be modelled. The *in surfo* structure 3RBH, in this region, shows very little difference from AlgE_2.4 and AlgE_2.8 and so is used as a "negative" representative for the analysis of citrate binding site, as shown in Figure 4.16.

A.



B.

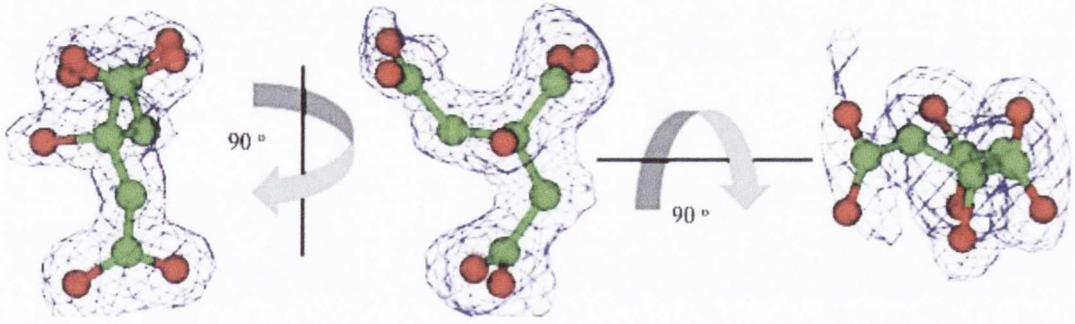


Figure 4.15 The citrate molecule modelled into AlgE_1.9. In panel A, chemical structure of citrate is shown. In panel B, The citrate is shown with the final electron density. Citrate is drawn in ball-and-stick style and coloured in atom types: oxygen is red and carbon is green. The blue mesh around citrate molecule represents $2F_o$-F_c electron density map and it is contoured at 1σ. Three views are taken from three angles by rotating the model 90° vertically or horizontally. It is clear that all atoms from the citrate are inside the density and the shape of density fits the shape of citrate very well.

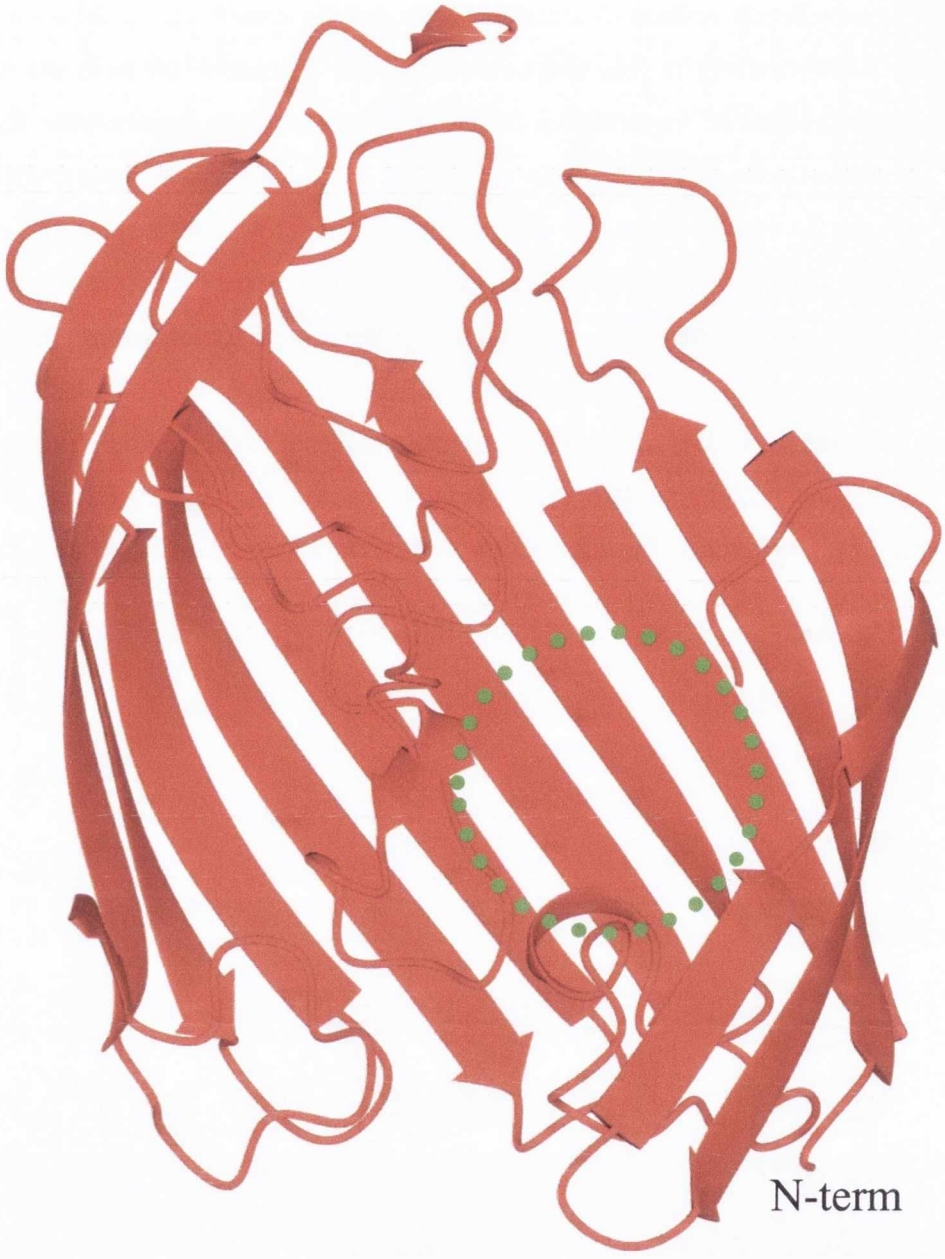


Figure 4.16. Slab view of AlgE_in_surfo from membrane shows no citrate molecule inside the barrel region. The protein molecule is drawn in ribbon style and coloured in red. The dashed green circle shows where citrate should be when AlgE_in_surfo superposed with AlgE_1.9.

In the structure model AlgE_1.9, the citrate is coordinating with the following residues and waters (as W) by H-bonding within distance of 3.2 Å as shown in Table 4.5.

Table 4.5 Contacts present between citrate and protein residues or water molecules[a].

Atom from protein or water	Atom from citrate	Distance (Å)	Bond type
Arg74: NH1	OA1	2.76	Hydrogen bond
Arg74: NH2	OA1	2.88	Hydrogen bond
Thr103: N	OA2	2.73	Hydrogen bond
Thr103: OG1	OA2	3.00	Hydrogen bond
Arg152: NH1	OB2	2.81	Hydrogen bond
Arg152: NH2	OB1	2.96	Hydrogen bond
W98: O	OB1	3.10	Hydrogen bond
W152: O	OHB1	2.75	Hydrogen bond

a. All bond distances and types are calculated by online program PISA: http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html.

The above H-bondings agree well with some of the citrate binding sites discovered in other proteins, where one or multiple Arg residues are always involved: a) In the human ATP-citrate lyase (PDB ID: 3MWD), the carboxyl groups of citrate interact with sidechain nitrogen of Arg379, backbone nitrogen and the sidechain oxygen of Thr348 and backbone nitrogen of Gly309 (Sun *et al.* 2010); b) In the citrate synthase of *Pyrococcus furiosus* (PDB ID: 1AJ8), the carboxyl groups of citrate interact with sidechain nitrogen of Arg271, Arg312, Arg337, Arg356, His223 and His262 (Russell *et al.* 1997); c) In the ferric citrate transporter of *E. coli* (PDB ID: 1PO3), the carboxyl groups and hydroxyl group of citrate interact with sidechain nitrogen of Arg155, Arg365, Arg380, Arg418 and Arg438 (Yue *et al.* 2003). d) In the protruding domain of norovirus GII.10 Vietnam026 (PDB ID: 3ONY), the carboxyl groups and hydroxyl group of citrate interact with sidechain nitrogen of Arg356, Asn355 and the backbone nitrogen of Gly451 (Hansman *et al.* 2012).

By looking at the residues interacting with citrate in structure model AlgE_1.9, the highly conserved residues Arg74 and Arg152 (Table 4.5) catch my attention immediately because they are among the residues that are proposed to be forming alginate secretion portal (Whitney *et al.* 2011). A summary of previous work shows how these residues are discov-

ered by Professor Lynne Howell's group (Whitney *et al.* 2011): the previous study is done based on AlgE_in_surfo structure model, in which the occluding loops, L2 and T8, are removed completely and then HOLLOW program reveals a possible secretion pathway. After examining the pore, ten highly conserved charged residues are found in this region in their study: Lys47, Arg74, Arg129, Arg152, Asp162, Asn164, Arg353, Arg362, Arg459 and Asp485. These residues are considered to form the secretion pathway because earlier study on alginate lyase from *Sphingomonas sp.* shows that a number of charged residues are involved to form the substrate binding centre for the enzyme (PDB ID: 3EVL and publication not available). A detailed view of this hypothetical pore is shown in Figure 4.18, where the ten conserved residues are surrounding the pore with a diameter of around 8.0 Å (Whitney *et al.* 2011).

Because Arg74 and Arg152 are inside the proposed secretion portal, I decide to look at all the objects that are close to citrate molecule within 4.0 Å in the hope to find more residues that are proposed to form alginate secretion portal. The following residues and waters (as W) are then identified: Lys47, Arg74, Arg152, Arg362, Thr101, Asp102, Thr103, W80, W98, W152 and W204. Among the residues, Lys47, Arg74, Arg152 and Arg362 are found in the proposed secretion portal, with Thr101 and Asp102 highly conserved and Thr103 less conserved, as shown in Table 4.6. Lys47 and Arg362 also form very weak hydrogen bonding with the citrate molecule as calculated by PISA program (distance as 3.26 Å and 3.85 Å, respectively). This finding shows that citrate molecule is definitely located right inside this proposed secretion pathway as shown in Figure 4.19.

Thus I propose that, in AlgE_1.9 structure model, citrate resembles the native substrate of AlgE, alginate, providing direct experimental evidence to the existence of the hypothetical secretion pathway proposed by previous modelling work. Although the chemical structures of citrate and alginate are very different, they share some basic common features: a). one citrate molecule has three carboxyl groups and mannuronate or guluronate molecule has one carboxyl group; b). citrate has one hydroxyl group and mannuronate or guluronate has four; c) these carboxyl and hydroxyl groups can form H-bonding with surrounding residues or waters in similar ways; d) The size of a citrate molecule is close to that of a mannuronate molecule, a monomeric unit of alginate, as shown in Figure 4.20. Connoly solvent excluded volume of citrate is calculated by Chem3D Ultra program to be 135.78 $Å^3$ and that of mannuronate is 139.55 $Å^3$, which proves that the sizes of both molecules are quite

similar, although the shape of them are different. In addition, it is not unusual that citrate is considered to be mimicry of some sugar molecules: docking studies show that citrate molecule can enter the fucose binding sites of *Anguilla anguilla* agglutinin, *Aleuria aurantia* lectin, *Streptococcus pneumoniae* virulence factor SpGH98 and *P. aeruginosa* PA-IIL lectin, the sialyllactose binding site of parainfluenza virus 5 hemagglutininneuraminidase and also the lactose binding site of porcine adenovirus type 4 galectin domain (Hansman *et al.* 2012). The modelled sugar molecules overlap with citrate molecule found in the structure to some extent

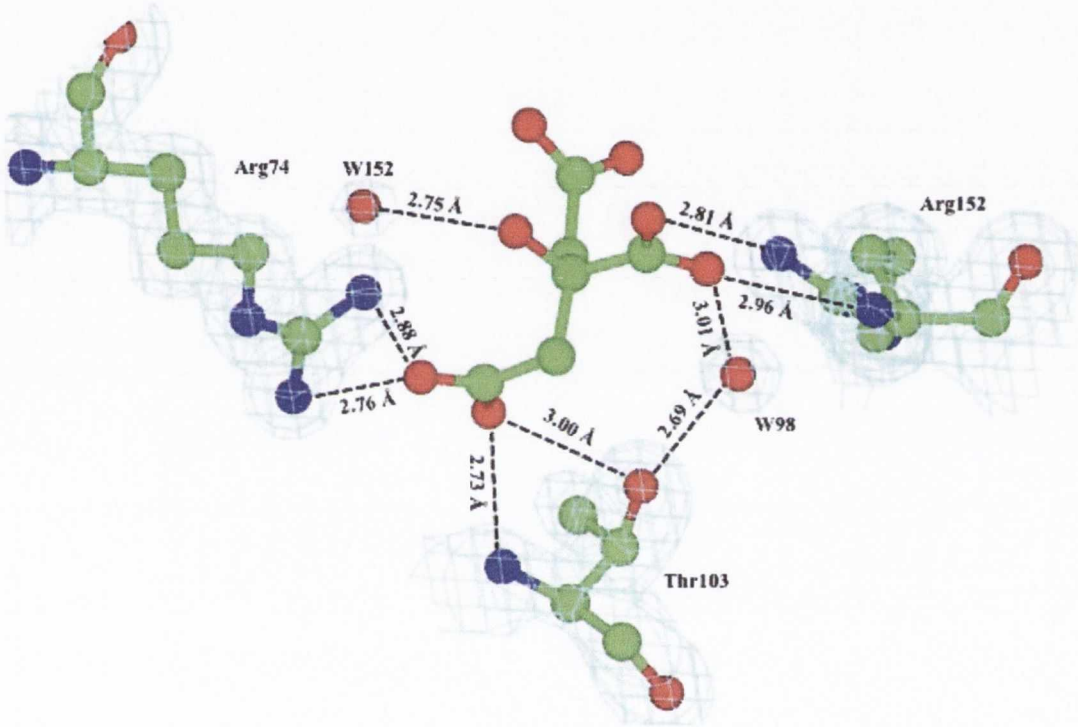


Figure 4.17 The citrate binding site in AlgE_1.9 forming H-bonding with surrounding residues and waters. All objects are drawn in ball-and-stick style and coloured in atom types: nitrogen is blue, oxygen is red and carbon is green. Hydrogen bondings between atoms are labelled as dotted line with the distance listed. The blue mesh around residues and waters represents electron density map and it is contoured at 1σ.

Table 4.6 Alignment of the residues shown in Figure 4.17 that have close contact with citrate molecule.

Species name	Maximum Identity to *P. aeruginosa* (%)	Residue					
		47	74	101	102	103	152
Pseudomonas. aeruginosa	100	K	R	T	D	T	R
Pseudomonas. putida	68	K	R	T	D	T	R
Pseudomonas sp.GM84	67	K	R	T	D	T	R
Pseudomonas entomophila	66	K	R	T	D	T	R
Pseudomonas mendocina	66	K	R	T	D	T	R
Pseudomonas fluoresescens	60	K	R	T	D	P[a]	R
Pseudomonas chlororaphis	59	K	R	T	D	T	R
Pseudomonas synxantha	56	K	R	T	D	T	R
Pseudomonas protegens	56	K	R	T	D	T	R
Pseudomonas brassicacearum	56	K	R	T	D	T	R
Pseudomonas extremaustralis	56	K	R	T	D	T	R
Pseudomonas savastanoi	56	K	R	T	D	T	R
Pseudomonas syringae	51	K	R	T	D	P[a]	R
Acotobacter vinelandii	34	K	R	T	D	P[a]	R

a. In these bacteria, Thr103 is replaced by Pro103.

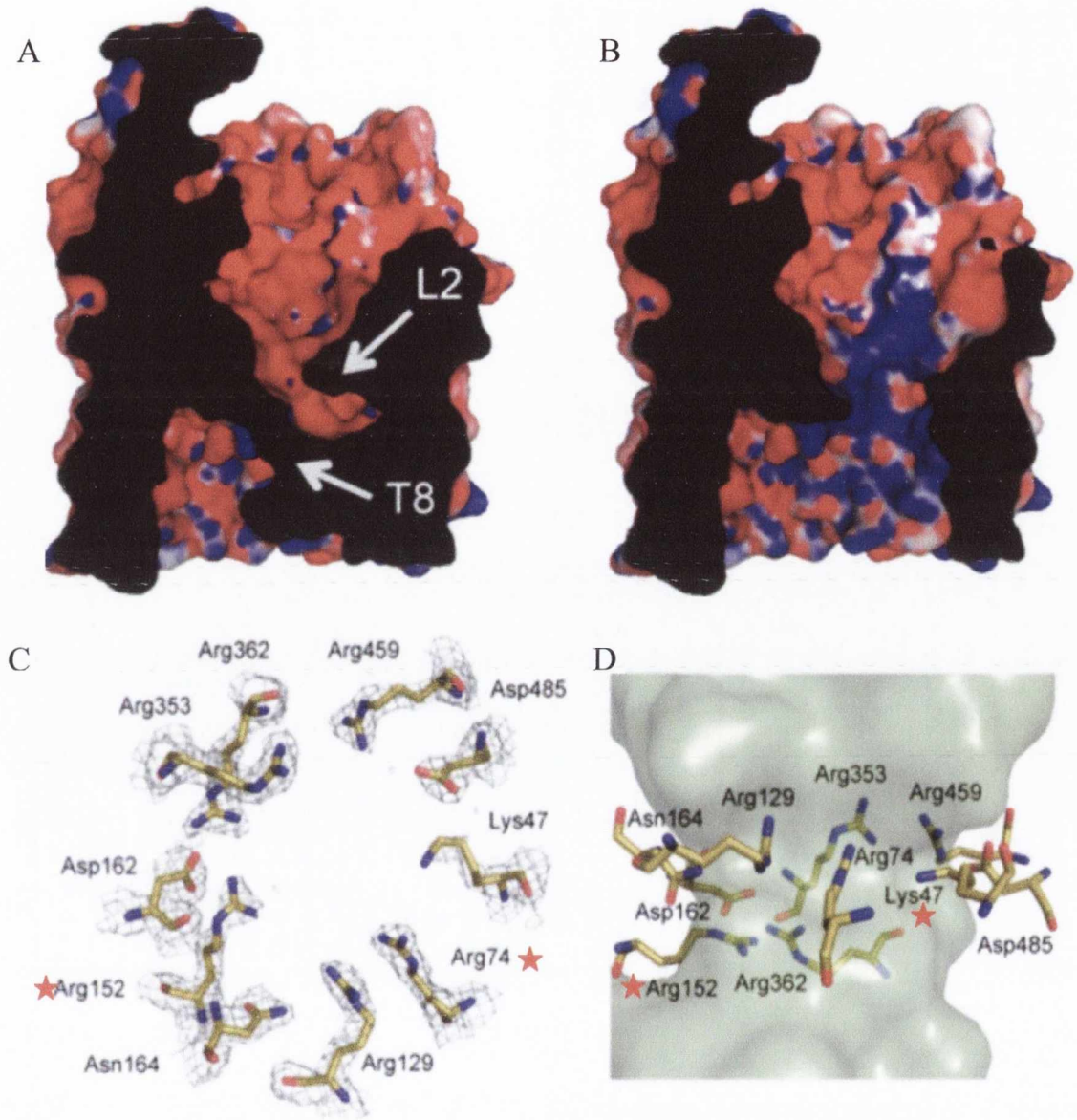


Figure 4.18 The hypothetical secretion pathway of AlgE. Panel A is the slab view of AlgE_in_surfo model, with electrostatic surface representation. Positive charges are coloured in blue and negative in red. In panel B, the occluding loops L2 and T8 are removed to reveal the proposed secretion pore. In panel C, stick representation of the pore-forming residues is shown as viewed from the extracellular space. The 2*Fo* - *Fc* electron density map is contoured at 1.2σ and displayed as a grey mesh. Carbon, nitrogen, and oxygen atoms are coloured yellow, blue, and red, respectively. In panel D, side view of the pore restriction shows a surface representation of interior of the pore generated by HOLLOW. The stars highlight residues that have close contact with citrate in AlgE_1.9. Copied from (Whitney *et al.* 2011).

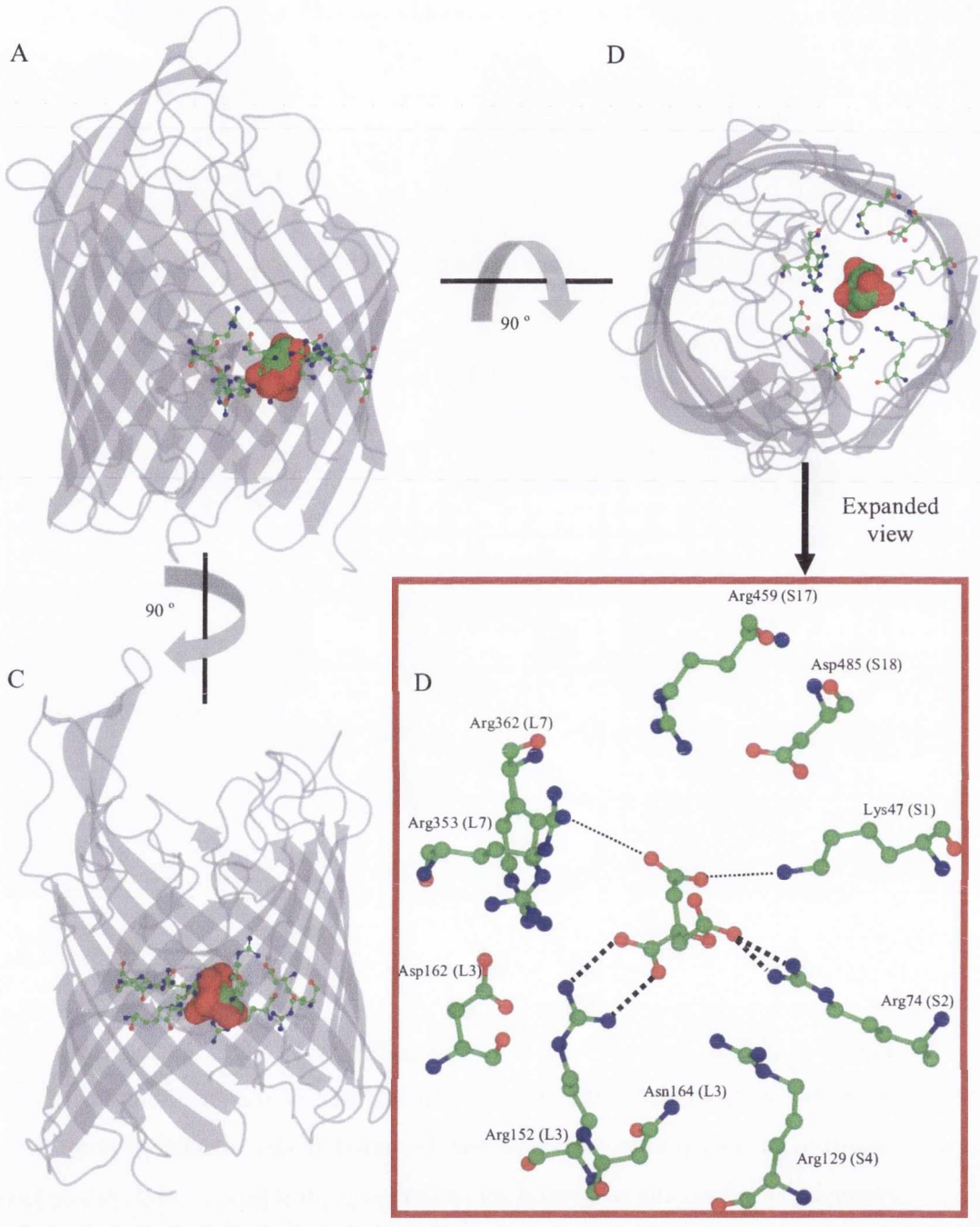


Figure 4.19 The citrate molecule in secretion portal in AlgE_1.9. AlgE molecule is drawn in ribbon style and coloured in grey. Citrate molecules in A, B and C are drawn in surface-filling style and in D is drawn in ball and stick style. The residues around portal are drawn in ball-and-stick style and coloured by atom type. A: view from membrane. B: view from extracellular side. C: another view from membrane. D: expanded view of details in the pore region with hydrogen bonds shown as dashed lines. The thicker the dashed line is, the closer the atoms are to each other.

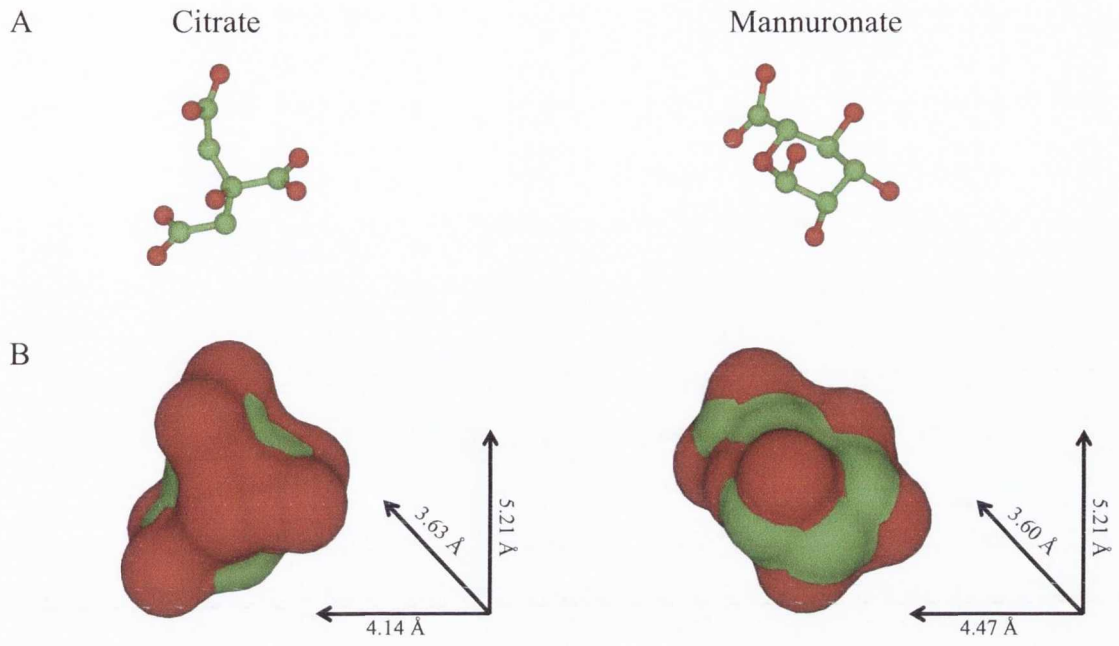


Figure 4.20 A comparison between the size of citrate and size of mannuronate. In panel A, both citrate and mannuronate molecules are drawn in ball-and-stick style and coloured in atom type. The citrate molecule in this case adopts the conformation as in AlgE1.9 structure model. In panel B, both molecules are drawn in surface-representation style. Dimensions in three directions are calculated using CCP4mg tool and shown on the axis in panel B. Guluronate is not shown as it is different from mannuronate only in the position of hydroxyl group on C5.

After identifying the proposed citrate molecule inside hypothetical secretion pathway, an immediate attention is paid to T8 and L2, which are removed on computer in the *in surfo* structure to reveal the pore for secretion. T8 is completely disordered in AlgE_1.9 and can't be modelled, however, it is complete and folded inside pore region of AlgE_in_surfo. 17 residues are disordered and can't be modelled on L2 in AlgE_in_surfo but only 8 residues are missing for the same reason in AlgE_1.9. When AlgE_1.9 and AlgE_in_surfo are superposed together, it is clear that T8 and L2 from AlgE_in_surfo will clash with the citrate from AlgE_1.9, as shown in Figure 4.21. The distance between citrate and Asp452 from T8 or Asp102 from L2 is as short as 0.80 Å or 0.60 Å as calculated by COOT. This means if AlgE adopts the conformation as in AlgE_in_surfo, there is no way for citrate to enter the hypothetical secretion pathway from the periplasmic side because the space for binding is occupied by L2 and T8, as shown in Figure 4.22. So it will be the same for algi-

nate as the volume of citrate is around the same as that of the monomeric unit of alginate as shown in Figure 4.20. It can now be confirmed again that AlgE_in_surfo structure model is in a closed state because of the positions of L2 and T8 as said in previous study (Whitney *et al.* 2011). In contrast, in AlgE_1.9, because T8 is completely disordered and no residues from L2 clash with citrate, as shown in Figure 4.23, it could mean that AlgE_1.9 is in an open state. However, when examining the citrate and L2 together with the exceptionally long loop L7, which also folds into the barrel region, it is clear that there is no space for the citrate to diffuse through to the extracellular side. The space on the extracellular side is completely blocked by both L2 and L7, as shown in Figure 4.23.

Now, it is reasonable to conclude that AlgE_1.9 is an inward open state, which can receive substrate from periplasmic side but doesn't allow the substrate to pass through on the extracellular side. Furthermore, to form an outward open state or completely open state, it is possible that the conformation of L7 or L2 will be important because they together block the pathway on extracellular side. Maybe L2 or L7 needs a conformational change that exposes the ligand to the extracellular side to fulfil the secretion function. But from both *in meso* and *in surfo* structures, little knowledge about L2 can be obtained, because none of the four AlgE structure models have a complete L2. As said above, even the best model, AlgE_1.9, has 8 residues missing on L2. Instead, analysis on L7 will be discussed in more details in the next section.

In addition, in previous study, it was suggested that AlgK interacts with T8 from AlgE and this interaction is important for AlgE to form an open state for alginate transportation (Whitney *et al.* 2011). However, based on the information provided by AlgE_1.9, it is possible that at least the inward open state can be obtained without the existence of AlgK. Thus, the function of T8 and T8 related protein, AlgK, needs to be readdressed.

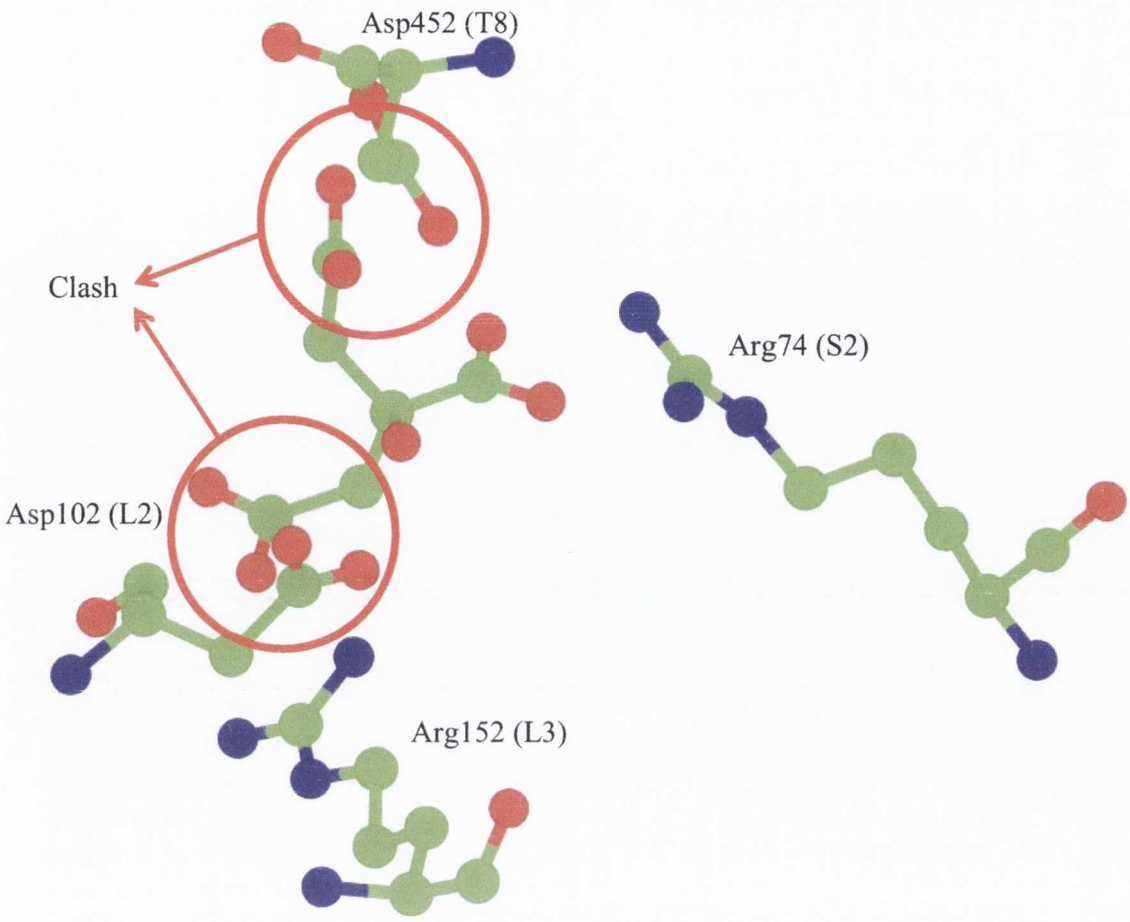


Figure 4.21 Clashing sites around the citrate site when AlgE_1.9 and AlgE_in_surfo are superposed. All residues and citrate are drawn in ball-and-stick style and coloured in atom type. Citrate molecule is from AlgE_1.9 and the residues are from AlgE_in_surfo. Clashing sites are labelled in red circles.

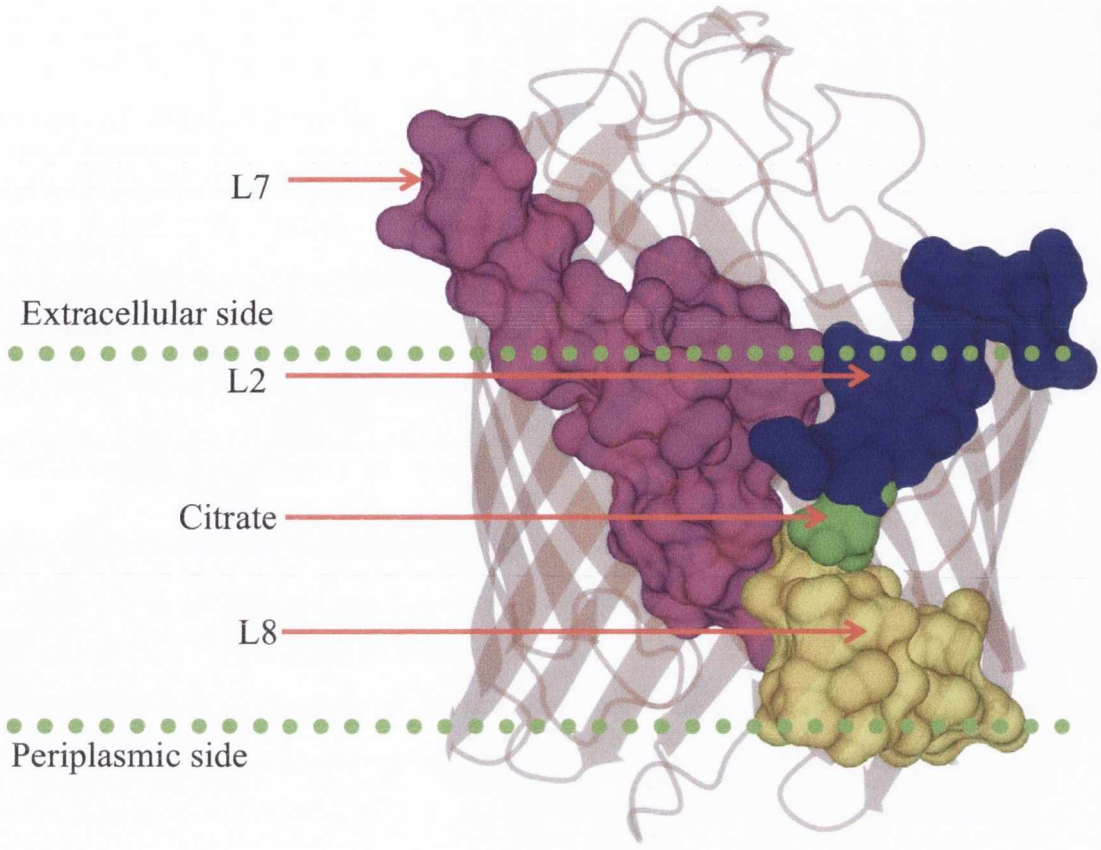


Figure 4.22 Space-filling model that shows AlgE_in_surfo is possibly in a closed state. The model is created by superposing AlgE_in_surfo with AlgE_1.9. The loops shown here are all from AlgE_in_surfo and the citrate molecule is from AlgE_1.9. AlgE protein molecule is drawn in ribbon style and coloured in red in transparency. L2 is coloured in blue, L7 in magenta, T8 in yellow and citrate in green, as indicated by arrows. Green dotted lines represent the interface between hydrophobic and hydrophilic region. It is clear that L2, L7 and T8 block the barrel region from both extracellular and periplasmic sides of structure model AlgE_in_surfo and in the meantime, L2 and T8 clash with the citrate molecule that is from AlgE_1.9. This means that citrate molecule won't be able to diffuse in or out of the secretion pathway when AlgE adopts the conformation of AlgE_in_surfo. As discussed earlier, this leads to the conclusion that mannuronate or guluronate won't be able to diffuse in or out either, as they are considered to be similar to citrate molecule and that may share a same binding site. So this analysis may suggest that AlgE_in_surfo is a closed state.

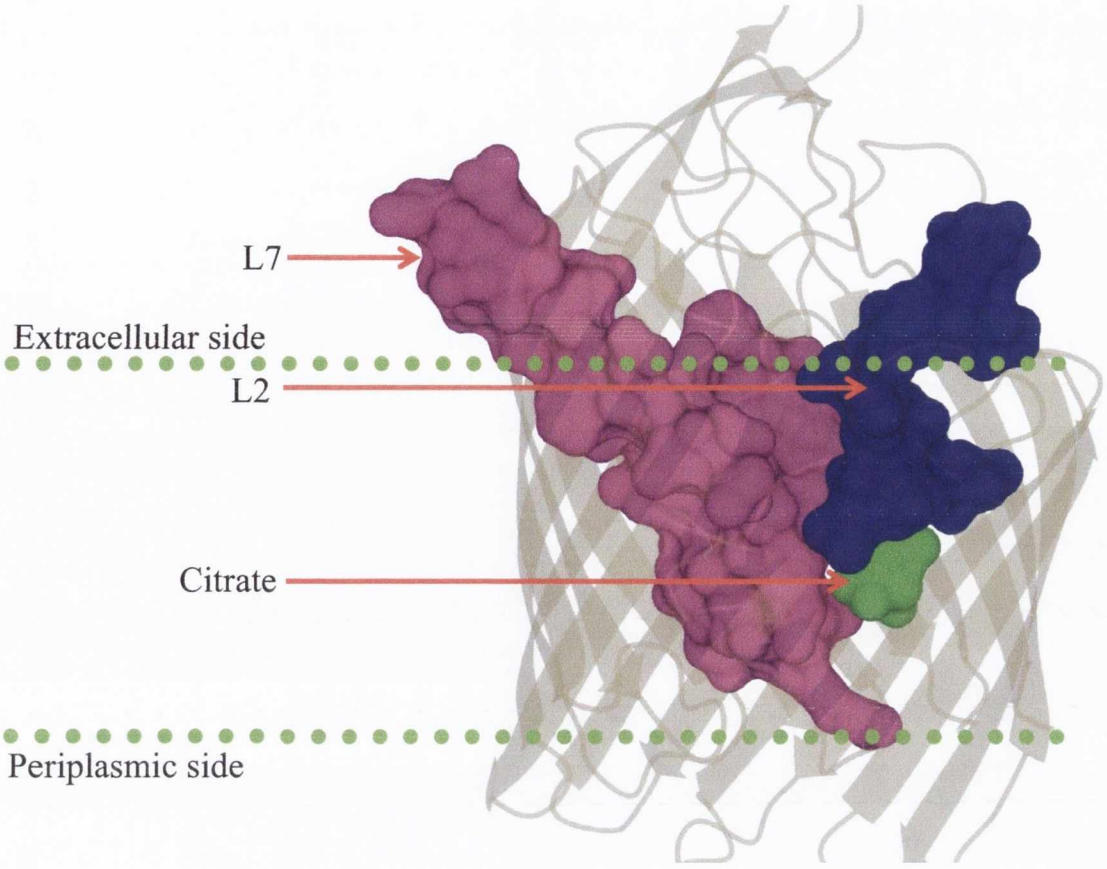


Figure 4.23 Space-filling model that shows AlgE_1.9 is possibly in an inward open state. All loops shown here and the citrate molecule are from AlgE_1.9. AlgE protein molecule is drawn in ribbon style and coloured in yellow in transparency. L2 is coloured in blue, L7 in cyan and citrate in green, as indicated by arrows. Blue dotted lines represent the interface between hydrophobic and hydrophilic region. As a comparison to Figure 4.22, citrate molecule can diffuse into the binding site because T8 is completely disordered when protein adopts the conformation of AlgE_1.9. Although L7 still blocks the citrate from going out to extracellular side, it is possible that the AlgE_1.9 represents an inward conformation of AlgE protein that can receive substrate from periplasmic side.

4.4.4 L7 and unidentified molecule

Among all loops in AlgE structure, L7 is the longest one and it folds into barrel cavity. It has been noted that there are five arginine residues on this L7, as shown in Table 4.7. Three of them are highly conserved and two are relatively conserved. In addition, Arg353

and Arg362 are among the proposed alginate secretion portal. It is also proven that L7 is very important for the function of AlgE as either deletion mutation or insertion mutation on L7 will completely fail the secretion of alginate in *P. aeruginosa* cells (Hay *et al.* 2010). In the previous *in surfo* structure study, it has been proposed that L7, together with L3, forms a constriction that is required for substrate specificity (Whitney *et al.* 2011).

Table 4.7 Alignment of arginine residues found on loop L7 of AlgE protein.

Species name	Maximum identity to *P. aeruginosa* (%)	Residue				
		353	360	362	365	370
Pseudomonas. aeruginosa	100	R	R	R	R	R
Pseudomonas. putida	68	R	R	R	R	R
Pseudomonas sp.GM84	67	R	R	R	R	R
Pseudomonas entomophila	66	R	R	R	R	R
Pseudomonas mendocina	66	R	R	R	R	R
Pseudomonas fluoresescens	60	R	E	R	R	R
Pseudomonas chlororaphis	59	R	R	R	R	R
Pseudomonas synxantha	56	R	Q	R	R	R
Pseudomonas protegens	56	R	R	R	R	R
Pseudomonas brassicacearum	56	R	R	R	R	R
Pseudomonas extremaustralis	56	R	Q	R	R	R
Pseudomonas savastanoi	56	R	R	R	R	Q
Pseudomonas syringae	51	R	R	R	R	Q
Acotobacter vinelandii	34	R	Q	R	R	R

Because of the importance of L7, the molecules that have close contact with L7 are then carefully examined. One interesting observation is that, when AlgE_in_surfo and AlgE_1.9 are superposed altogether, where a PEG fragment is modelled in AlgE_1.9 overlaps exact-

ly with where two 1,2-ethanediol (EDO) molecules are modelled in AlgE_in_surfo, as shown in Figure 4.24. PEG fragment is modelled into AlgE_1.9 because PEG400 is used in the crystallisation condition and the EDO molecules are modelled into 3RBH because EDO is used as cryo-protectant. However, when the same regions in AlgE_2.4 and AlgE_2.8 are examined, continuous unbranched positive electron densities are found, which are very close to the shape of PEG fragment or EDO molecules, as shown in Figure 4.25. It seems clear that this electron density comes from a common object. Because the purification methods and crystallisation conditions differ quite much (chemicals are shown in the Table 4.8), the common molecule could only one of the followings: Tris or sodium chloride. However, neither of these chemicals has "sausage" shape as suggested by the electron density. Thus, I propose that this molecule can only be picked up by protein molecule from *E. coli* cells during expression process.

Although there is little clue as to what this molecule is, it may be important. As discussed in Section 4.4.3, L2 and L7 together block the secretion pathway from extracellular side. For the protein to become outward open or completely open, either L2 or L7 or both will have to move away from current positions. It can't be ruled out the possibility that this unidentified molecule alters the conformation of L7 or limits the movement of L7 in all AlgE structure models, and in turn, L7 blocks the alginate pathway from both sides. The removal or alteration of this molecule may then reveal the "functional" confirmation of L7. To test this idea, a native AlgE prep from *P. aeruginosa* is needed and the purified protein then used for structure study. In that case, all molecules will come from *P. aeruginosa* cells except necessary components in the media for cell culture. Hopefully by doing this, the unidentified molecule will be excluded from the protein prep.

Table 4.8 Chemicals used for the determination of all four AlgE structure models.

	Chemicals used for purification	Chemicals used for crystallisation
AlgE_ in_surfo	Tris-HCl, NaCl, Urea, imidazole, DM, C_8E_4	PEG3350, ammonium formate, EDO
AlgE_1.9	Tris-HCl, KCl, NaCl, NH_4Cl, β-OG, LDAO	PEG400, sodium citrate, Tris-HCl
AlgE_2.4	Tris-HCl, KCl, NaCl, NH_4Cl, β-OG, LDAO	Sodium acetate, sodium cacodylate
AlgE_2.8	Tris-HCl, KCl, NaCl, NH_4Cl, β-OG, LDAO	Sodium acetate, sodium cacodylate

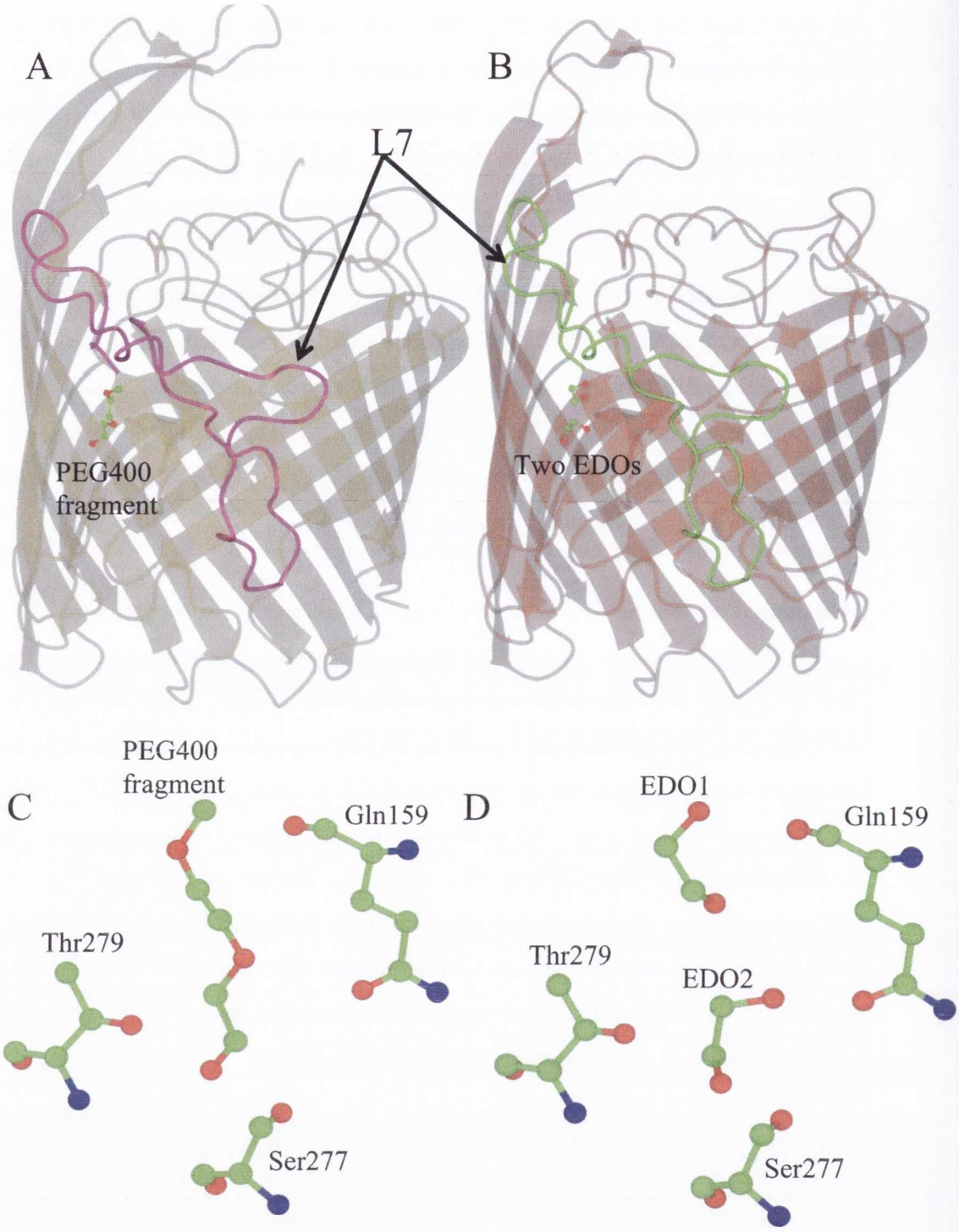


Figure 4.24 Cartoon representation of L7 and unidentified molecule. AlgE_1.9 and AlgE_in_surfo is shown in panel A and B respectively. L7 is coloured in magenta in panel A and green in panel B. PEG400 fragment and EDO molecules are shown as well. Detailed views of the residues around either PEG400 fragment or EDOs are shown in panel C and D.

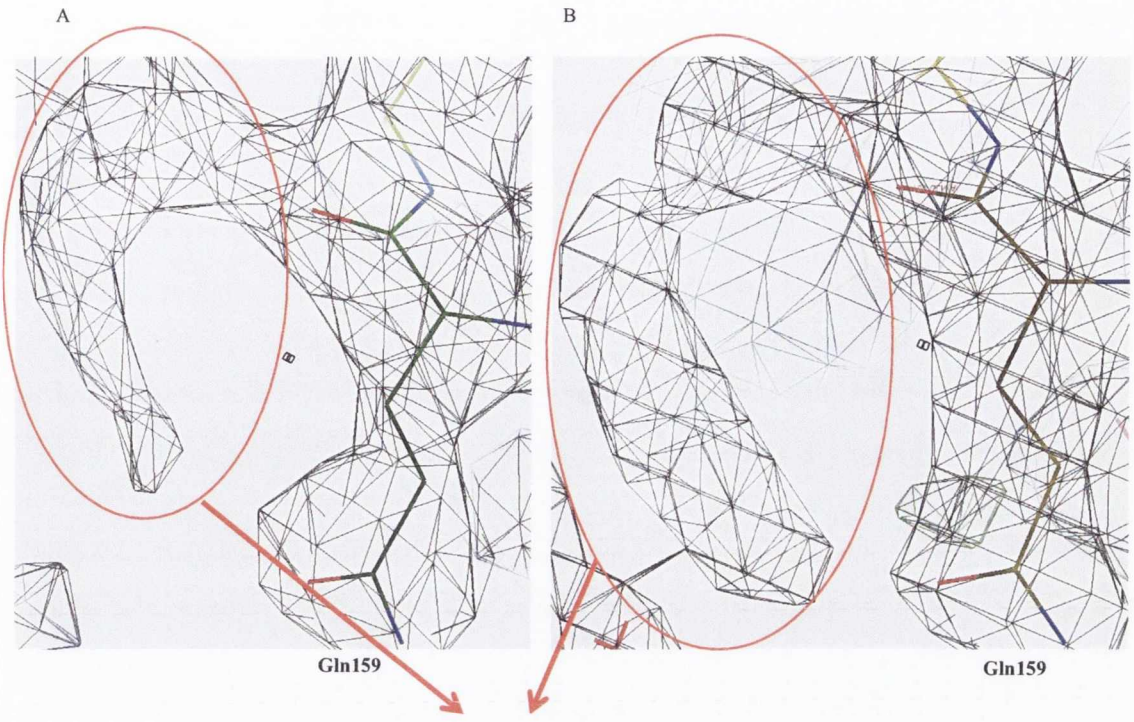


Figure 4.25 The electron density of the unidentified molecule. Panel A and B represent AlgE_2.4 and AlgE_2.8, respectively. Gln159 from L7, beside the electron density map, is labelled and shown on the right-hand side in each panel. The dark mesh represents electron density map and it is contoured at 1σ. It is clear that the unassigned electron density beside Gln159 has similar linear shape as the PEG400 fragment in AlgE_1.9 and EDO molecules in AlgE_in_surfo (Figure 4.24). The residues Gln159 are in the same orientation as the Gn159 residues shown in Figure 4.24.

4.4.5 Lipid binding sites

Not surprisingly, as result of the method, a number of lipid molecules (7.8 MAG) and detergent molecules (LDAO) are found and modelled onto the hydrophobic surface of AlgE molecules in all *in meso* structure models. As stated in Table 4.3 and shown in Figure 4.6 to Figure 4.8, there are 6 7.8 MAG molecules in AlgE_1.9, 5 in AlgE_2.4 and 5 in AlgE_2.8. These lipid molecules are reasonably well ordered and their head groups can be modelled and are coordinating with either residues from protein molecules or water molecules. In all cases, the hydrophobic region of lipids is residing on the hydrophobic surface of protein molecules while the head groups are either sticking out of or right at the boundary between hydrophobic and hydrophilic regions, which is calculated by PPM server as described in Figure 4.10. The headgroups are stabilised by hydrogen bonds with either protein residues or water molecules and the hydrocarbon chains are stabilised by Van der Waal's interactions. Overall distribution of complete lipid molecules can be seen in Figure 4.26 and the electron density for one example of 7.8 MAG is shown in Figure 4.27.

Besides these well ordered 7.8 MAG molecules, many hydrocarbon chains are modelled as well. Although not sure about the identity, the electron densities into which these carbon acyl chains are modelled also reside on the hydrophobic surface of protein molecules and have unbranched linear shape (sometimes slightly curved), suggesting that they may be the tail part of less well ordered lipids. Another possibility is that they are less well ordered LDAO molecules (Figure 3.3), because the tail of LDAO has similar hydrophobicity and shape. However, because the head group is disordered, it is not possible to decide whether these carbon acyl chains come from 7.8 MAG or LDAO, unless the length of the carbon acyl chain is longer than 12 atoms, which is the maximum number for LDAO tail.

These lipids or hydrocarbon chains appear in various sites on protein surface but two overlapped lipid binding sites are found by superimposing all *in meso* structures together. As in Figure 4.28, a complete 7.8 MAG (name as 7.8 MAG-1 in this thesis) molecule is modelled into all *in meso* structure models. In all models, the lipid headgroup oxygen forms hydrogen bond with sidechain nitrogen of Trp81. Also the neck and tail of 7.8 MAG-1 in all structures overlaps and they are probably stabilised by Van der Wall forces. In Figure 4.29, a complete 7.8 MAG molecule (name as 7.8 MAG-2 in this thesis) is modelled into AlgE_1.9 and AlgE_2.8, while a 13-atom hydrocarbon chain is modelled in AlgE_2.4. In

AlgE_1.9 and AlgE_2.8, the headgroup oxygen forms hydrogen bond with the sidechain nitrogen atoms of both Asn54 and His480. And, same as 7.8 MAG-1, the neck and tail of 7.8 MAG-2 in all structures overlap and are stabilised by Van der Waal's forces. Expanded views of these two lipid binding sites are shown in both Figure 4.28 and Figure 4.29.

According to the sequence alignment shown in Figure 1.7, His480 is highly conserved among *Pseudomonas* species but Trp81 or Asn54 are not. Arg or Ser replaces Trp81 and Asp replaces Asn54 in other *Pseudomonas* species. It is possible that His480 is involved in the interaction between native *Pseudomonas* lipids and AlgE molecules in cells. Thus, the interaction between 7.8 MAG-2 and His480 may have helped to stabilise the protein during crystallisation process by resembling the native environment for AlgE.

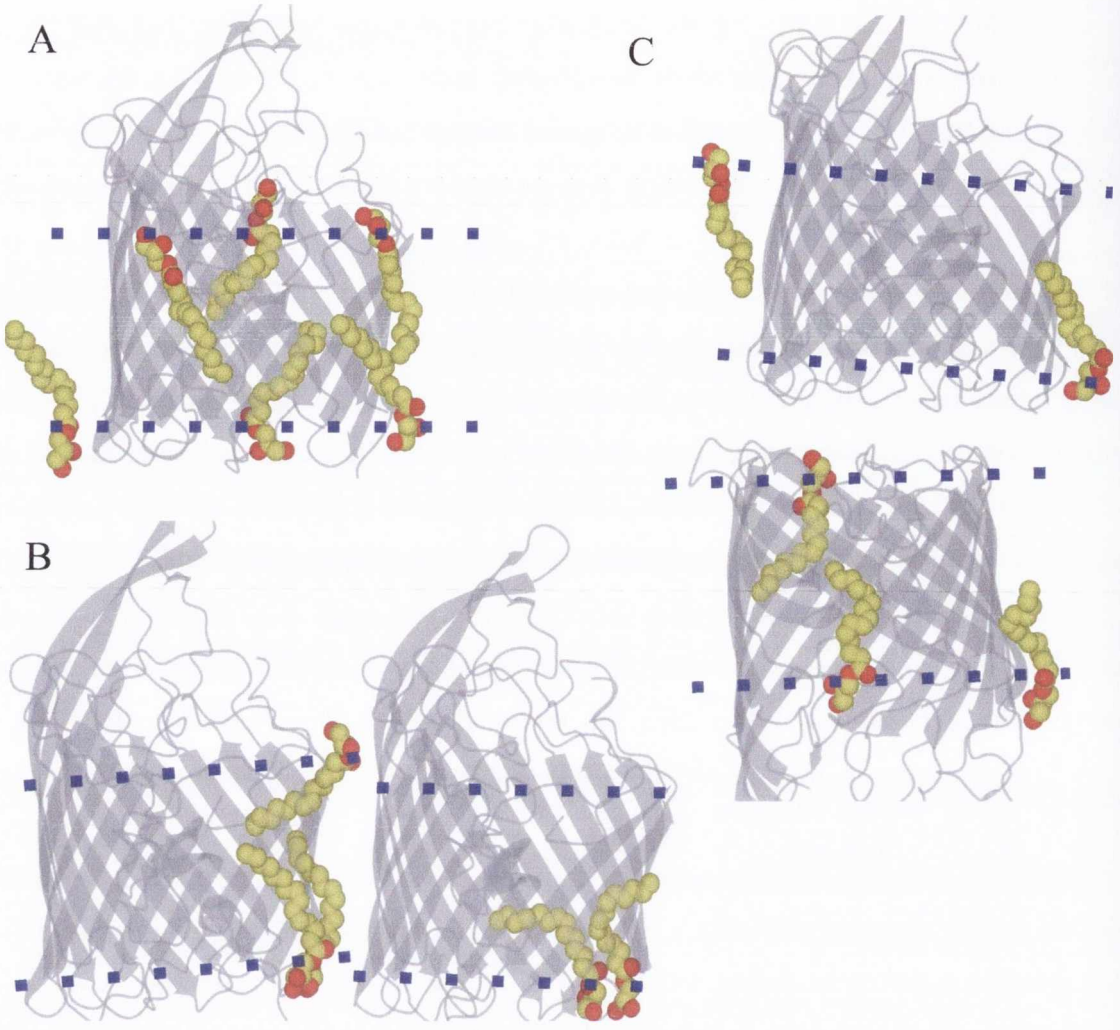


Figure 4.26 Overall distributions of complete lipid molecules in each *in meso* AlgE structure model. AlgE_1.9, AlgE_2.4 and AlgE_2.8 is shown in panel A, B and C, respectively. Protein molecules are drawn in ribbon style and coloured in grey in transparency. 7.8 MAG molecules are drawn as spheres and coloured by atom type: carbon in yellow and oxygen in red. Blue dotted lines represent interface between hydrophobic region and hydrophilic region. It is clear that all head groups of all lipid molecules are located right on or very close to the interface and the hydrocarbon chains are located on the protein hydrophobic surface.

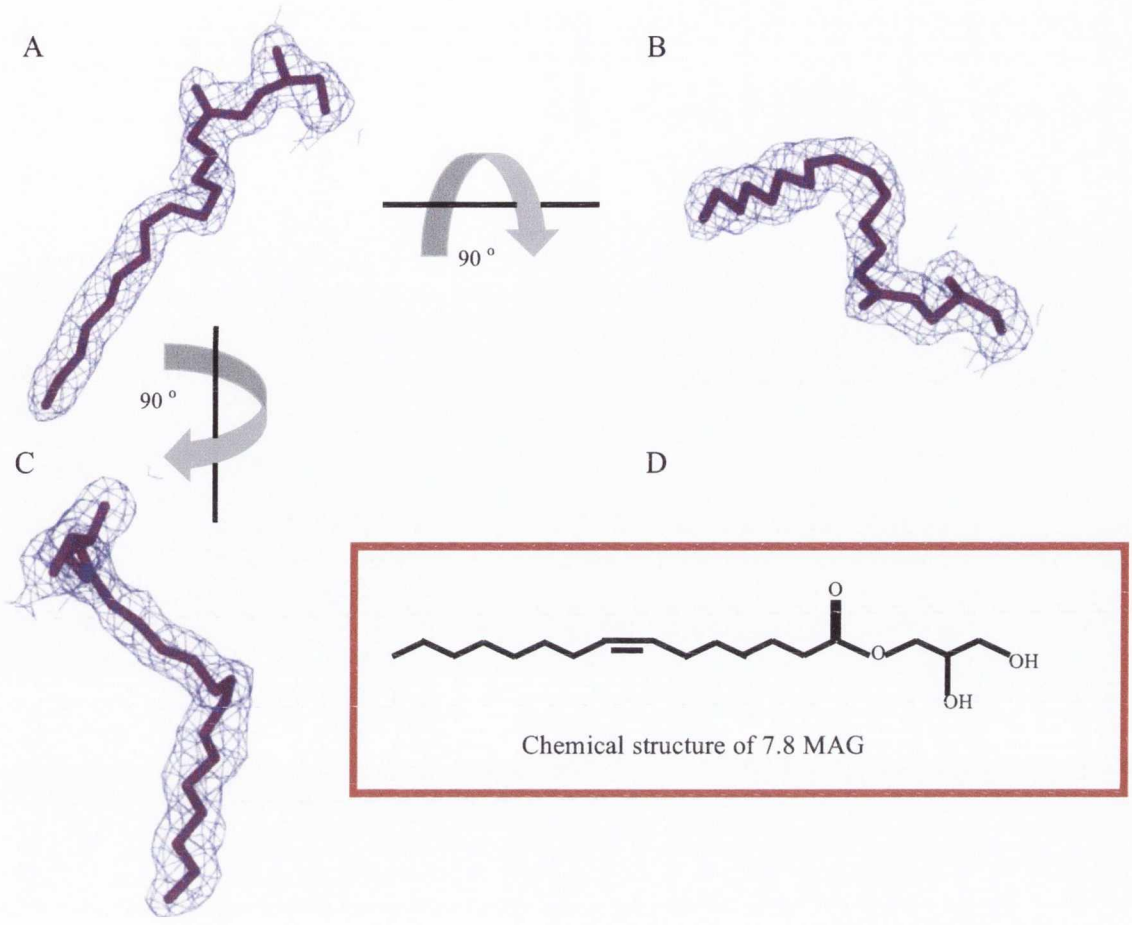


Figure 4.27 An example of 7.8 MAG (7.8 MAG-1) molecule with electron density from *in meso* AlgE_1.9 structure model. The lipid is drawn in cylinder style and coloured in purple. Three views are taken from three angles by rotating the model 90° vertically or horizontally as in panel A, B and C. The blue mesh around lipid molecule represents electron density and the map is contoured at 1σ. A chemical structure of 7.8 MAG is shown in panel D.

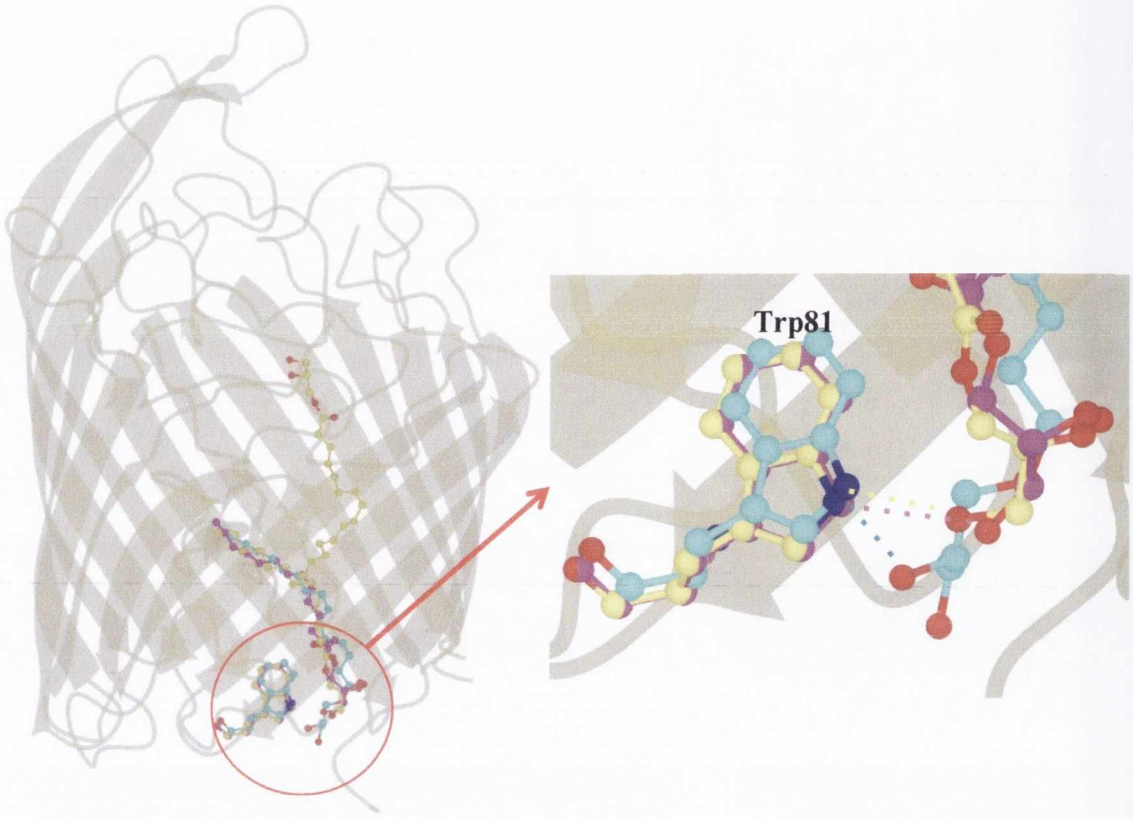


Figure 4.28 7.8 MAG-1 binding site in *in meso* AlgE structure models. The protein molecule is drawn in ribbon and coloured in transparent grey. The carbon atoms on 7.8 MAG molecules and Trp81 from AlgE_1.9, AlgE_2.4 and AlgE_2.8 are coloured yellow, cyan and magenta, respectively Nitrogen atoms are coloured in blue and oxygen in red. Where 7.8 MAG interacts with Trp81 is circled in red and the expanded view is shown on the right-hand side. Hydrogen bondings between 7.8 MAG and the sidechain nitrogen atom of Trp81 are labelled by dotted lines.

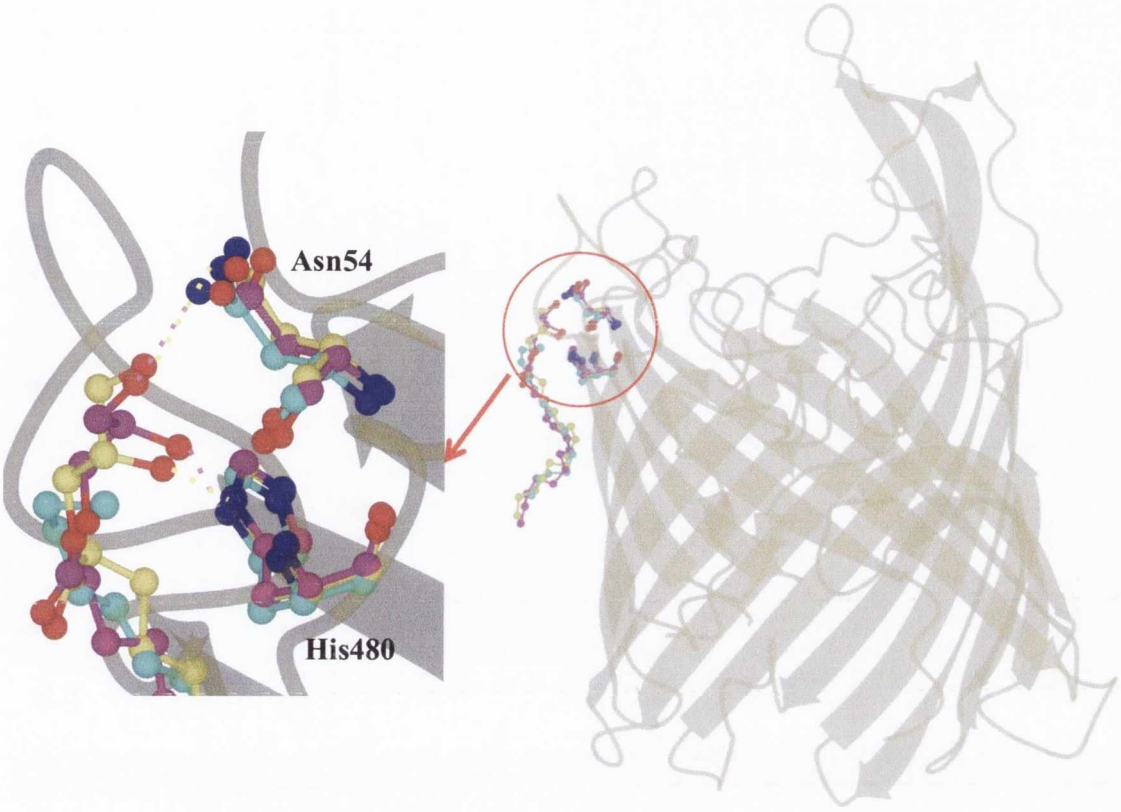


Figure 4.29 7.8 MAG-2 lipid binding site in *in meso* AlgE structure models. The protein molecule is drawn in ribbon and coloured in transparent grey. The carbon atoms on 7.8 MAG molecules, Asn54 and His480 from AlgE_1.9, AlgE_2.4 and AlgE_2.8 are coloured yellow, cyan and magenta, respectively. Nitrogen atoms are coloured in blue and oxygen in red. Where 7.8 MAG interacts with Asn54 and His480 is circled in red and the expanded view is shown on the left-hand side. Hydrogen bondings between 7.8 MAG and the sidechain nitrogen atom of Asn54 and His480 are labelled by dotted lines. Because no head group is modelled in AlgE_2.4, no hydrogen bonding is shown but the hydrocarbon chain for 7.8 MAG in this model overlaps with the other two models.

4.4.6 Metal ions

As in the Table 4.3, there are various metal ions in each *in meso* AlgE model. Details of these ions are listed in Table 4.9.

Table 4.9 List of metal ions in one asymmetric unit of all *in meso* structure models.

	AlgE_1.9	AlgE_2.4	AlgE_2.8	AlgE_in_surfo
Ca^{2+}	2 per 1 protein molecule	0	2 per 2 protein molecules	4 per 4 protein molecules
Na^{+}	1 per 1 protein molecule	0	0	0
Mg^{2+}	0	1 per 2 protein molecule	0	0
Cu^{2+}	0	4 per 2 protein molecules	0	0

The position of Ca^{2+} ions found in both AlgE_1.9 and AlgE_2.8 agrees with that of Ca^{2+} ions found in AlgE_in_surfo, as shown in panel A and C of Figure 4.29. This Ca^{2+} ion is located on L1 and coordinating with the following residues: Asp55, Asp57, Ala61 and Gly63. It is thought to stabilize the conformation of L1 instead of contributing directly to the channel specificity: mutation on L1 results in a slower migration speed of *Δ*AlgE on SDS-PAGE than wild-type AlgE, which means L1 is possibly unfolded in *Δ*AlgE, but doesn't change the ability of *Δ*AlgE to transport alginate by a large degree, which means it didn't affect the crucial part of the protein (Hay *et al.* 2010, Whitney *et al.* 2011).

It is not known whether Na^{+} and Mg^{2+} ions are important to the function of AlgE as alginate transporter but given that they are all located on short periplasmic turns (Figure 4.30), it is unlikely.

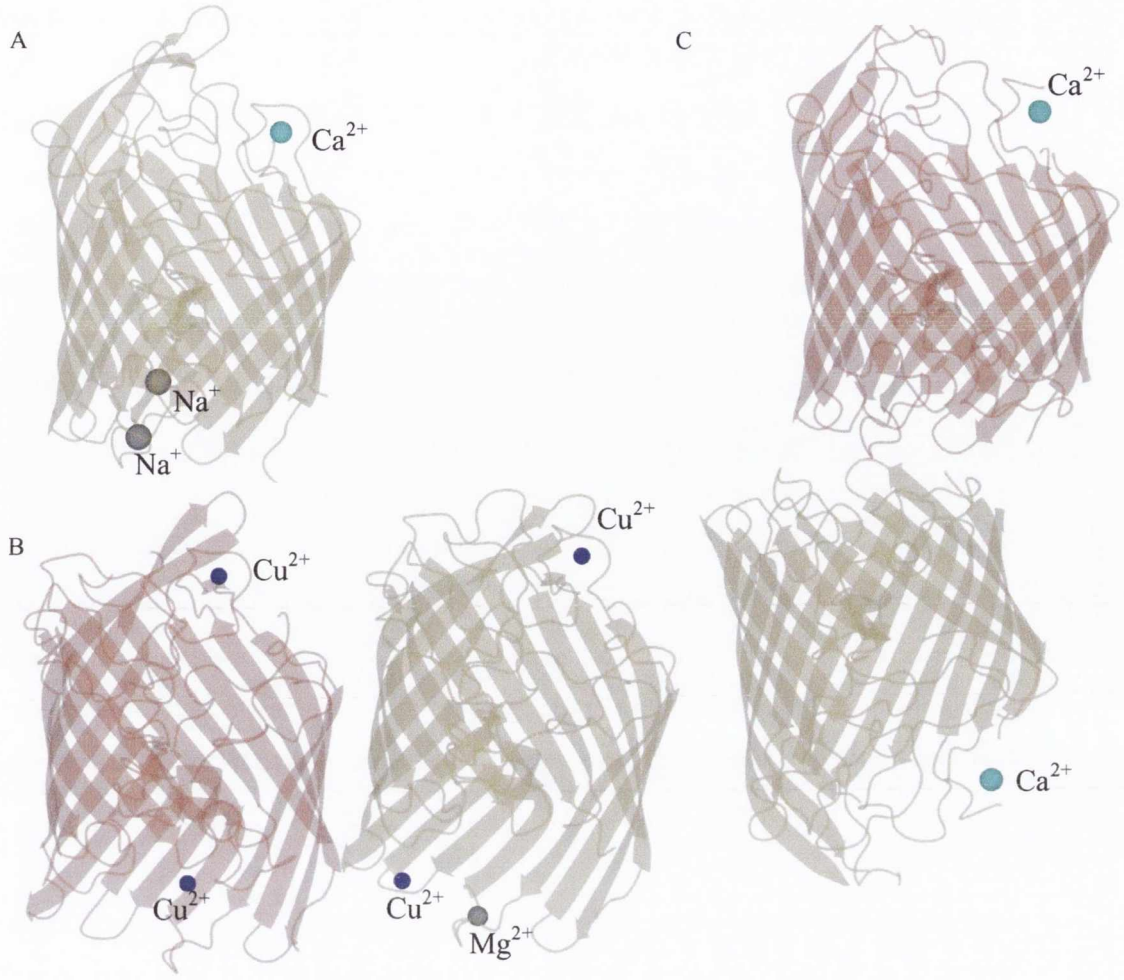


Figure 4.30 Ions modelled in *in meso* AlgE structure models. In panel A, AlgE_1.9 structure model is shown in transparency. Ca^{2+} ion is drawn as cyan sphere and Na^{+} ions are drawn in grey spheres. In panel B, AlgE_2.4 structure model is shown in transparency. Cu^{2+} ions are drawn as cyan spheres and Mg^{2+} ion is drawn in grey sphere. In panel C, AlgE_2.8 structure model is shown in transparency. Ca^{2+} ions are drawn as cyan spheres.

Unlike other models, when Ca^{2+} ion was first placed in the region of L1 of AlgE_2.4, positive electron density appeared around the metal ions after refinement by PHENIX program, suggesting that the ions in this position in AlgE_2.4 is bigger than Ca^{2+}. This also happens to two other sites in AlgE_2.4. As a result, As^{3+}, Cu^{2+} and Zn^{2+} ions have been modelled one by one into these positions because it is possible that AlgE picks up them during expression, purification or crystallisation process. After refining with As^{3+}, negative electron density appears around the ions suggesting As^{3+} is too big for the electron density. In addi-

tion, the average B-factor for the ions goes up to ~120.0 $Å^2$, which is too high compared to an average B-factor of ~40.0 $Å^2$ from the surrounding objects. As a result, As^{3+} is ruled out. After refining with Cu^{2+} and Zn^{2+}, neither negative electron density nor positive electron density is observed around the ions. In addition, Cu^{2+} and Zn^{2+} gave similar average B-factor as ~58.0 $Å^2$ and ~59.0 $Å^2$, both of which are reasonable when compared to the surrounding objects. As a result, these two ions are considered possible choices for the AlgE_2.4 model. In order to determine which ions are actually in the crystal, X-ray spectroscopy experiment was carried out by kind help from Dr. Ruslan Sanishvili at APS. However, because the energy of X-ray emission for Zn and Cu overlap with each other from 8.6 keV to 8.9 keV ($Zn_{K\alpha 1}$ = 8.638 kev and $Cu_{K\beta 1}$ = 8.905 keV) (Thompson *et al.* 2001), it was not able differentiate the two ions from each other. Considering that AlgE is purified by Cu^{2+}-IMAC while there is no obvious source for Zn^{2+} except trace amount in *E. coli* cells, Cu^{2+} is decided to be a more reasonable choice. But different from typical Cu^{2+} binding sites found in other copper containing proteins, where usually multiple His, Asp or Cys residues are involved (Ducros *et al.* 1998, Katoka *et al.* 1999, Van Den Berg et al. 2004), each Cu^{2+} ion is only interacting with one His in the case of AlgE_2.4.

The details of Cu^{2+} ions with surrounding environment are shown both in Table 4.10 and Figure 4.31. In panel A and B, the two Cu^{2+} ions in panel A and B seem to exactly replace the Ca^{2+} ions found in other AlgE structure models and probably play the same role to stabilize L1. The other two Cu^{2+} ions in panel C and D of Figure 4.31 coordinate with two histidine residues: His218 in molecule A and His145 in molecule B. The existence of these two Cu^{2+} ions may help to explain why AlgE binds to Cu^{2+}-IMAC during purification process as His145 and His 218 are both available to the local environment and can be used for copper binding. But it is not sure whether they are important to the function of AlgE as alginate transporter. In addition, it is still not known why Cu^{2+} is only found in AlgE_2.4 although all AlgE preps follow exactly the same protocol (section 2.3.11) and the same crystallisation condition generated the data sets used for AlgE_2.4 and AlgE_2.8 (section 3.4.2.3).

Table 4.10 Contacts present in between residues, acetate or water molecules and the four Cu^{2+} ions in AlgE2.4 structure model[a].

Atom from residue, acetate or water molecules	Cu2+ ion ID	Distance	Bond type
A-Asp55: OD1	1	2.86	Not identified
A-Asp55: OD2	1	2.37	Not identified
A-Asp57: OD1	1	2.36	Not identified
A-Asp57: OD2	1	2.44	Not identified
A-Ala61: O	1	2.39	Not identified
A-Ala63: O	1	2.26	Not identified
B-Asp55: OD1	2	2.51	Not identified
B-Asp55: OD2	2	2.14	Not identified
B-Asp57: OD1	2	2.24	Not identified
B-Asp57: OD2	2	2.55	Not identified
B-Ala61: O	2	2.77	Not identified
B-Ala63: O	2	2.35	Not identified
Act1: OXT	3	3.07	Not identified
Act1: O	3	1.96	Not identified
W197: O	3	2.46	Not identified
B-His145: NE	3	2.03	Not identified
A-His218: NE	4	2.07	Not identified
W198: O	4	3.10	Not identified
W199: O	4	2.09	Not identified
W204: O	4	2.45	Not identified

a. All bond distances and types are calculated using online program PISA: http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html. However, the program is not able to identify the type for any bond to Cu^{2+} ions.

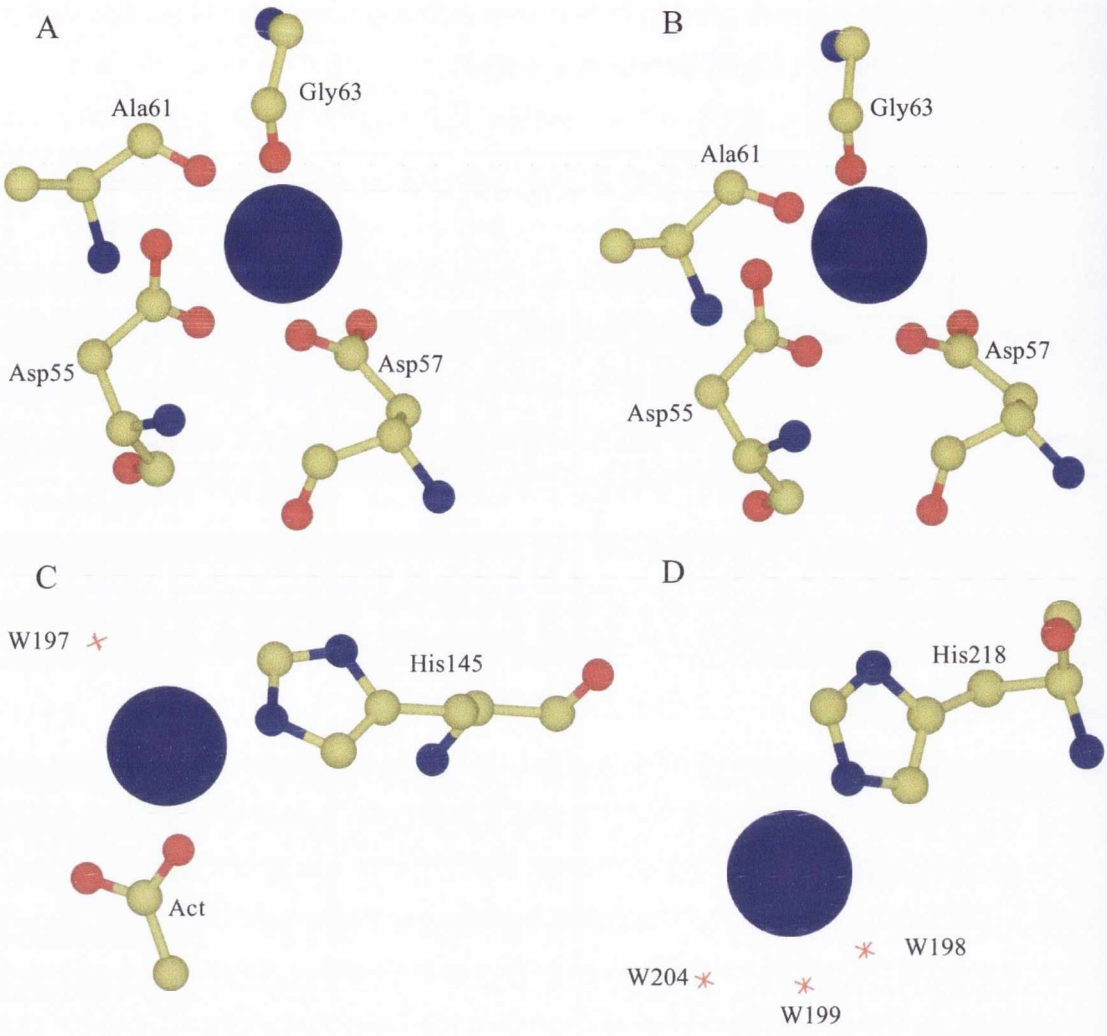


Figure 4.31 The Cu^{2+} ions in AlgE_2.4 with surrounding environment. Cu^{2+} ions are drawn as blue spheres. Residues and acetate (Act) molecule are drawn in ball-and-stick style and coloured according to atom type: carbon in yellow, oxygen in red and nitrogen in blue. Water molecules are drawn in small red stars. From panel A to panel D, the ID of Cu^{2+} ions is from 1 to 4 as in Table 4.8.

4.5 Conclusion

4.5.1 By molecular replacement method using AlgE in surfo structure as starting model, three in meso structure models are generated. Before deposition into PDB as 4AFK (1.9 Å), 4B61 (2.4 Å) and 4AZL (2.8 Å), they have been fully refined and assessed by online programs. In one asymmetric unit of AlgE_1.9, there are 1 protein molecule, 6 complete 7.8 MAG molecules, 6 complete LDAO molecules, 9 hydrocarbon chains, 1 sodium ion, 2 calcium ions, 1 citrate molecule, 2 PEG400 fragments and 242 water molecules. In one asymmetric unit of AlgE_2.4, there are 2 protein molecules, 5 complete 7.8 MAG molecules, 1 complete LDAO molecules, 19 hydrocarbon chains, 1 magnesium ion, 4 copper ions, 1 acetate molecule and 205 water molecules. In one asymmetric unit of AlgE_2.8, there are 2 protein molecules, 5 complete 7.8 MAG molecules, 8 hydrocarbon chains, 2 calcium ions and 50 water molecules. These *in meso* structure models all contain 18 transmembrane β-strands, 9 extracellular loops and 9 periplasmic turns. They share similar features with the in surfo structure 3RBH but differ in the loop regions. Some of these differences are biologically important but others are only related to crystal packing that happened during crystallisation process.

4.5.2 In previous study, a hypothetical secretion pathway was proposed based on modelling and bioinformatics work. Ten highly conserved charged residues are identified as members of an alginate secretion portal. In my project, it is observed that a citrate molecule sits right inside this proposed pore and coordinating with two of the ten highly conserved charged residues. Because citrate shares some basic common properties with the natural ligand, alginate, it is then proposed that citrate in the structure model AlgE_1.9 resembles the natural substrate of AlgE in the way of interacting with portal residues and this finding may have provided important experimental evidence to the existence of the proposed secretion portal. In addition, by analyzing both loop L2 and turn T8 from *in surfo* and *in meso* AlgE structure models, it suggests that AlgE_in_surfo is a closed state. In comparison, due to the conformation of L2, T8 and L7, AlgE_1.9 is then considered to be in an inward open state that can receive ligand from periplasmic side but doesn't allow ligand to diffuse out to extracellular side. In the case of inward open state, L2 and L7 together block the outward side of secretion channel.

4.5.3 As described above, the AlgE_1.9 is proposed to be in an inward open state but what triggers outward open state or completely open state remains unclear. To form outward open or completely open conformation, either L2 or L7 or both will have to move away from current position to allow citrate or alginate to diffuse out. As L2 in all AlgE structure models are incomplete, little information can be derived from the structures. More attention is paid to L7 in order to find out what can possibly trigger the formation of a completely open state of AlgE. One unidentified molecule is then discovered in all four AlgE structures at the same position. In AlgE_1.9 and AlgE_in_surfo, PEG400 fragment and EDO molecules are modeled based on limited information at the stage of refinement. Furthermore, at the position of PEG400 fragment or EDO molecules, the same type of unbranched linear positive electron density is observed in AlgE_2.4 and AlgE_2.8 too and this leads to the conclusion that the density may be from a common molecule. However, throughout all four structures, the only common chemicals used for purification and crystallisation work are Tris and NaCl, neither of which has unbranched linear shape. Thus, it is proposed that this common molecule may be picked up from cells during expression process and should be something other than either PEG400 or EDO. Although it is still unknown what this molecule might be, it is possible that the molecule has changed conformation of L7 and the change causes the blockage of secretion pathway to the extracellular side, considering that the unidentified molecule is located inside the barrel region and beside L7.

4.5.3 Various lipids, detergents and ions are modeled into different structure models. Two conserved lipid binding sites are observed: one coordinating with Trp81 by head group and the other coordinating with Asn54 and His480 by head group. According to sequence alignment result, His480 is highly conserved among *Pseudomonas* species but Trp81 and Asn54 are not. It is possible that His480 is involved in the interaction between native *Pseudomonas* lipids and AlgE protein. So the interaction between His480 and 7.8 MAG may have helped stablise the protein during *in meso* crystallisation process by resembling the native environment of AlgE. Ca^{2+} ions are modeled into extracellular loop L1 of AlgE_1.9 and AlgE_2.8 but Cu^{2+} ions are modeled into AlgE_2.4. The copper ions in AlgE_2.4 may help explain why it is possible to use Cu-IMAC for AlgE purification. However, overall, it is not conclusive whether these lipids, detergents or ions are important to the function of AlgE.

Chapter 5 General conclusions and future work

Overall, the project is successful in a few aspects: 1) An outer membrane protein from *P. aeruginosa*, AlgE has been cloned, expressed and purified. 2) *In meso* method once again proved to be successful in membrane protein crystallisation. 3) A high resolution structure of AlgE protein has been obtained. 4) Some important new information has been suggested by the high resolution structure.

In chapter 2, AlgE gene is amplified from *P. aeruginosa* PAO1 strain. The protein is then cloned into *E. coli* expression host and expressed in two forms: inclusion body (following published protocol (Whitney *et al.* 2009), this work was repeated by Dr. Toufic El Arnaout and me but not included in the thesis) in cytoplasm and folded mature protein on the membrane. The folded mature AlgE is then solubilised from membrane by detergents and purified by Cu-IMAC and SEC. The purified protein shows at least 90 % purity and typical heat modifiability as most other beta-barrel proteins.

In chapter 3, purified AlgE is used for crystallisation purpose. Both traditional *in surfo* and novel *in meso* methods have been applied but only in meso method yielded promising crystal hits. Among all lipids used for in meso methods, 7.8 MAG is considered to be the best lipid for this project according to the X-ray diffraction experiment results. This later is considered to be proving the theory of "hydrophobic matching" in early studies. Optimization is carried out in order to obtain as good diffraction as possible. Three diffraction data sets are yielded and the resolution is 1.9 Å, 2.4 Å and 2.8 Å, respectively. These data sets have been indexed and processed in three space groups. Attempts to phase the structures have been carried out by Se-Met labelling (this work was done by me but not incuded in this thesis) and heavy atom soaking methods and yielded a very preliminary model for AlgE protein. It turns out later that the model agrees with the structure at least in that the shape of the barrel of AlgE is a square when viewed from membrane side and it adopts typical Type I packing (layered packing). But there is no way to trace the amino acid sequence based on limited phase information.

In chapter 4, molecular replacement is used to create three *in meso* structure models. Same as the previously determined *in surfo* structure, AlgE_in_surfo, used for molecular replacement work, the models all have 18 transmembrane beta-strands, 9 extracellular loops and 9 periplasmic turns. Various organic and inorganic molecules are modelled into these structures, including lipids, detergents, hydrocarbon chains, citrate, water and metal ions. All AlgE in meso structures adopt type I packing style as most protein structures solved by in meso structures and L5 or L6 or both play important role in crystal packing. What's more interesting is that a citrate molecule is proposed and modelled into the barrel region of the structure model AlgE_1.9. This citrate molecule coordinates with local residues Arg74, Arg152 and Thr103 by H-bond and arginine residues are among the residues that consists alginate secretion portal proposed by previous *in surfo* structure work. Due to the likeness between the chemical structure and size of citrate and monomeric unit of alginate, it is considered that this may have provided some evidence for existence of the proposed secretion portal. Furthermore, when superposing AlgE_1.9 with AlgE_in_surfo, it is shown that AlgE_in_surfo may be in a closed state and AlgE_1.9 is in an inward open state. Although how AlgE can form an outward open or completely open state is unknown, the movement of L2 and L7 is considered to be involved as the loops work together to block the pathway from extracellular side. An unidentified molecule is found right beside the loop L7 in all *in meso* and *in surfo* AlgE structures. It is not known if this molecule has changed conformation of L7 but the possibility can't be excluded and needs further investigation. A number of 7.8 MAG molecules are modelled into each structure on the hydrophobic surface of protein and two common lipid binding sites are observed throughout three *in meso* structures. Interaction by H-bond is observed between 7.8 MAG and Trp81, Asn54 and His480. Due to the high conservation score of His480 among all *Pseudomonas* species, one of these sites are thought to be able to stabilise the protein when reconstituted into lipid bilayer by resembling native environment for AlgE. One *in meso* structure, AlgE_2.4, contains special metal ions as copper and this may help explain the binding affinity of AlgE to copper that is used for AlgE purification.

As stated in the beginning, this project has achieved success but exactly how AlgE transports alginate needs to be addressed. Experimental work can be carried out to obtain a ligand bound structure of AlgE. Co-crystallisation and soaking using alginate or alginate de-

rivatives can both possibly be applied to achieve the goal. This structure may be able to provide a clear insight to the functioning state of AlgE. In addition, whether AlgE interacts with other proteins encoded by algD operon, especially AlgK, can be explored possibly by the use of microscale thermophoresis technique. This may become a basis to pursue the further goal as to understand whether AlgE needs other factors to perform complete function. Besides structure work, docking study can be used to prove the feasibility of using citrate as mimicry of alginate and may reveal more undiscovered residues for the secretion pathway. How the flexible loops or turns of AlgE move during transportation process can be calculated by molecular dynamics study and may provide insight as to what is needed for AlgE to adopt outward or completely open state. Some preliminary work has been carried out but the results are not conclusive.

References

Adams, P. D., P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D, **66**:213–221.

Ai, X. and M. Caffrey (2000). Membrane Protein Crystallization in Lipidic Mesophases: Detergent Effects. Biophys. J., **79**:394 – 405.

Bakkevig, K., H. Sletta, M. Gimmestad, R. Aune, H. Ertesvåg, K. Degnes, B. E. Christensen, T. E. Ellingsen, and S. Valla (2005). Role of the Pseudomonas fluorescens Alginate Lyase (AlgL) in Clearing the Periplasm of Alginates Not Exported to the Extracellular Environment. J. Bacteriol., **187**:8375–8384.

Bentzmann, S. d. and P. Plésiat (2011). The Pseudomonas aeruginosa opportunistic pathogen and human infections. Environ. Microbiol., **13**:1655 - 1665.

Bianciotto, V., S. Andreotti, R. Balestrini, P. Bonfante, and S. Perotto (2001). Mucoid mutants of the biocontrol strain Pseudomonas fluorescens CHA0 show increased ability in biofilm formation on mycorrhizal and nonmycorrhizal carrot roots. Mol. Plant Microbe Interact., **14**:255–260.

Bill, R. M., P. J. Henderson, S. Iwata, E. R. Kunji, H. Michel, R. Neutze, S. Newstead, B. Poolman, C. G. Tate, and H. Vogel (2011). Overcoming barriers to membrane protein structure determination. Nat. Biotechnol., **29**:335-340.

Briggs, J., H. Chung, and M. Caffrey (1996). The temperature-composition phase diagram and mesophase structure characterization of the monoolein/water system. J. Phys. II France, **6**:723 - 751.

Brunger, A. T. (1997). Free R value: cross-validation in crystallography. Methods Enzymol., **277**:366-396.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, and G. L. Warren (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr. D, **54**:905-921.

Byrd, M. S., I. Sadovskaya, E. Vinogradov, H. Lu, A. B. Sprinkle, S. H. Richardson, L. Ma, B. Ralston, M. R. Parsek, E. M. Anderson, J. S. Lam, and D. J. Wozniak (2009). Genetic and biochemical analyses of the Pseudomonas aeruginosa Psl exopolysaccharide reveal overlapping roles for polysaccharide synthesis enzymes in Psl and LPS productionmmi. Mol. Microbiol., **73**:622 - 638.

Caffrey, M. (1989). The study of lipid phase transition kinetics by time-resolved X-ray diffraction. Annu. Rev. Biophys. Biophys. Chem., **18**:159-186.

Caffrey, M. (2009). Crystallizing membrane proteins for structure determination: use of lipidic mesophases. Annu. Rev. Biophys. Biomol. Struct., **38**:29 - 51.

Caffrey, M. and V. Cherezov (2009). Crystallizing membrane proteins using lipidic mesophases. Nat. Protoc., **4**:706 - 731.

Caffrey, M., D. Li, and A. Dukkipati (2012). Membrane Protein Structure Determination Using Crystallography and Lipidic Mesophases: Recent Advances and Successes. Biochemistry.

Caffrey, M. and C. Porter (2010). Crystallizing Membrane Proteins for Structure Determination using Lipidic Mesophases. Journal of Visualized Experiments.

Carpentier, B. and O. Cerf (1993). Biofilms and their consequences, with particular reference to hygiene in the food industry. J. Appl. Bacteriol., **75**:499 - 511.

Cecchini, G. (2003). Function and structure of complex II of the respiratory chain. Annu. Rev. Biochem., **72**:77-109.

Chen, V. B., W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson, and D. C. Richardson (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D, **66**:12-21.

Cheng, A., B. Hummel, H. Qiu, and M. Caffrey (1998). A simple mechanical mixer for small viscous lipid-containing samples. Chem. Phys. Lipids, **95**:11–21.

Cherezov, V. and M. Caffrey (2005). A simple and inexpensive nanoliter-volume dispenser for highly viscous materials used in membrane protein crystallization. J. Appl. Cryst, **38**:398–400.

Cherezov, V., M. A. Hanson, M. T. Griffith, M. C. Hilgart, R. Sanishvili, V. Nagarajan, S. Stepanov, R. F. Fischetti, P. Kuhn, and R. C. Stevens (2009). Rastering strategy for screening and centring of microcrystal samples of human membrane proteins with a sub-10 microm size X-ray synchrotron beam. J. R. Soc Interface, **5**:587 - 597.

Cherezov, V., A. Peddi, L. Muthusubramaniam, Y. F. Zheng, and M. Caffrey (2004). A robotic system for crystallizing membrane and soluble proteins in lipidic mesophases. Acta Crystallogr. D, **60**:1795-1807.

Cooper, M., G. R. Tavankar, and H. D. Williams (2003). Regulation of expression of the cyanide-insensitive terminal oxidase in Pseudomonas aeruginosa. Microbiology, **149**:1275–1284.

Cowie, D. B. and G. N. Cohen (1957). Biosynthesis by E. coli of active altered proteins containing selenium instead of sulfur. Biochim. Biophys. Acta., **26**:252–261.

Cystic Fibrosis Association Ireland (CFAI) (2011). About Cystic Fibrosis. http://www.cfireland.ie/articles.php/about_cf.

David, E. N., E. O. Dennis, W. Jessica, and J. F. Michael (2001). Role of Alginate and Its O Acetylation in Formation of Pseudomonas aeruginosa Microcolonies and Biofilms. J. Bacteriol., **183**:1047–1057.

Drake, D. and T. C. Montie (1988). Flagella, Motility and Invasive Virulence of Pseudomonas aeruginosa. J. Gen. Microbiol., **134**:43-52.

Drenkard, E. and A. M. Frederick (2002). Pseudomonas biofilm formation and antibiotic resistance are linked to phenotypic variation. Nature, **416**:740 - 743.

Drenth, J. (2007). Principles of Protein X-Ray Crystallography. Springer, New York.

Ealick, S. E. (2000). Advances in multiple wavelength anomalous diffraction crystallography. Curr. Opin. Chem. Biol., **4**:495 - 499.

Emsley, P. and K. Cowtan (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D, **60**:2126-2132.

Esquisabel, C., M. C. Rodríguez, A. O. Campo-Sosa, C. Rodríguez, and L. Martínez-Martínez (2011). Mechanisms of resistance in clinical isolates of Pseudomonas aeruginosa less susceptible to cefepime than to ceftazidime. Clin. Microbiol. Infect., **4**:1469.

Faham, S. and J. U. Bowie (2002). Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. J. Mol. Biol., **316**:1-6.

Fazekas, d. S. G. S., R. G. Webster, and A. Datyner (1963). Two new staining procedures for quantitative estimation of proteins on electrophoretic strips. Biochim. Biophys. Acta, **71**:377 - 391.

Fetar, H., C. Gilmour, R. Klinoski, D. M. Daigle, C. R. Dean, and K. Poole (2011). mexEF-oprN Multidrug Efflux Operon of Pseudomonas aeruginosa: Regulation by the MexT Activator in Response to Nitrosative Stress and Chloramphenicol. Antimicrob. Agents Chemother., **55**:508 - 514.

Féthière, J. (2007). Three-Dimensional Crystallization of Membrane Proteins. Methods Mol. Biol., **363**:191 - 223.

Fischetti, R. F., S. Xu, D. W. Yoder, M. Becker, V. Nagarajan, R. Sanishvili, M. C. Hilgart, S. Stepanov, O. Makarov, and J. L. Smitha (2009). Mini-beam collimator enables

microcrystallography experiments on standard beamlines. J. Synchrotron Radiat., **16**:217–225.

Franklin, M. J., C. E. Chitinis, P. Gacesa, A. Sonesson, D. C. White, and D. E. Ohman (1994). Pseudomonas aeruginosa AlgG Is a Polymer Level Alginate C5-Mannuronan Epimerase. J. Bacteriol., **176**:1821-1830.

Franklin, M. J., D. Nivens, E. , J. T. Weadge, and P. L. Howell (2011). Biosynthesis of the Pseudomonas aeruginosa extracellular polysaccharides, alginate, Pel, and Psl. Front. Microbiol., **2**:1 - 16.

Franklin, M. J. and D. E. Ohman (2002). Mutant Analysis and Cellular Localization of the AlgI, AlgJ, and AlgF Proteins Required for O Acetylation of Alginate in Pseudomonas aeruginosa. J. Bacteriol., **184**:3000–3007.

Fromme, P. (2003). Crystallization of Photosystem I. Methods and Results in Crystallization of Membrane Proteins. International University Line, San Diego.

Garman, E. F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? Acta Crystallogr. D, **66**:339-351.

Gonzalez, A. (2002). Faster data-collection strategies for structure determination using anomalous dispersion. Acta Crystallogr. D, **59**:325-322.

Govan and Deretic (1996). Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. Microbiol. Rev., **60**:539-574.

Grabert, E., J. Wingender, and U. K. Winkler (1990). An outer membrane protein characteristic of mucoid strains of Pseudomonas aeruginosa. FEMS Microbiol. Lett., **68**.

Grabert, E., J. Wingender, and U. K. Winkler (1990). An outer membrane protein characteristic of mucoid strains of Pseudomonas aeruginosa. FEMS Microbiol. Lett., **56**:83 - 87.

Hall-Stoodley, L., J. W. Costerton, and P. Stoodley (2004). Bacterial biofilms: from the natural environmen to infectious diseases. Nature Reviews, **2**:95 - 108.

Hansman, G. S., S. Shahzad-Ul-Hussan, J. S. McLellan, G. Y. Chuang, I. Georgiev, T. Shimoike, K. Katayama, C. A. Bewley, and P. D. Kwong (2012). Structural basis for norovirus inhibition and fucose mimicry by citrate. J. Virol., **86**:284-292.

Harker, D. (1956). The determination of the phases of the structure factors of noncentrosymmetric crystals by the method of double iomorphous replacement. Acta Cryst., **9**:1-9.

Hassett, D. J. (1996). Anaerobic production of alginate by Pseudomonas aeruginosa: alginate restricts diffusion of oxygen. J. Bacteriol., **178**:7322 - 7325.

Hay, I. D., Z. U. Rehman, and B. H. A. Rehm (2010). Membrane Topology of Outer Membrane Protein AlgE, Which Is Required for Alginate Production in Pseudomonas aeruginosa. Appl. Environ. Microbiol., **76**:1806–1812.

Helenius, A. and K. Simons (1975). Solubilization of membranes by detergents. Biochim. Biophys. Acta, **415**:29-79.

Hendrickson, W. A., J. R. Horton, and D. M. LeMaster (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J., **9**:1665–1672.

Hendrickson, W. A. and C. M. Ogata (1997). Phase determination from multiwavelength anomalous diffraction measurements. Methods Enzymol., **276**:494 - 523.

Hilgart, M. C., R. Sanishvili, C. M. Ogata, M. Becker, N. Venugopalan, S. Stepanov, O. Makarov, J. L. Smith, and R. F. Fischetti (2011). Automated sample-scanning methods for radiation damage mitigation and diffraction-based centering of macromolecular crystals. J. Synchrotron Radiat, **18**:717 - 722.

Imamura, Y., J. Chandra, P. K. Mukherjee, and e. al (2008). Fusarium and Candida albicans biofilms on soft contact lenses: model development, influence of lens type, and susceptibility to lens care solutions. Antimicrob. Agents Chemother., **52**:171 - 182

Jain, S. and D. Ohman (1998). Deletion of algK in mucoid Pseudomonas aeruginosa blocks alginate polymer formation and results in uronic acid secretion. J. Bacteriol., **180**:634-641.

Jain, S. and D. E. Ohman (2005). Role of an Alginate Lyase for Alginate Transport in Mucoid Pseudomonas aeruginosa. Infect. immun., **73**:6429–6436.

Jumpertz, T., B. Tschapek, N. Infed, S. H. Smits, R. Ernst, and L. Schmitt (2011). High-throughput evaluation of the critical micelle concentration of detergents. Anal. Biochem., **408**:64-70.

Keiski, C.-L., M. Harwich, S. Jain, A. M. Neculai, P. Yip, H. Robinson, J. C. Whitney, L. Riley, L. L. Burrows, D. E.Ohman, and P. L. Howell (2010). AlgK is a TPR-containing protein and the periplasmic component of a novel exopolysaccharide secretin. Structure, **10**:265–273.

Kendrew, J. C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature, **181**:662.

King, E. O., M. K. Ward, and D. E. Raney (1954). Two simple media for the demonstration of pyocyanin and fluorescein. J. Lab. Clin. Med., **44**:301 - 307.

Kloppmann, E., M. Puntaand, and B. Rost (2012). Structural genomics plucks high-hanging membrane proteins. Curr. Opin. Struct. Biol., **22**:326 – 332.

Landau, E. M. and J. P. Rosenbusch (1996). Lipidic cubic phases: a novel concept for the crystallization of membrane proteins. Proc. Natl. Acad. Sci. USA., **93**:14532–14535.

Leid, J. G., C. J. Willson, M. E. Shirtliff, D. J. Hassett, M. R. Parsek, and A. K. Jeffers (2005). The exopolysaccharide alginate protects Pseudomonas aeruginosa biofilm bacteria from IFN-gamma-mediated macrophage killing. J. Immunol., **175**:7512-7518.

Lewis, K. (2011). Riddle of biofilm resistance. Antimicrob. Agents Chemother., **45**:999–1007.

Li, D., C. Boland, D. Aragao, K. Walsh, and M. Caffrey (2012). Harvesting and Cryo-cooling Crystals of Membrane Proteins Grown in Lipidic Mesophases for Structure Determination by Macromolecular Crystallography. Journal of Visualized Experiments.

Li, D., C. Boland, K. Walsh, and M. Caffrey (2012). Use of a Robot for High-throughput Crystallization of Membrane Proteins in Lipidic Mesophases. Journal of Visualized Experiments.

Li, D., J. Lee, and M. Caffrey (2011). Crystallizing Membrane Proteins in Lipidic Mesophases. A Host Lipid Screen. Cryst. Growth Des., **11**:530-537.

Loll, P. J., A. Tretiakova, and E. Soderblom (2003). Compatibility of detergents with the microbatch-under-oil crystallization method. Acta Crystallogr. D, **59**:1114–1116.

Luzzati, V. (1997). Biological significance of lipid polymorphism: the cubic phases. Curr. Opin. Chem. Biol., **7**:661-668.

Luzzati, V., A. Tardieu, T. Gulik-Krzywicki, E. Rivas, and F. Reiss-Husson (1968). Structure of the cubic phases of lipid-water systems. Nature, **220**:485-488.

Lyons, J. A. (2012). Towards the rational design of lipids for use in in Meso crystallisation of membrane proteins: a case study with a cytochrome oxidase and a G-protein coupled receptor. Chemical and Environmental Sciences Department, **Doctor of Philosophy** 416.

Mann, E. E. and D. J. Wozniak (2011). Pseudomonas biofilm matrix composition and niche biology. FEMS Microbiol. Rev., **36**:893 - 916.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read (2007). Phaser crystallographic software. J. Appl. Cryst., **40**:658-674.

Merighi, M., V. T. Lee, M. Hyodo, Y. Hayakawa, and S. Lory (2007). The second messenger bis-(3'-5')-cyclic-GMP and its PilZ domain-containing receptor Alg44 are required for alginate biosynthesis in Pseudomonas aeruginosa. Mol. Microbiol., **65**:876-895.

Meyer, T. S. and B. L. Lamberts (1965). Use of coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips. Biochim. Biophys. Acta., **107**:144 -145

Misquitta, L. V., Y. Misquitta, V. Cherezov, O. Slattery, J. M. Mohan, D. Hart, M. Zhalnina, W. A. Cramer, and M. Caffrey (2004). Membrane protein crystallization in lipidic mesophases with tailored bilayers. Structure, **12**:2113-2124.

Misquitta, Y. and M. Caffrey (2003). Detergents Destabilize the Cubic Phase of Monoolein: Implications for Membrane Protein Crystallization. Biophys. J., **85**:3084–3096.

Misquitta, Y., V. Cherezov, F. Havas, S. Patterson, J. M. Mohan, A. J. Wells, D. J. Hart, and M. Caffrey (2004). Rational design of lipid for membrane protein crystallization. J. Struct. Biol., **148**:169-175.

Moller, J. V. and L. M. M. (1993). Detergent binding as a measure of hydrophobic surface area of integral membrane proteins. J. Biol. Chem., **268**:18659 - 18672.

Morikawa, M. (2006). Beneficial biofilm formation by industrial bacteria Bacillus subtilis and related species. J. Biosci. Bioeng., **101**:1 - 8.

Morita, Y., Y. Komori, T. Mima, T. Kuroda, T. Mizushima, and T. Tsuchiya (2001). Construction of a series of mutants lacking all of the four major mex operons for multidrug efflux pumps or possessing each one of the operons from Pseudomonas aeruginosa PAO1: MexCD-OprJ is an inducible pump. FEMS Microbiol. Lett., **202**:139 - 143.

Mortensen, K. P. and S. N. Conley (1994). Film fill fouling in counterflow cooling towers: mechanisms and design. CTI. J., **15**:10–25.

Nakamura, K. and S. Mizushima (1976). Effects of heating in dodecyl sulfate solution on the conformation and electrophoretic mobility of isolated major outer membrane proteins from Escherichia coli K-12. J. Biochem., **80**:1411 - 1422.

Newstead, S., S. Ferrandon, and S. Iwata (2008). Rationalizing alpha-helical membrane protein crystallization. Protein Sci., **17**:466-472.

Newstead, S., J. Hobbs, D. Jordan, E. P. Carpenter, and S. Iwata (2008). Insights into outer membrane protein crystallization. Mol. Membr. Biol., **25**:631-638.

Ng, J. D., J. A. Gavira, and J. M. Garcia-Ruiz (2003). Protein crystallization by capillary counterdiffusion for applied crystallographic structure determination. J. Struct. Biol., **142**:218–231.

Ng, J. D., R. C. Stevens, and P. Kuhn (2008). Protein crystallization in restricted geometry: advancing old ideas for modern times in structural proteomics. Methods Mol. Biol., **426**:363–376.

National Institue of Health (NIH) (2002). Research on microbial biofilms http://grants.nih.gov/grants/guide/pa-files/PA-03-047.html.

Oglesby, L. L., S. Jain, and D. E. Ohman (2008). Membrane topology and roles of Pseudomonas aeruginosa Alg8 and Alg44 in alginate polymerization. Microbiology, **154**:1605–1615.

Ostermeier, C. and H. Michel (1997). Crystallization of membrane proteins. Curr.1 Opin. Struct. Biol., **7**:697-701.

Pace, C. N., F. Vajdos, L. Fee, G. Grimsley, and T. Gray (1995). How to measure and predict the molar absorption coefficient of a protein. Protein Sci., **4**:2411 - 2423.

Palmer, K. L., S. A. Brown, and M. Whiteley (2007). Membrane-bound nitrate reductase is required for anaerobic growth in cystic fibrosis sputum. J. Bacteriol., **189**:4449 - 4455.

Parsek, M. R. and P. K. Singh (2003). Bacterial biofilms: an emerging link to disease pathogenesis. Annu. Rev. Microbiol., **57**:677 - 701.

Patton, J. S. and M. C. Carey (1979). Watching fat digestion. Science, **204**:145 - 148.

Peter, H., I. Ylla, C. Gudasz, A. M. Romaní, S. Sabater, and L. J. Tranvik (2011). Multifunctionality and diversity in bacterial biofilms. PLoS One, **6**:e23225.

Peyton, B. M. and W. G. Characklis (1995). Microbial biofilms and biofilm reactors. Bioprocess Technol., **20**:187 - 231.

Pier (1998). Pseudomonas aeruginosa: a key problem in cystic fibrosis. ASM News., **6**:339-347.

Pindar, D. F. and C. Bucke (1975). The Biosynthesis of Alginiic Acid by Azotobacter vinelandii. Biochem. J., **152**:617 - 622.

Qiu, H. and M. Caffrey (2000). The phase diagram of the monoolein/water system: metastability and equilibrium aspects. Biomaterials, **2000**:223-234.

Regni, C., P. A. Tipton, and L. J. Beamer (2002). Crystal Structure of PMM/PGM: An Enzyme in the Biosynthetic Pathway of P. aeruginosa Virulence Factors. Structure, **10**:269–279.

Rehm, B. H. A., G. Boheim, J. Tommasen, and U. K. Winkler (1994). Overexpression of algE in Escherichia coli: Subcellular Localization, Purification, and Ion Channel Properties. J. Bacteriol., **176**:5639-5647.

Rehm, H. A., G. Boheim, J. Tommassen, and U. K. Winkler (1994). Overexpression of algE in Escherichia coli: Subcellular Localization, Purification, and Ion Channel Properties. J Bacteriol., **176**:5639-5647.

Remminghorst, U. and H. A. Rehm (2006). Alg44, a unique protein required for alginate biosynthesis in Pseudomonas aeruginosa. FEBS Lett., **580**:3883–3888.

Remminghorst, U. and H. A. Rehm (2006). Bacterial alginates: from biosynthesis to applications. Biotechnol. Lett., **28**:1701–1712.

Remminghorst, U. and H. A. Rehm (2006). In Vitro Alginate Polymerization and the Functional Role of Alg8 in Alginate Production by Pseudomonas aeruginosa. Appl. Environ. Microbiol., **72**:298–305.

Remminghorst, U. and H. A. Rehm (2006). In vitro alginate polymerization and the functional role of Alg8 in alginate production by Pseudomonas aeruginosa. Appl. Environ. Microbiol., **72**:298 - 305.

Rogers, A. H. (2008). Molecular Oral Microbiology. Caister Academic Press.

Rummel, G., A. Hardmeyer, C. Widmer, M. L. Chiu, P. Nollert, K. P. Locher, I. I. Pedruzzi, E. M. Landau, and J. P. Rosenbusch (1998). Lipidic Cubic Phases: New Matrices for the Three-Dimensional Crystallization of Membrane Proteins. J. Struct. Biol., **121**:82 - 91.

Rupp, B. (2010). Biomolecular Crystallography: Principles, Practice and Application to Structural Biology. Garland Science, New York.

Russell, R. J., J. M. Ferguson, D. W. Hough, D. M.J., and G. L. Taylor (1997). The crystal structure of citrate synthase from the hyperthermophilic archaeon pyrococcus furiosus at 1.9 A resolution. Biochemistry, **36**:9983-9994.

Sambrook and Russell (2011). Molecular Cloning: A Laboratory Manual (3rd ed). Cold Spring Harbor Laboratory Press.

Siegel, D. P., V. Cherezov, D. V. Greathouse, R. E. Koeppe, J. A. Killian, and M. Caffrey (2006). Transmembrane peptides stabilize inverted cubic phases in a biphasic length-dependent manner: implications for protein-induced membrane fusion. Biophys. J., **90**:200 - 211.

Simpson, J. A. (1989). Scavenging by alginate of free radicals released by macrophages Source. Free Rad. Biol. Med., **6**:347 - 353.

Smyth, M. S. and J. H. Martin (2000). X Ray crystallography. J. Clin. Pathol., **53**:8 - 14.

Snook, C. F., P. Tipton, A. , and L. J. Beamer (2003). Crystal Structure of GDP-Mannose Dehydrogenase: A Key Enzyme of Alginate Biosynthesis in P. aeruginosa. Biochemistry, **42**:4658-4668.

Sun, T., K. Hayakawa, K. S. Bateman, and M. E. Fraser (2010). Identification of the Citrate-binding Site of Human ATP-Citrate Lyase Using X-ray Crystallography. J. Biol. Chem., **285**:27418–27428.

Sutherland, I. (2001). Biofilm exopolysaccharides: a strong and sticky framework. Microbiology, **147**:3 - 9.

Sutton, B. J. and M. K. Sohi (1994). Crystallization of membrane proteins for X-ray analysis. Methods Mol. Biol., **27**:1–18.

Taylor, G. (2003). The phase problem. Acta Crystallogr. D, **59**:1881-1890.

Tickle, I. J. (1996). MIR Tutorial Bath. Isomorphous replacement and anomalous scattering, http://www.ysbl.york.ac.uk/~mgwt/CCP4/ccp4i/help/modules/appendices/ir-bathtutorial/irbath98.html.

Todar, K. (2011). Pseudomonas aeruginosa, Todar's Online Textbook of Bacteriology. http://www.textbookofbacteriology.net/pseudomonas.html.

Usón, I. and G. M. Sheldrick (1999). Advances in direct methods for protein crystallography. Curr. Opin. Chem. Biol., **9**:643–648.

Whitney, J. C., I. D. Hay, C. Li, H. Robinson, M. F. Amaya, L. F. Wood, D. E. Ohman, C. E. Bear, B. H. A. Rehm, and L. H. P. (2011). Structural basis for alginate secretion across the bacterial outer membrane. Proc. Natl. Acad. Sci. USA., **108**:13083 - 13088.

Whitney, J. C., A. M. Neculai, D. E. Ohman, and P. L. Howell (2009). Expression, refolding, crystallization and preliminary X-ray analysis of Pseudomonas aeruginosa AlgE. Acta Crystallogr. F, **65**:463 - 465.

Williams, H. D., J. E. Zlosnik, and B. Ryall (2007). Oxygen, cyanide and energy generation in the cystic fibrosis pathogen Pseudomonas aeruginosa. Adv. Microb. Physiol., **52**:61 - 71.

Winsor, G. L., R. T. Van, R. Lo, B. Khaira, M. D. Whiteside, R. E. Hancock, and F. S. Brinkman (2009). Pseudomonas genome database: facilitating user-friendly, comprehensive comparisons of microbial genomes. Nucleic. Acids Res., **37**:D483 - 488.

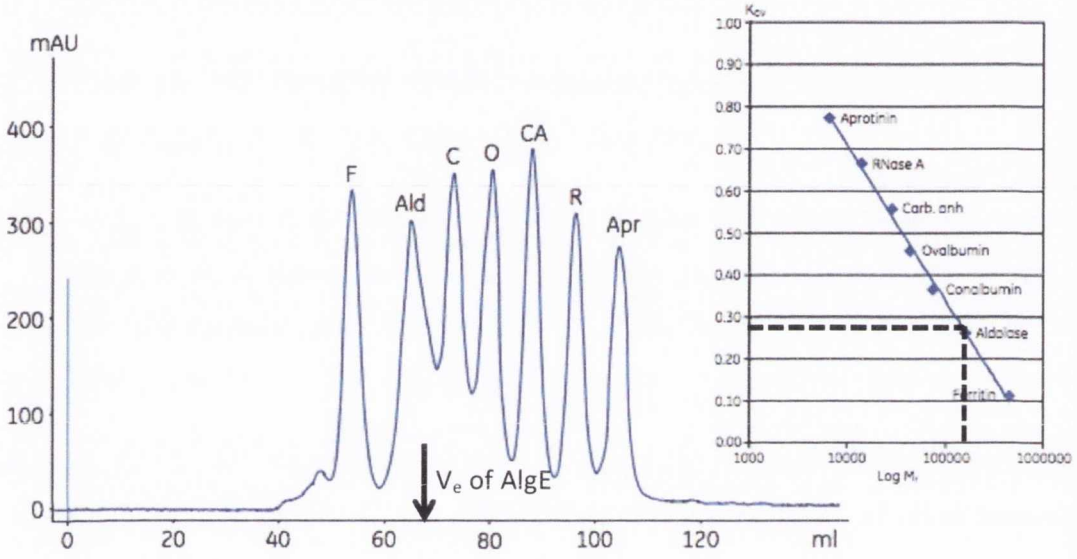
Worlitzsch, D., R. Tarran, and M. Ulrich (2002). Effects of reduced mucus oxygen concentration in airway Pseudomonas infections of cystic fibrosis patients. J. Clin. Invest., **109**:317–325.

Yoshimura, F. and H. Nikaido (1982). Permeability of Pseudomonas aeruginosa outer membrane to hydrophilic solutes. J Bacteriol., **152**:636-642.

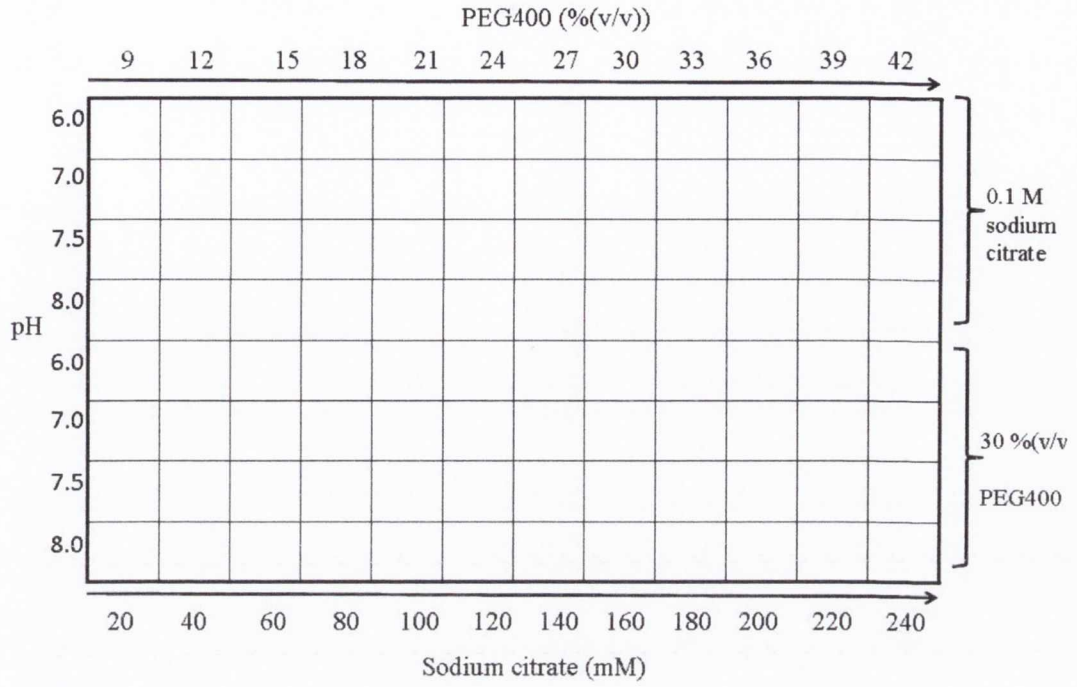
Yue, W. W., S. Grizot, and S. K. Buchanan (2003). Structural evidence for iron-free citrate and ferric citrate binding to the TonB-dependent outer membrane transporter FecA. J. Mol. Biol., **332**:353-368.

Appendix

Appendix 1. Calibration curve of Superdex200 Hiload 16/60 gel filtration column. The figure is adapted from Figure 7 in the online document provided by GE Healthcare: https://www.gelifesciences.com/gehcls_images/GELS/Related%20Content/Files/1314750913712/litdoc28407384AA_20110831032125.pdf

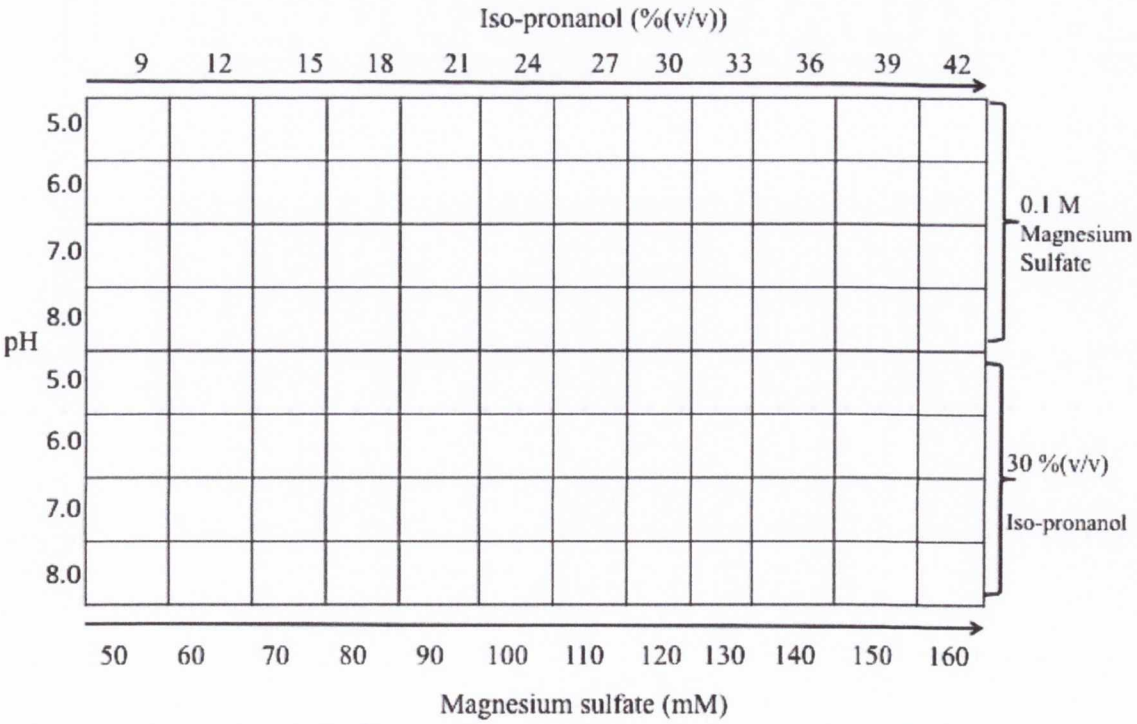


Appendix 2 Composition of AlgE screen kit 1 used for AlgE *in meso* crystallisation



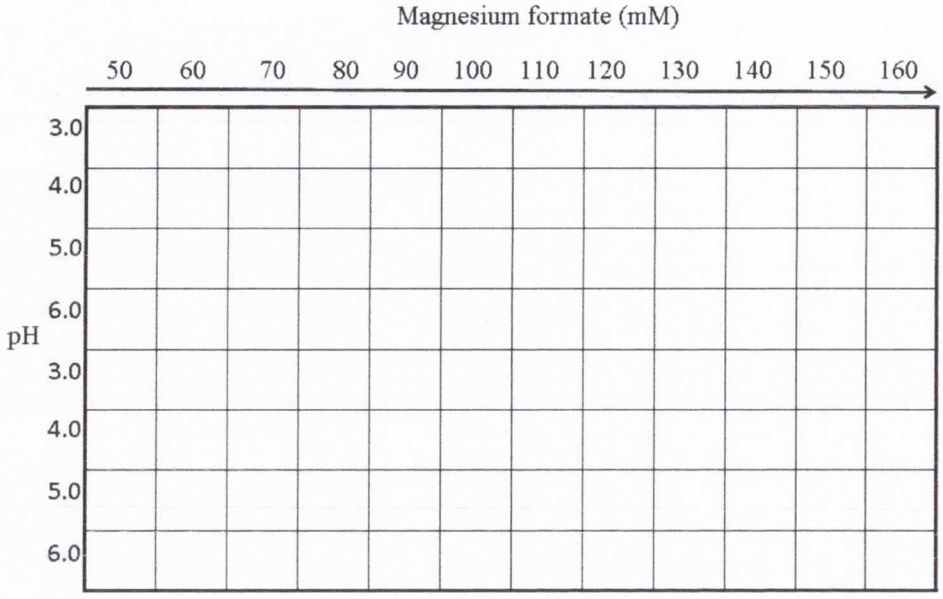
* For pH 6.0 and pH 7.0, Bis-Tris buffer is used. For pH 7.5 and pH 8.0, Tris-HCl buffer is used.

Appendix 3 Composition of AlgE screen kit 2 used for AlgE *in meso* crystallisation



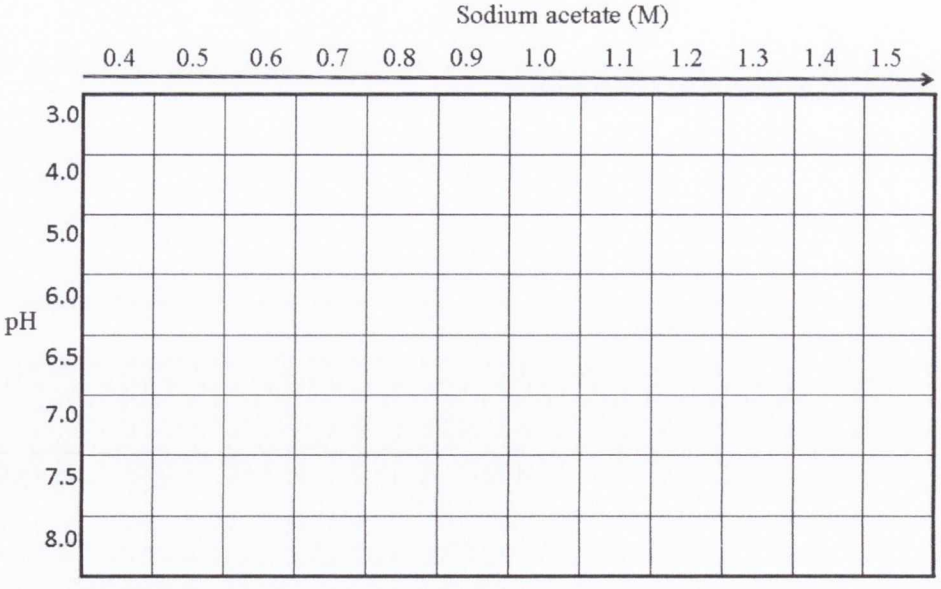
* For pH 5.0 and pH 6.0 Bis-Tris buffer is used. For pH 7.5 and pH 8.0, Tris-HCl buffer is used.

Appendix 4 Composition of AlgE screen kit 3 used for AlgE *in meso* crystallisation



* For pH 3.0 and pH 4.0, sodium citrate buffer is used. For pH 5.0 and pH 6.0, sodium acetate buffer is used.

Appendix 5 Composition of AlgE screen kit 5 used for AlgE *in meso* crystallisation



* For pH 3.0 and pH 4.0, sodium citrate buffer is used. For pH 5.0 and pH 6.0, sodium acetate buffer is used. For pH 6.5, sodium cacodylate buffer is used. For pH 7.0, Bis-Tris buffer is used. For pH 7.5 and pH 8.0, Tris-HCl buffer is used.

Appendix 6 PDB header for AlgE_1.9 (4AFK in PDB database)

```
HEADER    MEMBRANE PROTEIN                        19-JAN-12
4AFK
TITLE     IN MESO STRUCTURE OF ALGINATE TRANSPORTER, ALGE, FROM
TITLE    2 PSEUDOMONAS AERUGINOSA, PAO1
COMPND    MOL_ID: 1;
COMPND    2 MOLECULE: ALGINATE PRODUCTION PROTEIN ALGE;
COMPND    3 CHAIN: A;
COMPND    4 FRAGMENT: RESIDUES 33-490;
COMPND    5 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: PSEUDOMONAS AERUGINOSA;
SOURCE    3 ORGANISM_TAXID: 208964;
SOURCE    4 STRAIN: PAO1;
SOURCE    5 ATCC: 15692;
SOURCE    6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE    7 EXPRESSION_SYSTEM_TAXID: 469008;
SOURCE    8 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE    9 EXPRESSION_SYSTEM_VARIANT: STAR;
SOURCE   10 EXPRESSION_SYSTEM_VECTOR: PET200 D TOPO;
SOURCE   11 EXPRESSION_SYSTEM_PLASMID: ALGE_PET200/D-TOPO;
SOURCE   12 OTHER_DETAILS: HOLLOWAY COLLECTION
KEYWDS    MEMBRANE PROTEIN, OUTER MEMBRANE, ALGINATE TRANS-
PORTER, IN MESO,
KEYWDS    2 LIPIDIC CUBIC PHASE, BETA BARREL, PSEUDOMONAS AERU-
GINOSA
EXPDTA    X-RAY DIFFRACTION
AUTHOR    J.TAN,V.E.PYE,D.ARAGAO,M.CAFFREY
REVDAT    1    20-JAN-12 4AFK      0
JRNL        AUTH    J.TAN,V.E.PYE,D.ARAGAO,M.CAFFREY
JRNL        TITL    IN MESO STRUCTURE OF ALGINATE TRANSPORTER,
ALGE
JRNL        REF     TO BE PUBLISHED
JRNL        REFN
REMARK    2
REMARK    2 RESOLUTION.      1.90 ANGSTROMS.
REMARK    3
REMARK    3 REFINEMENT.
REMARK    3    PROGRAM      : PHENIX (PHENIX.REFINE)
REMARK    3    AUTHORS      : PAUL ADAMS,PAVEL AFONINE,VICENT
CHEN,IAN
REMARK    3                 : DAVIS,KRESHNA GOPAL,RALF GROSSE-
KUNSTLEVE,
REMARK    3                 : LI-WEI HUNG,ROBERT IMMORMINO,TOM
IOERGER,
REMARK    3                 : AIRLIE MCCOY,ERIK MCKEE,NIGEL MORI-
ARTY,
REMARK    3                 : REETAL PAI,RANDY READ,JANE RICHARDSON,
REMARK    3                 : DAVID RICHARDSON,TOD ROMO,JIM SAC-
```

```
CHETTINI,
REMARK   3                    : NICHOLAS SAUTER,JACOB SMITH,LAURENT
REMARK   3                    : STORONI,TOM TERWILLIGER,PETER ZWART
REMARK   3
REMARK   3     REFINEMENT TARGET : ML
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 1.897
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 19.770
REMARK   3   MIN(FOBS/SIGMA_FOBS)                : 0.00
REMARK   3   COMPLETENESS FOR RANGE        (%) : 94.51
REMARK   3   NUMBER OF REFLECTIONS               : 35580
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   R VALUE       (WORKING + TEST SET) : 0.1656
REMARK   3   R VALUE           (WORKING SET) : 0.1633
REMARK   3   FREE R VALUE                    : 0.2086
REMARK   3   FREE R VALUE TEST SET SIZE   (%) : 5.0
REMARK   3   FREE R VALUE TEST SET COUNT     : 1788
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT (IN BINS).
REMARK   3   BIN  RESOLUTION RANGE  COMPL.     NWORK NFREE
RWORK  RFREE
REMARK   3     1 19.7712 -  4.4456    1.00     2814   150  0.1763 0.1942
REMARK   3     2  4.4456 -  3.5352    1.00     2770   155  0.1543 0.1976
REMARK   3     3  3.5352 -  3.0902    1.00     2736   160  0.1601 0.2039
REMARK   3     4  3.0902 -  2.8085    0.99     2734   141  0.1646 0.1917
REMARK   3     5  2.8085 -  2.6077    0.99     2688   143  0.1633 0.2182
REMARK   3     6  2.6077 -  2.4543    0.98     2719   138  0.1547 0.1984
REMARK   3     7  2.4543 -  2.3315    0.97     2674   161  0.1566 0.2297
REMARK   3     8  2.3315 -  2.2302    0.97     2647   137  0.1582 0.2298
REMARK   3     9  2.2302 -  2.1444    0.95     2615   130  0.1533 0.1961
REMARK   3    10  2.1444 -  2.0705    0.94     2591   122  0.1574 0.2280
REMARK   3    11  2.0705 -  2.0059    0.91     2526   124  0.1611 0.2108
REMARK   3    12  2.0059 -  1.9486    0.86     2306   129  0.1739 0.2219
REMARK   3    13  1.9486 -  1.8973    0.71     1972    98  0.2185 0.3114
REMARK   3
REMARK   3  BULK SOLVENT MODELLING.
REMARK   3   METHOD USED          : FLAT BULK SOLVENT MODEL
REMARK   3   SOLVENT RADIUS       : 0.90
REMARK   3   SHRINKAGE RADIUS     : 0.61
REMARK   3   K_SOL                : 0.400
REMARK   3   B_SOL                : 59.825
REMARK   3
REMARK   3  ERROR ESTIMATES.
REMARK   3   COORDINATE ERROR (MAXIMUM-LIKELIHOOD BASED)      :
0.20
REMARK   3   PHASE ERROR (DEGREES, MAXIMUM-LIKELIHOOD BASED) :
20.17
REMARK   3
```

```
REMARK   3  NUMBER OF NON-HYDROGEN ATOMS USED IN REFINEMENT.
REMARK   3   PROTEIN ATOMS                : 3489
REMARK   3   NUCLEIC ACID ATOMS           : 0
REMARK   3   HETEROGEN ATOMS              : 342
REMARK   3   SOLVENT ATOMS                : 242
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT                (A**2) : 15.42
REMARK   3   MEAN B VALUE         (OVERALL, A**2) : 23.505
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) : 1.0359
REMARK   3    B22 (A**2) : 2.2185
REMARK   3    B33 (A**2) : -3.2544
REMARK   3    B12 (A**2) : 0.0000
REMARK   3    B13 (A**2) : -1.7468
REMARK   3    B23 (A**2) : 0.0000
REMARK   3
REMARK   3  TWINNING INFORMATION.
REMARK   3   FRACTION: NULL
REMARK   3   OPERATOR: NULL
REMARK   3
REMARK   3  DEVIATIONS FROM IDEAL VALUES.
REMARK   3                      RMSD              COUNT
REMARK   3   BOND         :   0.010              3898
REMARK   3   ANGLE        :   1.344              5199
REMARK   3   CHIRALITY :   0.099               508
REMARK   3   PLANARITY :   0.005               669
REMARK   3   DIHEDRAL    : 19.379              1523
REMARK   3
REMARK   3  TLS DETAILS
REMARK   3   NUMBER OF TLS GROUPS  : 4
REMARK   3   TLS GROUP : 1
REMARK   3    SELECTION: (CHAIN A AND RESID 37:118)
REMARK   3    ORIGIN FOR THE GROUP (A):    3.6627   -8.1154   32.6708
REMARK   3    T TENSOR
REMARK   3       T11:    0.0106 T22:    0.0250
REMARK   3       T33:    0.0705 T12:   -0.0739
REMARK   3       T13:   -0.0291 T23:    0.0097
REMARK   3    L TENSOR
REMARK   3       L11:    0.2982 L22:    0.3297
REMARK   3       L33:    0.0512 L12:    0.0388
REMARK   3       L13:   -0.0508 L23:    0.0809
REMARK   3    S TENSOR
REMARK   3       S11:    0.1473 S12:    0.0640 S13:   -0.0011
REMARK   3       S21:   -0.0113 S22:   -0.1318 S23:    0.1254
REMARK   3       S31:   -0.1201 S32:    0.0579 S33:   -0.0085
REMARK   3   TLS GROUP : 2
REMARK   3    SELECTION: (CHAIN A AND RESID 119:232)
REMARK   3    ORIGIN FOR THE GROUP (A):   20.1401 -14.2823   33.0461
REMARK   3    T TENSOR
```

```
REMARK   3        T11:    0.0580 T22:    0.0504
REMARK   3        T33:    0.0567 T12:   -0.0023
REMARK   3        T13:   -0.0100 T23:    0.0004
REMARK   3     L TENSOR
REMARK   3        L11:    0.3597 L22:    0.2654
REMARK   3        L33:    0.1383 L12:    0.1055
REMARK   3        L13:   -0.1073 L23:   -0.1465
REMARK   3     S TENSOR
REMARK   3        S11:    0.0236 S12:    0.0117 S13:    0.0035
REMARK   3        S21:    0.0486 S22:   -0.0222 S23:   -0.0050
REMARK   3        S31:   -0.0067 S32:   -0.0020 S33:   -0.0070
REMARK   3    TLS GROUP : 3
REMARK   3     SELECTION: (CHAIN A AND RESID 233:313)
REMARK   3     ORIGIN FOR THE GROUP (A):   24.8815   -6.1799   19.1723
REMARK   3     T TENSOR
REMARK   3        T11:    0.0736 T22:    0.1458
REMARK   3        T33:    0.0708 T12:   -0.0201
REMARK   3        T13:    0.0094 T23:   -0.0165
REMARK   3     L TENSOR
REMARK   3        L11:    0.1973 L22:    0.2544
REMARK   3        L33:    0.1049 L12:    0.0740
REMARK   3        L13:   -0.1401 L23:   -0.1089
REMARK   3     S TENSOR
REMARK   3        S11:    0.0194 S12:    0.1120 S13:   -0.0215
REMARK   3        S21:    0.0175 S22:   -0.0262 S23:   -0.0344
REMARK   3        S31:    0.1142 S32:   -0.1331 S33:   -0.0194
REMARK   3    TLS GROUP : 4
REMARK   3     SELECTION: (CHAIN A AND RESID 314:490)
REMARK   3     ORIGIN FOR THE GROUP (A):   16.5885    7.3036   28.9381
REMARK   3     T TENSOR
REMARK   3        T11:    0.0380 T22:    0.0174
REMARK   3        T33:    0.0244 T12:   -0.0105
REMARK   3        T13:    0.0045 T23:    0.0071
REMARK   3     L TENSOR
REMARK   3        L11:    0.5584 L22:    0.3755
REMARK   3        L33:    0.1618 L12:    0.0253
REMARK   3        L13:   -0.0431 L23:    0.0740
REMARK   3     S TENSOR
REMARK   3        S11:    0.0174 S12:    0.0205 S13:    0.0147
REMARK   3        S21:   -0.0290 S22:   -0.0070 S23:    0.0145
REMARK   3        S31:   -0.0268 S32:   -0.0100 S33:   -0.0082
REMARK   3
REMARK   3   NCS DETAILS
REMARK   3    NUMBER OF NCS GROUPS : NULL
REMARK   3
REMARK   3   OTHER REFINEMENT REMARKS: NULL
REMARK   4
REMARK   4 4AFK COMPLIES WITH FORMAT V. 3.30, 13-JUL-11
REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY PDBE ON 20-JAN-12.
```

REMARK 100 THE PDBE ID CODE IS EBI-50989.
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200 EXPERIMENT TYPE : X-RAY DIFFRACTION
REMARK 200 DATE OF DATA COLLECTION : 22-APR-08
REMARK 200 TEMPERATURE (KELVIN) : 100
REMARK 200 PH : 7.5
REMARK 200 NUMBER OF CRYSTALS USED : 1
REMARK 200
REMARK 200 SYNCHROTRON (Y/N) : Y
REMARK 200 RADIATION SOURCE : APS
REMARK 200 BEAMLINE : 23-ID-B
REMARK 200 X-RAY GENERATOR MODEL : NULL
REMARK 200 MONOCHROMATIC OR LAUE (M/L) : M
REMARK 200 WAVELENGTH OR RANGE (A) : 0.97934
REMARK 200 MONOCHROMATOR : DOUBLE CRYSTAL
CRYO-COOLED
REMARK 200 OPTICS : NULL
REMARK 200
REMARK 200 DETECTOR TYPE : CCD
REMARK 200 DETECTOR MANUFACTURER : MARRESEARCH
REMARK 200 INTENSITY-INTEGRATION SOFTWARE : HKL2000
REMARK 200 DATA SCALING SOFTWARE : SCALEPACK
REMARK 200
REMARK 200 NUMBER OF UNIQUE REFLECTIONS : 37703
REMARK 200 RESOLUTION RANGE HIGH (A) : 1.90
REMARK 200 RESOLUTION RANGE LOW (A) : 113.18
REMARK 200 REJECTION CRITERIA (SIGMA(I)) : -3.0
REMARK 200
REMARK 200 OVERALL.
REMARK 200 COMPLETENESS FOR RANGE (%) : 100.0
REMARK 200 DATA REDUNDANCY : 6.4
REMARK 200 R MERGE (I) : 0.14
REMARK 200 R SYM (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR THE DATA SET : 15.7
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) : 1.90
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE LOW (A) : 1.97
REMARK 200 COMPLETENESS FOR SHELL (%) : 84.9
REMARK 200 DATA REDUNDANCY IN SHELL : 2.5
REMARK 200 R MERGE FOR SHELL (I) : 0.35
REMARK 200 R SYM FOR SHELL (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR SHELL : 2.9
REMARK 200
REMARK 200 DIFFRACTION PROTOCOL: SINGLE WAVELENGTH
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE: MOLECULAR
REPLACEMENT
REMARK 200 SOFTWARE USED: PHASER
REMARK 200 STARTING MODEL: PDB ENTRY 3RBH

REMARK 200
REMARK 200 REMARK: NONE
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS (%): 47.96
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 2.36
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: 0.1 M TRIS-HCL, PH 7.5, 0.1 M
REMARK 280 SODIUM CITRATE, 18 %(W/V) PEG 400. PROTEIN WAS MIXED
WITH
REMARK 280 7.8 MAG AT RATIO 1:1 (W:W) TO FORM CUBIC PHASE.
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: C 1 2 1
REMARK 290
REMARK 290 SYMOP SYMMETRY
REMARK 290 NNNMMM OPERATOR
REMARK 290 1555 X,Y,Z
REMARK 290 2555 -X,Y,-Z
REMARK 290 3555 X+1/2,Y+1/2,Z
REMARK 290 4555 -X+1/2,Y+1/2,-Z
REMARK 290
REMARK 290 WHERE NNN -> OPERATOR NUMBER
REMARK 290 MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE AT-
OM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPH-
ICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290 SMTRY1 1 1.000000 0.000000 0.000000 0.00000
REMARK 290 SMTRY2 1 0.000000 1.000000 0.000000 0.00000
REMARK 290 SMTRY3 1 0.000000 0.000000 1.000000 0.00000
REMARK 290 SMTRY1 2 -1.000000 0.000000 0.000000 0.00000
REMARK 290 SMTRY2 2 0.000000 1.000000 0.000000 0.00000
REMARK 290 SMTRY3 2 0.000000 0.000000 -1.000000 0.00000
REMARK 290 SMTRY1 3 1.000000 0.000000 0.000000 28.62650
REMARK 290 SMTRY2 3 0.000000 1.000000 0.000000 37.20800
REMARK 290 SMTRY3 3 0.000000 0.000000 1.000000 0.00000
REMARK 290 SMTRY1 4 -1.000000 0.000000 0.000000 28.62650
REMARK 290 SMTRY2 4 0.000000 1.000000 0.000000 37.20800
REMARK 290 SMTRY3 4 0.000000 0.000000 -1.000000 0.00000
REMARK 290
REMARK 290 REMARK: NULL
REMARK 300
REMARK 300 BIOMOLECULE: 1
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PRO-
GRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE

IN
REMARK 300 THIS ENTRY. THE REMARK MAY ALSO PROVIDE INFORMATION
ON
REMARK 300 BURIED SURFACE AREA.
REMARK 350
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING
THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF
THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANS-
FORMATIONS
REMARK 350 GIVEN BELOW. BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: MONO-
MERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.00000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.00000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.00000
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME;
C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465 M RES C SSSEQI
REMARK 465 ALA A 33
REMARK 465 ASN A 34
REMARK 465 SER A 35
REMARK 465 GLY A 36
REMARK 465 ASP A 109
REMARK 465 ASP A 110
REMARK 465 GLY A 111
REMARK 465 ASN A 112
REMARK 465 ASN A 113
REMARK 465 SER A 114
REMARK 465 ARG A 115
REMARK 465 ASN A 116
REMARK 465 LYS A 439
REMARK 465 GLN A 440
REMARK 465 GLY A 441
REMARK 465 LEU A 442
REMARK 465 LEU A 443
REMARK 465 PRO A 444
REMARK 465 ALA A 445

REMARK 465 SER A 446
REMARK 465 MET A 447
REMARK 465 SER A 448
REMARK 465 GLN A 449
REMARK 465 TYR A 450
REMARK 465 VAL A 451
REMARK 465 ASP A 452
REMARK 465 GLU A 453
REMARK 465 PRO A 454
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CLOSE CONTACTS
REMARK 500
REMARK 500 THE FOLLOWING ATOMS THAT ARE RELATED BY CRYSTALLOGRAPHIC
REMARK 500 SYMMETRY ARE IN CLOSE CONTACT. AN ATOM LOCATED WITHIN 0.15
REMARK 500 ANGSTROMS OF A SYMMETRY RELATED ATOM IS ASSUMED TO BE ON A
REMARK 500 SPECIAL POSITION AND IS, THEREFORE, LISTED IN REMARK 375
REMARK 500 INSTEAD OF REMARK 500. ATOMS WITH NON-BLANK ALTERNATE
REMARK 500 LOCATION INDICATORS ARE NOT INCLUDED IN THE CALCULATIONS.
REMARK 500
REMARK 500 DISTANCE CUTOFF:
REMARK 500 2.2 ANGSTROMS FOR CONTACTS NOT INVOLVING HYDROGEN ATOMS
REMARK 500 1.6 ANGSTROMS FOR CONTACTS INVOLVING HYDROGEN ATOMS
REMARK 500
REMARK 500 ATM1 RES C SSEQI ATM2 RES C SSEQI SSYMOP DISTANCE
REMARK 500 NE2 GLN A 303 NE2 GLN A 303 2655 2.17
REMARK 500
REMARK 500 REMARK: NULL
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: TORSION ANGLES
REMARK 500
REMARK 500 TORSION ANGLES OUTSIDE THE EXPECTED RAMACHANDRAN REGIONS:
REMARK 500 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 500 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT:(10X,I3,1X,A3,1X,A1,I4,A1,4X,F7.2,3X,F7.2)
REMARK 500
REMARK 500 EXPECTED VALUES: GJ KLEYWEGT AND TA JONES (1996).

PHI/PSI-
REMARK 500 CHOLOGY: RAMACHANDRAN REVISITED. STRUCTURE 4, 1395 - 1400
REMARK 500
REMARK 500 M RES CSSEQI PSI PHI
REMARK 500 ASN A 54 110.38 -170.29
REMARK 500 ASN A 67 84.43 -154.44
REMARK 500 ASP A 162 80.97 -162.91
REMARK 500 ASP A 321 -157.85 -150.75
REMARK 500 ASN A 355 39.81 -85.24
REMARK 500 ARG A 389 -129.31 45.50
REMARK 500 ASN A 416 78.44 -109.00
REMARK 500
REMARK 500 REMARK: NULL
REMARK 600
REMARK 600 HETEROGEN
REMARK 600 3,6,9,12,15,18,21,24-OCTAOXAHEXACOSAN-1-OL (PE5): PEG 400
REMARK 600 LAURYL DIMETHYLAMINE-N-OXIDE (LDA): SOME LDA MOLECULES
REMARK 600 HAVE NOT HAD HEAD GROUPS FITTED AS NO ELECTRON DENSITY,
REMARK 600 ALKANE CHAINS WERE MODELLED WHERE WE WERE UNABLE TO
REMARK 600 DISTINGUISH BETWEEN LADO AND 7.8 MAG.
REMARK 600 INCLUSION OF THE TAILS FIT INTO DENSITY AND REDUCES
REMARK 600 FREE R FACTOR BY 1.2 PERCENT.
REMARK 610
REMARK 610 MISSING HETEROATOM
REMARK 610 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 610 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 610 I=INSERTION CODE):
REMARK 610 M RES C SSEQI
REMARK 610 LDA A 1495
REMARK 610 LDA A 1497
REMARK 610 LDA A 1498
REMARK 610 LDA A 1499
REMARK 610 LDA A 1500
REMARK 610 LDA A 1501
REMARK 610 LDA A 1502
REMARK 610 LDA A 1503
REMARK 610 LDA A 1504
REMARK 610 LDA A 1505
REMARK 610 LDA A 1506
REMARK 610 PE5 A 1516
REMARK 610 PE5 A 1517
REMARK 620
REMARK 620 METAL COORDINATION
REMARK 620 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFI-

ER;
REMARK 620 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 NA A1507 NA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ALA A 136 O
REMARK 620 2 GLU A 144 O 101.7
REMARK 620 3 GLU A 144 OE1 97.8 77.8
REMARK 620 4 THR A 139 O 78.4 157.3 79.7
REMARK 620 5 TYR A 141 O 153.3 89.4 108.4 100.6
REMARK 620 6 HOH A2064 O 74.7 112.2 168.3 89.9 78.6
REMARK 620 N 1 2 3 4 5
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 NA A1508 NA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 LYS A 325 NZ
REMARK 620 2 ARG A 317 NH2 153.7
REMARK 620 3 ASN A 273 OD1 148.5 57.8
REMARK 620 4 ARG A 317 NE 124.9 44.2 76.5
REMARK 620 5 ASN A 319 OD1 81.2 116.4 77.5 85.8
REMARK 620 N 1 2 3 4
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CA A1509 CA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 GLY A 63 O
REMARK 620 2 ALA A 61 O 112.7
REMARK 620 3 ASP A 57 OD2 105.6 90.5
REMARK 620 4 ASP A 55 OD1 95.3 151.0 74.2
REMARK 620 5 ASP A 55 OD2 107.2 125.6 112.8 46.1
REMARK 620 6 ASP A 57 OD1 147.7 83.9 44.5 67.9 81.4
REMARK 620 N 1 2 3 4 5
REMARK 700
REMARK 700 SHEET
REMARK 700 DETERMINATION METHOD: DSSP
REMARK 700 THE SHEETS PRESENTED AS "AA" IN EACH CHAIN ON SHEET RECORDS
REMARK 700 BELOW IS ACTUALLY AN 18-STRANDED BARREL THIS IS REP-RESENTED BY
REMARK 700 A 19-STRANDED SHEET IN WHICH THE FIRST AND LAST STRANDS
REMARK 700 ARE IDENTICAL.
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1510
REMARK 800

REMARK 800 SITE_IDENTIFIER: AC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1511
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1512
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1513
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1514
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1515
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1492
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1493
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC9
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1494
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1495
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1496
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1497
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1498
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC5
REMARK 800 EVIDENCE_CODE: SOFTWARE

REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1499
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1500
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1501
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1502
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC9
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1503
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1505
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1506
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE FLC A1491
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE NA A1507
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE NA A1508
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CA A1509
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE PE5 A1516
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE PE5 A1517
DBREF 4AFK A 33 490 UNP P18895 ALGE_PSEAE 33 490

Appendix 7 PDB header for AlgE_2.4 (4B61 in PDB database)

```
HEADER      MEMBRANE PROTEIN                                08-AUG-12
4B61
TITLE       IN MESO STRUCTURE OF ALGINATE TRANSPORTER, ALGE, FROM
PSEUDOMOAS
TITLE     2 AERUGINOSA, PAO1. CRYSTAL FORM 3.
COMPND      MOL_ID: 1;
COMPND    2 MOLECULE: ALGINATE PRODUCTION PROTEIN ALGE;
COMPND    3 CHAIN: A, B;
COMPND    4 FRAGMENT: RESIDUES 33-490;
COMPND    5 SYNONYM: ALGE;
COMPND    6 ENGINEERED: YES
SOURCE      MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: PSEUDOMONAS AERUGINOSA PAO1;
SOURCE    3 ORGANISM_TAXID: 208964;
SOURCE    4 ATCC: 15692;
SOURCE    5 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE    6 EXPRESSION_SYSTEM_TAXID: 469008;
SOURCE    7 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE    8 EXPRESSION_SYSTEM_VARIANT: STAR;
SOURCE    9 EXPRESSION_SYSTEM_VECTOR: PET200-D-TOPO;
SOURCE   10 EXPRESSION_SYSTEM_PLASMID: ALGE-PET200/D-TOPO;
SOURCE   11 OTHER_DETAILS: HOLLOWAY COLLECTION
KEYWDS      MEMBRANE PROTEIN, OUTER MEMBRANE, IN MESO CRYSTAL-
LISATION, LIPIDIC
KEYWDS    2 CUBIC PHASE
EXPDTA      X-RAY DIFFRACTION
AUTHOR      J.TAN,V.E.PYE,D.ARAGAO,M.CAFFREY
REVDAT    1   22-AUG-12 4B61      0
JRNL          AUTH    J.TAN,V.E.PYE,D.ARAGAO,M.CAFFREY
JRNL          TITL    IN MESO STRUCTURE OF ALGINATE TRANSPORTER,
ALGE
JRNL          REF     TO BE PUBLISHED
JRNL          REFN
REMARK    2
REMARK    2 RESOLUTION.      2.40 ANGSTROMS.
REMARK    3
REMARK    3 REFINEMENT.
REMARK    3    PROGRAM       : PHENIX (PHENIX.REFINE)
REMARK    3    AUTHORS       : PAUL ADAMS,PAVEL AFONINE,VICENT
CHEN,IAN
REMARK    3                  : DAVIS,KRESHNA GOPAL,RALF GROSSE-
KUNSTLEVE,
REMARK    3                  : LI-WEI HUNG,ROBERT IMMORMINO,TOM
IOERGER,
REMARK    3                  : AIRLIE MCCOY,ERIK MCKEE,NIGEL MORI-
ARTY,
REMARK    3                  : REETAL PAI,RANDY READ,JANE RICHARDSON,
REMARK    3                  : DAVID RICHARDSON,TOD ROMO,JIM SAC-
```

```
CHETTINI,
REMARK   3                    : NICHOLAS SAUTER,JACOB SMITH,LAURENT
REMARK   3                    : STORONI,TOM TERWILLIGER,PETER ZWART
REMARK   3
REMARK   3     REFINEMENT TARGET : ML
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.402
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 48.865
REMARK   3   MIN(FOBS/SIGMA_FOBS)              : 1.35
REMARK   3   COMPLETENESS FOR RANGE        (%) : 99.25
REMARK   3   NUMBER OF REFLECTIONS             : 45596
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   R VALUE       (WORKING + TEST SET) : 0.2217
REMARK   3   R VALUE           (WORKING SET) : 0.2205
REMARK   3   FREE R VALUE                    : 0.2436
REMARK   3   FREE R VALUE TEST SET SIZE   (%) : 5.0
REMARK   3   FREE R VALUE TEST SET COUNT     : 2299
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT (IN BINS).
REMARK   3   BIN  RESOLUTION RANGE  COMPL.     NWORK NFREE
RWORK  RFREE
REMARK   3     1 48.8754 -  6.0491    0.96      2844   149   0.2393 0.2570
REMARK   3     2  6.0491 -  4.8027    0.99      2753   151   0.2067 0.2056
REMARK   3     3  4.8027 -  4.1960    0.99      2724   143   0.1826 0.1963
REMARK   3     4  4.1960 -  3.8125    0.99      2694   145   0.2056 0.2303
REMARK   3     5  3.8125 -  3.5393    0.99      2707   147   0.1937 0.2048
REMARK   3     6  3.5393 -  3.3307    0.99      2717   142   0.1946 0.2222
REMARK   3     7  3.3307 -  3.1639    0.99      2694   150   0.2095 0.2309
REMARK   3     8  3.1639 -  3.0262    0.99      2686   135   0.2221 0.2787
REMARK   3     9  3.0262 -  2.9097    1.00      2691   139   0.2406 0.2649
REMARK   3    10  2.9097 -  2.8094    1.00      2687   144   0.2362 0.2899
REMARK   3    11  2.8094 -  2.7215    1.00      2712   146   0.2435 0.2819
REMARK   3    12  2.7215 -  2.6437    1.00      2664   142   0.2582 0.3137
REMARK   3    13  2.6437 -  2.5741    1.00      2655   151   0.2641 0.2987
REMARK   3    14  2.5741 -  2.5113    1.00      2714   125   0.2717 0.3181
REMARK   3    15  2.5113 -  2.4542    1.00      2688   135   0.2885 0.3270
REMARK   3    16  2.4542 -  2.4020    1.00      2667   155   0.3062 0.3410
REMARK   3
REMARK   3  BULK SOLVENT MODELLING.
REMARK   3   METHOD USED          : FLAT BULK SOLVENT MODEL
REMARK   3   SOLVENT RADIUS       : 1.00
REMARK   3   SHRINKAGE RADIUS     : 0.60
REMARK   3   K_SOL                : 0.351
REMARK   3   B_SOL                : 44.62
REMARK   3
REMARK   3  ERROR ESTIMATES.
REMARK   3   COORDINATE ERROR (MAXIMUM-LIKELIHOOD BASED)      :
0.30
```

```
REMARK   3   PHASE ERROR (DEGREES, MAXIMUM-LIKELIHOOD BASED) :
24.56
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT                (A**2) : 35.72
REMARK   3   MEAN B VALUE         (OVERALL, A**2) : 40.72
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) : 11.17
REMARK   3    B22 (A**2) : -5
REMARK   3    B33 (A**2) : -6
REMARK   3    B12 (A**2) : 0
REMARK   3    B13 (A**2) : 0
REMARK   3    B23 (A**2) : 0
REMARK   3
REMARK   3  TWINNING INFORMATION.
REMARK   3   FRACTION: NULL
REMARK   3   OPERATOR: NULL
REMARK   3
REMARK   3  DEVIATIONS FROM IDEAL VALUES.
REMARK   3                      RMSD              COUNT
REMARK   3   BOND         :   0.019              7479
REMARK   3   ANGLE        :   1.109             10050
REMARK   3   CHIRALITY :   0.081               999
REMARK   3   PLANARITY :   0.004               1322
REMARK   3   DIHEDRAL    : 15.298              2769
REMARK   3
REMARK   3  TLS DETAILS
REMARK   3   NUMBER OF TLS GROUPS    : 8
REMARK   3   TLS GROUP : 1
REMARK   3    SELECTION: (CHAIN A AND RESID 39:125)
REMARK   3    ORIGIN FOR THE GROUP (A): -49.4241 -22.5869 -35.1300
REMARK   3    T TENSOR
REMARK   3       T11:     0.4117 T22:     0.4032
REMARK   3       T33:     0.3110 T12:    -0.1246
REMARK   3       T13:     0.1207 T23:     0.0115
REMARK   3    L TENSOR
REMARK   3       L11:     1.8710 L22:     1.5203
REMARK   3       L33:     0.8283 L12:    -0.9552
REMARK   3       L13:     1.2124 L23:    -0.8409
REMARK   3    S TENSOR
REMARK   3       S11:    -0.0776 S12:    -0.0059 S13:    -0.1791
REMARK   3       S21:     0.5478 S22:    -0.0227 S23:     0.1363
REMARK   3       S31:    -0.2440 S32:    -0.0306 S33:     0.1039
REMARK   3   TLS GROUP : 2
REMARK   3    SELECTION: (CHAIN A AND RESID 126:236)
REMARK   3    ORIGIN FOR THE GROUP (A): -42.1951    -4.4699 -38.6512
REMARK   3    T TENSOR
REMARK   3       T11:     0.3370 T22:     0.2211
REMARK   3       T33:     0.1613 T12:     0.0241
REMARK   3       T13:     0.0272 T23:     0.0016
```

```
REMARK   3     L TENSOR
REMARK   3       L11:     2.0982 L22:     1.4721
REMARK   3       L33:     0.1782 L12:     0.4247
REMARK   3       L13:     0.2673 L23:     0.2484
REMARK   3     S TENSOR
REMARK   3       S11:     0.0793 S12:     0.0575 S13:    -0.0023
REMARK   3       S21:     0.3764 S22:    -0.1114 S23:     0.1499
REMARK   3       S31:    -0.2967 S32:    -0.2692 S33:     0.1191
REMARK   3    TLS GROUP : 3
REMARK   3     SELECTION: (CHAIN A AND RESID 237:354)
REMARK   3     ORIGIN FOR THE GROUP (A): -31.1113    -8.6569 -28.4093
REMARK   3     T TENSOR
REMARK   3       T11:     0.3859 T22:     0.1618
REMARK   3       T33:     0.1734 T12:    -0.0896
REMARK   3       T13:    -0.0465 T23:     0.0423
REMARK   3     L TENSOR
REMARK   3       L11:     1.8007 L22:     1.5425
REMARK   3       L33:     1.2949 L12:    -0.5328
REMARK   3       L13:    -0.6194 L23:    -0.1248
REMARK   3     S TENSOR
REMARK   3       S11:     0.0397 S12:    -0.2600 S13:    -0.0649
REMARK   3       S21:     0.4062 S22:    -0.1281 S23:     0.0156
REMARK   3       S31:    -0.3090 S32:    -0.1534 S33:     0.0247
REMARK   3    TLS GROUP : 4
REMARK   3     SELECTION: (CHAIN A AND RESID 355:490)
REMARK   3     ORIGIN FOR THE GROUP (A): -30.1369 -25.1415 -33.5328
REMARK   3     T TENSOR
REMARK   3       T11:     0.4063 T22:     0.1011
REMARK   3       T33:     0.2045 T12:    -0.0501
REMARK   3       T13:    -0.0704 T23:     0.0990
REMARK   3     L TENSOR
REMARK   3       L11:     1.8835 L22:     2.1292
REMARK   3       L33:     1.7462 L12:     0.4148
REMARK   3       L13:    -0.6332 L23:    -0.1551
REMARK   3     S TENSOR
REMARK   3       S11:     0.1321 S12:    -0.2252 S13:    -0.1703
REMARK   3       S21:     0.5737 S22:    -0.2899 S23:    -0.1151
REMARK   3       S31:     0.0649 S32:    -0.0712 S33:    -0.0398
REMARK   3    TLS GROUP : 5
REMARK   3     SELECTION: (CHAIN B AND RESID 38:117)
REMARK   3     ORIGIN FOR THE GROUP (A): -19.9343    13.0478 -34.7253
REMARK   3     T TENSOR
REMARK   3       T11:     0.2845 T22:     0.1871
REMARK   3       T33:     0.1850 T12:    -0.0680
REMARK   3       T13:    -0.0646 T23:     0.0162
REMARK   3     L TENSOR
REMARK   3       L11:     0.8869 L22:     1.9232
REMARK   3       L33:     0.3222 L12:    -0.0054
REMARK   3       L13:    -0.4411 L23:     0.0531
REMARK   3     S TENSOR
```

```
REMARK   3      S11:    0.0372 S12:   -0.1015 S13:   -0.0292
REMARK   3      S21:    0.2723 S22:   -0.0642 S23:   -0.1632
REMARK   3      S31:   -0.0603 S32:    0.2481 S33:    0.1410
REMARK   3    TLS GROUP : 6
REMARK   3     SELECTION: (CHAIN B AND RESID 118:249)
REMARK   3     ORIGIN FOR THE GROUP (A): -16.9183   30.7374 -33.5954
REMARK   3     T TENSOR
REMARK   3       T11:    0.3479 T22:    0.1345
REMARK   3       T33:    0.1310 T12:   -0.0425
REMARK   3       T13:   -0.0601 T23:   -0.0383
REMARK   3     L TENSOR
REMARK   3       L11:    1.5273 L22:    1.7954
REMARK   3       L33:    0.3950 L12:    0.4262
REMARK   3       L13:    0.7528 L23:    0.4626
REMARK   3     S TENSOR
REMARK   3       S11:    0.0887 S12:   -0.1241 S13:    0.0837
REMARK   3       S21:    0.4994 S22:   -0.1626 S23:   -0.1805
REMARK   3       S31:   -0.0778 S32:    0.0753 S33:    0.0327
REMARK   3    TLS GROUP : 7
REMARK   3     SELECTION: (CHAIN B AND RESID 250:358)
REMARK   3     ORIGIN FOR THE GROUP (A):   -1.3343   29.3209 -31.5512
REMARK   3     T TENSOR
REMARK   3       T11:    0.2214 T22:    0.2616
REMARK   3       T33:    0.2621 T12:   -0.1469
REMARK   3       T13:   -0.1423 T23:    0.0227
REMARK   3     L TENSOR
REMARK   3       L11:    2.5221 L22:    2.1416
REMARK   3       L33:    0.9769 L12:   -0.3526
REMARK   3       L13:   -1.2075 L23:   -0.0481
REMARK   3     S TENSOR
REMARK   3       S11:    0.0982 S12:   -0.0713 S13:    0.0703
REMARK   3       S21:    0.4037 S22:   -0.1432 S23:   -0.4429
REMARK   3       S31:   -0.0825 S32:    0.1787 S33:    0.0639
REMARK   3    TLS GROUP : 8
REMARK   3     SELECTION: (CHAIN B AND RESID 359:490)
REMARK   3     ORIGIN FOR THE GROUP (A):   -2.5059   13.9566 -34.7518
REMARK   3     T TENSOR
REMARK   3       T11:    0.2867 T22:    0.2699
REMARK   3       T33:    0.3578 T12:    0.0223
REMARK   3       T13:   -0.1045 T23:    0.0092
REMARK   3     L TENSOR
REMARK   3       L11:    2.4855 L22:    2.8971
REMARK   3       L33:    0.6751 L12:    1.5190
REMARK   3       L13:    0.1116 L23:   -0.1562
REMARK   3     S TENSOR
REMARK   3       S11:    0.0248 S12:   -0.0061 S13:   -0.1173
REMARK   3       S21:    0.2733 S22:   -0.1498 S23:   -0.4513
REMARK   3       S31:    0.0326 S32:    0.3049 S33:    0.1124
REMARK   3
REMARK   3   NCS DETAILS
```

REMARK 3 NUMBER OF NCS GROUPS : NULL
REMARK 3
REMARK 3 OTHER REFINEMENT REMARKS: NULL
REMARK 4
REMARK 4 4B61 COMPLIES WITH FORMAT V. 3.30, 13-JUL-11
REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY PDBE ON 09-AUG-12.
REMARK 100 THE PDBE ID CODE IS EBI-53545.
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200 EXPERIMENT TYPE : X-RAY DIFFRACTION
REMARK 200 DATE OF DATA COLLECTION : 17-JUN-09
REMARK 200 TEMPERATURE (KELVIN) : 293
REMARK 200 PH : 6.5
REMARK 200 NUMBER OF CRYSTALS USED : 1
REMARK 200
REMARK 200 SYNCHROTRON (Y/N) : Y
REMARK 200 RADIATION SOURCE : APS
REMARK 200 BEAMLINE : 23-ID-B
REMARK 200 X-RAY GENERATOR MODEL : NULL
REMARK 200 MONOCHROMATIC OR LAUE (M/L) : M
REMARK 200 WAVELENGTH OR RANGE (A) : 0.97934
REMARK 200 MONOCHROMATOR : DOUBLE CRYSTAL
CRYO-COOLED
REMARK 200 OPTICS : NULL
REMARK 200
REMARK 200 DETECTOR TYPE : CCD
REMARK 200 DETECTOR MANUFACTURER : MARRESEARCH
REMARK 200 INTENSITY-INTEGRATION SOFTWARE : MOSFLM
REMARK 200 DATA SCALING SOFTWARE : SCALA
REMARK 200
REMARK 200 NUMBER OF UNIQUE REFLECTIONS : 45680
REMARK 200 RESOLUTION RANGE HIGH (A) : 2.40
REMARK 200 RESOLUTION RANGE LOW (A) : 55.64
REMARK 200 REJECTION CRITERIA (SIGMA(I)) : 1.8
REMARK 200
REMARK 200 OVERALL.
REMARK 200 COMPLETENESS FOR RANGE (%) : 98.6
REMARK 200 DATA REDUNDANCY : 9.3
REMARK 200 R MERGE (I) : 0.09
REMARK 200 R SYM (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR THE DATA SET : 6.90
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) : 2.40
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE LOW (A) : 2.53
REMARK 200 COMPLETENESS FOR SHELL (%) : 99.9
REMARK 200 DATA REDUNDANCY IN SHELL : 9.4
REMARK 200 R MERGE FOR SHELL (I) : 0.55
REMARK 200 R SYM FOR SHELL (I) : NULL

REMARK 200 <I/SIGMA(I)> FOR SHELL : 1.80
REMARK 200
REMARK 200 DIFFRACTION PROTOCOL: SINGLE WAVELENGTH
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE: MOLECULAR REPLACEMENT
REMARK 200 SOFTWARE USED: PHASER
REMARK 200 STARTING MODEL: PDB ENTRY 3RBH
REMARK 200
REMARK 200 REMARK: NONE
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS (%): 56.26
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 2.81
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: 0.1 M SODIUM CACODYLATE, PH 6.5,
REMARK 280 1.1 M SODIUM ACETATE IN SOLUTION; PROTEIN RECONSTITUTED
REMARK 280 IN LIPIDIC CUBIC PHASE USING 7.8 MAG AS THE HOSTING LIPID.
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 21 21 21
REMARK 290
REMARK 290 SYMOP SYMMETRY
REMARK 290 NNNMMM OPERATOR
REMARK 290 1555 X,Y,Z
REMARK 290 2555 -X+1/2,-Y,Z+1/2
REMARK 290 3555 -X,Y+1/2,-Z+1/2
REMARK 290 4555 X+1/2,-Y+1/2,-Z
REMARK 290
REMARK 290 WHERE NNN -> OPERATOR NUMBER
REMARK 290 MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290 SMTRY1 1 1.000000 0.000000 0.000000 0.00000
REMARK 290 SMTRY2 1 0.000000 1.000000 0.000000 0.00000
REMARK 290 SMTRY3 1 0.000000 0.000000 1.000000 0.00000
REMARK 290 SMTRY1 2 -1.000000 0.000000 0.000000 30.83600
REMARK 290 SMTRY2 2 0.000000 -1.000000 0.000000 0.00000
REMARK 290 SMTRY3 2 0.000000 0.000000 1.000000 120.14500
REMARK 290 SMTRY1 3 -1.000000 0.000000 0.000000 0.00000
REMARK 290 SMTRY2 3 0.000000 1.000000 0.000000 38.73150
REMARK 290 SMTRY3 3 0.000000 0.000000 -1.000000 120.14500
REMARK 290 SMTRY1 4 1.000000 0.000000 0.000000 30.83600

```
REMARK 290     SMTRY2    4  0.000000 -1.000000  0.000000        38.73150
REMARK 290     SMTRY3    4  0.000000  0.000000 -1.000000         0.00000
REMARK 290
REMARK 290 REMARK: NULL
REMARK 300
REMARK 300 BIOMOLECULE: 1, 2
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PRO-
GRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE
IN
REMARK 300 THIS ENTRY. THE REMARK MAY ALSO PROVIDE INFORMATION
ON
REMARK 300 BURIED SURFACE AREA.
REMARK 350
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING
THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF
THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANS-
FORMATIONS
REMARK 350 GIVEN BELOW.  BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: MONO-
MERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A
REMARK 350     BIOMT1    1  1.000000  0.000000  0.000000         0.00000
REMARK 350     BIOMT2    1  0.000000  1.000000  0.000000         0.00000
REMARK 350     BIOMT3    1  0.000000  0.000000  1.000000         0.00000
REMARK 350
REMARK 350 BIOMOLECULE: 2
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: MONO-
MERIC
REMARK 350 APPLY THE FOLLOWING TO CHAINS: B
REMARK 350     BIOMT1    1  1.000000  0.000000  0.000000         0.00000
REMARK 350     BIOMT2    1  0.000000  1.000000  0.000000         0.00000
REMARK 350     BIOMT3    1  0.000000  0.000000  1.000000         0.00000
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME;
C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)
REMARK 465
REMARK 465     M RES C SSSEQI
REMARK 465       ALA A      33
```

REMARK 465 ASN A 34
REMARK 465 SER A 35
REMARK 465 GLY A 36
REMARK 465 GLU A 37
REMARK 465 ALA A 38
REMARK 465 SER A 106
REMARK 465 ASP A 107
REMARK 465 THR A 108
REMARK 465 ASP A 109
REMARK 465 ASP A 110
REMARK 465 GLY A 111
REMARK 465 ASN A 112
REMARK 465 ASN A 113
REMARK 465 SER A 114
REMARK 465 ARG A 115
REMARK 465 GLY A 441
REMARK 465 LEU A 442
REMARK 465 LEU A 443
REMARK 465 PRO A 444
REMARK 465 ALA A 445
REMARK 465 SER A 446
REMARK 465 MET A 447
REMARK 465 SER A 448
REMARK 465 GLN A 449
REMARK 465 TYR A 450
REMARK 465 VAL A 451
REMARK 465 ASP A 452
REMARK 465 GLU A 453
REMARK 465 ALA B 33
REMARK 465 ASN B 34
REMARK 465 SER B 35
REMARK 465 GLY B 36
REMARK 465 GLU B 37
REMARK 465 SER B 106
REMARK 465 ASP B 107
REMARK 465 THR B 108
REMARK 465 ASP B 109
REMARK 465 ASP B 110
REMARK 465 GLY B 111
REMARK 465 ASN B 112
REMARK 465 ASN B 113
REMARK 465 SER B 114
REMARK 465 ARG B 115
REMARK 465 ASN B 116
REMARK 465 GLY B 472
REMARK 465 PRO B 473
REMARK 465 GLY B 474
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: CLOSE CONTACTS

REMARK 500
REMARK 500 THE FOLLOWING ATOMS THAT ARE RELATED BY CRYSTAL-
LOGRAPHIC
REMARK 500 SYMMETRY ARE IN CLOSE CONTACT. AN ATOM LOCATED
WITHIN 0.15
REMARK 500 ANGSTROMS OF A SYMMETRY RELATED ATOM IS ASSUMED
TO BE ON A
REMARK 500 SPECIAL POSITION AND IS, THEREFORE, LISTED IN REMARK 375
REMARK 500 INSTEAD OF REMARK 500. ATOMS WITH NON-BLANK ALTER-
NATE
REMARK 500 LOCATION INDICATORS ARE NOT INCLUDED IN THE CALCU-
LATIONS.
REMARK 500
REMARK 500 DISTANCE CUTOFF:
REMARK 500 2.2 ANGSTROMS FOR CONTACTS NOT INVOLVING HYDROGEN
ATOMS
REMARK 500 1.6 ANGSTROMS FOR CONTACTS INVOLVING HYDROGEN AT-
OMS
REMARK 500
REMARK 500 ATM1 RES C SSEQI ATM2 RES C SSEQI SSYMOP
DISTANCE
REMARK 500 OD2 ASP B 83 O20 78N A 1506 1565 2.20
REMARK 500
REMARK 500 REMARK: NULL
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: COVALENT BOND ANGLES
REMARK 500
REMARK 500 THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING
RESIDUES
REMARK 500 HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY
MORE
REMARK 500 THAN 6*RMSD (M=MODEL NUMBER; RES=RESIDUE NAME;
C=CHAIN
REMARK 500 IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,3(1X,A4,2X),12X,F5.1)
REMARK 500
REMARK 500 EXPECTED VALUES PROTEIN: ENGH AND HUBER, 1999
REMARK 500 EXPECTED VALUES NUCLEIC ACID: CLOWNEY ET AL 1996
REMARK 500
REMARK 500 M RES CSSEQI ATM1 ATM2 ATM3
REMARK 500 PRO A 473 C - N - CA ANGL. DEV. = 9.4 DE-
GREES
REMARK 500 PRO B 39 C - N - CA ANGL. DEV. = 10.8 DE-
GREES
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: TORSION ANGLES

REMARK 500
REMARK 500 TORSION ANGLES OUTSIDE THE EXPECTED RAMACHANDRAN REGIONS:
REMARK 500 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 500 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT:(10X,I3,1X,A3,1X,A1,I4,A1,4X,F7.2,3X,F7.2)
REMARK 500
REMARK 500 EXPECTED VALUES: GJ KLEYWEGT AND TA JONES (1996). PHI/PSI-
REMARK 500 CHOLOGY: RAMACHANDRAN REVISITED. STRUCTURE 4, 1395 - 1400
REMARK 500
REMARK 500 M RES CSSEQI   PSI   PHI
REMARK 500   ASN A 54   110.60   -170.86
REMARK 500   ASN A 67   83.26   -154.40
REMARK 500   ASP A 162   80.82   -162.89
REMARK 500   ASP A 230   20.37   -147.33
REMARK 500   ASP A 321   -158.14   -151.05
REMARK 500   ASN A 355   39.67   -85.60
REMARK 500   ARG A 389   -130.03   45.98
REMARK 500   ASN A 416   78.60   -109.44
REMARK 500   ASN B 54   110.61   -170.70
REMARK 500   ASN B 67   83.25   -154.12
REMARK 500   ASP B 162   80.83   -162.92
REMARK 500   ASP B 242   -158.62   -113.15
REMARK 500   ASP B 294   -116.29   48.82
REMARK 500   ASP B 321   -158.08   -150.84
REMARK 500   ASN B 355   39.71   -85.55
REMARK 500   ARG B 389   -129.67   45.54
REMARK 500   ASN B 416   78.15   -112.61
REMARK 500   SER B 455   -169.11   -161.16
REMARK 500
REMARK 500 REMARK: NULL
REMARK 600
REMARK 600 HETEROGEN
REMARK 600 (2S)-2,3-DIHYDROXYPROPYL(7Z)-PENTADEC-7-ENOATE (78M): BOTH
REMARK 600 2S AND 2R PRESENT. LIPIDS WITHOUT HEADGROUP SEEN MODELLED
REMARK 600 AS ALKYL CHAINS - BUT THESE COULD BE DETERGENT.
REMARK 620
REMARK 620 METAL COORDINATION
REMARK 620 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 620 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL

REMARK 620 CU A1504 CU
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ACT A1508 OXT
REMARK 620 2 HIS A 218 NE2 97.6
REMARK 620 3 HOH A2112 O 149.3 110.3
REMARK 620 N 1 2
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CU A1505 CU
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASP A 55 OD2
REMARK 620 2 ASP A 57 OD1 80.8
REMARK 620 3 ASP A 57 OD2 120.6 53.7
REMARK 620 4 ALA A 61 O 137.3 80.9 76.1
REMARK 620 5 GLY A 63 O 118.7 124.7 73.6 103.4
REMARK 620 N 1 2 3 4
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CU B1499 CU
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASP B 57 OD2
REMARK 620 2 GLY B 63 O 73.4
REMARK 620 3 ASP B 55 OD1 83.7 91.9
REMARK 620 4 ASP B 55 OD2 131.7 125.9 55.1
REMARK 620 5 ASP B 57 OD1 53.7 127.0 79.5 91.2
REMARK 620 N 1 2 3 4
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CU B1500 CU
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 HIS B 145 NE2
REMARK 620 2 HOH B2090 O 164.5
REMARK 620 3 HOH B2027 O 107.5 84.2
REMARK 620 N 1 2
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 MG A1503 MG
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ALA A 136 O
REMARK 620 2 THR A 139 O 79.2
REMARK 620 3 TYR A 141 O 159.3 107.9
REMARK 620 4 GLU A 144 OE1 89.9 77.2 110.5
REMARK 620 5 HOH A2019 O 76.4 101.1 83.1 166.2
REMARK 620 6 GLU A 144 O 91.2 149.3 91.3 73.6 104.9
REMARK 620 N 1 2 3 4 5
REMARK 700
REMARK 700 SHEET
REMARK 700 DETERMINATION METHOD: DSSP
REMARK 700 THE SHEETS PRESENTED AS "AA" IN EACH CHAIN ON SHEET RECORDS

REMARK 700 BELOW IS ACTUALLY AN 18-STRANDED BARREL THIS IS REP-
RESENTED BY
REMARK 700 A 19-STRANDED SHEET IN WHICH THE FIRST AND LAST
STRANDS
REMARK 700 ARE IDENTICAL.
REMARK 700
REMARK 700 THE SHEETS PRESENTED AS "BA" IN EACH CHAIN ON SHEET
RECORDS
REMARK 700 BELOW IS ACTUALLY AN 18-STRANDED BARREL THIS IS REP-
RESENTED BY
REMARK 700 A 19-STRANDED SHEET IN WHICH THE FIRST AND LAST
STRANDS
REMARK 700 ARE IDENTICAL.
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1492
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1493
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1494
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1491
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1492
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1493
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1495
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1496
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC9
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1498

REMARK 800
REMARK 800 SITE_IDENTIFIER: BC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1494
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1495
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1496
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1497
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1500
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1498
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1501
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE LDA A1502
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC9
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE MG A1503
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CU A1504
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CU A1505
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CU B1499
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC4

REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CU B1500
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1501
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1502
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1503
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1506
REMARK 800
REMARK 800 SITE_IDENTIFIER: CC9
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1507
REMARK 800
REMARK 800 SITE_IDENTIFIER: DC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE ACT A1508
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 RELATED ID: 4AFK RELATED DB: PDB
REMARK 900 IN MESO STRUCTURE OF ALGINATE TRANSPORTER, ALGE,
FROM
REMARK 900 PSEUDOMONAS AERUGINOSA, PAO1
REMARK 900 RELATED ID: 4AZL RELATED DB: PDB
REMARK 900 IN MESO STRUCTURE OF ALGINATE TRANSPORTER, ALGE,
FROM
REMARK 900 PSEUDOMOAS AERUGINOSA, PAO1, CRYSTAL FORM 2.
REMARK 999
REMARK 999 SEQUENCE
REMARK 999 STRUCTURE CONTAINS FOLLOWING RESIDUES 38-105, 117-471,
475-
REMARK 999 490 DATABASE PSEUDOMONAS GENOME DATABASE (V2)
REMARK 999 STRUCTURE CONTAINS FOLLOWING RESIDUES 39-105, 116-440,
454-
REMARK 999 490 DATABASE PSEUDOMONAS GENOME DATABASE (V2)
DBREF 4B61 A 33 490 UNP P18895 ALGE_PSEAE 33 490
DBREF 4B61 B 33 490 UNP P18895 ALGE_PSEAE 33 490

Appendix 8 PDB header for AlgE_2.8 (4AZL in PDB database)

```
HEADER    MEMBRANE PROTEIN                        26-JUN-12
4AZL
TITLE     IN MESO STRUCTURE OF ALGINATE TRANSPORTER, ALGE, FROM
TITLE    2 PSEUDOMOAS AERUGINOSA, PAO1, CRYSTAL FORM 2.
COMPND    MOL_ID: 1;
COMPND   2 MOLECULE: ALGINATE PRODUCTION PROTEIN ALGE;
COMPND   3 CHAIN: A, B;
COMPND   4 SYNONYM: ALGE;
COMPND   5 ENGINEERED: YES
SOURCE    MOL_ID: 1;
SOURCE   2 ORGANISM_SCIENTIFIC: PSEUDOMONAS AERUGINOSA;
SOURCE   3 ORGANISM_TAXID: 208964;
SOURCE   4 STRAIN: PAO1;
SOURCE   5 ATCC: 15692;
SOURCE   6 EXPRESSION_SYSTEM: ESCHERICHIA COLI;
SOURCE   7 EXPRESSION_SYSTEM_TAXID: 469008;
SOURCE   8 EXPRESSION_SYSTEM_STRAIN: BL21(DE3);
SOURCE   9 EXPRESSION_SYSTEM_VARIANT: STAR;
SOURCE  10 EXPRESSION_SYSTEM_VECTOR_TYPE: PLASMID;
SOURCE  11 EXPRESSION_SYSTEM_VECTOR: PET200-D-TOPO;
SOURCE  12 EXPRESSION_SYSTEM_PLASMID: ALGE-PET200/D-TOPO;
SOURCE  13 OTHER_DETAILS: HOLLOWAY COLLECTION
KEYWDS    MEMBRANE PROTEIN, OUTER MEMBRANE
EXPDTA    X-RAY DIFFRACTION
AUTHOR    J.TAN,V.E.PYE,D.ARAGAO,M.CAFFREY
REVDAT   1   27-JUN-12 4AZL      0
JRNL        AUTH    J.TAN,V.E.PYE,D.ARAGAO,M.CAFFREY
JRNL        TITL    IN MESO STRUCTURE OF ALGINATE TRANSPORTER,
ALGE
JRNL        REF     TO BE PUBLISHED.
JRNL        REFN
REMARK   2
REMARK   2 RESOLUTION.      2.80 ANGSTROMS.
REMARK   3
REMARK   3 REFINEMENT.
REMARK   3    PROGRAM      : PHENIX (PHENIX.REFINE)
REMARK   3    AUTHORS      : PAUL ADAMS,PAVEL AFONINE,VICENT
CHEN,IAN
REMARK   3                 : DAVIS,KRESHNA GOPAL,RALF GROSSE-
KUNSTLEVE,
REMARK   3                 : LI-WEI HUNG,ROBERT IMMORMINO,TOM
IOERGER,
REMARK   3                 : AIRLIE MCCOY,ERIK MCKEE,NIGEL MORI-
ARTY,
REMARK   3                 : REETAL PAI,RANDY READ,JANE RICHARDSON,
REMARK   3                 : DAVID RICHARDSON,TOD ROMO,JIM SAC-
CHETTINI,
REMARK   3                 : NICHOLAS SAUTER,JACOB SMITH,LAURENT
```

```
REMARK   3                  : STORONI,TOM TERWILLIGER,PETER ZWART
REMARK   3
REMARK   3    REFINEMENT TARGET : ML
REMARK   3
REMARK   3  DATA USED IN REFINEMENT.
REMARK   3   RESOLUTION RANGE HIGH (ANGSTROMS) : 2.800
REMARK   3   RESOLUTION RANGE LOW  (ANGSTROMS) : 44.889
REMARK   3   MIN(FOBS/SIGMA_FOBS)              : 1.34
REMARK   3   COMPLETENESS FOR RANGE        (%) : 93.35
REMARK   3   NUMBER OF REFLECTIONS             : 23738
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT.
REMARK   3   R VALUE       (WORKING + TEST SET) : 0.2393
REMARK   3   R VALUE           (WORKING SET) : 0.2371
REMARK   3   FREE R VALUE                     : 0.2803
REMARK   3   FREE R VALUE TEST SET SIZE   (%) : 5.1
REMARK   3   FREE R VALUE TEST SET COUNT      : 1215
REMARK   3
REMARK   3  FIT TO DATA USED IN REFINEMENT (IN BINS).
REMARK   3   BIN  RESOLUTION RANGE  COMPL.     NWORK NFREE
RWORK  RFREE
REMARK   3     1 44.8951 -  5.8205    0.89      2382   136   0.2738 0.2968
REMARK   3     2  5.8205 -  4.6214    0.92      2454   148   0.2083 0.2277
REMARK   3     3  4.6214 -  4.0377    0.93      2509   125   0.1894 0.2176
REMARK   3     4  4.0377 -  3.6687    0.94      2488   152   0.2136 0.2739
REMARK   3     5  3.6687 -  3.4058    0.94      2504   122   0.2188 0.2630
REMARK   3     6  3.4058 -  3.2051    0.95      2558   125   0.2336 0.2752
REMARK   3     7  3.2051 -  3.0446    0.95      2521   135   0.2805 0.3620
REMARK   3     8  3.0446 -  2.9121    0.95      2551   151   0.3077 0.3937
REMARK   3     9  2.9121 -  2.8000    0.95      2556   121   0.3225 0.3654
REMARK   3
REMARK   3  BULK SOLVENT MODELLING.
REMARK   3   METHOD USED          : FLAT BULK SOLVENT MODEL
REMARK   3   SOLVENT RADIUS       : 1.00
REMARK   3   SHRINKAGE RADIUS     : 0.73
REMARK   3   K_SOL                : 0.296
REMARK   3   B_SOL                : 14.245
REMARK   3
REMARK   3  ERROR ESTIMATES.
REMARK   3   COORDINATE ERROR (MAXIMUM-LIKELIHOOD BASED)      :
0.47
REMARK   3   PHASE ERROR (DEGREES, MAXIMUM-LIKELIHOOD BASED) :
29.34
REMARK   3
REMARK   3  B VALUES.
REMARK   3   FROM WILSON PLOT           (A**2) : 36.86
REMARK   3   MEAN B VALUE      (OVERALL, A**2) : 44.24
REMARK   3   OVERALL ANISOTROPIC B VALUE.
REMARK   3    B11 (A**2) : -3.3652
REMARK   3    B22 (A**2) : 5.2082
```

```
REMARK   3      B33 (A**2) : -1.8429
REMARK   3      B12 (A**2) : 0.0000
REMARK   3      B13 (A**2) : -0.6168
REMARK   3      B23 (A**2) : 0.0000
REMARK   3
REMARK   3  TWINNING INFORMATION.
REMARK   3   FRACTION: NULL
REMARK   3   OPERATOR: NULL
REMARK   3
REMARK   3  DEVIATIONS FROM IDEAL VALUES.
REMARK   3                      RMSD              COUNT
REMARK   3   BOND         :   0.006             7048
REMARK   3   ANGLE        :   1.218             9488
REMARK   3   CHIRALITY    :   0.085              952
REMARK   3   PLANARITY    :   0.005             1252
REMARK   3   DIHEDRAL     : 15.279              2555
REMARK   3
REMARK   3  TLS DETAILS
REMARK   3   NUMBER OF TLS GROUPS   : NULL
REMARK   3
REMARK   3  NCS DETAILS
REMARK   3   NUMBER OF NCS GROUPS : 1
REMARK   3   NCS GROUP : 1
REMARK   3    NCS OPERATOR : 1
REMARK   3     REFERENCE SELECTION: CHAIN A AND RESSEQ 41:70 OR
CHAIN A AND
REMARK   3                          RESSEQ 72:98 OR CHAIN A AND
RESSEQ 104:
REMARK   3                          231 OR CHAIN A AND RESSEQ
240:264 OR CHAIN
REMARK   3                          A AND RESSEQ 269:297 OR CHAIN A
AND RESSEQ
REMARK   3                          306:318 OR CHAIN A AND RESSEQ
323:333 OR
REMARK   3                          CHAIN A AND RESSEQ 343:388 OR
CHAIN A AND
REMARK   3                          RESSEQ 390:408 OR CHAIN A AND
RESSEQ 416:
REMARK   3                          438 OR CHAIN A AND RESSEQ
456:464 OR CHAIN
REMARK   3                          A AND RESSEQ 469:470 OR CHAIN A
AND RESSEQ
REMARK   3                          476:498
REMARK   3     SELECTION            : CHAIN B AND RESSEQ 41:70 OR
CHAIN B AND
REMARK   3                          RESSEQ 72:98 OR CHAIN B AND
RESSEQ 104:
REMARK   3                          231 OR CHAIN B AND RESSEQ
240:264 OR CHAIN
REMARK   3                          B AND RESSEQ 269:297 OR CHAIN B
```

AND RESSEQ
REMARK 3 306:318 OR CHAIN B AND RESSEQ
323:333 OR
REMARK 3 CHAIN B AND RESSEQ 343:388 OR
CHAIN B AND
REMARK 3 RESSEQ 390:408 OR CHAIN B AND
RESSEQ 416:
REMARK 3 438 OR CHAIN B AND RESSEQ
456:464 OR CHAIN
REMARK 3 B AND RESSEQ 469:470 OR CHAIN B
AND RESSEQ
REMARK 3 476:498
REMARK 3 ATOM PAIRS NUMBER : 2774
REMARK 3 RMSD : 0.014
REMARK 3
REMARK 3 OTHER REFINEMENT REMARKS: NULL
REMARK 4
REMARK 4 4AZL COMPLIES WITH FORMAT V. 3.30, 13-JUL-11
REMARK 100
REMARK 100 THIS ENTRY HAS BEEN PROCESSED BY PDBE ON 27-JUN-12.
REMARK 100 THE PDBE ID CODE IS EBI-53066.
REMARK 200
REMARK 200 EXPERIMENTAL DETAILS
REMARK 200 EXPERIMENT TYPE : X-RAY DIFFRACTION
REMARK 200 DATE OF DATA COLLECTION : 05-MAR-10
REMARK 200 TEMPERATURE (KELVIN) : 293
REMARK 200 PH : 6.5
REMARK 200 NUMBER OF CRYSTALS USED : 1
REMARK 200
REMARK 200 SYNCHROTRON (Y/N) : Y
REMARK 200 RADIATION SOURCE : APS
REMARK 200 BEAMLINE : 23-ID-B
REMARK 200 X-RAY GENERATOR MODEL : NULL
REMARK 200 MONOCHROMATIC OR LAUE (M/L) : M
REMARK 200 WAVELENGTH OR RANGE (A) : 0.97934
REMARK 200 MONOCHROMATOR : DOUBLE CRYSTAL
CRYO-COOLED
REMARK 200 OPTICS : NULL
REMARK 200
REMARK 200 DETECTOR TYPE : CCD
REMARK 200 DETECTOR MANUFACTURER : MARRESEARCH
REMARK 200 INTENSITY-INTEGRATION SOFTWARE : MOSFLM
REMARK 200 DATA SCALING SOFTWARE : SCALA
REMARK 200
REMARK 200 NUMBER OF UNIQUE REFLECTIONS : 23738
REMARK 200 RESOLUTION RANGE HIGH (A) : 2.80
REMARK 200 RESOLUTION RANGE LOW (A) : 44.90
REMARK 200 REJECTION CRITERIA (SIGMA(I)) : 2.0
REMARK 200
REMARK 200 OVERALL.

REMARK 200 COMPLETENESS FOR RANGE (%) : 93.2
REMARK 200 DATA REDUNDANCY : 2.7
REMARK 200 R MERGE (I) : 0.22
REMARK 200 R SYM (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR THE DATA SET : 5.80
REMARK 200
REMARK 200 IN THE HIGHEST RESOLUTION SHELL.
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE HIGH (A) : 2.80
REMARK 200 HIGHEST RESOLUTION SHELL, RANGE LOW (A) : 2.95
REMARK 200 COMPLETENESS FOR SHELL (%) : 95.3
REMARK 200 DATA REDUNDANCY IN SHELL : 2.7
REMARK 200 R MERGE FOR SHELL (I) : 0.91
REMARK 200 R SYM FOR SHELL (I) : NULL
REMARK 200 <I/SIGMA(I)> FOR SHELL : 2.00
REMARK 200
REMARK 200 DIFFRACTION PROTOCOL: SINGLE WAVELENGTH
REMARK 200 METHOD USED TO DETERMINE THE STRUCTURE: MOLECULAR REPLACEMENT
REMARK 200 SOFTWARE USED: PHASER
REMARK 200 STARTING MODEL: PDB ENTRY 3RBH
REMARK 200
REMARK 200 REMARK: NONE
REMARK 280
REMARK 280 CRYSTAL
REMARK 280 SOLVENT CONTENT, VS (%): 52.51
REMARK 280 MATTHEWS COEFFICIENT, VM (ANGSTROMS**3/DA): 2.59
REMARK 280
REMARK 280 CRYSTALLIZATION CONDITIONS: 0.1 M SODIUM CACODYLATE, PH 6.5,
REMARK 280 1.2 M SODIUM ACETATE
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY
REMARK 290 SYMMETRY OPERATORS FOR SPACE GROUP: P 1 21 1
REMARK 290
REMARK 290 SYMOP SYMMETRY
REMARK 290 NNNMMM OPERATOR
REMARK 290 1555 X,Y,Z
REMARK 290 2555 -X,Y+1/2,-Z
REMARK 290
REMARK 290 WHERE NNN -> OPERATOR NUMBER
REMARK 290 MMM -> TRANSLATION VECTOR
REMARK 290
REMARK 290 CRYSTALLOGRAPHIC SYMMETRY TRANSFORMATIONS
REMARK 290 THE FOLLOWING TRANSFORMATIONS OPERATE ON THE ATOM/HETATM
REMARK 290 RECORDS IN THIS ENTRY TO PRODUCE CRYSTALLOGRAPHICALLY
REMARK 290 RELATED MOLECULES.
REMARK 290 SMTRY1 1 1.000000 0.000000 0.000000 0.00000
REMARK 290 SMTRY2 1 0.000000 1.000000 0.000000 0.00000

REMARK 290 SMTRY3 1 0.000000 0.000000 1.000000 0.00000
REMARK 290 SMTRY1 2 -1.000000 0.000000 0.000000 0.00000
REMARK 290 SMTRY2 2 0.000000 1.000000 0.000000 122.88000
REMARK 290 SMTRY3 2 0.000000 0.000000 -1.000000 0.00000
REMARK 290
REMARK 290 REMARK: NULL
REMARK 300
REMARK 300 BIOMOLECULE: 1, 2
REMARK 300 SEE REMARK 350 FOR THE AUTHOR PROVIDED AND/OR PROGRAM
REMARK 300 GENERATED ASSEMBLY INFORMATION FOR THE STRUCTURE IN
REMARK 300 THIS ENTRY. THE REMARK MAY ALSO PROVIDE INFORMATION ON
REMARK 300 BURIED SURFACE AREA.
REMARK 350
REMARK 350 COORDINATES FOR A COMPLETE MULTIMER REPRESENTING THE KNOWN
REMARK 350 BIOLOGICALLY SIGNIFICANT OLIGOMERIZATION STATE OF THE
REMARK 350 MOLECULE CAN BE GENERATED BY APPLYING BIOMT TRANSFORMATIONS
REMARK 350 GIVEN BELOW. BOTH NON-CRYSTALLOGRAPHIC AND
REMARK 350 CRYSTALLOGRAPHIC OPERATIONS ARE GIVEN.
REMARK 350
REMARK 350 BIOMOLECULE: 1
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: MONOMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 APPLY THE FOLLOWING TO CHAINS: A
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.00000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.00000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.00000
REMARK 350
REMARK 350 BIOMOLECULE: 2
REMARK 350 AUTHOR DETERMINED BIOLOGICAL UNIT: MONOMERIC
REMARK 350 SOFTWARE DETERMINED QUATERNARY STRUCTURE: MONOMERIC
REMARK 350 SOFTWARE USED: PISA
REMARK 350 APPLY THE FOLLOWING TO CHAINS: B
REMARK 350 BIOMT1 1 1.000000 0.000000 0.000000 0.00000
REMARK 350 BIOMT2 1 0.000000 1.000000 0.000000 0.00000
REMARK 350 BIOMT3 1 0.000000 0.000000 1.000000 0.00000
REMARK 465
REMARK 465 MISSING RESIDUES
REMARK 465 THE FOLLOWING RESIDUES WERE NOT LOCATED IN THE
REMARK 465 EXPERIMENT. (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 465 IDENTIFIER; SSSEQ=SEQUENCE NUMBER; I=INSERTION CODE.)

```
REMARK 465
REMARK 465   M RES C SSSEQI
REMARK 465     ALA A     33
REMARK 465     ASN A     34
REMARK 465     SER A     35
REMARK 465     GLY A     36
REMARK 465     GLU A     37
REMARK 465     ALA A     38
REMARK 465     PRO A     39
REMARK 465     GLY A     63
REMARK 465     GLY A     64
REMARK 465     LEU A    104
REMARK 465     GLN A    105
REMARK 465     SER A    106
REMARK 465     ASP A    107
REMARK 465     THR A    108
REMARK 465     ASP A    109
REMARK 465     ASP A    110
REMARK 465     GLY A    111
REMARK 465     ASN A    112
REMARK 465     ASN A    113
REMARK 465     SER A    114
REMARK 465     ARG A    115
REMARK 465     ASN A    116
REMARK 465     ASP A    117
REMARK 465     GLY A    118
REMARK 465     ARG A    119
REMARK 465     GLU A    120
REMARK 465     GLU A    240
REMARK 465     VAL A    241
REMARK 465     ASP A    242
REMARK 465     ASN A    243
REMARK 465     LEU A    244
REMARK 465     ASP A    245
REMARK 465     LYS A    246
REMARK 465     THR A    291
REMARK 465     THR A    292
REMARK 465     VAL A    293
REMARK 465     ASP A    294
REMARK 465     ASP A    295
REMARK 465     ARG A    296
REMARK 465     ARG A    297
REMARK 465     ALA B     33
REMARK 465     ASN B     34
REMARK 465     SER B     35
REMARK 465     GLY B     36
REMARK 465     GLU B     37
REMARK 465     ALA B     38
REMARK 465     GLY B     63
REMARK 465     GLY B     64
```

REMARK 465 GLN B 105
REMARK 465 SER B 106
REMARK 465 ASP B 107
REMARK 465 THR B 108
REMARK 465 ASP B 109
REMARK 465 ASP B 110
REMARK 465 GLY B 111
REMARK 465 ASN B 112
REMARK 465 ASN B 113
REMARK 465 SER B 114
REMARK 465 ARG B 115
REMARK 465 ASN B 116
REMARK 465 ASP B 117
REMARK 465 GLY B 118
REMARK 465 ARG B 119
REMARK 465 ASP B 242
REMARK 465 ASN B 243
REMARK 465 LEU B 244
REMARK 465 THR B 291
REMARK 465 THR B 292
REMARK 465 VAL B 293
REMARK 465 ASP B 294
REMARK 465 ASP B 295
REMARK 465 ARG B 296
REMARK 465 ARG B 297
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: COVALENT BOND ANGLES
REMARK 500
REMARK 500 THE STEREOCHEMICAL PARAMETERS OF THE FOLLOWING RESIDUES
REMARK 500 HAVE VALUES WHICH DEVIATE FROM EXPECTED VALUES BY MORE
REMARK 500 THAN 6*RMSD (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN
REMARK 500 IDENTIFIER; SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT: (10X,I3,1X,A3,1X,A1,I4,A1,3(1X,A4,2X),12X,F5.1)
REMARK 500
REMARK 500 EXPECTED VALUES PROTEIN: ENGH AND HUBER, 1999
REMARK 500 EXPECTED VALUES NUCLEIC ACID: CLOWNEY ET AL 1996
REMARK 500
REMARK 500 M RES CSSEQI ATM1 ATM2 ATM3
REMARK 500 ASP A 68 CB - CG - OD1 ANGL. DEV. = -5.5 DE-GREES
REMARK 500 PRO A 473 C - N - CA ANGL. DEV. = 10.5 DE-GREES
REMARK 500 LEU B 104 CA - CB - CG ANGL. DEV. = 15.0 DE-GREES

REMARK 500 PRO B 473 C - N - CA ANGL. DEV. = 10.4 DEGREES
REMARK 500
REMARK 500 REMARK: NULL
REMARK 500
REMARK 500 GEOMETRY AND STEREOCHEMISTRY
REMARK 500 SUBTOPIC: TORSION ANGLES
REMARK 500
REMARK 500 TORSION ANGLES OUTSIDE THE EXPECTED RAMACHANDRAN REGIONS:
REMARK 500 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 500 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE).
REMARK 500
REMARK 500 STANDARD TABLE:
REMARK 500 FORMAT:(10X,I3,1X,A3,1X,A1,I4,A1,4X,F7.2,3X,F7.2)
REMARK 500
REMARK 500 EXPECTED VALUES: GJ KLEYWEGT AND TA JONES (1996). PHI/PSI-
REMARK 500 CHOLOGY: RAMACHANDRAN REVISITED. STRUCTURE 4, 1395 - 1400
REMARK 500
REMARK 500 M RES CSSEQI PSI PHI
REMARK 500 ASN A 54 111.58 -170.46
REMARK 500 ASN A 67 85.20 -153.74
REMARK 500 ASP A 162 80.83 -162.84
REMARK 500 ASP A 321 -159.12 -149.80
REMARK 500 ASN A 355 39.31 -85.03
REMARK 500 ARG A 389 -130.37 45.81
REMARK 500 ASN A 416 78.64 -109.05
REMARK 500 SER A 455 -171.62 -177.13
REMARK 500 ASN B 54 111.66 -170.53
REMARK 500 ASN B 67 84.78 -153.87
REMARK 500 ASP B 162 80.71 -162.88
REMARK 500 ASP B 321 -158.15 -151.03
REMARK 500 ASN B 355 39.21 -85.20
REMARK 500 ARG B 389 -128.67 43.05
REMARK 500 ASN B 416 78.88 -108.97
REMARK 500
REMARK 500 REMARK: NULL
REMARK 600
REMARK 600 HETEROGEN
REMARK 600 (2RS)-2,3-DIHYDROXYPROPYL(7Z)-PENTADEC-7-ENOATE (78M/78N)
REMARK 600 ALKANE CHAINS WERE MODELLED AS 78M, BUT COULD POSSIBLY BE
REMARK 600 OTHER DETERGENT.
REMARK 610
REMARK 610 MISSING HETEROATOM
REMARK 610 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL

NUMBER;
REMARK 610 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 610 I=INSERTION CODE):
REMARK 610 M RES C SSEQI
REMARK 610 78M A 1491
REMARK 610 78M A 1492
REMARK 610 78M A 1493
REMARK 610 78M B 1491
REMARK 610 78M B 1492
REMARK 610 78M B 1493
REMARK 610 78M B 1494
REMARK 610 78M B 1495
REMARK 620
REMARK 620 METAL COORDINATION
REMARK 620 (M=MODEL NUMBER; RES=RESIDUE NAME; C=CHAIN IDENTIFIER;
REMARK 620 SSEQ=SEQUENCE NUMBER; I=INSERTION CODE):
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CA A1494 CA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASP A 55 OD1
REMARK 620 2 ASP A 55 OD2 40.4
REMARK 620 3 ASP A 57 OD2 71.6 97.1
REMARK 620 4 ALA A 61 O 118.7 105.1 65.6
REMARK 620 N 1 2 3
REMARK 620
REMARK 620 COORDINATION ANGLES FOR: M RES CSSEQI METAL
REMARK 620 CA B1496 CA
REMARK 620 N RES CSSEQI ATOM
REMARK 620 1 ASP B 55 OD2
REMARK 620 2 ALA B 61 O 114.2
REMARK 620 N 1
REMARK 700
REMARK 700 SHEET
REMARK 700 DETERMINATION METHOD: DSSP
REMARK 700 THE SHEETS PRESENTED AS "AA" IN EACH CHAIN ON SHEET RECORDS
REMARK 700 BELOW IS ACTUALLY AN 18-STRANDED BARREL THIS IS REPRESENTED BY
REMARK 700 A 19-STRANDED SHEET IN WHICH THE FIRST AND LAST STRANDS
REMARK 700 ARE IDENTICAL.
REMARK 700 THE SHEETS PRESENTED AS "BA" IN EACH CHAIN ON SHEET RECORDS
REMARK 700 BELOW IS ACTUALLY AN 18-STRANDED BARREL THIS IS REPRESENTED BY
REMARK 700 A 19-STRANDED SHEET IN WHICH THE FIRST AND LAST STRANDS

REMARK 700 ARE IDENTICAL.
REMARK 800
REMARK 800 SITE
REMARK 800 SITE_IDENTIFIER: AC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1491
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1492
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M A1493
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CA A1494
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1495
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1496
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC7
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78N A1497
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC8
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1491
REMARK 800
REMARK 800 SITE_IDENTIFIER: AC9
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1492
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC1
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1493
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC2
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1494
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC3
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1495

REMARK 800
REMARK 800 SITE_IDENTIFIER: BC4
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE CA B1496
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC5
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1497
REMARK 800
REMARK 800 SITE_IDENTIFIER: BC6
REMARK 800 EVIDENCE_CODE: SOFTWARE
REMARK 800 SITE_DESCRIPTION: BINDING SITE FOR RESIDUE 78M B1498
REMARK 900
REMARK 900 RELATED ENTRIES
REMARK 900 RELATED ID: 4AFK RELATED DB: PDB
REMARK 900 IN MESO STRUCTURE OF ALGINATE TRANSPORTER, ALGE,
FROM
REMARK 900 PSEUDOMONAS AERUGINOSA, PAO1
DBREF 4AZL A 33 490 UNP P18895 ALGE_PSEAE 33 490
DBREF 4AZL B 33 490 UNP P18895 ALGE_PSEAE 33 490