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.

STRUCTURAL AND FUNCTIONAL STUDIES OF *Pseudomonas aeruginosa* STRAIN PAO1 LYSINE SPECIFIC PERMEASE (LysP)

By

EMMANUEL DOH NJI M. Sc.

A THESIS SUBMITTED TO

TRINITY COLLEGE DUBLIN

FOR THE DEGREE
OF
DOCTOR OF PHILOSOPHY

SUPERVISOR: PROF. MARTIN CAFFREY

SCHOOL OF BIOCHEMISTRY AND IMMUNOLOGY
TRINITY COLLEGE DUBLIN

April 2013

TRINITY COLLEGE
1 9 JUN 2013
LIBRARY DUBLIN
Thesis 10154

Abstract

The means by which substrates are shuttled across prokaryotic membranes by membrane transporters and the conformational changes involved are unique to each transporter. To decipher the underlying mechanism of substrate selectivity and the conformational changes required for transport, we sought to establish the high resolution crystal structure of a lysine specific transporter from *Pseudomonas aeruginosa* strain PAO1, hereafter referred to as LysP. Also, we aim to show that the protein binds and transports L-lysine. To this end, LysP has been cloned into the expression vector pWaldo-GFPd and expressed in *Escherichia coli* C43 (DE3) cells as a His-tagged Green Fluorescent Protein (GFP) fusion protein. The GFP addition proved particularly useful as its fluorescence could be easily monitored to quantify protein yield at every step in the purification process and to screen for membrane solubilizing detergents in which the transporter is monodispersed. To date, the fused protein has been successfully over-expressed and solubilised in n-dodecyl-β-D–maltoside (DDM) and lauryl maltose neopentyl glycol (LMNG). Following purification on a nickel affinity resin it has been cleaved with TEV protease to remove the GFP. Further purification of free LysP has been carried out by size exclusion chromatography. The final yield of purified LysP was ~0.5 mg/L of cell culture medium. Purified LysP after size-exclusion chromatography (SEC) was identified by both in-gel tryptic digestion and total mass spectrometry. Blue native page suggest that LysP is a monomer in LMNG micelle. The oligomerization state was supported by glutaraldehyde cross-linking. Also, circular diachroism (CD) spectroscopy on purified LysP revealed 79 % α-helical content, 21 % random coil and no β strand. Stability studies of LysP purified in LMNG showed that LysP was still monodisperse for 7 days at 4°C on SEC after concentrating the protein to 10 mg/mL. Crystallization trials of TEV cleaved LysP were set up by in surfo and in meso methods. Several conditions gave crystals by the in surfo method only that diffracted to 7 Å with reflections seen to 6 Å after exposure to synchrotron X-radiation after several optimization experiments were performed. To improve the diffraction quality as well as the stability of LysP, nanobodies provided through collaboration were used for co-crystallization and crystals were obtained that diffracted to 8 Å. Bicelle crystallization was also carried out but no crystal was obtained in the conditions tested. TEV cleaved LysP purified in DDM was shown to transport [^{3}H] L-lysine by radioactivity assay reconstituted into proteoliposomes with the transport of [^{3}H] L-lysine being inhibited L-4-thialysine. We also showed that, the

transport of L-lysine by reconstituted proteoliposomes of LysP was sodium independent. Microscale thermophoresis (MST) analysis confirmed that TEV cleaved LysP purified in LMNG binds to L-lysine, D-lysine, L-L-4-thialysine, L-arginine and L-histidine with a K_d of 15 μM, 124 μM, 31 μM, 21 mM and 4.5 mM, respectively. No binding was seen for L-leucine, L-isoleucine, L-threonine, L-methionine, glycine, L-valine, L-alanine, L-glutamate, L-glutamine and L-proline by MST analysis.

Lastly, whole cell radioactivity assay was carried out on the cloned LysP-GFP fusion to see if it was functionally active. Inhibition studies were also carried out in whole cell to investigate other possible amino acid substrate of LysP-GFP and thus LysP. The whole cell transport assay was inhibited by D-lysine, L-4-thialysine and cold L-lysine but not with L-histine, L-arginine, L-glutamate, L-phenylalanine, L-leucine, L-proline, L-asparagine and L-serine. However, there are weaknesses in the whole cell radioactivity assay. They are described in Chapter 4, Section 4.1. Thus, we could not conclude based on the results of the whole cell radioactivity assay that LysP-GFP was functionally active.

Acknowledgements

Special thanks to God almighty for providing me with the strength and grace to complete this PhD.

This research is supported by grants from Science Foundation Ireland 07-IN.1-B1836, The USA National Institutes of Health GM75915 and Trinity College Dublin Postgraduate Award (Code: 1252) and Trinity College Dublin School of Biochemistry and Immunology - PhD enhancement scheme.

Thanks to;

Prof. Martin Caffrey for supervising this project.

Dr. Abhiram Dukkipati, Dr. Declan Doyle and Dr. Dianfan Li for useful discussions on this project.

Dr. Valery Pye, Dr. David Aragao and Dr. Declan Doyle for guidance during data collection.

Dr. Syed Tasadaque for synthesizing 7.7 MAG and 7.8 MAG lipids used for this project.

Other members of the Membrane Structural and Functional Biology Group, Trinity College Dublin (Toufic El-Arnaout, Dr. Joseph Lyons, Jinquan Tan, Dr. Colin Boland, Ashley Bray and Dr. Nicole Howe) for their support and useful discussions.

Dr. Amir Khan and Humberto Fernandes for his help on data collection on LysP crystals.

Prof. Peter Henderson and Dr. Preethi Sukumar at the Faculty of Biological Sciences, University of Leeds, for their guidance and help on the radioactivity assays.

Dr. David Drew at the Division of Molecular Biosciences, Imperial College London for supplying me with pWaldo GFPd and TEV protease vectors and for teaching me how to express, purify and to crystallize membrane proteins.

Dr. Moran Jerabek, Dr. Stephen Blanke and Anna Muench at NanoTemper Technologies GmbH, Germany for their help and guidance on the microscale thermophoresis analysis.

Mary-Pat O'Sullivan and Miriam Wilson for providing me with the letters used for visa applications and residence permits.

Kate Louise Gildea-Byrne and Martha Motherway Gildea for their help in purchasing reagents used for this projects.

My family members: Mr. Clement Tambendong Nji, Mrs Florence Ikwe Nji, Talia Anim Nji, Magdalin Anim Nji, Nji Valentine Aben, Juliana Okwa Nji, Nyah Terry Nji,

Manka Henrietta Nji, Micheal Nji, Sylvie Aben, Kate Kimberly Aben, Dean Aben Nji, Emmanuel Devante, Kieran Aben and Jada Kourtney Aboubacar Kindo.

Susan Tang, James Emmanuel Wanki and Momo Remi, for their support and prayers.

Pastor Zak Eroje, Pastor Christy Eroje, Joy Olowo, Angela Hayes and all the members of Christ Embassy Dublin for their prayers and support.

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. Where used has been made to the work of other people it has been fully acknowledged and referenced.

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

....

Table of Contents

ABSTRACT ..... **II**

ACKNOWLEDGEMENTS ..... **IV**

DECLARATION ..... **VI**

TABLE OF CONTENTS ..... **VII**

LIST OF FIGURES ..... **X**

LIST OF TABLES ..... **XV**

LIST OF MATERIALS ..... **XVI**

ABBREVIATIONS ..... **XXIV**

CHAPTER 1: GENERAL INTRODUCTION **27**

1.1. LYSINE SPECIFIC PERMEASE (LYSP) 27
1.2. AMINO ACID POLYAMINE ORGANOCATION (APC) TRANSPORTER FAMILY 31
1.3. AIMS OF THE PROJECT 38
1.4. *PSEUDOMONAS AERUGINOSA* 39
1.5. MEMBRANE TRANSPORT 40
1.5.1. Secondary Active Transporters 42
1.6. MEMBRANE PROTEIN STRUCTURAL AND FUNCTIONAL BIOLOGY 43

CHAPTER 2: BIOINFORMATICS, BIOCHEMICAL, BIOPHYSICAL AND FUNCTIONAL CHARACTERIZATION OF A LYSINE SPECIFIC PERMEASE (LYSP) FROM *PSEUDOMONAS AERUGINOSA* STRAIN PAO1 ..... **52**

2.1. INTRODUCTION 52
2.2. BIOINFORMATICS OF LYSP 53
2.2.2. Amino acid Sequence (fasta format) 54
2.2.3. Amino acid composition and other parameters of LysP 54
2.2.4 Signal Peptide Prediction 55
2.2.5. Secondary structure prediction 57
2.2.6. Topology map of LysP 60
2.2.7. Disorder predictions 65
2.2.7.1. FoldIndex© (Prilusky *et al.* 2005) 65
2.2.7.2. RONN (Regional Order Neural Network) prediction (Yang *et al.* 2005). 66
2.2.8. Post translational modification 68
2.3. EXPERIMENTAL PROCEDURES 71
2.3.1. Cloning of LysP 71
2.3.2. Recombinant Expression of LysP-GFP fusion and In-gel fluorescence 71
2.3.3. Isolation of Membranes containing LysP-GFP fusion 72
2.3.4. Detergent screening of LysP-GFP fusion. 73
2.3.5. Purification of LysP from LysP-GFP fusion. 74
2.3.6. Purity evaluation and Determination of Oligomerization State of TEV Cleaved GFP free LysP 75
2.3.7. Biochemical/Biophysical Characterization of LysP. 76
2.3.7.1. Total Mass Spectrometry of LysP 76
2.3.7.2. In-gel Tryptic Digestion of LysP 76
2.3.7.3. Glutaraldehyde Crosslinking Attempts. 77
2.3.7.4. Stability Studies 77
2.3.7.5. Circular Dichroism (CD) Measurements 78

2.3.7.6. [^{3}H] L-lysine Transport, Transport Inhibition and Transport pH Screening in Whole Cell. 78
2.3.7.7. Reconstitution and Counterflow Assay 80
2.3.7.8. Microscale Thermophoresis (MST) Analysis. 84
2.3.7.9. Tryptophan Fluorescence Change 85
2.4. RESULTS AND DISCUSSIONS 86
2.4.1. Recombinant Expression of LysP-GFP fusion and In-gel Fluorescence 86
2.4.2. Detergent screening of LysP-GFP 89
2.4.3. Purification of LysP from LysP-GFP Fusion 91
2.3.6. Purity Evaluation and Determination of Oligomerization State of GFP free LysP. 95
2.4.4. Biochemical, Biophysical and Functional Characterization of LysP 96
2.4.4.1. Total Mass Spectrometry of LysP 96
2.4.4.2. In-gel Tryptic Digestion of LysP 96
2.4.4.3. Glutaraldehyde (GA) Crosslinking to Investigate the Oligomerization States of GFP Free LysP 99
2.4.4.4. Stability studies of TEV Cleaved Free LysP 100
2.4.4.5. CD Spectroscopy for Secondary Structural Determination of GFP free LysP 101
2.4.4.6. [^{3}H] L-lysine transport in whole cell 102
2.4.4.7. Reconstitution and Counterflow Assays. 106
2.4.5.3. Microscale thermophoresis (MST) analysis of GFP free LysP. 109
2.4.5.4. Tryptophan Fluorescence Change 113
2.5. CONCLUDING REMARKS 115

CHAPTER 3: CRYSTALLIZATION AND PRELIMINARY X-RAY DIFFRACTION DATA OF A LYSINE SPECIFIC PERMEASE (LYSP) FROM *PSEUDOMONAS AERUGINOSA* STRAIN PAO1 116

3.1. INTRODUCTION 116
3.2. EXPERIMENTAL PROCEDURES 117
3.2.1. *In surfo* Crystallization Trials of LysP 117
3.2.2. *In Surfo* Crystallization Optimization of LysP 117
3.2.2.1. Precipitant Concentration Screening 117
3.2.2.2. Precipitant/Protein Drop Size Screening 118
3.2.2.3. Organic and Detergent Additive Screening 118
3.2.2.4. Addition of Polar *E. coli* Phospholipids 118
3.2.2.5. Optimization of LysP Using Similar Published Condition for Wild Type LacY. 119
3.2.3. *In Meso* Crystallization Trials of LysP 119
3.2.4. Bicelle Crystallization Trials of LysP/CA5755/L-lysine 120
3.2.6. *In Surfo* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764. 127
3.2.7. *In Surfo* Crystallization Optimization of LysP with Nanodody Clone CA5755 127
3.2.7.1. PEG 2000 MME Screening. 127
3.2.7.2. PEG 2000 MME and pH Screening. 127
3.2.7.3. $CaCl_2$ Concentration Screening 128
3.2.7.4. Precipitant/Protein Drop Size Screening. 128
3.2.7.5. Dehydration of LysP/CA5755/L-lysine Crystals in PEG 400. 128
3.2.7.6. Seeding of LysP/CA5755/L-lysine Crystals 129
3.2.8. *In Meso* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764. 129

3.2.9. X-ray Diffraction Studies 130
3.3. RESULTS 131
3.3.1. *In surfo* Crystallization Trials of LysP 131
3.3.2. *In surfo* Crystallization Optimization LysP 135
3.3.2.1. PRECIPITANT CONCENTRATION SCREENING 135
3.3.2.2. PRECIPITANT/PROTEIN DROP SIZE SCREENING 136
3.3.2.3. ORGANIC AND DETERGENT ADDITIVE SCREENING 136
3.3.2.4. ADDITION OF POLAR *E. COLI* PHOSPHOLIPID 137
3.3.2.5. OPTIMIZATION OF LYSP USING SIMILAR PUBLISHED CONDITION FOR WILD TYPE LACY 137
3.3.3. *In Meso* Crystallization Trials of LysP 138
3.3.4. Bicelle Crystallization Trials of LysP/CA5755/L-lysine 142
3.3.5. Nanobody Production and Screening for Tight Binders to LysP 142
3.3.6. *In Surfo* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764 145
3.3.7. *In Surfo* Crystallization Optimization of LysP with Nanodody Clone CA5755 146
3.3.7.1. PEG 2000 MME Screening 146
3.3.7.2. PEG 2000 MME and pH Screening 147
3.3.7.3. $CaCl_2$ Concentration Screening 147
3.3.7.4. Precipitant/Protein Drop Size Screening 150
3.3.7.5. Dehydration of LysP/CA5755/L-lysine Crystals in PEG 400 155
3.3.7.6. Seeding of LysP/CA5755/L-lysine Crystals 155
3.3.8. *In Meso* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764 156
3.3.9. X-ray Diffraction Studies of LysP with and without Nanobodies 156
3.4. CONCLUDING REMARKS 159

CHAPTER 4: GENERAL CONCLUSIONS AND DISCUSSIONS AND FUTURE PLANS 162

4.1. GENERAL CONCLUSIONS AND DISCUSSIONS 162
4.2. FUTURE PLANS 171

REFERENCES 173

List of Figures

Figure 1.1. Chemical structure of L-lysine and the analog, L-4-thialysine at pH 7.0 28

Figure 1.2. Transmembrane Helical Hidden Markov Model (TMHMM) prediction for LysP (http://www.cbs.dtu.dk/services/TMHMM/). 28

Figure 1.3. Sequence alignment of *Pseudomonas aeruginosa* PAO1 LysP with different transporter homologues from both prokaryote and eukaryote. 34

Figure 1.4. Phylogenetic tree of the APC family transporters. 37

Figure 1.5. Sequence alignment of *M. Jannaschii* ApcT with different homologues from the APC family transporter in both prokaryote and eukaryote. 38

Figure 1.6. *Pseudomonas aeruginosa*. 40

Figure 1.7. "Schematic representation of a biological membrane showing the lipid bilayer in and on which are situated a variety of proteins" 41

Figure 1.8. Mechanism of action of a secondary transporter 42

Figure 1.9. Classes of secondary transporters. 43

Figure 1.10. "Successful detergents used for a-helical membrane protein crystallization 45

Figure 1.11. Schematic representation of the LysP-GFP fusion construct. 47

Figure 1.12. Flow diagram showing the purification methods used for the GFP-fusion approach 48

Figure 1.13. Schematic phase diagram illustrating the solubility of a soluble protein as a function of precipitant concentration (McPherson 1999). 49

Figure 1.14. Cartoon representation of how membrane proteins crystallizes by the *in meso* method. 51

Figure 2.1. Signal peptide prediction using Neural Network (NN, http://www.cbs.dtu.dk). 56

Figure 2.2. Signal peptide prediction using Hidden Markov Models (HMM) (http://www.cbs.dtu.dk). 57

Figure 2.3. TopPred for LysP. 58

Figure 2.4. TMHMM prediction for LysP. 59

Figure 2.6. Topological diagram of LysP protein. 61

Figure 2.7. Topological diagram of LysP protein showing the distribution of methionines. 62

Figure 2.8. Topological diagram of LysP protein showing the distribution of positive residues.....63

Figure 2.9. Topological diagram of LysP protein showing the distribution of negative residues. 64

Figure 2.10. Topological diagram of LysP protein showing the distribution of tryptophans.....65

Figure 2.11. Screenshot for FoldIndex prediction of LysP. 66

Figure 2.12. RONN prediction for LysP with disordered regions: 1-11, 341–341..67

Figure 2.13. Homology modelling plot of LysP using ApcT as a template. 69

Figure 2.14. Sequence of LysP showing regions of similar secondary structures with ApcT.....70

Figure 2.13. 12 % NuPAGE SDS PAGE gels of GFP free purified LysP used for in-gel tryptic digestion.....77

Figure 2.14. Flow chart illustrating the steps involved in the radioactivity whole cell transport assay.....80

Figure 2.15. Flow chart showing the steps involved in the acetone-ether washing of polar *E. coli* lipid.....83

Figure 2.16. Flow chart showing the different steps involved in the reconstitution of LysP into proteoliposomes and the entrance counterflow assay.....84

Figure 2.17. Accessing the integrity of overexpressed LysP-GFP by in-gel fluorescence and GelCode Blue Safe protein staining. 87

Figure 2.18. Bicinchoninic acid assay (BCA assay) for the determination of GFP concentration. 88

Figure 2.19. GFP-His_8 standard curve (fluorescence excitation at 485 nm and emission at 512 nm versus mg/mL concentration of GFP-His_8). 88

Figure 2.20. Detergent solubilization yield of LysP-GFP in the different detergents tested.....89

Figure 2.21. Fluorescence detection size-exclusion chromatography (FSEC) traces of LysP-GFP fusion in different detergents.....90

Figure 2.22. FSEC of LysP-GFP in the top 4 detergents (DDM, NM, LMNG and DM).....91

Figure 2.23. Purification of LysP. SDS PAGE of samples at every step in the purification process (left) and in-gel fluorescence of samples at every step in the purification process. 92

Figure 2.24. SEC chromatogram of TEV cleaved free LysP.....93

Figure 2.25. Size-exclusion chromatography (SEC) of TEV cleaved free LysP.....94

Figure 2.26. Purity evaluation and molecular weight estimation of GFP free LysP. 96

Figure 2.27. Spectrum analyses of GFP free LysP from MALDI-TOF MS after in-gel tryptic digesting.....97

Figure 2.28. In-gel tryptic digestion of LysP followed by MALDI-TOF MS. 98

Figure 2.29. Glutaraldehyde crosslinking of TEV cleaved free LysP. 100

Figure 2.30. Stability studies of TEV cleaved free LysP.....101

Figure 2.31. CD spectroscopy of GFP free LysP.....102

Figure 2.32. Whole cell transport of [^{3}H] L-lysine into *E. coli* C43 (DE3) by LysP-GFP.....103

Figure 2.33. Whole cell transport inhibition of [^{3}H] L-lysine into *E. coli* C43 (DE3) by LysP-GFP. 104

Figure 2.34. Whole cell transport assay of LysP-GFP.....105

Figure 2.35. Whole cell assay pH screening of [^{3}H] L-lysine transport into induced *E. coli* C43 (DE3) harboring LysP-GFP.....106

Figure 2.36. [^{3}H] L-lysine counterflow uptake by reconstituted GFP free LysP. .107

Figure 2.37. Transport inhibition of LysP reconstituted into proteoliposomes. ...108

Figure 2.38. Sodium gradient test for GFP free LysP. 109

Figure 2.39. MST analysis of free LysP in Buffer B from Section 2.3.5.....110

Figure 2.42. Emission spectrum for tryptophan fluorescence change of LysP in detergent solution.....114

Fugure 3.1. Structure of an antibody and a nanobody.....121

Figure 3.2. Mechanism of biotinylation of GFP free LysP with EZ-link Sulfo NHS-LC-LC-Biotin. 123

Figure 3.3. A Schematic overview of strategies to clone and select the VHH genes against LysP from an immunised dromedary or llama.....124

Figure 3.4. Amino acid sequence of the different nanobodies raised against TEV cleaved GFP free LysP.....125

Figure 3.5. A flow chart summarizing the progress through the crystallogenesis process of GFP free LysP. 131

Figure 3.6. Initial *in surfo* crystallisation hits of LysP obtained by vapour diffusion using the MemGold screen kit. 133

Figure 3.7. Silver staining of *in surfo* crystals of LysP and SDS PAGE analysis of the effect of heating on purified GFP free LysP.....135

Figure 3.8. *In surfo* srystals of LysP in the presence of 0.5 %(v/v) 1-propane diol as an organic additive. 136

Figure 3.9. *In surfo* crystals of LysP obtained using similar published condition for wild type LacY. 138

Figure 3.10. *In meso* crystallization trials of GFP free LysP showing precipitated protein in the bolus of cubic phase. 139

Figure 3.11. Example of *in surfo* salt crystals obtained in the bicelle crystallization trials at 20 °C. 142

Figure 3.12. Purification of nanobodies against GFP free LysP. 143

Figure 3.13. SEC of LysP/nanobody complex screening. 144

Figure 3.14. SDS PAGE analysis of LysP/nanobody screening. 145

Figure 3.15. Initial *in surfo* crystallization hit of LysP in the presence of nanodody clone CA5755 and L-lysine. 146

Figure 3.16. *In surfo* crystallization optimization of LysP/CA5755/L-lysine by screening PEG 2000 MME concentration. 147

Figure 3.17. *In surfo* $CaCl_2$ screening of LysP/CA5755/L-lysine. 148

Figure 3.18. *In surfo* $CaCl_2$ screening of LysP/CA5755/L-lysine. 148

Figure 3.19. *In surfo* $CaCl_2$ screening of LysP/CA5755/L-lysine. 149

Figure 3.20. *In surfo* $CaCl_2$ screening of LysP/CA5755/L-lysine. 150

Figure 3.21. *In surfo* optimized crystals of LysP/nanobody CA5755 complex. 151

Figure 3.22. *In surfo* optimized crystals of LysP/nanobody CA5755 complex. 151

Figure 3.23. *In surfo* optimized condition of LysP/nanobody CA5755 complex. 152

Figure 3.24. *In surfo* optimized condition of LysP/nanobody CA5755 complex. 152

Figure 3.25. *In surfo* optimized condition of LysP/nanobody CA5755 complex. 153

Figure 3.26. *In surfo* optimized crystals of LysP/nanobody CA5755 complex. 153

Figure 3.27. *In surfo* optimized crystals of LysP /nanobody CA5755 complex. 154

Figure 3.28. *In surfo* optimized condition of LysP/nanobody CA5755 complex. 154

Figure 3.29. *In surfo* optimized condition of LysP/nanobody CA5755 complex. 155

Figure 3.30. X-ray diffraction pattern of LysP at 7 Å with reflections to 6 Å. 157

Figure 3.31. Diffraction pattern of LysP/CA5755/L-4-thialysine crystals. 158

Figure 4.1. Chemical structure of the maltose neopentyl glycol detergent and the conventional detergent counterpart 163

Figure 4.2. "SDS-PAGE and SEC analysis of M_2-T4L (human M_2 muscarinic acetylcholine receptor. 164

Figure 4.3. "Correlation between structural resolution and the type of crystallization detergents" (Sonoda *et al.* 2010). 167

List of Tables

Table 1.I. Known structures of amino acid transporters. 30

Table 1.II. Related proteins of LysP. 32

Table 2.I. Amino acid composition of LysP 55

Table 2.II. NetPhosphorylation site for LysP. 68

Table 2.III. Whole cell GFP fluorescence count of LysP-GFP. 86

Table 2.IV. List of detergents used, their CMC in water and their aggregation number in water. 90

Table 2.V. Summary of dissociation constant (K_d) of the different amino acids tested for GFP free LysP by MST analysis. 112

Table 3.I. Initial sitting drop crystallization hits of LysP using the MemGold screen (Molecular Dimensions). 132

Table 3.II. Initial crystallization hits of LysP obtained by sitting drop on a MemGold crystallization screen kit. 134

Table 3.III. Summary of some *in meso* crystallization conditions which resulted in precipitation of LysP. 139

List of Materials

Materials	Company	Catalogue number	Lot / Batch number
1.5 mL Eppendorf tube	Sarstedt (Nümbrecht, Germany)	72.690.001	0000/0355001
15 mL Falcon tubes	Sarstedt (Nümbrecht, Germany)	62.559.001	1040601
-20 °C freezer	Liebherr (PA, USA)	GS 1301	N/A
20 °C plate incubator	Rumed (Leipzig, Germany)	Rubarth apparate	N/A
35 mm Petri dish	Greiner Bio One (Frickenhausen, Germany)	627161	N/A
4 ° C refrigerator	Microban (Staffordshire, UK)	Hotpoint RLA81	N/A
4 °C FPLC cold cabinet	Medical Supply (Ireland)	Ekofrigolab 1500	N/A
4 °C room	Aircold manufacturing (Ireland)	MasterCella - CAREL	N/A
42 °C water bath	Fisher scientific (MA, USA)	DMU5	N/A
7.7 MAG	N/A	N/A	JPL-4-92
7.8 MAG	N/A	N/A	TasA21
-80 °C freezer	Medical Supply (Ireland)	Platinum 500	
9.9 MAG (Monoolein)	Nuchek (MN, USA)	M239	M239-F15-L; M239-A21-Q
Acetone	Fisher scientific (MA, USA)	A/0606/17	959939
Agar	Merck (NJ, USA)	1.01613	VM135213
Agarose	Sigma (MO, USA)	A5304	108K1012
AKTA Explorer	GE Healthcare (Buckinghamshire , UK)	AKTA Explorer	N/A
AKTA FPLC	GE Healthcare (Buckinghamshire , UK)	AKTA FPLC	N/A
Autoclave	Rodwell (Essex, UK)	Crystal MP 25	N/A
Balance (0.1 mg – 64 g)	Sartorius (Goettingen, Germany)	CP64	N/A
Balance (0 – 3,100 g)	Mettler Toledo (OH, USA)	B300T-S	N/A
Balance (0.01 mg – 220 g)	Mettler Toledo (OH, USA)	AX205	N/A
Balance (for over 1 g)	Mettler Toledo (OH, USA)	B300T-S	N/A

List of Materials Continued

Benchtop minicentrifugre (0 – 16,300 g)	Spectrafuge (TX, USA)	24D	N/A
Biosafety cabinet	Labculture (Dorset, UK)	ESCO Class II Type A2	N/A
Bis-Tris	Fluka (MO, USA)	14880	1240589
Bottle	Duran Group (Main, Germany)	50 mL, 100 mL, 200 mL, 1 L, 2 L	N/A
Centricon 50 kDa MWCO	Millipore (MA, USA)	UFC905024	R0KA67379
Centrifuge	Beckman Coulter (Clare, Ireland)	Avanti J-E	N/A
Centrifuge rotor (0 – 10,000 rpm)	Beckman Coulter (Clare, Ireland)	C0650	N/A
Centrifuge rotor (0 – 9,000 rpm)	Beckman Coulter (Clare, Ireland)	JLA 9.1000	N/A
Chemical fume hood	Cameron (Longford, Ireland)	MACH-AIRE	N/A
Crystal Screen 1 crystallisation screen	Hampton Research (CA, USA)	HR2-110	211080
Crystal Screen 2 crystallisation screen	Hampton Research (CA, USA)	HR2-112	211251
Crystallisation plate imager	Formulatrix (MA, USA)	Rock Imager 1500	N/A
Culture flasks	Schott Duran (Wertheim, Germany)	Nalgen baffled shake flask: 250 mL, 3.5 L, 5 L	N/A
DDM	Anatrace (CA, USA)	D310S	4138324
DM	Anatrace (CA, USA)	D322	126862
Dnase I	Sigma (MO, USA)	ND25	056K7680
D-Lysine	Sigma (MO, USA)	L8021	034K5309
dNTPs mixture	Greiner Bio One (Frickenhausen, Germany)	201900	N/A
E. coli (DE3) C43	Gift from Dr. David Drew at Imperial College London	N/A	N/A
E. coli DH5α competent cells	NEB (MA, USA)	C2977H	1211103
E. coli Top10 competent cells	Invitrogen (Dun Laoghaire, Ireland)	C404003	825023
Electrophoresis chamber	Bio-Rad (Hertfordshire, UK)	XCell SureLock	N/A
Electrophoresis power supply	Biorad (Wicklow, Ireland)	Powerpac basic	N/A

List of Materials Continued

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
Freezer (- 20 °C)	Liebherr (PA, USA)	GS 1301	N/A
Freezer (- 80 °C)	Medical Supply (Ireland)	Platinum 500	N/A
French press	Fisher scientific (MA, USA)	FA-078A	N/A
Gel doc system	Alpha Innotech (CA, USA)	Fluorchem SP MultiImage II	N/A
Glycine	Merck (Darmstadt, Germany)	K32086201	1.04201.0100
Gelcode Blue Safe (protein stain)	Thermo scientific (MA, USA)	24594	LG147135
Glycerol	Melford (Suffolk, UK)	G1345	56-81-5
Hiload 16/60 S200 SEC column	GE Healthcare (Buckinghamshire, UK)	17-1069-01 (lot. 433047)	N/A
Homogenizer (probe)	Griffin George (Leicestershire, UK)	LC9	N/A
Homogenizer (tube)	Wheaton (NJ, USA)	30 Ml tube	N/A
HotStar Hifidelity DNA polymerase	Qiagen (Hamburg, Germany)	202602	1032734
HPLC fluorescence reader	Dionex (Dubin, Ireland)	RF-2000	N/A
Icemaker	Scotsman (Cumbria, UK)	AF 80	N/A
IMAC Sepharose High Performance Matrix	GE Healthcare (Buckinghamshire , UK)	17-0920-06	10029650
PBS	Sigma (MO, USA	79382	0001434519
Incubator (Room temperature to 60 °C)	Labnet (New Jersey, US)	Mini incubator	N/A
Incubator (0 °C to 60 °C)	Rumed (Laatzen, Germany)	1001	N/A
Incubator	INFORS HT (Bottmingen, Switzerland)	Multitron	N/A
Inoculation loop	Greiner Bio One (Frickenhausen, Germany)	1 µL	N/A

List of Materials Continued

IPTG	Melford (Suffolk, UK)	MB1008	A21345
Kanamycin	Melford (Suffolk, UK)	K0126	18022
L-Lysine	Sigma (Wicklow, Ireland)	62840	1370953
L-Leucine	Fluka (Steinheim, Germany)	61819	1347314
L-Prolin	Fluka (Steinheim, Germany)	81710	415931/1
L-Histidine	Fluka (Steinheim, Germany)	53320	1365390
L-Threonine	Fluka (Steinheim, Germany)	89180	1367095
L-Glutamic acid	Fluka (Steinheim, Germany)	494450	1280359
L-Valine	Fluka (Steinheim, Germany)	94620	1352709
L-Arginine	Fluka (Steinheim, Germany)	11010	1326295
L-Glutamine	Fluka (Steinheim, Germany)	49419	455300/1
L-Isoleucine	Fluka (Steinheim, Germany)	58879	1386777
L-Methionine	Sigma (Steinheim, Germany)	M5308	057K0088
L-4-thialysine	Sigma (Wicklow, Ireland)	A2636	BCBC0890V
LB High Salt Powder	Fluka (MO, USA)	51208	BCBC5118
LDAO	Sigma (MO, USA)	40236	BCBB8394
Lysozyme	Sigma (MO, USA)	L-6876	094K1454
Magnetic stirrer	Stuart (Staffordshire, UK)	SB 162	N/A
MNG DDM	Anatrace (CA, USA)	NG310	4145525

List of Materials Continued

MNG DM	Anatrace (CA, USA)	NG322	4134357
MNG OG	Anatrace (CA, USA)	NG311	4133790
Gel filtration calibration kit HMW	GE Healthcare	28-4038-42	377200
MembFac crystallization screen	Hampton Research (CA, USA)	HR2-114	211423
MemGold crystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-39	004-1-39
MemPlus crystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-45	BN002-1-45
MemStart crystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-21	011-1-21
MemSys crystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-25	010-1-25
MES	Sigma (MO, USA)	M3671	068K5421
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
$MgCl_2$	Sigma-Aldrich (MO, USA)	M2670	129K00891
Microscope	Nikon (Surrey, UK)	SMZ1500	N/A
MilliQ water (resistance at 18.2 mΩ)	Produced by UV5 Elix system	N/A	N/A
NaCl	Fisher scientific (MA, USA)	BP358-10	80786
NG	Anatrace (CA, USA)	N324	119057
NM	Anatrace (CA, USA)	N330	NM14
Nanodrop (spectrophotometer)	Mason Technology (Dublin, Ireland)	Nanodrop 1000 Spectrophoto meter	N/A
NaOH	Merck (NJ, USA)	8627	B382182 411
OG	Anatrace (CA, USA)	O311S	126780
PACT premier rystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-36	BN002
PCR machine	Eppendorf (Cambridge, UK)	Mastercycler gradient	N/A

List of Materials Continued

PCR tubes	Sarstedt (Nümbrecht, Germany)	72.737.002	52501
pH meter	Thermo scientific (MA, USA)	Orion 3 star	N/A
Pipettes	Gilson (WI, USA)	Pipetman kit (2 to 1000 µL pipettes)	N/A
PEG 400	Fluka (Steinheim, Germany)	81172	1421464
Primers and sequencing	MWG Biotech (Hamburg, Germany)	N/A	N/A
Probe sonicator	Bandelin (CT, USA)	Sonoplus HD 200	N/A
Progad filter for MilliQ Elix 5 system	Millipore (MA, USA)	PR0G00052	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
pWaldo GFPd vector	Gift from Dr. David Drew, Imperial College London	N/A	N/A
QiaPrep Spin mini prep kit	Qiagen (Hamburg, Germany)	27106	136231546
Refrigerator (4 °C)	Microban (Staffordshire, UK)	Hotpoint RLA81	N/A
Robot for *in meso* crystallisation	Sias AG (Hombrechtikon, Switzerland)	Xantus *in meso* 8-tip	N/A
Robot for *in surfo* crystallisation	TTP Labtech (Hertfordshire, UK)	Mosquito	N/A
SDS	Fluka (MO, USA)	71725	1344034

List of Materials Continued

SDS-PAGE apparatus	Invitrogen (Dun Laoghaire, Ireland)	Novex Mini-Cell	N/A
SDS-PAGE Bis-Tris gels	Thermo scientific (MA, USA)	84712	100407117
Shaking incubator (Room temperature to 80 °C)	AGB Scientific Ltd (Dublin, Ireland)	Sanyo Orbital Incubator	N/A
Simpack filter for MilliQ Elix 5 system	Millipore (MA, USA)	SIMPAK0RT	N/A
Simpack filter for MilliQ Elix 5 system	Millipore (MA, USA)	SIMPAK0RT (lot. F9HN06031)	N/A
SOC medium	NEB (MA, USA)	B9020S	1211007
Spacer	Saunders (CA, USA)	3M TAPE 9500PC 77MM	N/A
Spectrophotometer	Uvikon (Leeds, UK)	XL	N/A
Structure screen I rystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-30	BN016
Structure screen II rystallization screen	Molecular Dimensions (Suffolk, UK)	MD1-30	BN016
Sybersafe	Invitrogen (Dun Laoghaire, Ireland)	S33102	643459
Syringe for *in meso* crystallisation	Hamilton	Hamilton 100 µL syringe	N/A
TEV protease vector	Gift from David Drew at Imperial College London	N/A	N/A
Tabletop minicentrifuge	Spectrafuge (TX, USA)	24D	N/A
TOPO cloning kit	Invitrogen (Dun Laoghaire, Ireland)	450172	801495
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
Trizma® base	Sigma (MO, USA)	T87602	088K0118
Ultracentrifuge	Beckman Coulter (Clare, Ireland)	Optim L-100 XP	N/A
Ultracentrifuge rotor (0 – 70,000 rpm)	Beckman Coulter (Clare, Ireland)	70 Ti	N/A

List of Materials Continued

Vortex mixer	Nichel Electro Ltd (North Somerset, England)	Clifton Cyclone CM-1	N/A
Wizard I crystallisation screen	Emerald Biosystems (WA, USA)	EBS-WIZ-1	EBS000322 200499
Wizard II crystallisation screen	Emerald Biosystems (WA, USA)	EBS-WIZ-2	EBS000322 200499
Wizard III crystallisation screen	Emerald Biosystems (WA, USA)	EBS-WIZ-3	EBS000061 52009299

Abbreviations

Å	Årgström
APS	Advanced Photon Source
ATP	Adenosine Triphosphate
BCA	Bicinchoninic Acid Assay
BSA	Bovine Serum Albumin
CCD	Charge-Coupled Device
CMC	Critical Micelle Concentration
C-terminal	Carboxyl-terminal
DDM	n-dodecyl-β-D–maltoside
DM	n-decyl-β-D–maltoside
DMNG	Decyl maltose neopentyl glycol
DNA	Deoxyribonucleic Acid
dNTPs	Deoxy-Nucleotide-Triphosphates
E. coli	*Escherichia coli*
FPLC	Fast Protein Liquid Chromatography
FSEC	Fluorescent Size-Exclusion Chromatography
GFP	Green Fluorescent Protein
GPCR	G Protein-Coupled Receptor
HEPES	4-(2-Hydroxyethyl)-1- Piperazineethanesulfonic Acid
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl- β-D-Thiogalactopyranoside
K_d	Dissociation Constant
kDa	Kilo Daltons
LB	Luria Bertani
LDAO	N,N-Dimethyldodecylamine-N-Oxide
LMNG	Lauryl maltose neopentyl glycol
MAG	Monoacylglycerol
MALDI-TOF-MS	Matrix-Assisted Laser Desorption/Ionization–Time of Flight Mass Spectrometry
MES	2-(N-Morpholino) Ethanesulfonic Acid
MOPS	3-(N-Morpholino) Propanesulfonic Acid
MPDB	Membrane Protein Data Bank

MWCO	Molecular Weight Cut-Off
NG	n-nonyl-β-D-glucopyranoside
Ni^{2+}-NTA	Nickel-Nitrilotriacetic Acid
NM	n-nonyl-β-D-maltoside
NMR	Nuclear Magnetic Resonance Spectroscopy
NN	Neural Networks
N-terminal	Amino-terminal
OD	Optical Density
OG	n-octyl-β-D-glucopyranoside
OGNG	Octyl glucoside neopentyl glycol
OTG	Octylthioglucoside
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
P. aeruginosa	*Pseudomonas aeruginosa*
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
PEG MME	Polyethylene Glycol Monomethylether
PL	Phospholipid
RONN	Regional Order Neural Network
SDS	Sodium Dodecyl Sulphate
TEV	Tobacco Etch Virus
TMHMM	Transmembrane Helical Hidden Markov Model
Tris	Tris-(Hydroxymethyl) Aminomethane
UV	Ultra Violet
v/v	Volume per Volume
Vo	Void Volume
Vt	Included Volume
w/w	Weight per Weight

Amino Acid Symbols

Amino acid	One letter symbol	Three letter symbol
alanine	A	Ala
arginine	R	Arg
asparagine	N	Asn
aspartic acid	D	Asp
cysteine	C	Cys
glutamic acid	E	Glu
glutamine	Q	Gln
glycine	G	Gly
histidine	H	His
isoleucine	I	Ile
leucine	L	Leu
lysine	K	Lys
methionine	M	Met
phenylalanine	F	Phe
proline	P	Pro
serine	S	Ser
threonine	T	Thr
tryptophan	W	Trp
tyrosine	Y	Tyr
valine	V	Val

Chapter 1: General Introduction

1.1. Lysine Specific Permease (LysP)

P. aeruginosa strain PAO1 lysine specific permease (LysP) (Gene ID: 881192, Uniprot ID: Q9HVG3) belongs to the amino acid-polyamine-oganocation (APC) family of transporters. LysP is predicted to have 12 transmembrane segments joined by 5 perisplasmic and 6 cytoplasmic loops with both C and N terminus located in the cytoplasm (**Figure 1.2**). *P. aeruginosa* and other bacteria take up L-lysine from their surrounding and utilises it as their source of carbon and nitrogen (Weill-Thevenet *et al.* 1979). Reduced carbon is used to supply energy and nitrogen is one of the basic building block of biomolecules. There are two transport systems for lysine in *E. coli*: the lysine-arginine-ornitine (LAO) system (Rosen, 1971 & 1973) and the lysine specific permease (LysP) system (Steffes *et al.* 1992). The latter is inhibited by the S-(β-aminoethyl)-L-cysteine (L-4-thialysine) (Steffes *et al.* 1992, Tanaka & Soda 1974) which is an analogue of L-lysine (Figure 1.1. A) in which the γ methylene group in the side chain is replaced with a sulphur atom (**Figure 1.1. B**). This thesis describes the overexpression, detergent selection, purification, crystallization, biochemically, biophysical and functionally characterization of LysP from LysP-GFP fusion (**Figure 1.12**). This study will aid the pursuit of a high resolution crystal structure and a detailed functional analysis of LysP. The crystal structure of LysP could be used to solve the structures of other related transporters by homology modeling and also to provide insight into lysine transport and selectivity, the results of this study is germaine for the amino acid-polyamine-oganocation (APC) family transporters homologues in both prokaryote and eukaryote. The structures of these transporters solved by homology modeling will aid in the rational design of drugs to target these transporters. Also, the solved structures will be used as to solve other structures by molecular replacement.

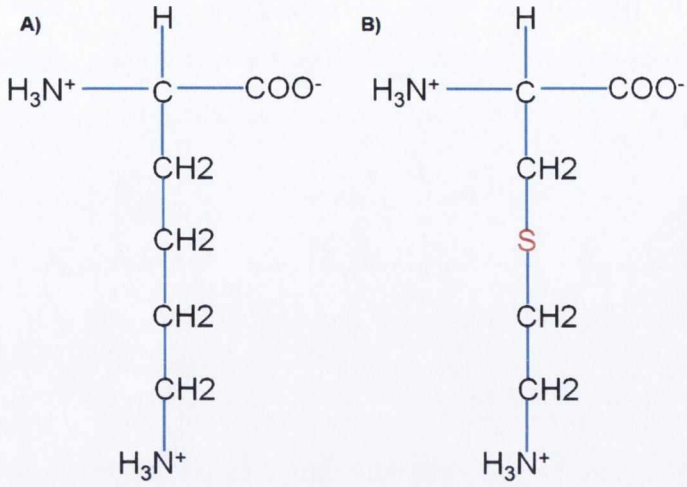


Figure 1.1. Chemical structure of L-lysine and the analog, L-4-thialysine at pH 7.0

A) Chemical structure of L-lysine with a series of unbranched methylene groups, B) Chemical structure of S-(β-aminoethyl)-L-cysteine (L-4-thialysine) (the γ-methylene group in the side chain of L-lysine is replaced with a sulphur atom).

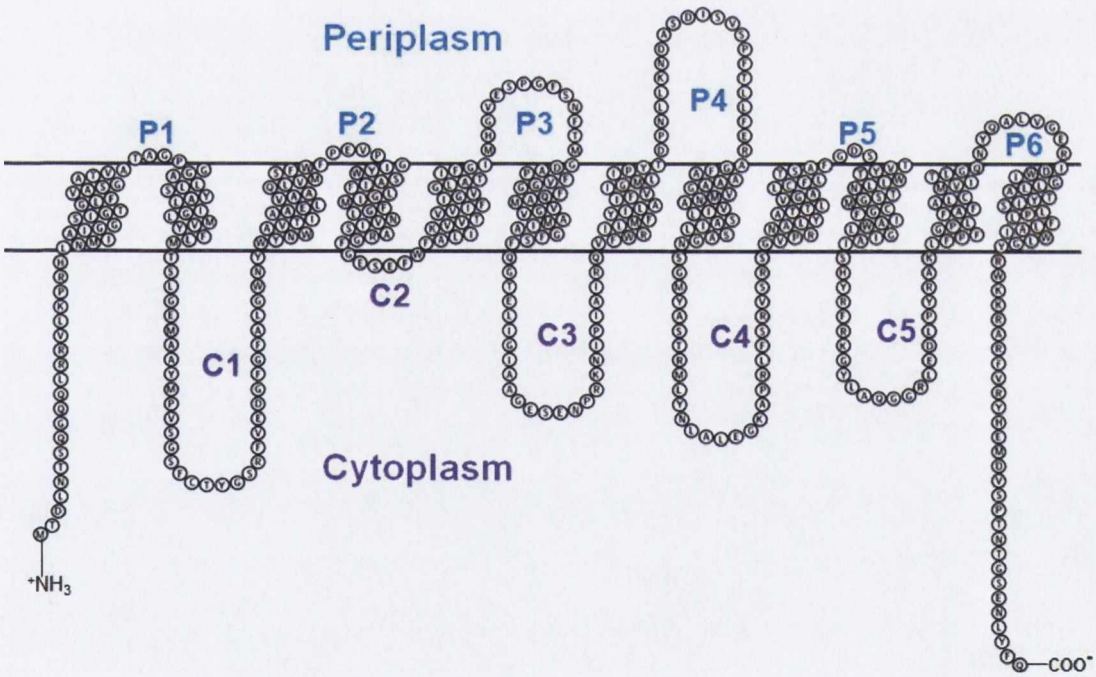


Figure 1.2. Transmembrane Helical Hidden Markov Model (TMHMM) prediction for LysP (http://www.cbs.dtu.dk/services/TMHMM/).

TMHMM predicts 12 transmembrane segments joined by 5 periplasmic and 6 cytoplasmic loops. Both C-and N-terminus are predicted to be in the cytoplasm (Krogh *et al.* 2001 and

Sonnhammer *et al.* 1998). Figure was drawn with the help of Dr. Lyons using T(E)Xtopo (http://homepages.uni-tuebingen.de/beitz/). T(E)Xtopo is a program used for plotting topology data from the SwissProt database files into shaded plots ready for publication using LaTex (Beitz 2000).

There are other amino acid transporters whose structures have already been solved. They include: the Na+-independent Amino Acid Transporter (ApcT) from *Methanocaldococcus jannaschii* (Shaffer *et al.* 2009), the leucine transporter (LeuT) from *Aquifex aeolicus* (Yamashita *et al.* 2005), the glutamate transporter (Glt) from *Pyrococcus horikoshii* (Yernool *et al.* 2004), the arginine-agmantine antiporter (Gao *et al.* 2009) and the aspartate transporter (Glt) from *P. horikoshii* (Boudker *et al.* 2007). Thus, lessons learned in solving the structures of these proteins may be applied to solving the structure of LysP. A database of amino acid transports of known structures has been created (**Table 1.I**).

Table 1.I. Known structures of amino acid transporters. The crystallization method for all was vapour diffusion. All tags were His tag except ApcT which had GFP-His tag. LeuT is from *Aquifex aeolicus*, GltPh is from *Pyrococcus horikoshii* and ApcT is from *M. jannaschii. M. jannaschii* is a thermophilic methane-producing bacterium first isolated from hot spring at Woods hole (Jeanthon *et al.* 1998, Bult *et al.* 1996). All other proteins are from *Escherichia coli*. All proteins have 12 transmembrane helices except GltPh which has 8. There is no human amino acid transporter solved at the time this table was created (1st of October 2012).

Protein	Organism	Solubilization detergent	Crystallization detergent	Precipitant used for crystallization	Protein concentration (mg/mL)	Phasing method	Space group	Diffraction (Å)	PDB ID	Reference
Aspartate transporter (GltPh)	*P. horikoshii*	DM	DM	PEG 1000	7-10	Se-Met, heavy atom soaking	P61	2.96	2nwl	1
Arginine:Agmatine Antiporter (AdiC)	*E. coli*	NG	NG	PEG 400	10	Se-Met	P212121	3.6	3h5m	2
apo Na+-independent Amino Acid Transporter (ApcT)	*M. jannaschii*	DDM	C8SG	PEG 550 MME	5-10	Se-Met	C2, P21	2.35	3gia	3
Glutamate transporter (GltPh)	*P. horikoshii*	DM	DM	PEG 1000	7-10	Se-Met, heavy atom soaking	P61	3.5	1xfh	4
Leucine transporter (LeuT)	*A. aeolicus*	DDM	OG	PEG 550 MME	7	Se-Met	C2	1.65	2a65	5

The references are: 1(Boudker *et al.* 2007), 2(Gao *et al.* 2009), 3(Shaffer *et al.* 2009), 4(Yernool *et al.* 2004) and 5(Yamashita *et al.* 2005).

1.2. Amino Acid Polyamine Organocation (APC) Transporter Family

The amino acid-polyamine-oganocation (APC) family of transporters are a group of transporters that belongs to a large family of secondary active transporters that catalyse the uniport, symport and antiport of a wide range of substrate across the membrane bilayer (Jack *et al.* 2000). They are very important in the uptake of nutrients, recycling of nutrients, ionic homeostasis and the regulation of cell volume (Shaffer *et al.* 2009). LysP belongs to the amino acid-polyamine-oganocation (APC) family (Jack *et al.* 2000, www.membranetransport.org). It shares sequence similarity with other proteins in both APC family and cationic amino acid transporter (CAT) family in both prokaryote and eukaryote (Reizer *et al.* 1992, Pi *et al.* 1993 and Shaffer *et al.* 2009). The APC family proteins include: lysine specific permease in *E. coli* (LysP-Ec) (Steffes *et al.* 1992), phenylalanine permease *E. coli* (PheP-Ec) (Pi *et al.* 1991), gamma aminobutyrate permease *E. coli* (GabP-Ec) (Bartsch *et al.* 1991), aromatic amino acid permease *E. coli* (AroP-Ec) (Honore & Cole 1990), probable transport protein in *E. coli* (YifK-Ec) (Daniels *et al.* 1992), arginine permease in *Saccharomyces cerevisiae* (Can1p-Sc) (Hoffmann 1985), proline permease in *S. cerevisiae* (PUT4-Sc) (Vandenbol *et al.* 1989), histidine permease in *S. cerevisiae* (Hip1-Sc) (Tanaka & Fink 1985), general amino acid permease in *S. cerevisiae* (GaP1-Sc) (Jauniaux & Grensen 1990), the sodium independent amino acid transporters from *Methanocaldococcus jannaschii* (ApcT-Mj) (Shaffer *et al.* 2009) and the CAT family proteins include: the high affinity cationic amino acid transporter in *Homo sapiens* (ecotropic retroviral leukamia receptor or human retroviral receptor (CAT1-Hs) (Albritton *et al.* 1989), low affinity cationic amino acid transporter 2 in *H. sapiens* or T-cell early activation factor protein (CAT2-Hs) (Macleod *et al.* 1990), the cationic amino acid transporter 4 in *H. sapiens* (CAT4-Hs) (Yoshimoto *et al.* 1991). The percentage identity of *P. aeruginosa* LysP compared with the above mentioned proteins are shown in **Table 1.I**, the sequence alignment to identified key residues that could be vital for function are shown **Figure 1.3** and the phylogenetic tree is given in **Figure 1.4.** This study was based on studies carried out by Pi *et al.* 1993, who showed that *E. coli* LysP share sequence similarities with other amino acid transporters of the APC family and also, Reizer *et al.* 1992 and Shaffer *et al.* 2009, showed that high affirnity cationic amino acid transporter 1 in *Homo sapiens* (ecotropic retroviral leukamia receptor or human retroviral receptor (CAT1-Hs) and low affinity cationic amino acid transporter 2 in *H. sapiens* also

known as T-cell early activation factor protein (CAT2-Hs) which are members of the CAT family share sequence similarity with amino acid transporters of the APC family which LysP belongs. Although their work was based on *E. coli* LysP, all the targets were combined and the sequence alignment was done with *P. aeruginosa* PAO1 LysP which is 70 % identical to *E. coli* LysP and also *M. jannaschii* ApcT was included. Reizer *et al.* suggested that the amino acid transporter of the APC family may also function as viral receptors probably because of their sequence similarity with the viral receptor protein (CAT1). The following residues were conserved in *P. aeruginosa* LysP but are located in different positions in the other homologues: Leucine 16, Glycine, 27, 28, 31, 33, 173 and 219, Glutamate 70, 155, 221, 229 and 233, Proline 264, Tyrosine 252 and Arginine 320. There was a conserved GWN motif at position 98, 99 and 100 respectively. Mutagenesis of these residues in the future will be carried out in these targets to investigate those that are vital for function.

Table 1.II. Related proteins of LysP.

Their sources are shown in column 2, their percentage sequence identity with P. aeruginosa PAO1 LysP are given in column 3 and their function is given in column 4. Percentage identity was calculated and arranged with **http://www.ebi.ac.uk** using ClustalW2 programme.

Protein	**Source**	**Percentage sequence identity with *P. aeruginosa* strain (PAO1) LysP**	**Function**
Lysine specific permease (LysP-Ps)	*P. aeruginosa*	100	Lysine transporter
Lysine specific permease (LysP-Ec)	*E. coli*	70.0	Lysine transporter
Arginine permease (Can1p-Sc)	*S. cerevisiae*	38.0	Arginine transporter
General amino acid permease (GAP1-Sc)	*S. cerevisiae*	36.0	General amino acid transporter

Table 1.II continued.

Aromatic amino acid permease (AroP-Ec)	*E. coli*	35.0	Aromatic amino acid transporter
Probable transport protein (YifK-Ec)	*E. coli*	35.0	Unknown
Phenylalanine permease (PheP-Ec)	*E. coli*	35.0	Phenylalanine transporter
Proline permease (PUT4-Sc)	*S. cerevisiae*	32.0	Proline transporter
Gamma aminobutyrate permease (GabP-Ec)	*E. coli*	31.0	Gamma aminobutyrate transporter
Histidine permease (HIP1-Sc)	*S. cerevisiae*	30.0	Histidine transporter
Sodium independent amino acid transporter (ApcT-Mj)	*M. jannaschii*	19	Broad range amino acid transporter
Cationic amino acid transporter 4 (CAT4-Hs)	*H. sapiens*	19.0	Cationic amino acid transporter
Low affinity cationic amino acid transporter 2 (T-cell early activation factor protein) (CAT2-Hs)	*H. sapiens*	18.0	Cationic amino acid transporter and T-cell early activation factor protein
High affinity cationic amino acid transporter (ecotropic retroviral leukamia receptor or human retroviral receptor (CAT1-Hs)	*H. sapiens*	16.0	Cationic amino acid transporter and ecotropic retroviral leukamia receptor or human retroviral receptor)

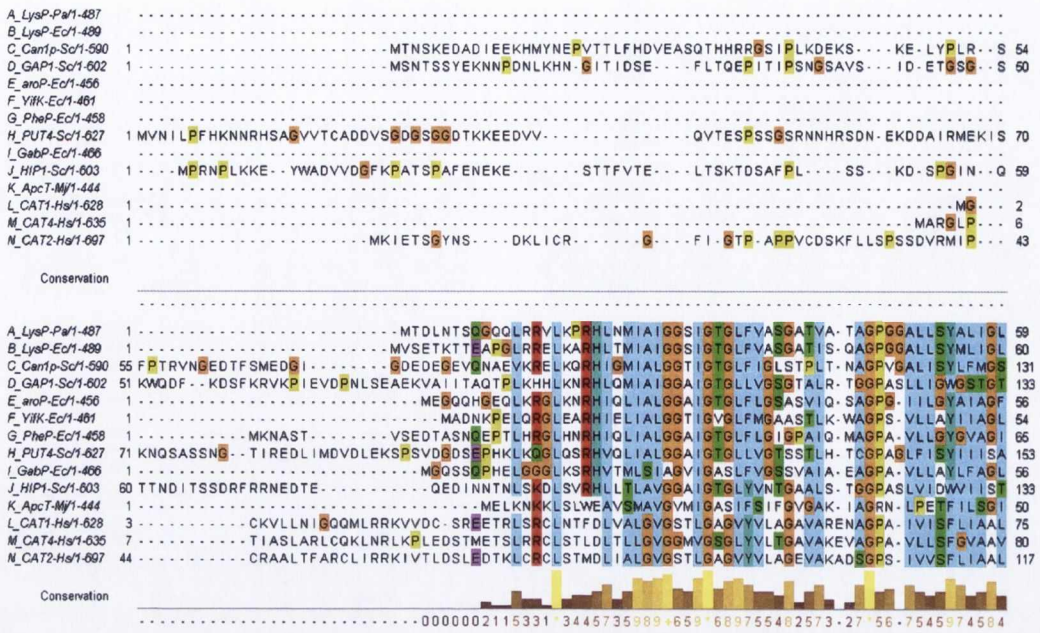


Figure 1.3. Sequence alignment of *Pseudomonas aeruginosa* PAO1 LysP with different transporter homologues from both prokaryote and eukaryote.

Alignment was done using the web link http://www.ebi.ac.uk using the programme called Clustal Omega. The dash lines represent gaps in the sequence as a result of the fact that some sequences are longer than others, the coloured amino acids represent those that are conserved in all species, the numbers at the sides are numberings of the position of the amino acids in the individual protein sequence. Protein name, source organisms and sequence length are shown on the left of the alignment. Protein names are listed from A to N to organize based on percentage identity to *P. aeruginosa* LysP. The following residues were conserved in *P. aeruginosa* LysP but are located in different positions in the other homologues: Leucine 16, Glycine, 27, 28, 31, 33, 173 and 219, Glutamate 70, 155, 221, 229 and 233, Proline 264, Tyrosine 252 and Arginine 320. There was a conserved GWN motif at position 98, 99 and 100 respectively. The conservation score are shown at the bottom of the alignment.

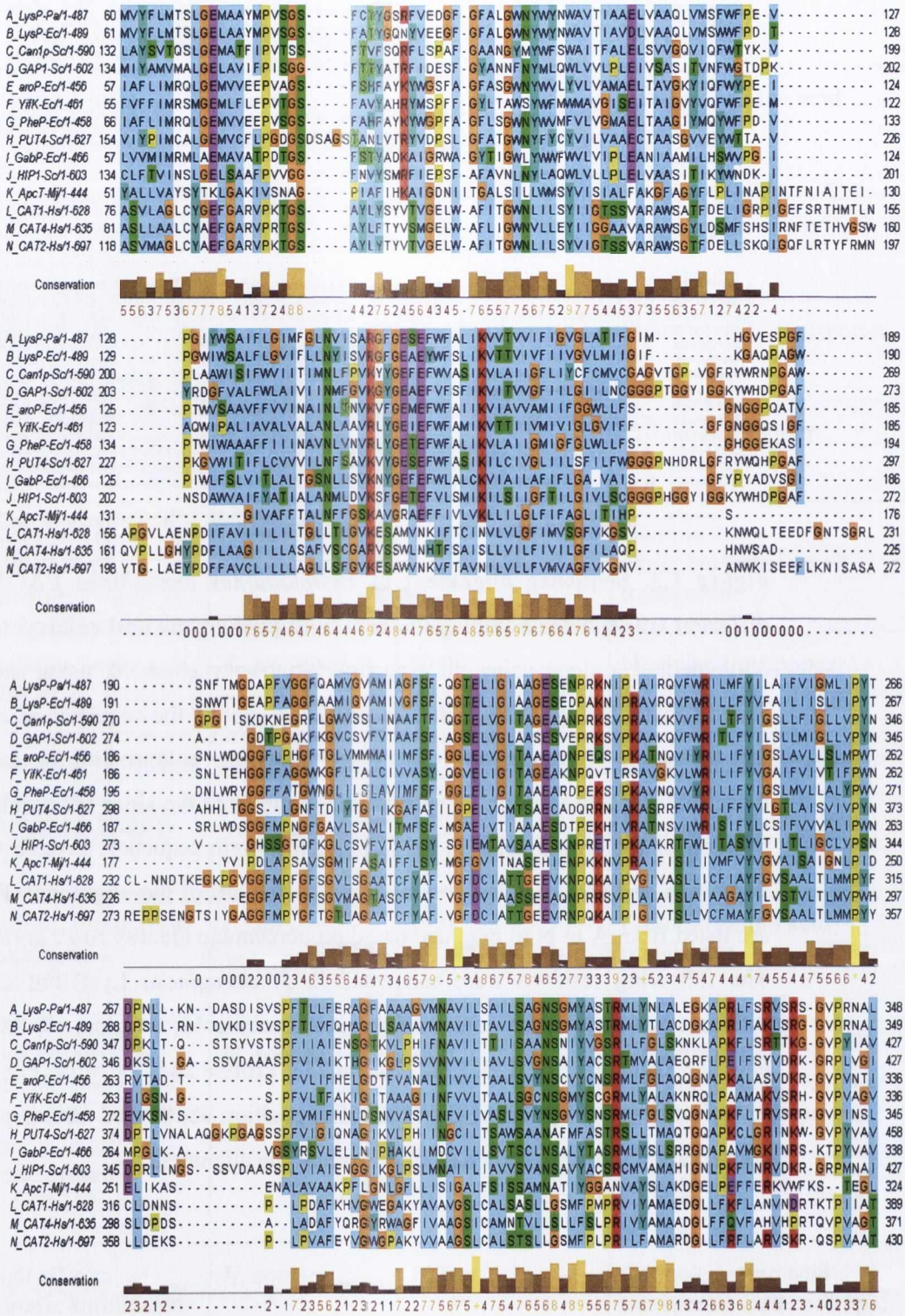


Figure 1.3. continued.

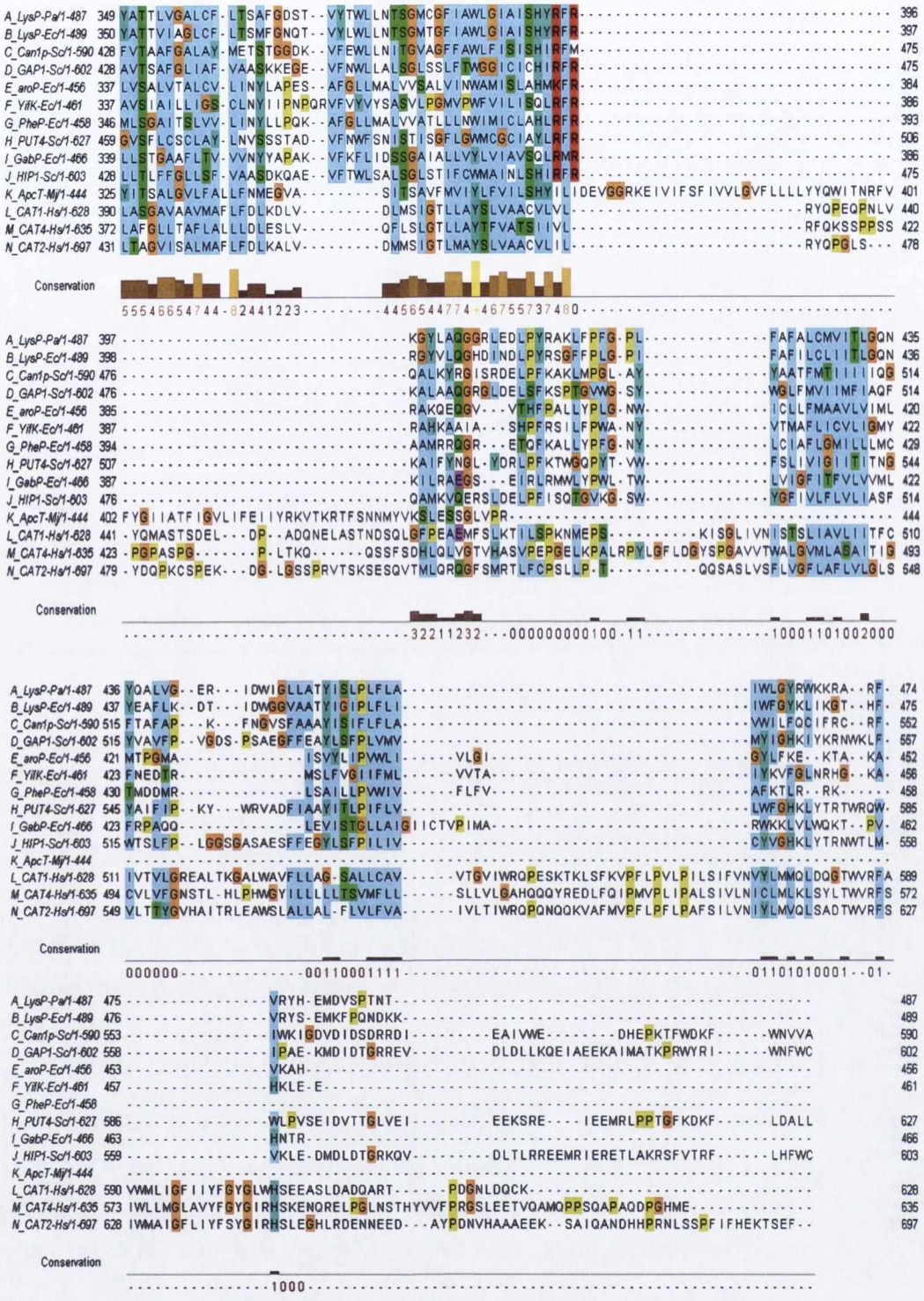


Figure 1.3. continued.

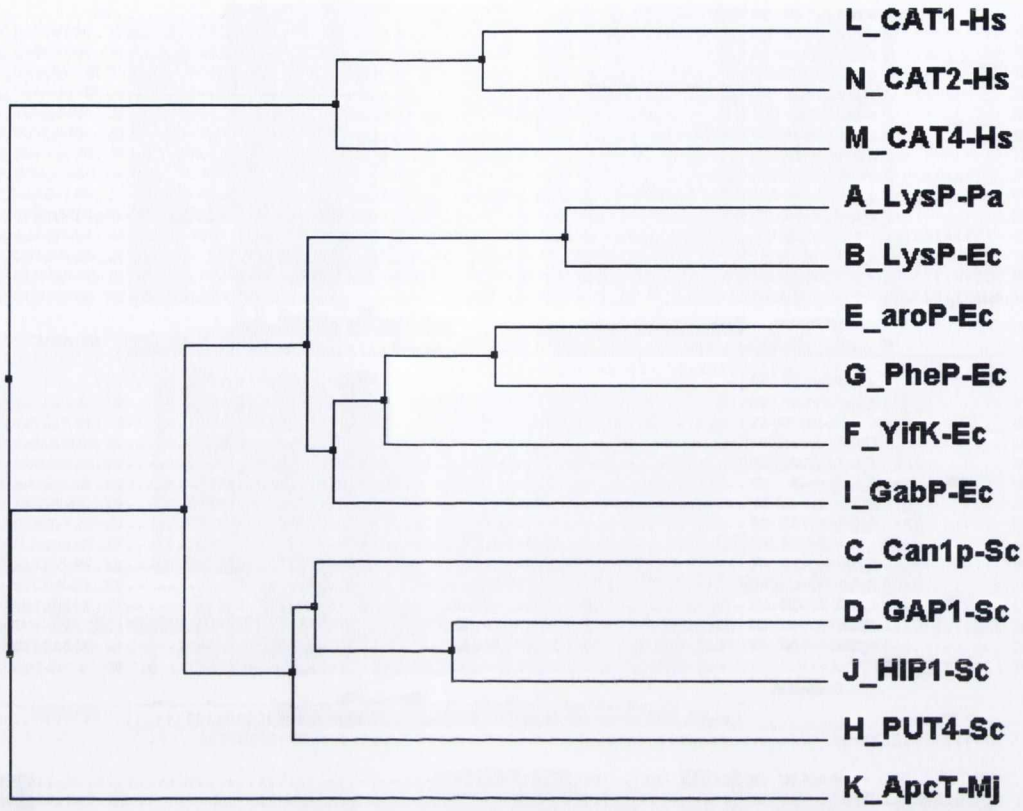


Figure 1.4. Phylogenetic tree of the APC family transporters.

The tree was calculated based on average distace using BLOSUM62 on Clustal Omega program.

The crystal structure of *M. Jannaschii* ApcT has been solved (Shaffer *et al.* 2009). The crystal structure revealed Lys158 in ApcT as the residue responsible for proton coupling to the transport of amino acids. A sequence alignment study was carried out in the APC family transporter to verify if this residue is conserved (**Figure 1.5**). This study was limited to the APC family transporter that has percentage identity of above 18 (**Table 1.I**). Lys 158 in *M. Jannaschii* ApcT was conserved in the other transporters and it corresponds to *P. aeruginosa* LysP Lys162, *E. coli* LysP Lys163, *S. cerevisiae* Can1P Lys234, *S. cerevisiae* GAP1 Lys237, *E. coli* AroP Lys159, *E. coli* YifK Lys157, *E. coli* PheP Lys168, *S. cerevisiae* Put4 Lys261, *E. coli* GabP Lys159, and *S. cerevisiae* HiP1 Lys236 and (**Figure 1.5**). Thus, we could conclude that LysP and the other APC family proteins in this study are symporters and coupled the transport of proton during substrate transport. This could be verified experimentally by reconstituting LysP into proteoliposomes having a neutral pH inside and a pH of 4 outside and see if the is a robust uptake of radioactive substrate when compare to

proteoliposomes having neutral pH both inside and outside as done by Shaffer *et al.* 2009. The difference in pH inside and outside the proteoliposomes creates a pH gradient that is required for substrate transport. Also, Shaffer *et al.* 2009 made a Lys to Ala substitution (K158A) and found that there was no measurable transport activity in this mutant but there was in the wild type.

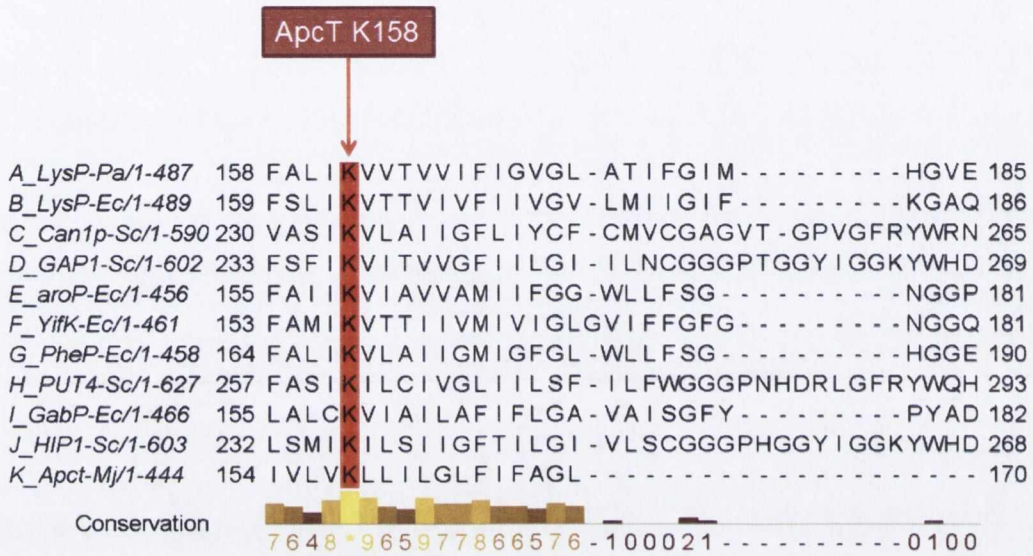


Figure 1.5. Sequence alignment of *M. Jannaschii* ApcT with different homologues from the APC family transporter in both prokaryote and eukaryote.

Alignment was done using the web link http://www.ebi.ac.uk using the programme called MUSCLE. The dash lines represent gaps in the sequence as a result of the fact that some sequences are longer than others, the coloured lysine residue represent those that are conserved in all species, the numbers at the sides are numberings of the position of the amino acids in the individual protein sequence. Protein name, source organisms and sequence length are shown at the left side of the alignment. Protein names are listed from A to K to organize based on percentage identity to *P. aeruginosa* LysP. The conservation score are shown at the bottom of the alignment. In this alignment only the region of interest was shown.

1.3. Aims of the Project

The goals of this project are to clone, express, purify, biochemically/biophysically and functionally characterize LysP from *Pseudomonas aeruginosa* strain PAO1 with the aim of crystallizing and solving its high resolution crystal structure. The high resolution

crystal structure and the functional characterization will provide insight into how the protein selectively transports lysine. Furthermore, structural data for LysP will help predict the structure and function of other metabolite transporters and provide potentially crucial information for novel antimicrobial drug target sites.

1.4. *Pseudomonas aeruginosa*

LysP is from *P. aeruginosa* strain PAO1 (http://www.pseudomonas.com/, Winsow *et al.* 2011). *P. aeruginosa* (**Figure 1.6**) is an opportunistic gram-negative pathogenic aerobic bacillus that belongs to the genus *Pseudomonas*, a γ subclass of the proteobacteria. Members of the genus *Pseudomonas* are characterised by simple growth requirements and can survive a wide temperature range (from 4 °C to 42 °C) (Bjarnsholt & Givskov 2006). The reasons why they can survive a wide temperature range could be because of it has a relatively larger genome size (6.3 million base pairs) and very complex genes (5570 open reading frames) such that there might be genes present or being synthesized that could be responsible for adaptation to increase in temperature of its environments (Stover *et al.* 2000). *P. aeruginosa* can cause infections in the urinary tract, respiratory system, soft tissue, bones and joints, and many systemic infections such as bacteremia and dermatitis, particularly in patients with severe burns, cancer and AIDS patients whose immune systems have been compromised (Lewenza *et al.* 2005). *P. aeruginosa* infection is a serious problem in patients hospitalized with cancer, cystic fibrosis, and severe burns. Indeed, 10 % of all hospital-acquired infections are caused by *P. aeruginosa* which is difficult to treat because of a naturally high antibiotic resistance (Lewenza *et al.* 2005). The resistance is linked with loss of outer membrane porin activity (OprD) that helps to uptake antibiotics, e.g., imipenem (El Amin *et al.* 2005). The resistance could be due to a number and/or combination of reasons: 1. The low cell permeability of the gram-negative bacteria outer membrane to antibiotics (Delcour 2009); 2. The ability to form biofilms makes it difficult to be accessed by therapeutic concentrations of antibiotics (Hoiby *et al.* 2010); 3. The fact that *P. aeruginosa* lives in the soil with bacilli, actinomycetes and molds have made it develop resistance against most of the naturally occurring antibiotics in these organisms (Davies & Davies 2010); 4. The presence of drug efflux proteins which helps to pump drugs away makes it difficult for therapeutic concentrations to be reached (Stover *et al.* 2000, Davies & Davies 2010), e.g., resistance to the antibiotics meropenem is as a result of the efflux protein MexB (El Amin *et al.* 2005); 5. The ability to undergo quorum sensing (QS) (Costerton *et al.*

1999). Today, infections caused by *Pseudomonas* are treated by administering 5 mg/kg/day of either tobramycin or gentamicin which are aminoglycoside antibiotics.

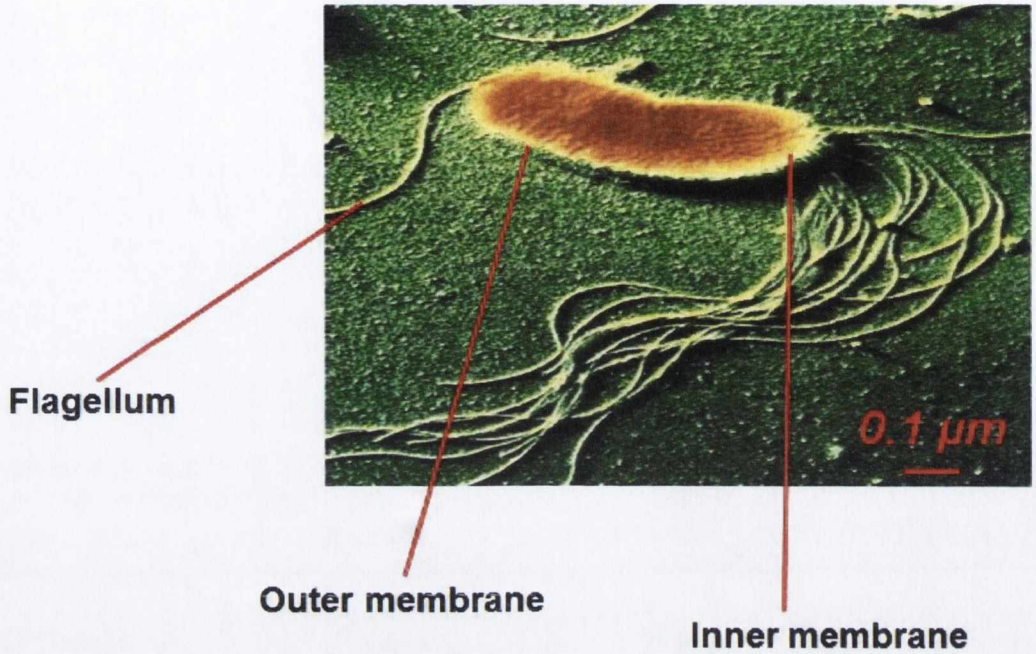


Figure 1.6. ***Pseudomonas aeruginosa.***

Figure was copied directly from http://activecelltech.com/images/pseudomonas.jpg.

1.5. Membrane Transport

The cell membrane of living organisms is made of lipids and proteins some of which are transporters and channels that are vital in cell physiology and highly relevant in diseases, e.g., depression, stroke and diabetes (Abramson *et al.* 2003). These proteins help transport materials and signals between cells, and their various compartments within. For example, two of the most prescribed drugs in the world-fluoxetine (Prozac) and omeprazole (Prilosec) are being transported to their destination by membrane transport proteins (Abramson *et al.* 2003). Prozac is an antidepressant and belongs to a group of drugs called selective serotonin reuptake inhibitors that prevent depression, panic, anxiety, or obsessive-compulsive symptoms (http://www.drugs.com/prozac.html) while Prilosec is a proton pump inhibitor that is used against symptoms of excessive acid in the stomach and in gastroesophageal reflux disease. Prilosec coupled with other antibiotics could also be used against stomach ulcers caused by *Helicobacter pylori* (http://www.drugs.com/prilosec.html).

The cell membrane bilayer of prokaryote (**Figure 1.7**) is composed of the phospholipid bilyer, different types of lipids as well as cholesterol and membrane proteins (integral membrane proteins that span the membrane and peripheral membrane proteins that are bound and exposed on one side of the membrane). The lipids have a hydrophobic fatty acid chain and a hydrophilic head group. The hydrophobicity of the phospholipid bilayer is due to the highly hydrophobic acid chains located at the interior. It allows only small uncharged molecules, such as carbon dioxide, ethanol, water, oxygen, to diffuse through it while polar or charged compounds cannot diffuse through the phospholipid bilayer because of its hydrophobicity. Membrane proteins that are covalently linked to carbohydrate groups are termed glycoproteins.

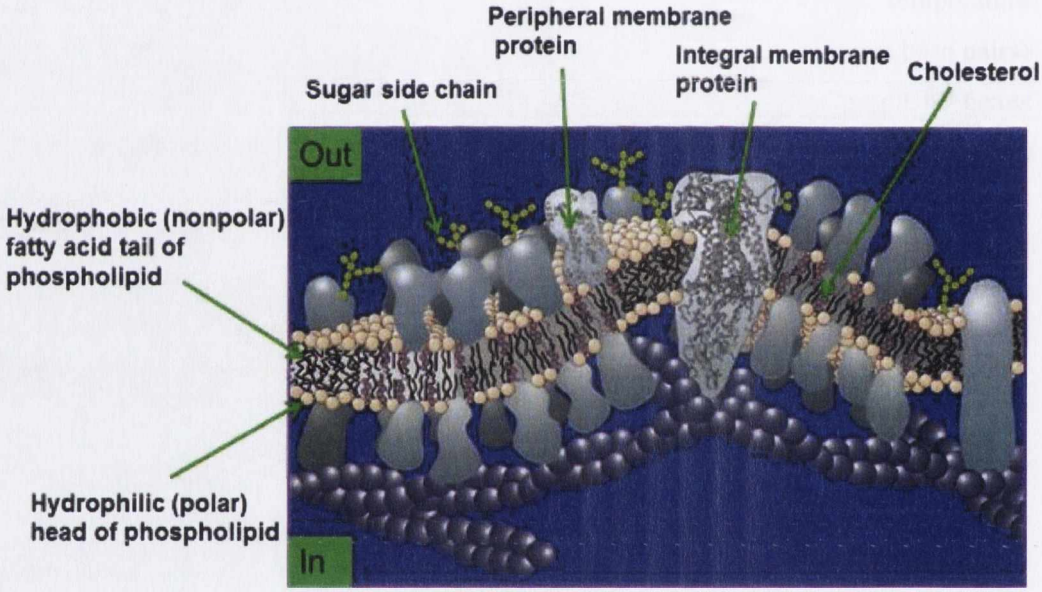


Figure 1.7. "Schematic representation of a biological membrane showing the lipid bilayer in and on which are situated a variety of proteins"
Figure copied directly from Li *et al.* 2012.

Integral membrane proteins perform a range of functions and one of them is to transport small molecules and ions across the membrane that could not be transported by passive diffusion. This mechanism is known as facilitated diffusion. Facilitated diffusion can be passive or active based on whether free energy is required, or not. Passive transport does not require free energy while active transport requires free energy to move molecules against their concentration gradient. Passive transport is also known as accelerated diffusion because the molecules are being transported at a faster

rate than normal diffusion to achieve equilibrium rapidly. Free energy is not needed for passive transport because the equilibrium process is energetically favored.

Ion channels span the membrane with α helices and require to form a membrane passage way. However, some transport proteins do not need channels or pores to transport molecules; instead they selectively bind the molecules on one side of the membrane that causes a conformational change to allow passage to the other side of the membrane. This class of transport proteins is also referred to as carrier protein.

1.5.1. Secondary Active Transporters

Secondary active transporters transport their substrate across the membrane using energy derived from an electrochemical gradient (Forrest *et al.* 2011). This process is called secondary because the source of energy, the electrochemical gradient must be derived from ATP dependent processes (primary transport) (Forrest *et al.* 2011). Secondary transporters cycle between conformations in which a solute binding site is accessible on one side of the membrane or the other (Forrest *et al.* 2011) (**Figure 1.8**).

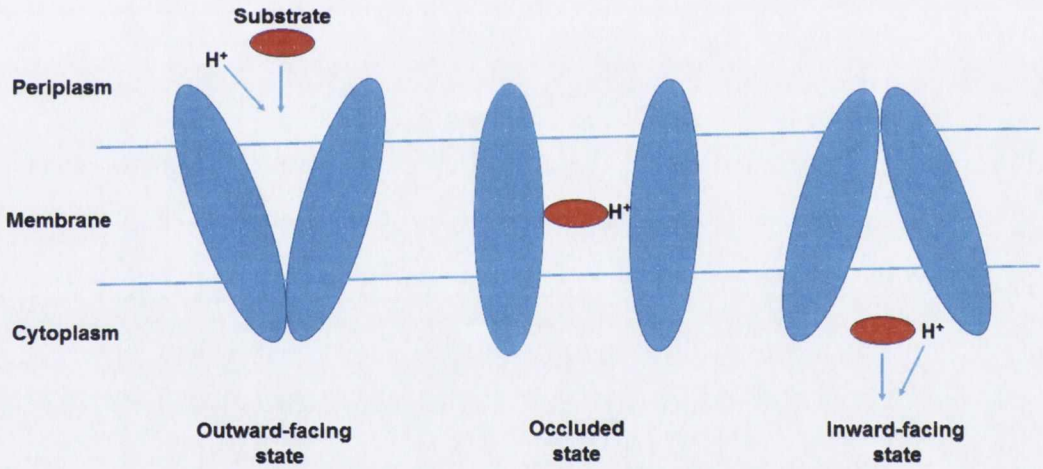


Figure 1.8. Mechanism of action of a secondary transporter

Secondary transporters cycle between conformations in which a solute binding site is accessible on one side of the membrane or the other

Secondary transporters can be classified as uniporters (transport a single solute across the membrane, e.g., the glucose transporter (GLUT1) (Mueckler *et al.* 1985) and the ionophore valinomycin), symporters (transport two dissimilar substrates in the same direction, e.g., the lactose/H^+ permease (LacY) which transports lactose and H^+ into the

cell (Abramson *et at.* 2003), antiporters (transport two dissimilar substrates in opposite direction), e.g., the arginine/agmantine antiporter (AdiC) which transports arginine into the cell and agmantine out of the cell (Fang *et al.* 2007) and Na^+/H^+ antiporter (NhaA) which transports Na^+ into the cell and H^+ out of the cell (Hunte *et al.* 2005) (**Figure 1.9**).

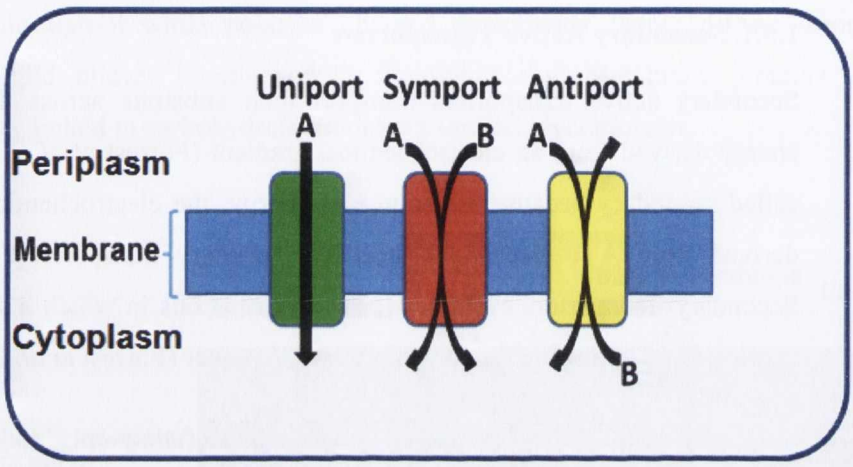


Figure 1.9. Classes of secondary transporters.
Uniport (transport of a single solute), symport (transport of two solutes in the same direction and antiport (transport of two solutes in opposite direction) (Wright *et al.* 2011).

1.6. Membrane Protein Structural and Functional Biology

Membrane proteins are very important in many cellular and physiological processes (Abramson *et al.* 2003). They transfer materials and information between cells and between the compartments within cells. They are targets for many drugs and are partly responsible for their uptake, metabolism and clearance. In other words, membrane proteins are essential for life and roughly one-quarter of all genome sequences code for membrane proteins (Wallin & Von Heijne 1998).

To best understand how membrane proteins function, biochemical, genetic and medical data should be complemented with high-resolution structures. X-ray crystallography is a powerful technique for determining atomic resolution structures of biological macromolecules (Hendrickson 2000, Friedmann *et al.* 2011). The resulting atomic structures, in turn, not only provide insight into the mechanisms of these

membrane proteins, but also they provide molecular details essential for the drug discovery process (Blundell *et al.* 2002, Kuhn *et al.* 2002, Giacomini *et al.* 2010). Despite the importance of membrane proteins, there is a large gap in the number of structures solved in comparison to soluble proteins. Only approximately 335 unique membrane protein structures have been solved to date (1st of September 2012) (http://blanco.biomol.uci.edu). Unique means only proteins from different species are considered, e.g., LysP from *E. coli* and LysP from *P. aeruginosa* will be considered as unique entries when solved (http://blanco.biomol.uci.edu). Proteins solved by mutagenesis and those solved with different substrates of a protein already in the database are not considered as unique entries (http://blanco.biomol.uci.edu). In short, the number of membrane proteins solved is still less than 1 % of all proteins solved.

One of the reasons for this difference is a consequence of the fact that the majority of them have naturally low abundance. This means membrane proteins overproduced in high yields will likely to facilitate structural studies. Over-expression of recombinant membrane proteins can be toxic to the cell, and many conditions need to be tested to optimize expression yields (Wagner *et al.* 2008). Such is the difficulty that for eukaryotic membrane proteins there have been only 10 structures solved using heterologously expressed material to date (www.mpdb.tcd.ie). Membrane proteins also can be extracted from the membranes using detergents. Detergents are used so as to provide an artificial environment to mimic the lipid bilayer, they pack into a monolayer or micelle of a spherical size at particular concentration called the critical micelle concentration (CMC), which is dependent on their head-group, and length of their acyl chain. Unfortunately, detergents that are the most successful in keeping the protein in a well-folded state, often have a large micelle size, which reduces protein-protein contact necessary for the formation of diffraction-quality crystals (Sonoda *et al.* 2011). This is because the membrane protein is surrounded by the detergents with large micelle size, thus the will be no contact from one protein molecule to another for ordered crystals to be formed (Sonoda *et al.* 2011). Finding the right detergents can be a difficult process, indeed, more than 20 different detergents have been successful for either solubilization (Raman *et al.* 2006; www.mpdb.tcd.ie) and/or obtaining high-resolution structures (**Figure 1.10**) (Packer & Newstead 2012). What is clear is that in order to increase the likelihood of obtaining well diffracting crystals of a membrane protein, it is necessary to find a detergent where the sample is monodisperse and pure (Takanishi *et al.* 2006).

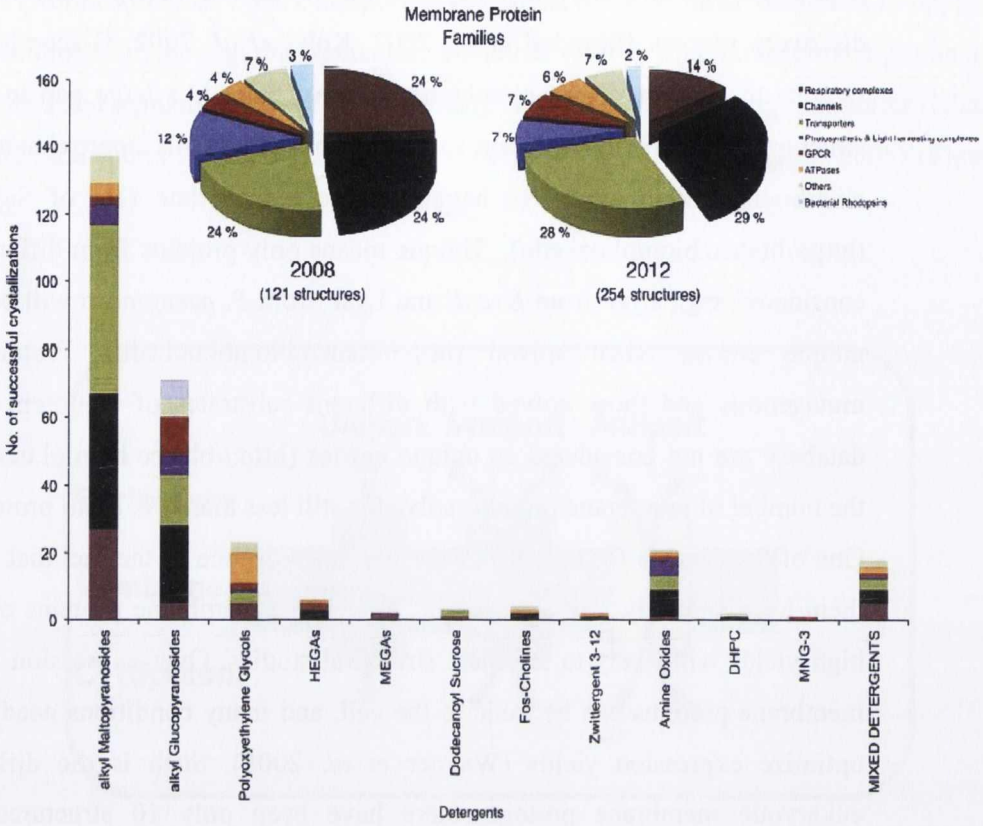


Figure 1.10. "Successful detergents used for a-helical membrane protein crystallization.

Numbers for each detergent class are shown, subdivided into the eight MP families used in the analysis. Respiratory complexes (brown), channels (black), transporters (green), photosynthetic and light-harvesting complexes (purple), GPCRs (red), ATPases (orange), bacterial rhodopsins (blue), and the ''others'' category (olive), those not fitting the seven main groupings. This color scheme is used throughout. Inset: Pie charts showing the change in the proportion of structures belonging to the eight MP families between 2008 and 2012, and the percentage contribution made by each family to the respective analyses is indicated" (Packer & Newstead 2012). Figure copied directly from (Packer & Newstead 2012).

Structural genomic groups over the last few years have started to address the problems associated with the process towards membrane protein structural determination, but have tended to focus on the overexpression bottleneck, and have, as yet, failed to deliver many high-resolution structures. However, one high throughput system has been developed that allows for empirical testing of the whole process, that is, from overexpression to crystals. This system in *E. coli* is simple and requires C-terminal

tagging of the membrane protein with Green Fluorescent Protein (Drew *et al.* 2001). If GFP is fused to the C-terminus of the membrane protein of interest, fluorescence can be monitored at every stage. During expression, GFP only folds and becomes fluorescent if the upstream membrane proteins folds and integrates into the membrane; therefore it is an accurate measurement of membrane-integrated material (Drew *et al.* 2001). GFP only folds if it is in the cytoplasm because all the chaperones required for folding are located in the cytoplasm and not in the periplasm. Protocols and vectors that ensure the C-terminus of the membrane protein is located in the cytoplasm are available (Hsieh *et al.* 2010). Not only can the GFP fluorescence be measured in whole-cells to evaluate how much protein is expressed, but because GFP remains folded even in SDS, full-length protein can be evaluated on SDS-PAGE with a CCD camera and exposure to blue-light, as an accurate and faster alternative to Western-blotting (Drew *et al.* 2006). To screen for a detergent that keeps the protein monodispersed, i.e., homogenous, membranes containing the over-expressed fusion protein can be solubilised in different detergents and then loaded separately onto a size-exclusion column. Although the UV trace records all proteins and provides no information as to the quality of the sample, the GFP-fluorescence is sensitive enough to be monitored as the protein migrates through the column. By using so-called Fluorescence-detection size exclusion chromatography (FSEC), detergents can be selected with little sample prior to purification (Kawate & Gauaux, 2006). The disadvantages of using the GFP fusion approach are: 1. The GFP may inhibit the function of the membrane protein (Fucile *et al.* 1999, Sonoda *et al.* 2010), 2. After the cleavage of GFP from the target membrane protein using TEV protease, the membrane protein may become unstable, and 3. Free GFP in the cell that may results from proteolysis of the fused target protein also accounts for the fluorescence reading, thus there will be false quantification of protein yield.

Lastly, once overexpression and detergent conditions have been empirically tested the protein can then be purified. In this system, there is an octa-histidine tag C-terminal to GFP facilitates binding to Ni2+-NTA resin (**Figure 1.11**) technically referred to as Immobilised Metal Affinity Chromatography (IMAC).

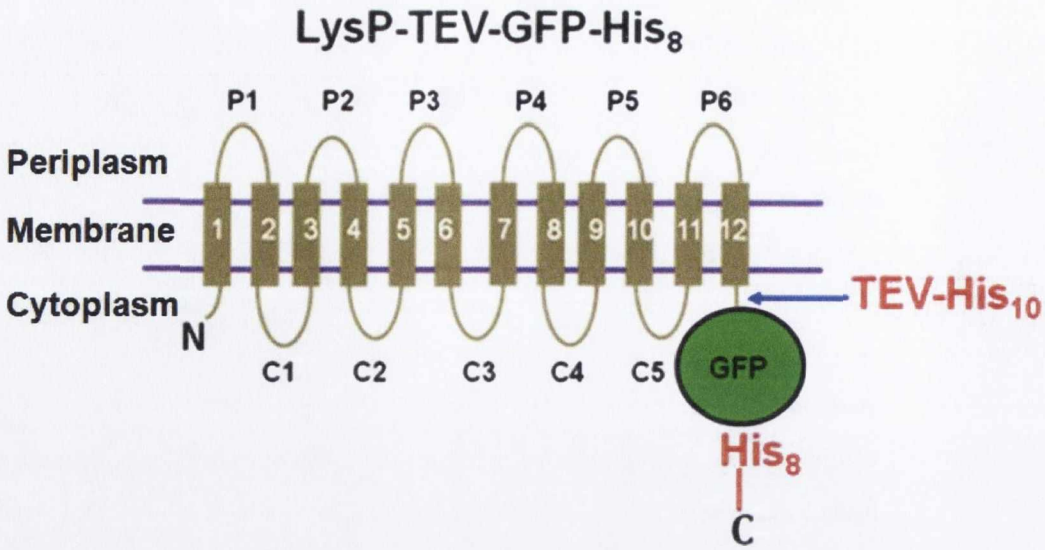


Figure 1.11. Schematic representation of the LysP-GFP fusion construct. Between LysP and GFP is a tobacco etch virus (TEV) protease cleavage site to separate LysP from GFP prior to crystallogenesis. The GFP and TEV protease has octa and deca-histidine tags respectively to further aid separation from LysP by Immobilized Metal Affinity Chromatography (IMAC).

After IMAC, the GFP can be removed by Tobacco Etch Virus (TEV) protease that recognizes an engineered cleavage site between the membrane protein and GFP-His_8 tag (**Figure 1.11**) to remove the GFP-His_8 tag and the TEV-His_{10} from the cleaved membrane protein, the digest can be passed back through the Ni^{2+}-NTA column and the flow-through containing the target membrane protein collected. Not only does this step remove the His-tagged GFP and TEV protease, but it also removes contaminating proteins that always bind to Ni^{2+}-NTA resin. Contaminants are a particular problem for membrane proteins as the His tags affinity to the resin is lowered by detergents, and means that washes with imidazole are kept to a relatively low concentration of ~ 30 – 50 mM (Mao *et al.* 2009). **Figure 1.12** summarizes the steps involved in the purification of a membrane protein using the GFP fusion approach (Drew *et al.* 2008).

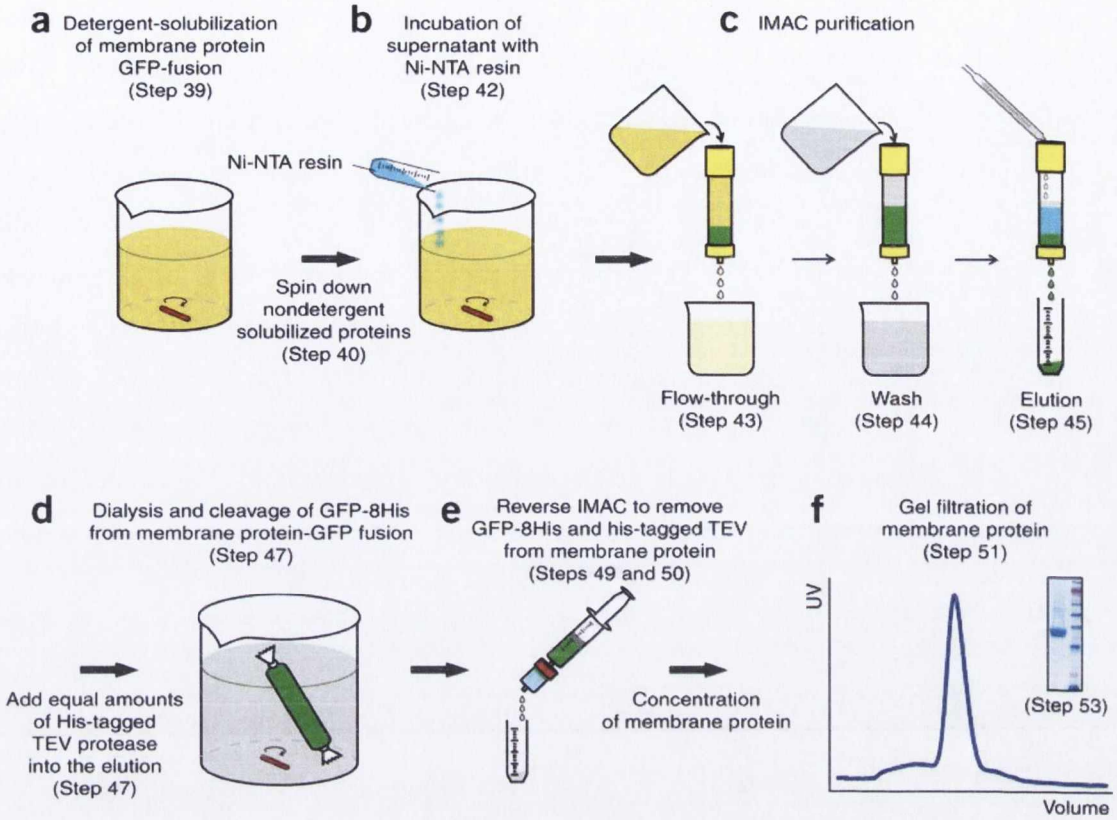


Figure 1.12. Flow diagram showing the purification methods used for the GFP-fusion approach

In step **a),** the membrane protein is first solubilised in an appropriate detergent and ultracentrifuged to remove non detergent solubilised proteins. In step **b),** the solubilised membrane is mixed with Ni-NTA for the LysP-GFP to bind. In **c),** the slurry is then poured onto an econo column and the LysP-GFP bound to Ni-NTA is trapped and the step washed with increasing amount of imidazol to remove contaminating proteins that may bind Ni-NTA and the sample is finally eluted with 250 mM imidazon in the wash buffer. In **d),** the eluted LysP-GFP is dialysed with an equimolar amount of TEV protease. This step removed excess imidazol that could inhibit peptide bond cleavage by TEV. In **e)**, the dialysed sample is passed through a His-trap column to remove the GFP and TEV and the flowthrough (free LysP) collected, concentrated and further purified by SEC and the purity evaluated on SDS PAGE gel in **f)**. Figure was copied directly from Drew *et al.* 2008.

There are five zones that have been established in the crystallization of soluble proteins and is shown in the phase diagram (**Figure 1.13**) (McPherson 1999). In the phase diagram, the precipitant and the soluble protein concentrations are related by the five

zones (McPherson 1999). The zones are; understuration zone, metastable zone, labile zone, precipitation zone and supersaturation zone (McPherson 1999). In the undersaturation and metastable zones the protein appears as a clear drop. The zone which nucleation and crystal growth can occur is the labile zone. However, there is a chance of getting crystals in the metastable zone if seeds from the labile zone are transferred to it. As a protein crystallizes its concentration decreases and causes one to move from the labile to the metastable zone. Precipitation occurs at very high protein and/or precipitant concentrations where the protein comes out of solution as an aggregate. The red line is called the solubility curve (McPherson 1999). It separates the undersaturated and the supersaturated zones and guides us during analysis of a crystallization experiment and also how to optimize a crystallization condition to either obtain crystals or to improve crystal quality or diffraction.

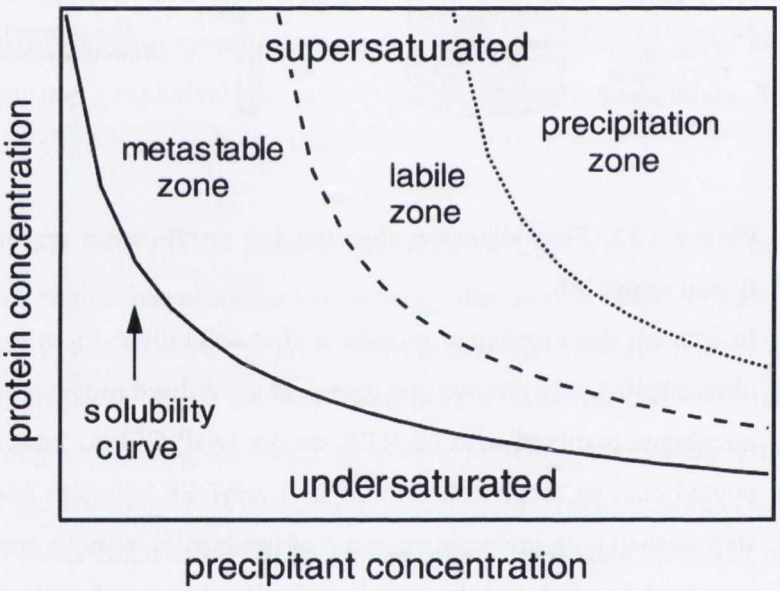


Figure 1.13. Schematic phase diagram illustrating the solubility of a soluble protein as a function of precipitant concentration (McPherson 1999).
In the undersaturation and metastable zones the protein appears as clear drops. Nucleation and crystal growth occurs in the labile zone. In the precipitation zone both the protein or precipitant concentration is very high and the protein comes out of solution as aggregate. Figure was copied directly from Asherie 2004.

There are two main methods that have been very successful for crystallizing membrane proteins; the traditional so called *in surfo* method and the lipidic cubic phase method

(*in meso* method) (Caffrey 2003). In the *in surfo* method, the protein is solubilised in a suitable detergent and crystallized as a soluble-protein surfactant complex while the in meso method involves reconstituting the membrane protein into the bilayers cubic phase using lipids and upon addition of a precipitant solution triggers nucleation and crystal growth (Caffrey 2003). The cubic phase provides an artificial stabilizing environment as in the lipid bilayer in the natural environment of the membrane protein. The methods have been demonstrated to be applicable to a range of membrane protein for example transmembrane peptides (Höfer *et al.* 2010), transporters (Cherezov *et al.* 2006), receptors (Luecke *et al.* 1999), complexes (Gordeliy *et al.* 2002) and G-protein coupled receptors (GPCRs) (Rosenbaum *et al.* 2011).

A mechanism for how this method works at the molecular level has been proposed (Caffrey 2000). Once the membrane protein is reconstituted into the bilayers of the cubic phase using a suitable lipid host, the addition of precipitant solution perturbs the cubic phase and local lamella phase is formed. Nucleation and crystal growth occurs via lamellar portal **(Figure 1.14).**

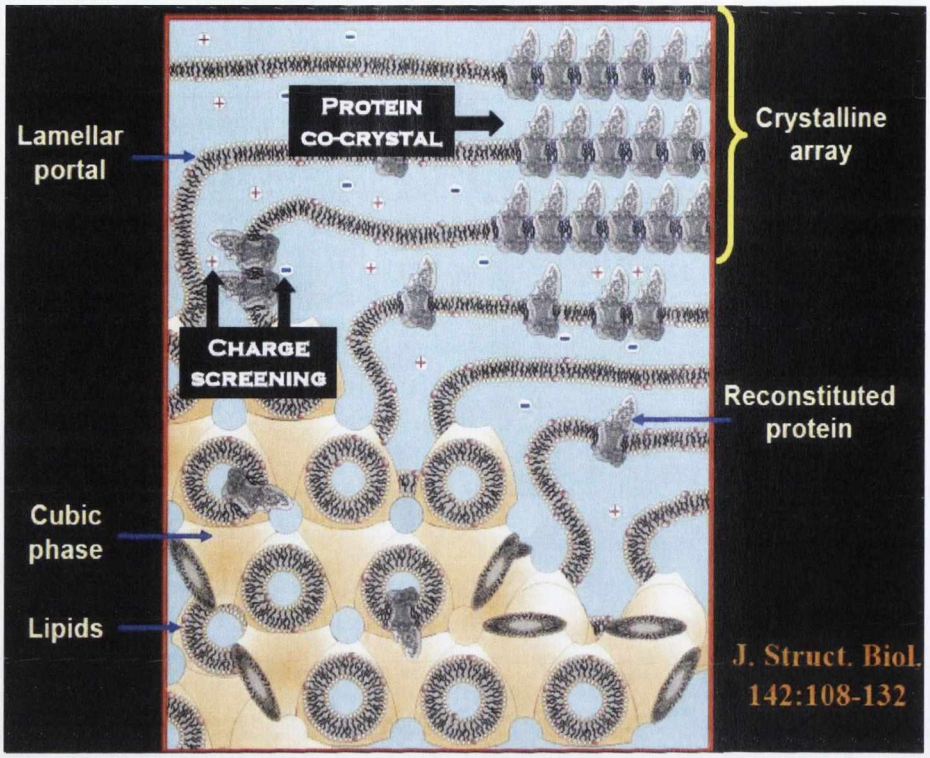


Figure 1.14. Cartoon representation of how membrane proteins crystallizes by the *in meso* method.

Figure copied directly from Caffrey 2009. The membrane protein is reconstituted into the bilayers of the cubic phase using a suitable lipid host and upon addition of precipitant solution disturbs the cubic phase causing nucleation and crystal growth. Crystal growth occurs via lamellar portal.

Chapter 2: Bioinformatics, Biochemical, Biophysical and Functional Characterization of a Lysine Specific Permease (LysP) from *Pseudomonas aeruginosa* Strain PAO1

2.1. Introduction

This chapter explains how lysP was cloned as a histidine tag LysP-GFP fusion using the expression vector pWaldo GFPd. It further explain how LysP-GFP was overexpressed in *E. coli* C43 (DE3) and how whole cell fluorescence was used to estimate protein yield and in-gel fluorescence to verify if full length LysP-GFP was overexpressed. The GFP tag was very useful in protein quantification as its fluorescence could be measured (excitation 485 nm and emission 512 nm). Also, since GFP remains folded in SDS gel, it made it possible for LysP-GFP samples ran on SDS PAGE gels to be exposed to blue light and image captured using a CCD camera. This procedure is termed in-gel fluorescence (Drew *et al.* 2006). Also, another importance of GFP tagging is that detergent screening could be carried out in with which membranes of LysP-GFP were isolated and solubilized in different detergents and the soluble fractions were analysed on fluorescence-detection size exclusion chromatography (FSEC) for monodispersity. After, a suitable detergent has been selected LysP-GFP was solubilized in this detergent, purified by IMAC and cleaved with TEV protease during dialysis and the free LysP was further purified by SEC. Also, in this chapter I showed how the purified free LysP was identified by total mass spectrometry and by in-gel tryptic digestion followed by MALDI-TOF mass spectrometry. The monomeric oligomerization state of free LysP was determined by Blue Native PAGE (BNP) and by gluteraldehyde crosslinking in this chapter. This chapter further highlights how a stability study was carried out on the free LysP after SEC run in Buffer B. This was done by concentrating the free LysP to 10 mg/mL in buffer containing 150 mM NaCl, 20 mM Tris pH 7.2, 0.003 %(w/v) LMNG and 5 %(v/v) glycerol and allowed to sit on ice for 7 days, aliquots were analysed on SEC for monodispersity on days 0, 1, 3, 5 and 7. 10 mg/mL was used because that will be the concentration to be used for crystallization trials, thus, it will be an indication of the monodispersity as the protein gets into crystallization trials. Also, CD spectroscopy was used to investigate the secondary structure of LysP in Buffer B. It was shown here that LysP binds to L-lysine, L-4-thialysine and D-lysine by microscale thermophoreses (MTS) analysis. Tryptophan fluorescence change on GFP free LysP in the presence of

L-lysine was carried and there was a percentage change of 10 upon the addition of 5 mM L-lysine.

Lastly, this chapter explains how purified LysP was shown to be transporting L-lysine by reconstituting into proteoliposomes and the transport of [^{3}H] L-lysine was measured. The transport was shown to be Na^+ independent in this chapter. Inhibition studies with L-4-thialysine and D-lysine was also carried out. The overexpressed LysP-GFP was shown to be active by whole cell radioactivity transport assay. An inhibition study was carried out with different amino acids.

2.2. Bioinformatics of LysP

Prior to the start of this project, and to enhance its success, LysP was analysed extensively by bioinformatics tools.

2.2.1. Nucleotide sequence

The sequence of nucleotide was obtained from http://www.pseudomonas.com and converted GCG format using the Readseq - biosequence conversion tool http://www.ebi.ac.uk as shown below. *lysP* is 1,464 bp in length.

```
   1 ATGACTGACC TGAACACCAG CCAGGGCCAG CAATTGCGGC GGGTGCTCAA GCCGCGCCAC
  61 CTGAACATGA TCGCCATCGG CGGCTCCATC GGCACCGGGC TGTTCGTCGC CTCCGGGGCC
 121 ACAGTGGCCA CCGCCGGGCC CGGCGGTGCG CTGCTTTCCT ACGCACTGAT CGGGCTGATG
 181 GTGTACTTCC TCATGACCAG TCTCGGCGAG ATGGCCGCCT ACATGCCCGT TTCCGGCTCA
 241 TTCTGTACCT ATGGCAGCCG GTTCGTCGAA GATGGCTTCG GTTTCGCCCT CGGCTGGAAC
 301 TACTGGTACA ACTGGGCGGT GACCATCGCC GCGGAGCTGG TGGCGGCGCA ACTGGTGATG
 361 AGTTTCTGGT TCCCCGAGGT GCCGGGCATC TACTGGAGCG CGATCTTCCT GGGCATCATG
 421 TTCGGCCTCA ACGTGATTTC CGCGCGCGGC TTCGGCGAGA GCGAATTCTG GTTCGCCCTG
 481 ATCAAGGTAG TCACGGTGGT GATCTTCATC GGCGTCGGCC TGGCGACCAT CTTCGGCATC
 541 ATGCATGGCG TGGAGTCGCC CGGCTTCAGC AACTTCACCA TGGGCGACGC GCCCTTCGTC
 601 GGCGGCTTCC AGGCGATGGT CGGGGTGGCG ATGATCGCCG GCTTCTCGTT CCAGGGCACC
 661 GAACTGATCG GCATCGCCGC CGGTGAATCG GAAAATCCGC GGAAGAACAT CCCGATCGCC
 721 ATTCGCCAGG TGTTCTGGCG CATCCTGATG TTCTACATCC TGGCGATCTT CGTGATCGGC
 781 ATGCTGATTC CCTACACCGA CCCGAACCTG CTGAAGAACG ATGCCAGCGA CATCAGCGTG
 841 TCGCCCTTCA CCCTGCTCTT CGAGCGCGCC GGCTTCGCCG CGGCGGCCGG GGTGATGAAC
 901 GCAGTGATCC TCTCGGCGAT CCTCTCCGCC GGCAACTCCG GCATGTACGC CTCGACCCGG
 961 ATGCTCTACA ACCTCGCCCT CGAGGGCAAG GCACCGCGCC TGTTCTCCCG CGTCTCGCGC
1021 AGCGGCGTGC CACGCAACGC GCTGTACGCC ACTACCCTGG TCGGCGCCCT GTGCTTCCTC
1081 ACTTCGGCCT TCGGCGACAG CACCGTCTAC ACATGGCTGC TGAACACCTC GGGCATGTGC
1141 GGCTTCATCG CCTGGCTGGG CATCGCCATC TCCCACTATC GCTTCCGCAA GGGTTACCTG
1201 GCCCAGGGCG GCCGCCTGGA AGACCTGCCC TACCGGGCCA AGCTGTTCCC CTTCGGGCCG
1261 CTGTTCGCCT TCGCCCTGTG CATGGTCATC ACCCTCGGGC AGAACTACCA GGCGCTGGTG
1321 GGCGAACGCA TCGACTGGAT CGGCCTCCTC GCCACCTATA TCAGCCTGCC GCTGTTCCTG
1381 GCGATCTGGC TGGGATACCG CTGGAAGAAG CGCGCACGCT TCGTTCGCTA CCACGAAATG
1441 GACGTCAGCC CGACCAATAC CTGA
```

This sequence was used for the designing of forward and reverse primers for the cloning of LysP into pWaldo GFPd vector and to select restriction enzymes that were used for the cloning. *Nde*I and *Bam*HI were used as there were no sites for these enzymes in the sequence that could promiscuously digested lysP during cloning. NdeI and BamHI sites were checked using the software GENtle (http://gentle.magnusmanske.de/). The % GC content is 64 (http://mbcf.dfci.harvard.edu/docs/oligocalc.html). Although, this was high there was no issue faced during PCR.

2.2.2. Amino acid Sequence (fasta format)

The sequence of nucleotide was obtained from http://www.pseudomonas.com and converted to Fasta format using http://www.compbio.dundee.ac.uk , as shown below. There are 487 amino acid residues in full length LysP.

>PA4628

MTDLNTSQGQ^{10}QLRRVLKPRH^{20}LNMIAIGGSI^{30}GTGLFVASGA^{40}TVATAGPGGA^{50}L
LSYALIGLM^{60}VYFLMTSLGE^{70}MAAYMPVSGS^{80}FCTYGSRFVE^{90}DGFGFALGWN^{100}YW
YNWAVTIA^{110}AELVAAQLVM^{120}SFWFPEVPGI^{130}YWSAIFLGIM^{140}FGLNVISARG^{150}
FGESEFWFAL^{160}IKVVTVVIFI^{170}GVGLATIFGI^{180}MHGVESPGFS^{190}NFTMGDAPFV
^{200}GGFQAMVGVA^{210}MIAGFSFQGT^{220}ELIGIAAGES^{230}ENPRKNIPIA^{240}IRQVFWRI
LM^{250}FYILAIFVIG^{260}MLIPYTDPNL^{270}LKNDASDISV^{280}SPFTLLFERA^{290}GFAAAA
GVMN^{300}AVILSAILSA^{310}GNSGMYASTR^{320}MLYNLALEGK^{330}APRLFSRVSR^{340}SGVP
RNALYA^{350}TTLVGALCFL^{360}TSAFGDSTVY^{370}TWLLNTSGMC^{380}GFIAWLGIAI^{390}SH
YRFRKGYL^{400}AQGGRLEDLP^{410}YRAKLFPFGP^{420}LFAFALCMVI^{430}TLGQNYQALV^{440}
GERIDWIGLL^{450}ATYISLPLFL^{460}AIWLGYRWKK^{470}RARFVRYHEM^{480}DVSPTNT

There was no TEV cleavage site in the gene prior to cloning into pWaldo GFPd that could cleave the protein during purification http://web.expasy.org/cgi-bin/peptide_cutter/peptidecutter.pl).

2.2.3. Amino acid composition and other parameters of LysP

The amino acid composition and the other parameters given below were obtained from http://www.expasy.ch/cgi-bin/protparam . **Table 2.I,** shows the amino acid composition of LysP.

Table 2.I. Amino acid composition of LysP obtained from http://www.expasy.ch/cgi-bin/protparam.

Amino acid	Number present	Percentage of total	Amino acid	Number present	Percentage of total
Ala (A)	53	10.9 %	Leu (L)	53	10.9 %
Arg (R)	23	4.7 %	Lys (K)	9	1.8 %
Asn (N)	17	3.5 %	Met (M)	20	4.1 %
Asp (D)	10	2.1 %	Phe (F)	38	7.8 %
Cys (C)	4	0.8 %	Pro (P)	19	3.9 %
Gln (Q)	10	2.1 %	Ser (S)	32	6.6 %
Glu (E)	15	3.1 %	Thr (T)	25	5.1 %
Gly (G)	54	11.1 %	Trp (W)	12	2.5 %
His (H)	4	0.8 %	Tyr (Y)	20	4.1 %
Ile (I)	36	7.4 %	Val (V)	33	6.8 %

LysP has 487 amino acid, a molecular weight of 53,098.2 Da, a theoretical pI of 9.15 and an estimated extinction coefficient of 95,800 M-1 cm-1.

There are 20 Met in full length GFP free LysP (~1 in every 50 amino acid). This is good because it can be exploited for crystallographic phasing using selenomethionine labelling. However, there is 1 Met at the N-terminus, which will not be useful for phasing. Also there are 12 Trp which will be important for absorbance and fluorescence measurements. It has 54 Gly which is a large number compared to soluble proteins and probable gives conformational flexibility to LysP. There are 4 Cys, this is not good as Cys in concentrated protein tends to form intermolecular disulfide bonds and cause protein aggregation. This information will be exploited during protein isolation to add reducing agent to avoid aggregation.

2.2.4 Signal Peptide Prediction

The function of a signal peptide is to direct a protein to the plasma membrane in prokaryotes. Different signal sequences encode for different organelles in eukaryotes, e.g., mitochondria, endoplasmic reticulum and nucleus. The signal sequence is an integral part of the structural gene for a protein but in some cases must be cleaved for the protein to become active by a signal peptidase found on specific organelles. Signal peptide prediction was done using the link; http://www.cbs.dtu.dk having two different

methods to calculate the probability of a signal peptide: neural network (NN) and hidden markov models (HMM) as shown below (Figure 2.1 & 2.2) respectively.

2.2.4.1. Using neural networks (NN)

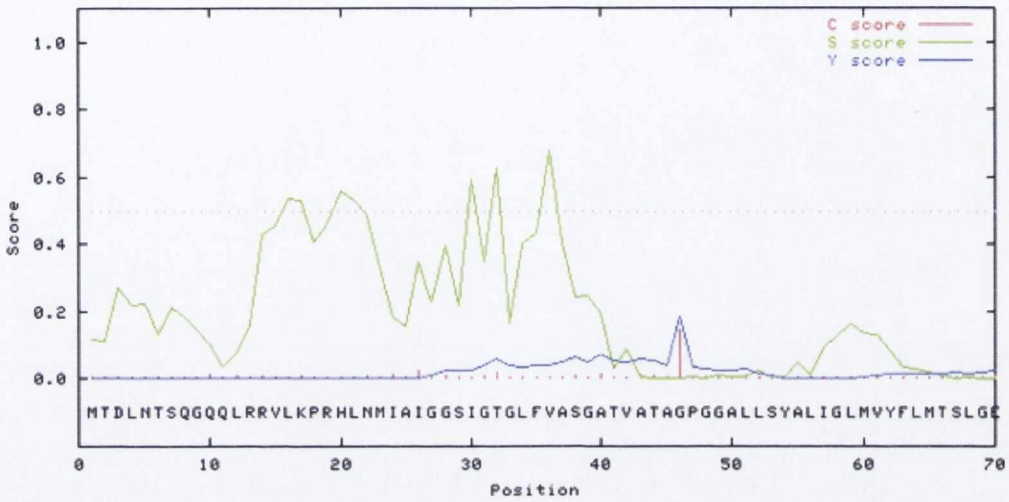


Figure 2.1. Signal peptide prediction using Neural Network (NN, http://www.cbs.dtu.dk).

The most important aspect of the NN is the Y score, which is a combination of the cleavage site score (C score, high at cleavage site) and the signal peptide prediction score (S-score). The smaller the S score is an indication that the amino acid belongs to a mature protein. If the Y score is less than 0.5, it means no signal sequence was predicted. Thus, NN did not predict any signal peptide for LysP.

2.2.4.2. Using Hidden Markov Models (HMM)

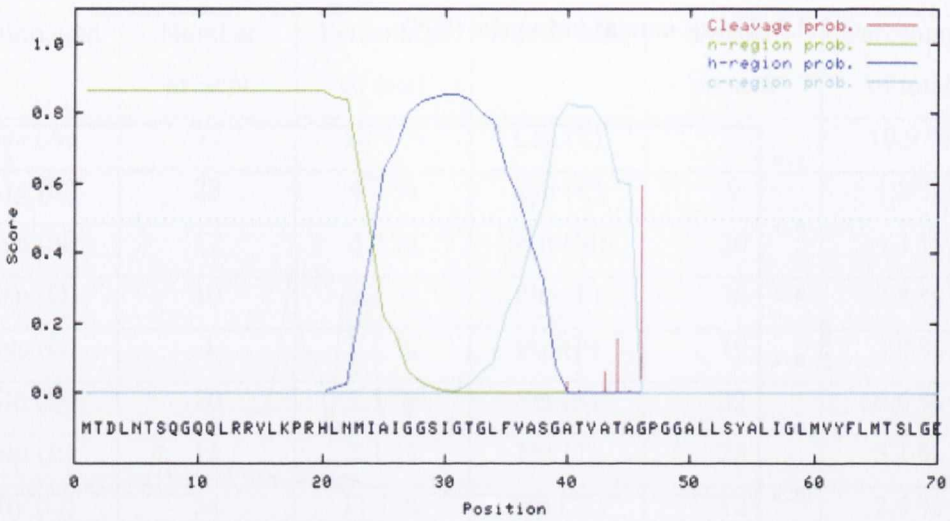


Figure 2.2. Signal peptide prediction using Hidden Markov Models (HMM) (http://www.cbs.dtu.dk).

The most important aspect HMM is the cleavage probability score. If the cleavage probability score is greater than 0.5, it implies there is a signal sequence at that region. Thus, HMM predicted a signal peptide cleavage between Ala 45 and Gly 46 but according to Ellis *et al.* 1995, "There is no signal sequence at the N-terminus of the LysP protein" in *E. coli.* both proteins have 70 % sequence identity (http://blast.ncbi.nlm.nih.gov). There is no signal sequence in LysP based on total mass spectrometry on purified TEV cleaved LysP from LysP-GFP in which a total mass of GFP free LysP was determined to be approximately 54 kDa which is the mass of full length GFP free LysP.

2.2.5. Secondary structure prediction

2.2.5.1. Topology prediction (TopPred) (von Heijne 1992 and Claros & von Heijne 1994)

TopPred predicts a protein membrane topology. It is based on the hydrophobicity analysis of a protein and the positive inside rule and was done using the link http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred.

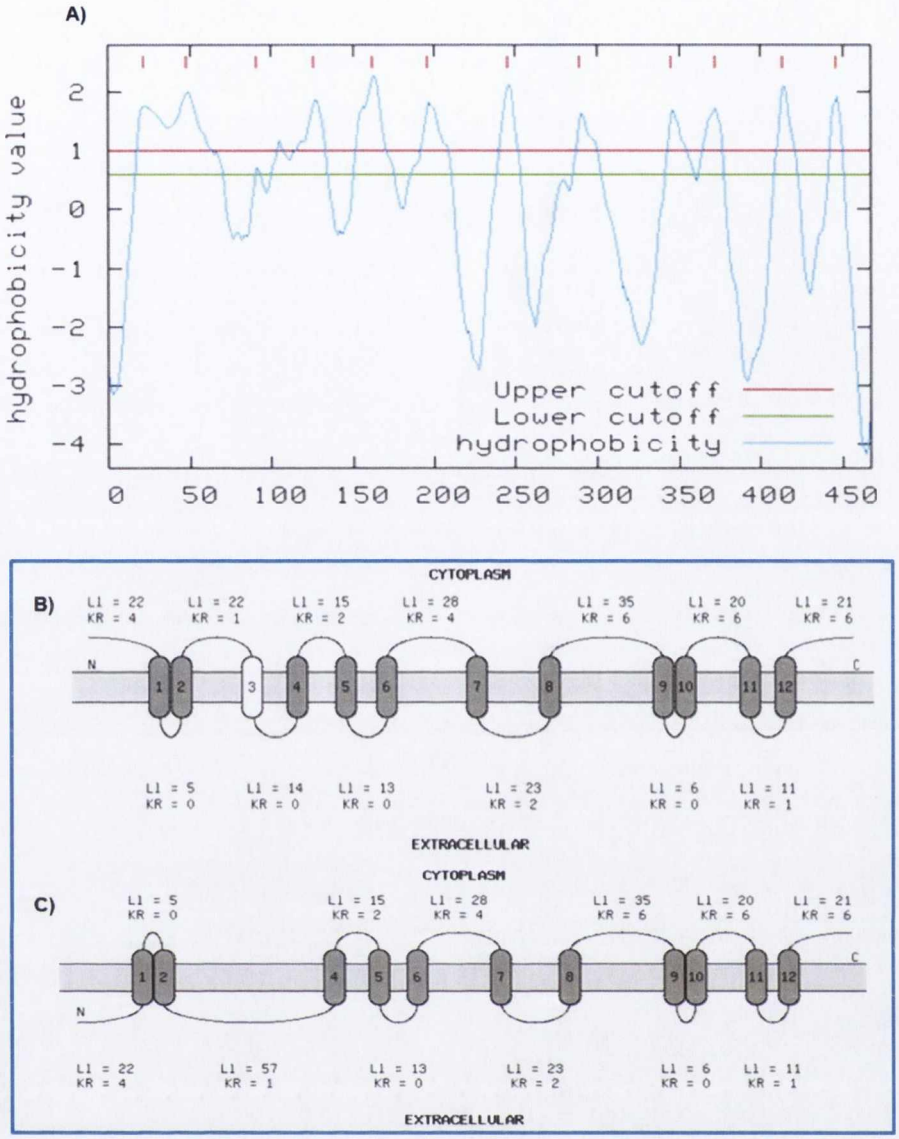


Figure 2.3. TopPred for LysP.

A) TopPred predicted 11 and 12 transmembrane helices with two different topology models for LysP. In B), there are 11 certain transmembrane segments and 1 putative transmembrane helix (segment 3) with both C and N terminus located in the cytoplasm while in C), there are 11 certain transmembrane segments but the N terminus is in the extracellular space. The red line in A) is the upper cutoff which has a hydrophobicity value (blue line) of one while the green line is the lower cut off and has a hydrophobicity value of 0.5. L1 stands for the loop length and KR stands for the number of lysine and arginine in that loop. The grey segments are certain segments while the white segments are putative segments. Y axis is the hydrophobicity value and the X axis is the residue number.

This model was not the best because it was not completely consistent with the solved structure of ApcT by Shaffer *et al.* 2009, which revealed that ApcT has 12 TM segments as well and that both the N- and C- termini located in the cytoplasm. Both LysP and ApcT belong to the APC transporter family (Shaffer *et al.* 2009). TopPred predicts instead 11 TM segments with 1 putative TM segment. It also predicts 12 TM segments but the N terminus was predicted to be located in the periplasm instead of the cytoplasm. This program predicted 11 TM segments for ApcT with both C and N terminus located in the cytoplasm.

2.2.5.2. TMHMM (Transmembrane Helical Hidden Markov Model) (Krogh *et al.* 2001 and Sonnhammer *et al.* 1998)

Since most of the bioinformatics tools are predictions it is important to search using many programs and compare the results. Another program used to predict transmembrane helices is the TMHMM http://www.cbs.dtu.dk/services/TMHMM/. It uses a different algorithm to predict the presences of transmembrane helices. TMHMM predicts 12 transmembrane helices (Figure 2.4).

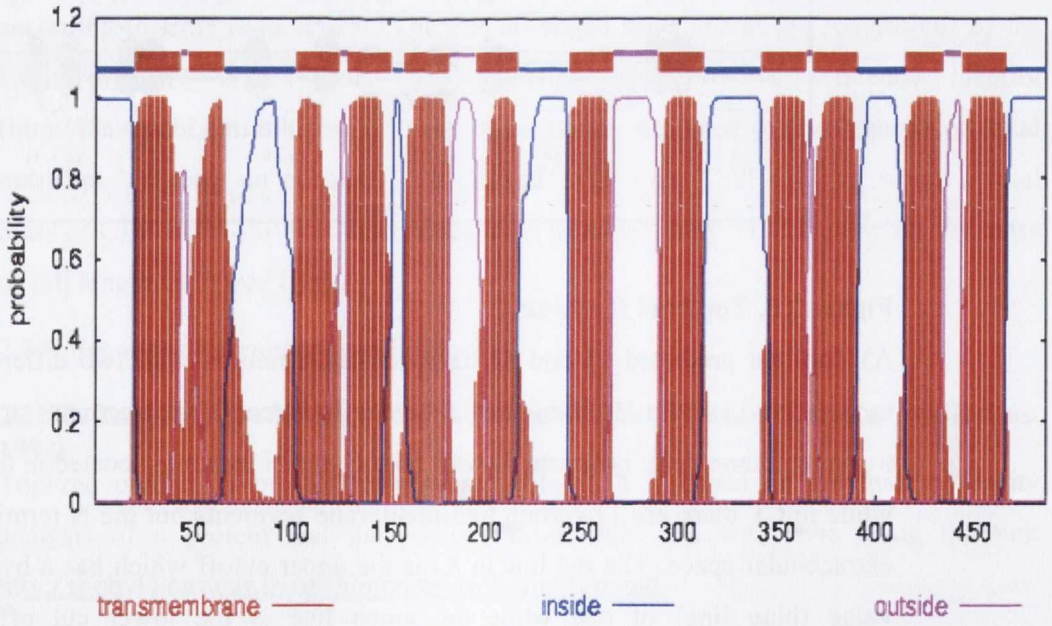


Figure 2.4. TMHMM prediction for LysP.

TMHMM displays the predicted transmembrane helices graphically with the Y axis being the probability that a transmembrane helix is present and the X axis is the residue

number. The red lines are the transmembrane segments, blue are the cytoplasmic loops and pink are the perisplasmic loops.

This model was the best because it is consistent with the solved structure of ApcT by Shaffer *et al.* 2009, which revealed that ApcT has 12 TM segments as well and that both the N- and C- termini located in the cytoplasm. Both LysP and ApcT belong to the APC transporter family (Shaffer *et al.* 2009). This program predicted 12 TM segments for ApcT with both C and N terminus located in the cytoplasm.

2.2.6. Topology map of LysP

The topology map of LysP was constructed based on the predicted result from http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred (**Figure 2.3**).

Figure 2.6, shows the topology map of LysP, **Figure 2.7**, shows the distribution of methionine, **Figure 2.8**, shows the distribution of positive residues and negatively charged cytoplasmic residues, **Figure 2.9**, shows the distribution of negative residues and **Figure 2.10**, the distribution of tryptophan residues.

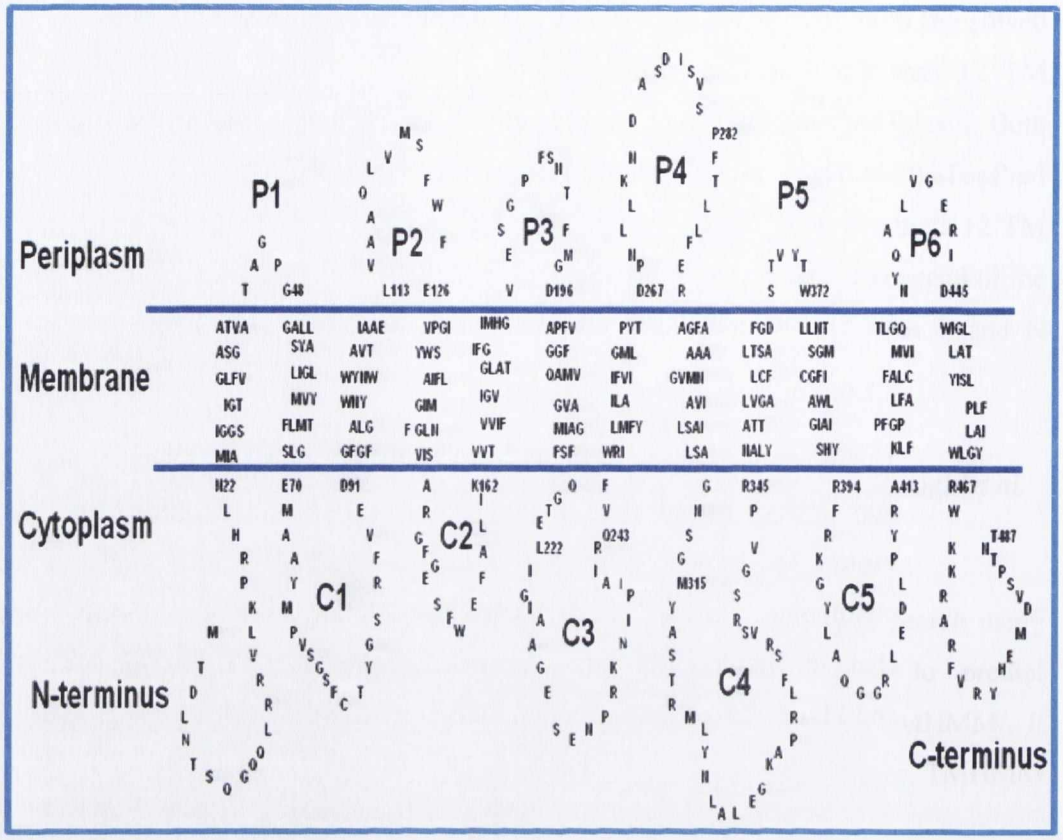


Figure 2.6. Topological diagram of LysP protein.

The model proposes 12 membrane-spanning α-helices joined by six periplasmic loops (Pl-P6) and five cytoplasmic loops (C1-C5). The N- and C-termini are suggested to be cytosolic. This figure was drawn based on TMHHM prediction because it predicts 12 TM segments with both the N- and C- termini located in the cytoplasm.

This prediction is consistent with the solved structure of ApcT (Shaffer *et al.* 2009) which revealed that ApcT has 12 TM segments as well and that both the N- and C-termini located in the cytoplasm. Both LysP and ApcT belong to the APC transporter family (Shaffer *et al.* 2009).

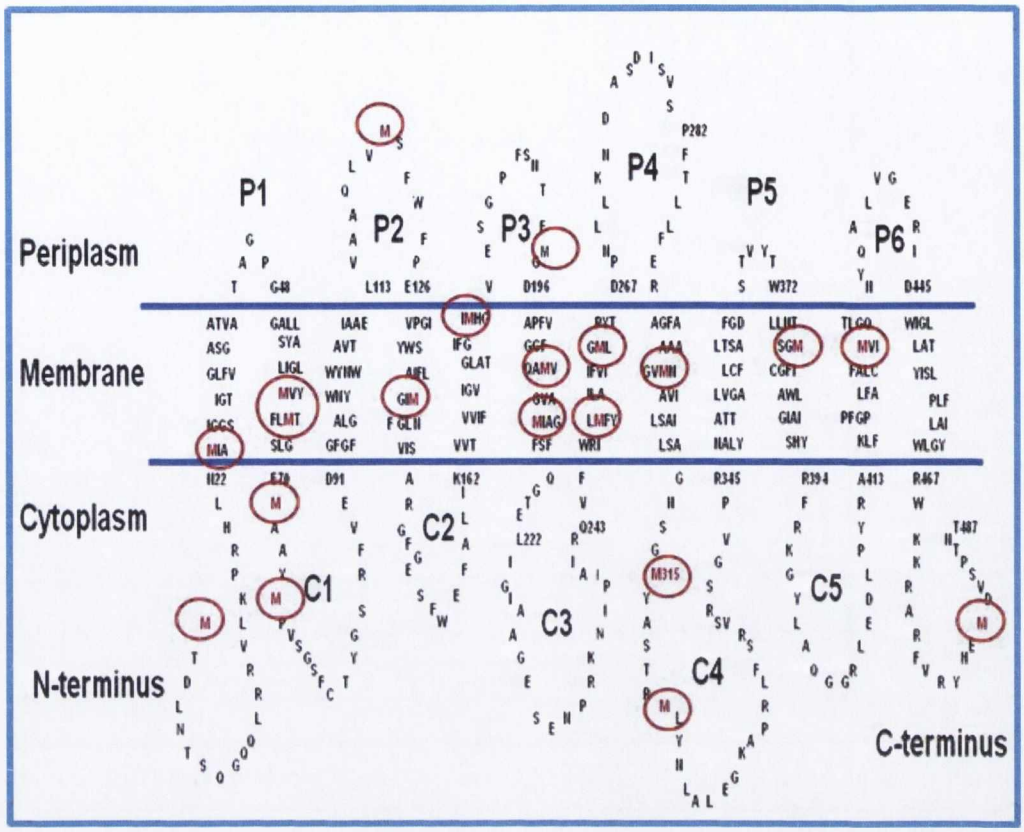


Figure 2.7. Topological diagram of LysP protein showing the distribution of methionines.

Met residues are indicated in red.

There are 20 methionines that are evenly distributed in the full length structure (~ 1 methionine in every 50 amino acid). This makes LysP a very good target as phasing by selenomethionine labelling could be possible.

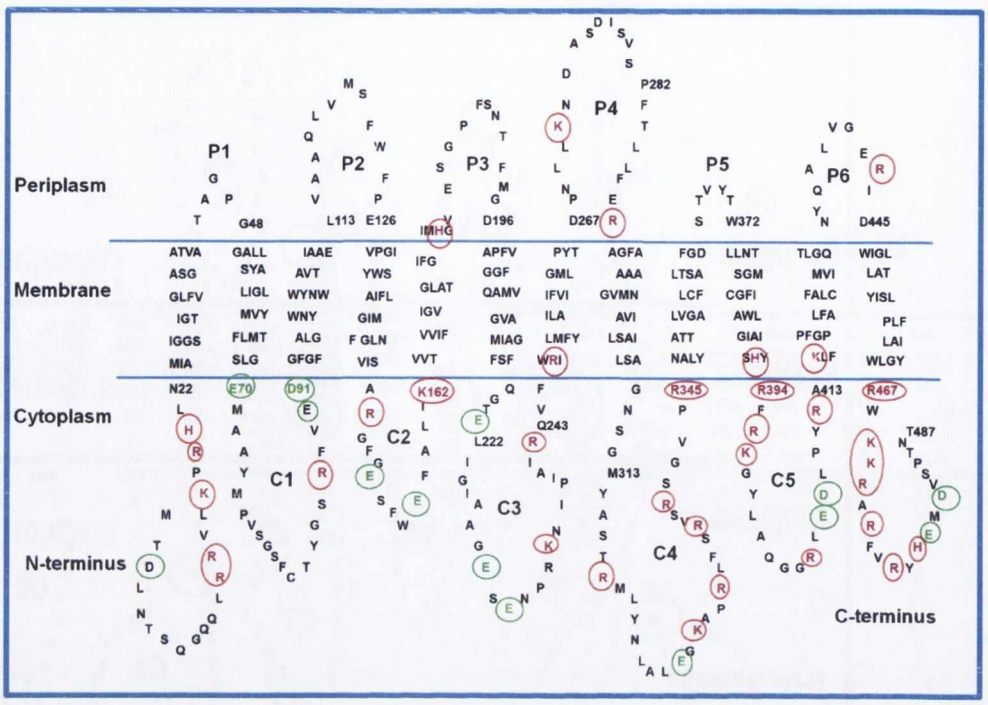


Figure 2.8. Topological diagram of LysP protein showing the distribution of positive residues.

Positive residues (lysine, histidine and arginine) are indicated in red and cytoplasmic negative charged residues (aspartate and glutamate) are in green. The model proposes 12 membrane-spanning α-helices joined by six periplasmic loops (Pl-P6) and five cytoplasmic loops (C1-C5). The N- and C-termini are suggested to be cytosolic.

There are 28 positively charged residues and 14 nagatively charged residues in the cytoplasm. Thus, there are 14 net positive charge residues in the cytoplasm which is consistent with the positive inside rule (von Heijne & Gavel, 1988). Positively charge residues tend to be located inside the cell like the cytoplasm and periplasm because it is hydrophilic and would love an aqueous environment.

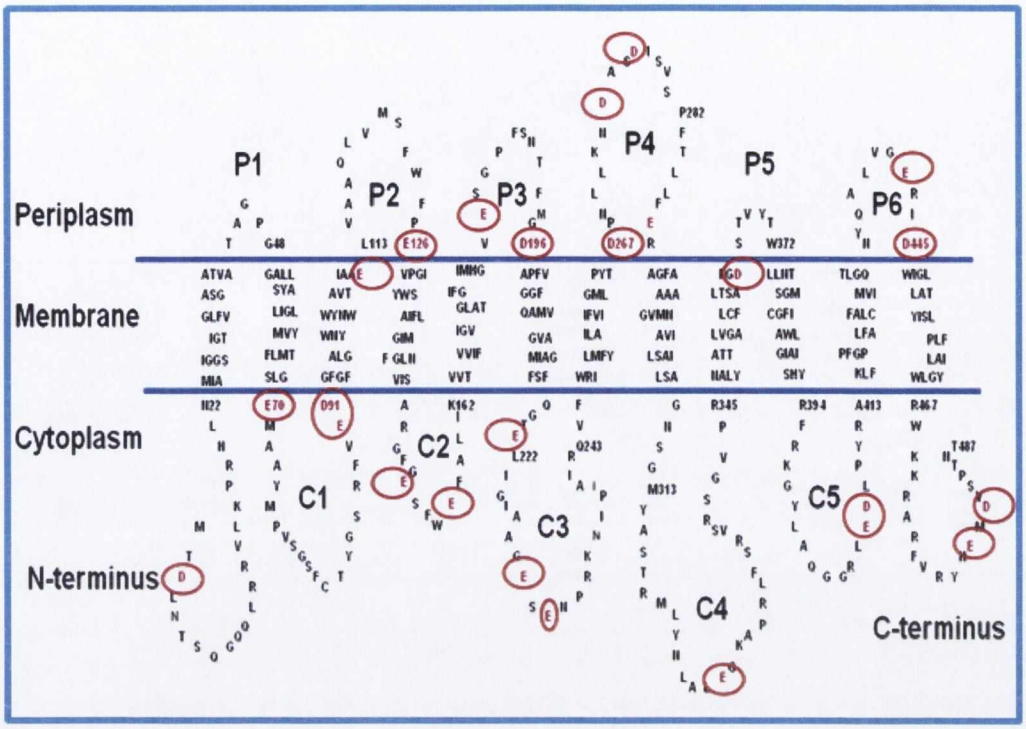


Figure 2.9. Topological diagram of LysP protein showing the distribution of negative residues.

Negative residues (aspartate and glutamate) are indicated in red.

There 14 negatively charged residues in the cytoplasm, 2 are located in the membranes and 8 are located in the periplasm.

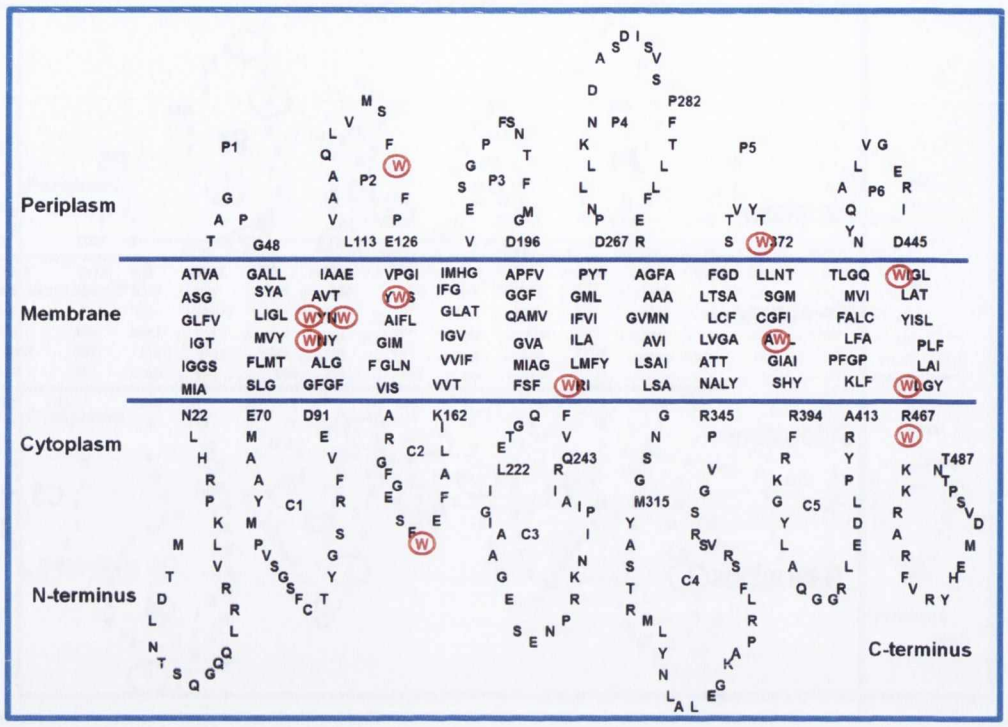


Figure 2.10. Topological diagram of LysP protein showing the distribution of tryptophans.

Tryptophans are indicated in red.

There are 12 tryptophans in LysP 10 located in the membrane, 2 are located in the cytoplasm and 2 in the periplasm. This is consistent with the fact that hypdrophobic residues tend to be located in the membrane because of its hydrophobic nature.

2.2.7. Disorder predictions

Prior to selecting a target for crystallographic studies, it is vital to know if the protein is stable or unstable as unstable proteins may be very difficult to produce and to crystallize. Although the stability of a protein is often well determined by biochemical stability assays, validating a protein's stability by bioinformatics is a good first choice.

2.2.7.1. FoldIndex© (Prilusky *et al.* 2005)

It is a simple tool to predict whether a given protein sequence is intrinsically unfolded. http://bioportal.weizmann.ac.il/fldbin/findex). FoldIndex is based on the average residue hydrophobicity and net charge of the sequence. **Figure 2.11,** shows a screenshot of the FoldIndex prediction.

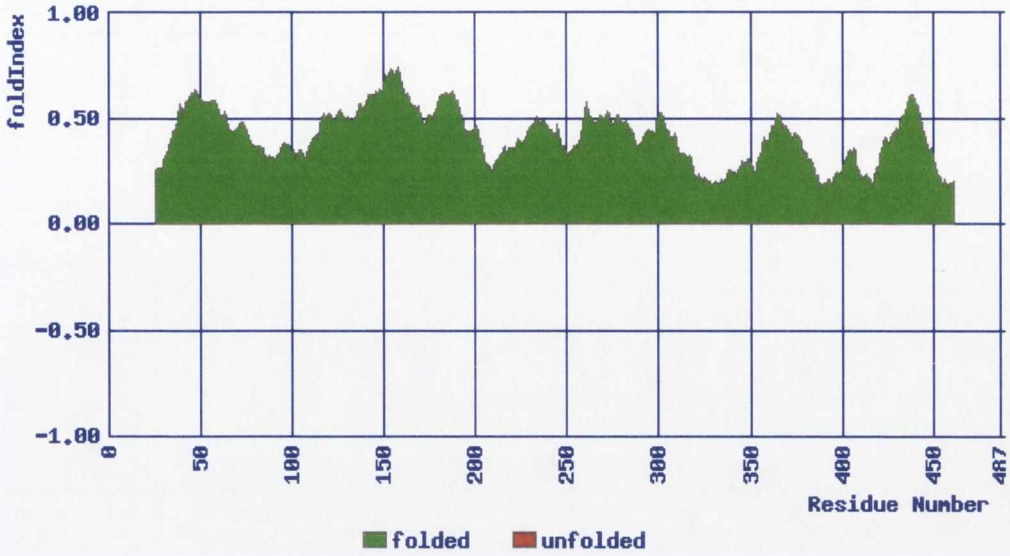


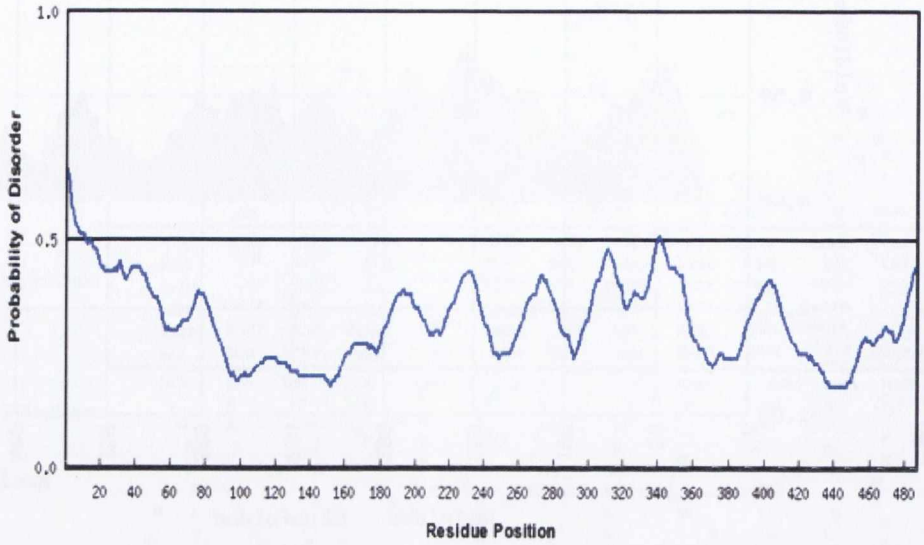
Figure 2.11. Screenshot for FoldIndex prediction of LysP.

FoldIndex predicts no disordered regions in the protein. It displays the presence of folded and unfolded regions graphically with the Y axis being the probability that a predicted folded or unfolded region is present and the X axis is the residue number. Green region shows the folded region.

2.2.7.2. RONN (Regional Order Neural Network) prediction (Yang *et al.* 2005).

Like fold index it gives the disorder of a protein but also predicts the probability of disorder of all the amino acid residues. http://www.strubi.ox.ac.uk/RONN. It is based on predicting proteolytic efficiency (Thormson *et al.* 2003).

The disordered regions predicted were between amino acids 1–11 and 341-341(**Figure 2.12**).




```
1      MTDLNTSQGQ QLRRVLKPRH LNMIAIGGSI GTGLFVASGA TVATAGPGGA   50
       ********** *_________ _________ _________ _________
51     LLSYALIGLM VYFLMTSLGE MAAYMPVSGS FCTYGSRFVE DGFGFALGWN  100
       _________ _________ _________ _________ _________
101    YWYNWAVTIA AELVAAQLVM SFWFPEVPGI YWSAIFLGIM FGLNVISARG  150
       _________ _________ _________ _________ _________
151    FGESEFWFAL IKVVTVVIFI GVGLATIFGI MHGVESPGFS NFTMGDAPFV  200
       _________ _________ _________ _________ _________
201    GGFQAMVGVA MIAGFSFQGT ELIGIAAGES ENPRKNIPIA IRQVFWRILM  250
       _________ _________ _________ _________ _________
251    FYILAIFVIG MLIPYTDPNL LKNDASDISV SPFTLLFERA GFAAAAGVMN  300
       _________ _________ _________ _________ _________
301    AVILSAILSA GNSGMYASTR MLYNLALEGK APRLFSRVSR SGVPRNALYA  350
       _________ _________ _________ _________ *_________
351    TTLVGALCFL TSAFGDSTVY TWLLNTSGMC GFIAWLGIAI SHYRFRKGYL  400
       _________ _________ _________ _________ _________
401    AQGGRLEDLP YRAKLFPFGP LFAFALCMVI TLGQNYQALV GERIDWIGLL  450
       _________ _________ _________ _________ _________
451    ATYISLPLFL AIWLGYRWKK RARFVRYHEM DVSPTNT
       _________ _________ _________ _____
```

Figure 2.12. RONN prediction for LysP with disordered regions: 1-11, 341–341. RONN predicts disordered regions in residues 1 to 11 and 341. It displays the presence of disordered regions graphically with the Y axis being the probability disorder region is present and the X axis is the residue position. It is important to note that FoldIndex only predicts whether a protein will fold while RONN is looking specifically for the

disordered regions. A protein may be disordered in a region but still folds, thus one cannot conclude that FoldIndex contradicts RONN.

2.2.8. Post translational modification

"Post translational modification (PTM) test for chemical modification of proteins after its translation, and have wide effects broadening its range of functionality" (Khoury *et al.* 2011). PTM was predicted by http://www.compbio.dundee.ac.uk/taro/cgi-taro/targpipe_home.pl. The only PTM predicted was NetPhosphorylation (NetPhos). NetPhos is used to predict serine, threonine and tyrosine phosphorylation sites. The NetPhos predictions with score of at least 0.7 are displayed in **Table 2.II**. When a protein is phosphorylated, it may become active or inactive (Scallen and Sanghvi 1983). PTM was carried out for LysP to facilitate the interpretation of processes that were used to identify LysP after purification, such as electrospray ionization (ESI)-MS and electrophoretic mobility on SDS PAGE. Based on ESI-MS result, LysP was not phosphorylated (**see section 2.4.4.1**).

Table 2.II. NetPhosphorylation site for LysP.

Residue	Confidence score
7	0.9
186	0.8
339	0.9
483	0.8

2.2.9. Homology Modelling of *P. aeruginosa* LysP using *M. janaschii* ApcT as a Template

Homology modelling was carried out to identify regions of conserved secondary structure of both ApcT and LysP. Although the sequence similarities between both protein was low (19 %), greater than 30 % is required for homology modelling to be carried out. However, proteins with low sequence identity may have high sequence identity at their binding pockets, in this case a homogy model could be carried out since we are interesting in the substrate binding sites of the proteins.

The sequence of LysP and ApcT was aligned using ClustalW and the resuting alignment was cut and paste onto the Swiss Model workspace (http://swissmodel.expasy.org) and LysP was modelled using ApcT as template. **Figure 2.13.** shows a plot of the resulting model obtained. The Y-axis of the plot

represents the energy for each amino acid (X-axis) of the protein chain. Negative energy values (in green) represent favourable energy environment whereas positive values (in red) unfavourable energy environment for a given amino acid.

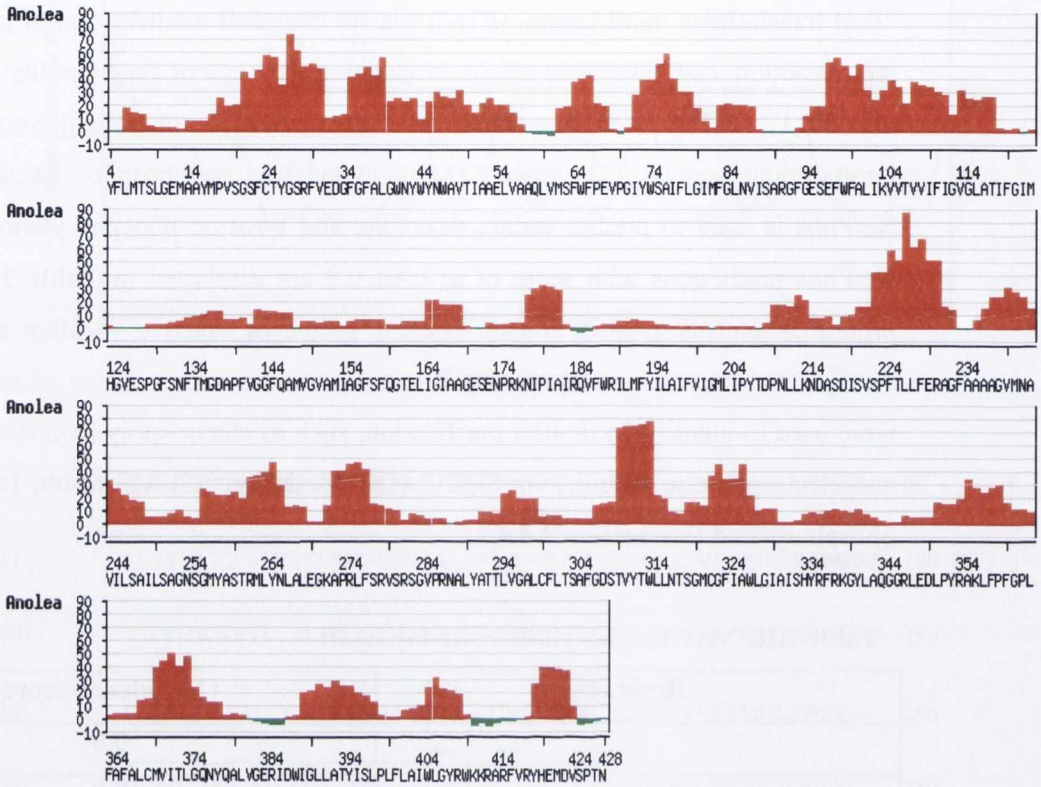


Figure 2.13. Homology modelling plot of LysP using ApcT as a template.

Modelling was perfomed using the Swiss Model workspace. The Y-axis of the plot represents the energy for each amino acid (X-axis) of the protein chain. Negative energy values (in green) represent favourable energy environment whereas positive values (in red) unfavourable energy environment for a given amino acid.

The regions of similar secondary structure (green in Figure 2.13) are highlighted in red in the sequence of LysP (**Figure 2.14**). Mutagenesis of these residues followed by functional studies will further validate those residues that are key for substrate transport.

```
MTDLNTSQGQQLRRVLKPRHLNMIAIGGSIGTGLFVASGATVATAGPGGALLSYALIGLMVYFLMTSLGE
MAAYMPVSGSFCTYGSRFVEDGFGFALGWNYWYNWAVTIAAELVAAQLVMSFWFPEVPGIYWSAIFLGIM
FGLNVISARGFGESEFWFALIKVVTVVIFIGVGLATIFGIMHGVESPGFSNFTMGDAPFVGGFQAMVGVA
MIAGFSFQGTELIGIAAGESENPRKNIPIAIRQVFWRILMFYILAIFVIGMLIPYTDPNLLKNDASDISV
SPFTLLFERAGFAAAAGVMNAVILSAILSAGNSGMYASTRMLYNLALEGKAPRLFSRVSRSGVPRNALYA
TTLVGALCFLTSAFGDSTVYTWLLNTSGMCGFIAWLGIAISHYRFRKGYLAQGGRLEDLPYRAKLFPFGP
LFAFALCMVITLGQNYQALVGERIDWIGLLATYISLPLFLAIWLGYRWKKRARFVRYHEMDVSPTNT
```

Figure 2.14. Sequence of LysP showing regions of similar secondary structures with ApcT.

The residues are highlighted in red. Model was performed using the Swiss Model Workspace.

2.3. Experimental procedures

The main buffers used in this section were:

Buffer A: 150 mM NaCl, 20 mM Tris pH 7.5, 0.03 %(w/v) DDM and 5 %(v/v) glycerol

Buffer B: 150 mM NaCl, 20 mM Tris pH 7.5, 0.003 %(w/v) LMNG and 5 %(v/v) glycerol.

The pH 7.5 is the pH is of Tris buffer before the other components were added. The pH of Tris buffer was adjusted using HCl and NaOH.

2.3.1. Cloning of LysP

The gene encoding the lysine specific permease (*lysP*) from *P. aeruginosa* strain PAO1 was cloned into pET200/D-TOPO (Invitrogen), digested with *Nde*I and *Bam*HI restriction enzymes and then sub-cloned into the standard pET 28(a) derived GFP-His_8-fusion vector (pWaldo-GFPd) (Drew *et al.* 2006) between *Nde*I and *Bam*HI restriction sites. The pWaldo-GFPd was a gift from Dr. David Drew at the Division of Molecular Biosciences, Imperial College, London. The forward and reverse primers used for the PCR/amplification were 5′-CACCCATATGACTGACCTGAACACCAGCCAG-3′ and 5′-GGATCCGGTATTGGTCGGGCTGACGTC-3′ respectively. The forward primer had *Nde*I site (highlighted in red) while the reverse had *Bam*HI site (highlighted in red). The template DNA source used for the PCR reaction was from *P. aeruginosa* strain PAO1 cells grown in Lauria Bertani (LB) broth medium (Fluka). The inclusion of the *lysP* gene into the pWaldo GFPd vector was validated by DNA sequencing, using T7 forward and reverses primers (Eurofins MWG GmbH, Anzinger Strasse 7a, D-85560 Ebersberg, Germany).

2.3.2. Recombinant Expression of LysP-GFP fusion and In-gel fluorescence

A single colony of freshly transformed *E. coli* (DE3) C43 cells with pWaldo GFPd vector harboring the *lysP* gene from *P. aeriginosa* strain PAO1 on LB agar/kanamycin plate was used to inoculate 50 mL of LB supplemented with 50 µg/mL of kanamycin and grown for 16 h at 37 °C with shaking at 200 r.p.m using the INFORS HT multitron standard shaker. Kanamycin was added because the pWaldo GFPd expression vector harbors a kanamycin resistance thus only cells that pick up the plasmid vector will grow in the LB/kanamycin plate or medium. The culture was then used to inoculate (1/50 dilution) 1 L of LB supplemented with 50 µg/mL of kanamycin and grown at 37 °C with shaking at 200 r.p.m for 2 h to an OD_{600}~0.3 using the INFORS HT multitron

standard shaker. The temperature of the shaker was then dropped from 37 °C to 25 °C, as it has been shown for most membrane protein that best yields were obtained by inducing at low temperature. After 3 h, an OD_{600} of 0.6 was reached and the cells were induced with 0.4 mM IPTG (1 M stock). The cells were then grown at 37 °C and shaking at 200 r.p.m for 22 h to an OD_{600} of 5. A control was used where the culture was inoculated with the expression vector without the *lysP* insert. GFP fluorescence (excitation 485 nm and emission 512 nm) was read on 100 µL samples taken 22 h after IPTG induction and also for the control using the SpectraMax Gemini EM plate reader (Molecular Devices) and the amount of protein expressed was determined as described in **Section 2.4.1** (Drew *et al.* 2006). For in-gel fluorescence, 50, 100 and 200 µL of both uninduced and induced cells were centrifuged at 16,000 g at 22 °C using a benchtop mini centrifuge (Labnet International, Inc.) for 3 min. The supernatants containing LB medium was discarded and the cell pellets resuspended with 20 µL of 1 x tricine SDS sample buffer was then added to the resuspended cells and all the samples were loaded into different wells of a 10-20 % gradient tricine SDS PAGE gels (Invitrogen) and run for 2 h at 160 V at 22 °C. The gel before staining was put under a 302 nm UV light source and the image captured using a CCD camera after 5 s exposure. After capturing the image, 10 mL of GelCode Blue Safe protein stain was added to the gel and stained for 1 h at 22 °C with gentle shaking at 70 r.p.m using a mini orbital shaker (Stuart). The GelCode Blue Safe protein stain was then discarded and the gel was destained with 20 mL of milliQ water for 3 h at 22 °C and the gel was then exposed to visible light and the image was captured using a CCD camera after 10 s exposure.

2.3.3. Isolation of Membranes containing LysP-GFP fusion

6 L of over-expressed cells were harvested immidiately by centrifugation (Beckman Coulter Avanti J-20 XPI centrifuge, Beckman JLA-8.100 rotor) at 4,000 x *g* for 10 minutes at 4 °C. Supernatants were discarded and cell pellets were resuspended in 50 mL of ice cold 1 x phosphate buffer saline (PBS) (Morris *et al.* 2001). Prior to cell lysis, 1 mM $MgCl_2$ (500 mM stock) was added to restore the activity of DNAse that will be added to to cells prior to breaking down of the cells, if deactivated by heat (Bickler *et al.* 1992), pefabloc (1 mg/mL final concentration to inhibit proteases from the lysed cells not to cleave LysP) and DNase (20-100 U/mL final concentration) to break down DNA and make the cell suspension to become less viscous) (Sigma) were added. Cells were broken immidiately by sonication at 4 °C at a power of 65 % for 5

min with 1 min pause every 1 min. The unbroken cells and debris (10 mL in total) were removed by centrifugation at 15,000 x *g* for 10 min at 4°C (Beckman Coulter Avanti J-20 XPI centrifuge, JLA-16,250 rotor). 40 mL of the supernatants containing the membrane fraction was transferred to Beckman ultracentrifuge tubes and subjected to ultracentrifugation for 1 h at 120, 000 x *g* (Beckman Optima L-100 XP ultracentrifuge, Beckman 45 Ti rotor). Membrane pellets were resuspended in 30 mL of ice cold 1 x PBS buffer by pipetting and the 30 mL of the resuspended membranes was rapidly frozen in liquid nitrogen and stored at -80 °C.

2.3.4. Detergent screening of LysP-GFP fusion.

Membrane suspensions of LysP-GFP were adjusted using 1 x PBS to a total protein concentration of 3.75 mg/mL on ice. 800 µL aliquots of the suspension were transferred into 1.5 mL Eppendorf tubes. n-dodecyl-β-D-maltoside (DDM), n-decyl-β-D-maltoside (DM), n-nonyl-β-D-maltoside (NM), n-octyl-β-D-glucopyranoside (OG), n-nonyl-β-D-glucopyranoside (NG) and lauryldimethylamine-oxide (LDAO), LMNG decyl maltose neopentyl glycol (DMNG) and octyl maltose neopentyl glycol (OMNG) (all from Anatrace) were tested for membrane solubilization. 1 %(w/v) of the selected detergents (each 10 %(w/v) stock, except LDAO that was 33 %(w/v)) was added to each of the 1.5 mL Eppendorf tubes containing 800 µL of membrane suspension. For the final percentage of each detergent, the final protein concentration was 3 mg/mL in 1 x PBS buffer. The mixtures were incubated at 4 °C for 1 h with gentle shaking. The samples were then ultracentrifuged at 50,000 r.p.m for 40 min at 4 °C to remove non-solubilized material (Beckman Benchtop Ultracentrifuge, TLA-55 rotor). 500 µL out of the 1 mL supernatant containing the soluble membrane fraction, were then injected onto a Superose 6, 10/300 connected to an Äkta Explorer at 4 °C at a flow rate of 0.2 mL/min. This was carried out to investigate the monodispersity of LysP-GFP in the different detergents. 200 uL fractions were collected on a 96-well black clear bottom plate (Nunc) and analysed my measuring the GFP fluorescence (excitation 485 nm and emission 512 nm). The relative percentage solubilization yield was calculated prior to SEC as follows: GFP count of the solubilised membrane supernatant after ultracentrifugation divided by the GFP count of the solubilised membrane before ultracentrifugation multiplied by 100.

2.3.5. Purification of LysP from LysP-GFP fusion.

Membrane suspensions (30 mL from -80 °C freezer) (**Section 2.3.3**) were thawed at room temperature for 30 mins and then diluted to a total protein concentration of 3.0 mg/mL in solubilization buffer (1 x PBS, 10 %(v/v) glycerol, 150 mM NaCl and either 1 %(w/v) of DDM or LMNG) and was further incubated for 1 h at 4 °C with gentle stirring.

Non-solubilized materials were removed by centrifuging the sample at 41, 000 r.p.m for 1 h at 4 °C using rotor type 70 Ti in the OptimaTM L-100 XP Ultracentrifuge (BECKMAN COULTER). The resulting supernatant (200 mL) containing the solubilized LysP-GFP was incubated with Ni^{2+}-NTA resin (Qiagen) for 2 h at 4 °C, i.e., for 1 mg of GFP, 1 mL of resin was added. To exclude unspecific binding, 10 mM imidazole pH 7.5 was added to the sample. The slurry (10 mL) was loaded on to a glass Econo gravity column (Bio-Rad) and washed with 20 column volumes of solubilization buffer containing 10 mM imidazole pH 7.5. The column was washed with further 20 column volumes each of solubilization buffer containing 20 and 35 mM imidazole pH 7.5 at 4 °C. The protein was then eluted from the column and the eluate collected using 40 mL elution buffer (solubilization buffer containing 250 mM imidazole pH 7.5) at 4 °C. The eluate (30 mL) was incubated overnight at 4 °C with an equal amount of his-tagged TEV protease (purified using methods according to Drew *et al.* 2006), i.e., for 1 mg of GFP, 1 mg of TEV protease was added into a dialysis tube (14 kDa MWCO) and the sample dialysis against 3 L of dialysis buffer containing 20 mM Tris HCl pH 7.5, 150 mM NaCl and either 0.03 % (w/v) DDM or 0.003 %(w/v) LMNG at 4 °C with gentle stiring of the dialysis buffer with a magnetic rod at the bottom and a magnetic stirrer. After dialysis, the sample (50 mL) was clear in the presence of LMNG but slightly cloudy in the presence of DDM. The 50 mL sample consisting of GFP free LysP, GFP, TEV and likely little amount of uncleave LysP-GFP was filtered using 0.22 μm Millipore filters to remove protein aggregates and was further subjected to the second round of IMAC to remove the free GFP, TEV protease and any uncleaved LysP-GFP. This step is refered to as reverse IMAC. In reverse IMAC, a 5 mL Ni^{2+}-NTA HisTrap fast flow column (Qiagen) was equilibrated with 10 mL of either Buffer A or B at a flow rate of 0.2 mL/min at 4 °C. The dialysed sample was then pass through the column using either Buffer A or B at a flow rate of 0.2 mL/min at 4 °C and the eluent containing GFP free LysP was collected and concentrated at 3,500 g at 4 °C using rotor C0650 in the Allegra™ X-22R centrifuge (BECKMANN COULTER,

Clare, Ireland) for 5 min at 4 °C using Centricon 50 kDa molecular weight cut off (MWCO) concentrator (Millipore: MA, USA, Lot R0KA67379). After the 5 min spin, the sample was resuspended by pipetting up and down. The 5 min spin and resuspension of the sample was carried out 10 times to get a final LysP concentration of 1.5 mg/mL in Buffer A or B (2 mL final volume). Ideally, a 100 kDa MWCO should have been used but because if used, LysP is being lost in the flowthrough when concentration the sample, thus a 50 kDa MWCO was used. The concentrated sample (1.5 mg/mL in Buffer A, 2.0 mL final volume) was loaded onto a HiLoad 16/60 superdex 200 column (GE Healthcare) at a flow rate of 1.0 mL/min and the absorbance at 280 nm was monitored. Prior to loading the sample, the column was pre-equilibrated with Buffer A or buffers containing 20 mM Tris HCl pH 7.5, 150 mM NaCl, 5 %(v/v) glycerol 0.2 %(w/v) DM or 0.3 %(w/v) NM. The samples were run in similar buffers. For the concentrated sample (1.5 mg/mL in B, 2.0 mL final volume), it was loaded onto a Superose 6, 10/300 column (GE Healthcare) at a flow rate of 0.2 mL/min and the absorbance at 280 nm was monitored. Prior to loading the sample, the column was pre-equilibrated in Buffer B and the sample was run with Buffer B. For the purification in LMNG (Buffer B), Samples were collected at every stage of the purification process and analysed on SDS PAGE gel by both Coommasie staining and in-gel fluorescence as described in **Section 2.3.2**. For samples in each of the detergents, LysP was concentrated using 50 kDa molecular weight cut off concentrator (Millipore) to 10 mg/mL. The concentration of purified GFP free LysP was determined from instant ultraviolet absorbance at 280 nm using extinction coefficient from calculated amino acid sequence.

The GFP, TEV and the trace amount of LysP-GFP were eluted from the reverse IMAC column using 5 mL of 1 M imidazole pH 7.5 and dialysed in buffer containing 150 mM NaCl and 20 mM Tris pH 7.2. The dialysed sample 20 mL was stored at -80 °C and was used as GFP standard for FSEC (Section 2.3.4).

2.3.6. Purity evaluation and Determination of Oligomerization State of TEV Cleaved GFP free LysP.

The purity and oligomerisation state of LysP was estimated on SDS PAGE and Blue Native PAGE. 12 % SDS PAGE gels (ThermoFisher SCIENTIFIC) were used for SDS-PAGE and Native PAGE 4-16 % Novex Bis-Tris gels (Invitrogen) were used for Blue Native PAGE. Blue native page was carried out exactly using protocols described in http://tools.invitrogen.com/content/sfs/manuals/nativepage_sampleprep_man.pdf.

2.3.7. Biochemical/Biophysical Characterization of LysP.

2.3.7.1. Total Mass Spectrometry of LysP

The total mass of purified GFP free LysP was determined by ESI MS. For ESI MS, LysP was precipitated and expressed mailed to Dr. James Ault (Faculty of Biological Sciences, University of Leeds) where the ESI MS analysis was carried out. The precipitation was carried out on the purified LysP after SEC from Section 2.3.5 in Buffer A. The buffer used for precipitation of LysP was prepared by mixing 1 mL of acetone (Fisher Scientific, Batch 0959939), 100 µL of aqueous ammonia (Sigma, Lot STBB1636L9) and 10 µL of trichloroacetic acid (prepared by adding 100 mg of TCA in 100 µL of MilliQ water) (Sigma, Batch 07028AJ). 200 µL of the precipitation buffer was added to 5 µL of LysP at 10 mg/mL in Buffer A (50 µg of total protein) and the mixture was vortexed for 2 min and incubated on ice for 30 min. The sample was then centrifuged for 3 min in a 1.5 mL Eppendorf tube at 4,000 g at 4 °C. 500 µL of ice cold acetone and 250 µL of hexane (Sigma, Lot S66430-339) were sequentially added to the precipitated protein pellet and then dried and express mailed to Dr. James Ault (Faculty of Biological Sciences, University of Leeds) (www.astbury.ac.uk).

2.3.7.2. In-gel Tryptic Digestion of LysP

In-gel tryptic digestion was carried out for further identification of purified GFP free LysP. 1 µL of purified LysP (10 mg/mL in Buffer A) from Section 2.3.5, was run on 12 %(w/v) SDS PAGE, stained with GelCode™ blue safe protein stain and the primary GFP free LysP band (~36.6 kDa) (**Figure 2.13**) was cut out using a sterile blade into a 1.5 mL Eppendorf tube and then sent for in-gel tryptic digestion to Protein Analytics (http://www.uni-giessen.de/cms/fbz/fb11/institute/biochemie/forschung/protein-analytik/c-elegans-lab). Protein Analytics digested the sample with trypsin and subjected the tryptic digested fractions to MALDI-TOF MS and the data analysed using BioTools version 3.1 (Dave *et al.* 2011).

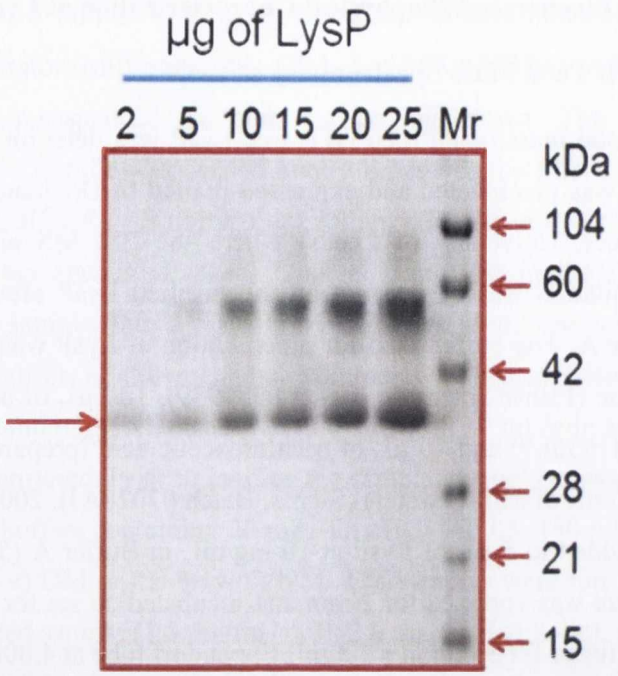


Figure 2.13. 12 % NuPAGE SDS PAGE gels of GFP free purified LysP used for in-gel tryptic digestion.

Different amounts of purified GFP free LysP were run using MES buffer for 1 h at room temperature. The gel was then stained and the primary bands were cut out using a sterile blade and sent for in-gel tryptic digestion to Protein Analytics (http://www.uni-giessen.de/cms/fbz/fb11/institute/biochemie/forschung/protein-analytik/c-elegans-lab) for analysis. The different bands seen may correspond to different folded forms of the purified GFP free LysP.

2.3.7.3. Glutaraldehyde Crosslinking Attempts.

Purified LysP (0.35 mg/mL in buffer containing 150 mM NaCl, 20 mM HEPES pH 7.5, 5 %(v/v) glycerol and 0.003 %(w/v) LMNG, was incubated with 0-200 mM glutaraldehyde (2.5 M stock) at 22 °C for two hours. Also, as a positive control, DgkA an integral membrane protein kinase (0.35 mg/mL in 150 mM NaCl, 20 mM HEPES pH 7.5 and 0.2 %(w/v) DM) was incubated with 0 and 16 mM glutaraldehyde at 22 °C for two hours. Samples were then analysed on 12 % NuPAGE SDS PAGE gels.

2.3.7.4. Stability Studies

Purified LysP concentrated to 10 mg/mL in Buffer B from Section 2.3.5 was kept on ice and aliquots (50 µg, i.e., 5 µL diluted with 500 µL of Buffer B) were analysed on

days 0, 1, 3, 5 and 7 by injecting onto a Superose 6, 10/300 column connected to an Äkta Explorer (GE Healthcare) at 4 °C and a flow rate of 0.2 mL/min using Buffer B. Prior to injecting the sample, the column was pre-equilibrated with Buffer B at flow rate of 0.2 mL/min at 4 °C.

2.3.7.5. Circular Dichroism (CD) Measurements

CD spectra of GFP free LysP at 0.6 mg/mL in Buffer A was recorded using Jasco J-815 CD Spectrophotometer from 190 nm to 260 nm using a scan speed of 100 nm/min, a band width of 1 nm, a path length of 1 mm and a sensitivity stand at 100 mdeg at 20 °C. Each measurement was done three time with different mixtures prepared each time. Raw data were processed using the K2d method in the online CD analysis webpage (DIOCROWEB) (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml).

2.3.7.6. [^{3}H] L-lysine Transport, Transport Inhibition and Transport pH Screening in Whole Cell.

This is to show that the overexpressed LysP-GFP is functionally active and transports L-lysine, we used whole cell radioactivity transport assay. This experiment was carried out in the Lab of Prof. Peter Henderson with the help and guidance of Dr. Preethi Sukumar, at The Department of Biological Sciences, University of Leeds, from 17th to 25th of July 2011. **Figure 2.14** shows a flow chart of the steps involved in the whole cell radioactivity transport assay of LysP-GFP. An overnight starter culture of the *E. coli* C43 (*DE3*) cells harbouring the *lysP*-GFP fusion was prepared by inoculation from LB agar/ 50 µg/mL kanamycin plate with a single colony of *E. coli* C43 (*DE3*) cells harbouring the *lysP*-GFP fusion into 50 mL of LB medium supplemented with 50 µg/mL kanamycin. Cells were grown using the INFORS multitron standard incubator shaker for 16 h at 37 °C with shaking at 200 r.p.m to and OD_{680} of 4. The cells were then diluted to a final OD_{680} of 0.08 using 100 mL minimal media supplemented with 50 µg/mL kanamycin (100 µL of 50 mg/mL in milliQ water). The cells were grown at 37 °C with shaking at 200 r.p.m to an OD_{680} of 0.6 and were either induced with 0.4 mM IPTG or not induced (this serve as a control) and grown further for 1.5 h. The cells (100 mL) were split into two 50 mL falcon tube were then harvested at 4,500 r.p.m for 10 min at 20 °C using AllegraTM X-22 R centrifuge (BEKMANN COULTER). The cells were then resuspended and washed five times in 50 mL of buffer containing 150 mM KCl and 5 mM MES pH 6.6 with spinning in between (4,500 r.p.m for 10 min at 20 °C), to get a final OD_{680} = 1.8-2.0. The transport assay was carried out by the

addition of 20 mM glycerol to 1447 μL of the cells (OD_{680} = 2.0) and the cells were then aerated for 3 min at 25 °C. Transport assay was immediately initiated by the addition of 37.5 μL of [^{3}H] L-lysine (stock 2 mM, 10 μCi) (Perkin Elmer, Catalog number 6013389). For transport inhibition 1398 μL of cells (OD_{680} = 2.0) were used instead of 1447 μL with the addition of 50 μL of each inhibitor (3.3 mM final concentration each of cold L-4-thialysine, D-lysine, L-histidine, L-arginine, L-lysine, L-leucine, L-phenylalanine, L-proline, L-asparagine, L-serine, L-glutamate in the reaction mix and with a control with no inhibitor added but instead 50 μL of buffer containing 150 mM KCl and 5 mM MES pH 6.6). For pH screening, the buffer used for cell resuspension consist of 150 mM KCl with one of the following (10 mM potassium acetate pH 4.01, 10 mM potassium acetate pH 4.97, 5 mM MES pH 6.06, 5 mM MES pH 6.6, 5 mM Tris pH 7.17, 5 mM Tris pH 7.93, 5 mM Tris pH 8.01, 5 mM glycine pH 9.0 and 5 mM glycine pH 10.0). 200 μL samples were taken from the reaction mix at different time frame and applied to a 0.4 μm pore size MF-Millipore Membrane Filter. The filters were immediately washed under vacuum with 4 mL of ice-cold buffer containing 150 mM KCl and 5 mM MES pH 6.6. The washed filter containing cells was transferred to a 20 mL scintillation vial and submerged in 10 mL Emulsifier-Safe™ Scintillation Cocktails (PerkinElmer) to measure the amount of hot L-lysine transported by the cells. Background counts were determined by washing filters under suction with counterflow buffer without radiolabelled substrate. Standards (to ensure that the [^{3}H] L-lysine used was radioactive), were done by applying 4 μL of the radiolabelled solution directly onto a pre-soaked filter without vacuum suction. The transport of radiolabelled substrate was calculated using the following equations and a step by step calculation of how the conversion was done using an example of the transport inhibition of L-lysine in the presence of L-4-thialysine after 15 sec as shown below.

Uptake (nmole/mg) = (average dpm of sample- average d.p.m of background) * [total volume / sample volume] * [1/no. of mg of cell] x [1/(average standard dpm/nmol)].

Number of mg of cells = total volume of reaction mix * A_{680} of the cells used for transport * 0.68 (1 A_{680} = 0.68 mg dry mass cells)

For example: dpm average of sample = 7392.2, dpm average of standard= 947432.5, average background dpm = 43.8, total volume= 1500 μL, sample volume = 200 μL, A_{680} was 2.09 = 2.13 mg of cells,

Standard no. of moles = concentration * volume = 0.002 mol/L * ($4x10^{-6}$)L = 8 x 10^-9 mole = 8 nmol.

Average standard dpm/nmol = 947432.5/8 nmol = 118429.1 $nmol^{-1}$

Therefore uptake (nmol/mg) = (7392.2-43.8)*(1500 µL /200 µL)*(1/2.13 mg)*(1/118429.1 nmol) = 0.22 nmol/mg of 3[H] L-lysine transported per mg of protein.

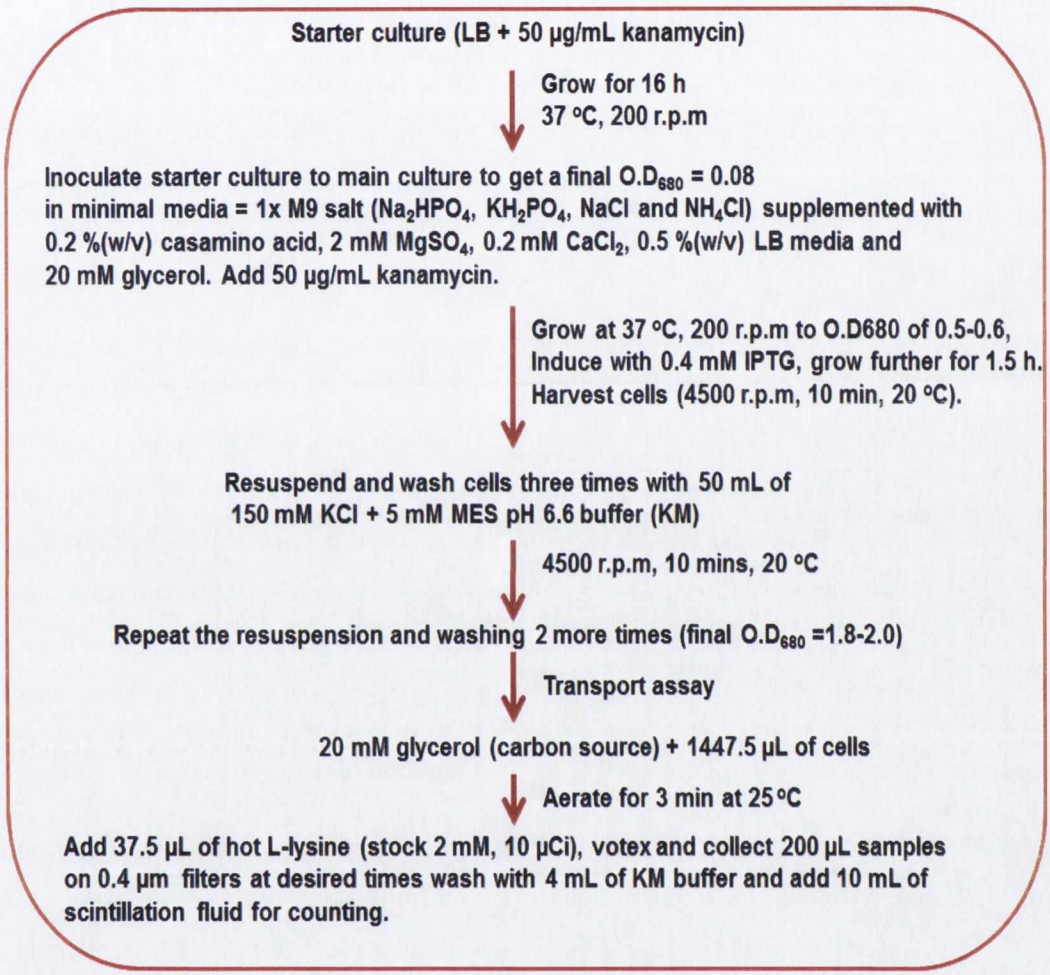


Figure 2.14. Flow chart illustrating the steps involved in the radioactivity whole cell transport assay.

2.3.7.7. Reconstitution and Counterflow Assay

The reconstitution and counterflow assay begins with the purification of the lipid to be used for reconstitution with acetone and ether. **Figure 2.15**, shows a flow chart of the acetone/ether washing of the lipid. 100 mg of *E. coli* polar lipid (Avanti) (white solid

powder) was dissolved in 2 mL of chloroform and mixed. 10 mL of acetone under argon containing 2 μL of β-mercaptoethanol (14.7 M stock) and the mixture was stirred for 16 h at 4 °C. The precipitated lipid was sedimented by centrifuging at 10, 000 x g for 12 min at room temperature and the supernatant discarded. The remaining lipid pellet was resuspended in 10 mL of diethyl-ether under argon containing 2 μL of β-mercaptoethanol (14.7 M stock stock) and centrifuged at 8,000 x g for 10 min at room temperature. The supernatant containing the dissolved lipid was retained and the ether removed by evaporation using argon, to produce a lipid film. The purified lipid was resuspended in chloroform at a final concentration of 10 mg/mL, under argon and stored at -20 °C. Glass tubes were used throughout at room temperature as chloroform dissolves plastic tubes. Purified lipid extract stored in chloroform was dried under argon and the lipid film rehydrated to a final concentration of 10 mg/mL using reconstitution buffer (50 mM potassium phosphate pH 7.6, 20 mM cold L-lysine, 1 mM DTT) by a sequential series (10 times) of vortexing (30 sec) and heating at 50 °C. To aid rehydration a series of incubations at 50 °C and mixing were performed until the lipid film had completely dissolved. Unilamellar vesicles of size, 0.1 μm were prepared using an extruder (Lipex Biomembranes, Inc. Vancouver, British Columbia) at a pressure of 300 psi to pass the lipid solution once through 0.4 μm polycarbonate Nucleopore™ Track Etch membrane double filters, it was then passed once through 0.2 μm polycarbonate Nucleopore™ Track Etch membrane double filters and 7 times through 0.1 μm pore size polycarbonate Nucleopore™ Track Etch membrane single filters. The extrusion was carried out with the extruder kept at 50 °C in a water bath. This produced a homogenous suspension of vesicles at 0.1 μm size. The liposomes were used immediately for reconstitution of membrane proteins. For reconstitution, β-octyl-glucoside (OG) was used to destabilise the preformed liposomes to allow the insertion LysP into liposomes. This method has been proven to be rapid and reliable (Racker *et al.* 1979). The protocol was performed as described by Ward *et al.* 2000. 1 mL of lipid at 10 mg/mL in reconstitution buffer were mixed at 4 °C for 15 min with 1.25 %(w/v) solution of β-OG (10 %(w/v) stock solution) and purified LysP from Section 2.3.5 at 10 mg/mL in Buffer A (100:1 (w/w), lipid:LysP respectively) in a 5 mL glass test tube. For the control (liposomes), equal volume of Buffer A (10 μL) was added in place of purified LysP (10 mg/mL in Buffer A). Excess β-OG was removed from the mixture by diluting with 70 mL of reconstitution buffer and the sample was centrifuged for 1 h at 100, 000 x g at 4 °C to pellet the proteoliposomes. The supernatant was discarded carefully and the proteoliposome pellets resuspended in 150

μL reconstitution buffer on ice after which they were used directly. Counterflow assays were performed as described by Ward *et al.* 2000. Proteoliposomes or liposomes (40 μL) were added to 920 μL of counterflow buffer (50 mM potassium phosphate pH 7.6 and 1 mM DTT) and the counterflow assay initiated by addition of 40 μL of radiolabelled [^{3}H] L-lysine (stock, 50 μM, 100 μCi/mL) (Perkin Elmer, Catalog number 6013389). After the addition of radiolabelled L-lysine samples (80 μL) were taken at 15 sec, 1 min, 2 min, 3 min, 5 min, 10 min, 15 min, 30 min, 1 h, 2 h, 3 h, 4 h and 5 h time period and applied to 0.2 μm pore size MF-Millipore Membrane Filter. The filters were immediately washed under vacuum with 4 mL of ice-cold counterflow buffer. The washed filter was transferred to a 20 mL scintillation vial and submerged in 10 mL Emulsifier-Safe^{TM} Scintillation Cocktails (PerkinElmer). For inhibition studies, 10 mM of cold substrate (20 standard amino acids, L-4-thialysine and D-lysine) was added to the outside of the proteoliposomes at internal pH 7.0 and external pH 4.0 and samples were collected after 3 min and 20 min. The inhibition study was done by Simon Patching in Prof. Peter Henderson's laboratory, University of Leeds. The level of radioactivity attained by the liposomes was measured by liquid scintillation counting. **Figure 2.16**, shows a flow chart of the different steps involved in the reconstitution of LysP into proteoliposomes and the entrance counterflow assay. Background counts were determined by washing filters under suction with counterflow buffer without radiolabelled substrate. Standards (to ensure that the [^{3}H] L-lysine used was radioactive), were done by applying 5 μL of the radiolabelled solution directly onto a pre-soaked filter without vacuum suction. The uptake of radiolabelled substrate was calculated using the following equation:

Uptake (nmol/mg) = (dpm-background) * total volume / sample volume * 1 / mg protein * No. moles Standard / dpm Standard

mg of LysP = ([LysP] * Volume) / total Volume

No. of moles Standard = moles of hot L-lysine * moles of cold L-lysine * Volume of Standard / total Volume.

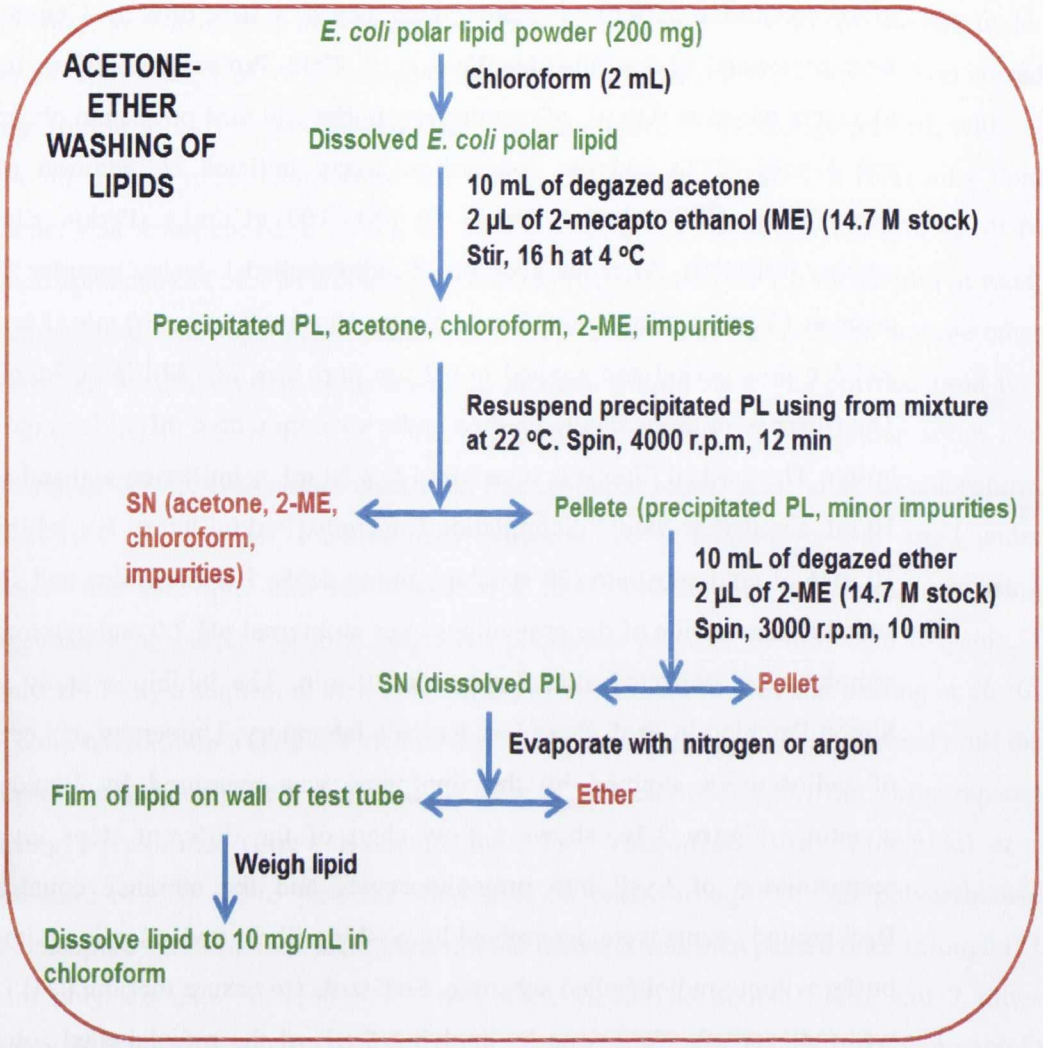


Figure 2.15. Flow chart showing the steps involved in the acetone-ether washing of polar *E. coli* lipid.

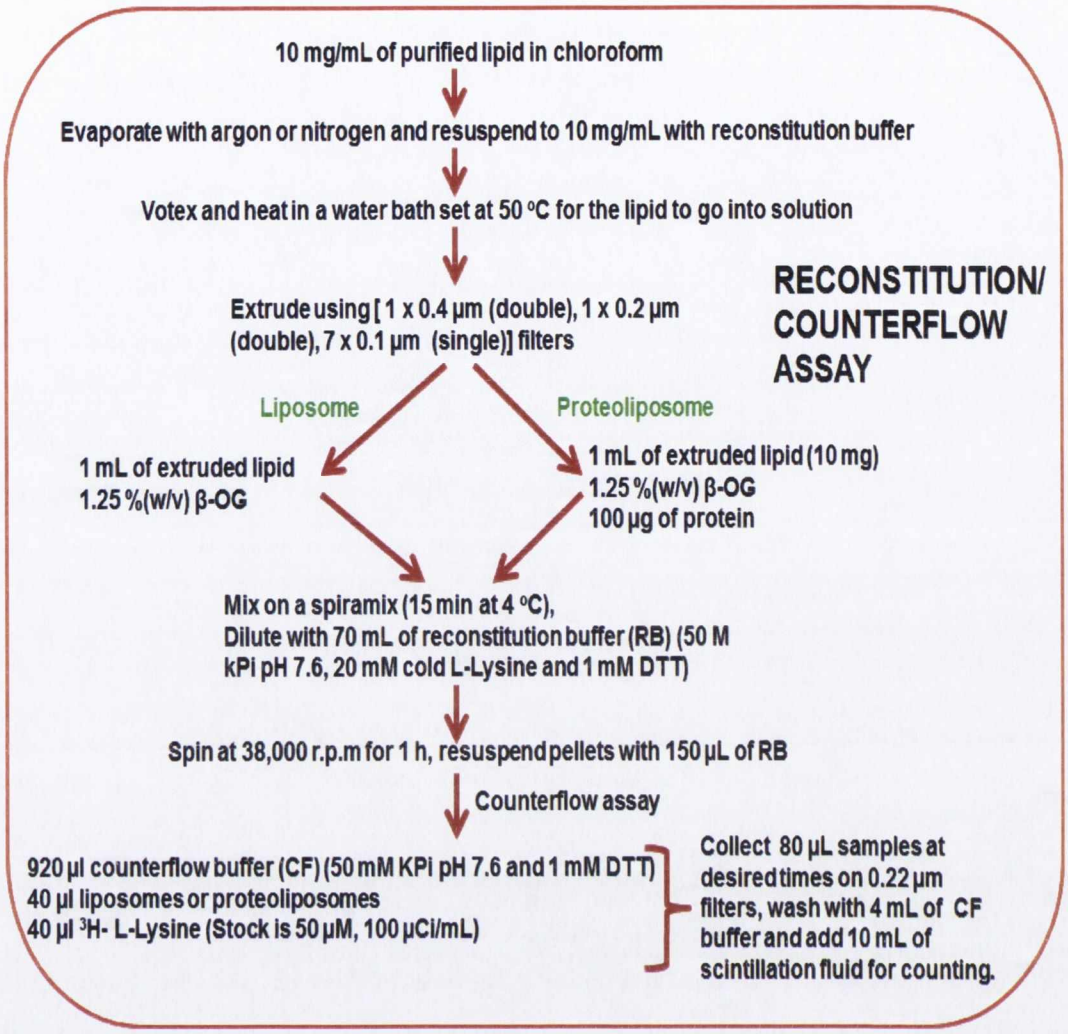


Figure 2.16. Flow chart showing the different steps involved in the reconstitution of LysP into proteoliposomes and the entrance counterflow assay.

2.3.7.8. Microscale Thermophoresis (MST) Analysis.

MST was carried out as described by Jerabek-Willemsen *et al.* 2011 and Wienken *et al.* 2010). This experiment was carried out at NanoTemper Technologies GmbH, Floβergasse 4, 81369 Munich, Germany from 11th to 12th of October 2012 under the direction of Anna Muench. Serial dilutions of each ligand were prepared starting at 800 µM for L-lysine, 6.4 mM for L-4-thialysine, 50 mM D-lysine, 200 mM of L-arginine and 100 mM each of L-histidine, glycine, L-proline, L-glutamine, L-glutamate, L-methionine, L-isoleucine, L-leucine, L-threonine, L-valine and L-alanine to obtained 16 different concentrations of ligand at 22 °C. In the serial dilutions, the 16 Eppendorf tubes (200 µL volume) had 30 µL of the highest concentration of L-lysine in the first

and 15 μL was transfer to the next having 15 μL of Buffer B. The two solutions were mixed by pipetting up and down and 15 μL was then transfered to the next also having 15 μL of Buffer B, the procedure was repeated to the last tube and 15 μL from the last dilution was discarded to leave a 15 μL mixture of L-lysine and Buffer B. All ligands were prepared in Buffer B. All concentrations of ligands were mixed with protein in a 1:1 v/v ratio (15 μL of 1.2 μM LysP from Section 2.3.5 in Buffer B was added to the 15 μL of the 16 different dilutions) and 4 μL were loaded into a glass capillaries (5 cm long and 200 μm in internal diameter) (http://www.nanotemper-technologies.com/products/consumables/capillaries/). The final protein concentration was 0.6 μM. MST-analysis was carried out using Monolith LabelFree instrument (NanoTemper) using a Laser power and LED sensitivity of 20 % at 22 °C. All ligands were prepared in Buffer B.

2.3.7.9. Tryptophan Fluorescence Change

Purified LysP from **Section 2.3.5** was diluted into Buffer A to a final concentration of 0.26 μM (3 mL final volume) and pipetted into a 5 mL quartz cuvette and allowed to thermally equilibrate for 5 min at 20 °C. The emission spectrum was read instantly from 470 nm to 300 nm with excitation set at 280 nm. The integration time was 1 sec, the increment was 1 nm and the excitation and emission slits was 2 nm. The same procedure was repeated with the addition of 10 mM L-Lysine (Fluka, Lot 1370953) to a final volume of 3 mL. The mole ratio of LysP to L-lysine was 1:26 and the change in volume upon addition of 10 mM L-Lysine was 1 %. As controls, 10 mM L-lysine was added into Buffer A with no LysP, the spectrum was recorded as described above and also the spectrum was recorded for Buffer A.

2.4. Results and Discussions

2.4.1. Recombinant Expression of LysP-GFP fusion and In-gel Fluorescence

The pWaldo-GFPd expression vector harboring the *lysP* gene was transformed and over expressed in *E. coli* (DE3) C43 cells as described in the Materials and Methods **section 2.3.2.** In-gel fluorescence was used to investigate if the whole cell GFP counts (**Table 2.III**) was coming from LysP-GFP and not free GFP, the uninduced and induced cells were analysed. **Figure 2.17,** showed that there was no free GFP detected (the instrument can detect up to less than 5 ng of GFP). Typical yield was estimated at approximately 44 mg/L of cell culture medium as follows: Firstly, the concentration of GFP was determined from (BSA) standard curve (**Figure 2.18**) from which a standard curve of fluorescence (excitation @ 485 nm and emission @ 512 nm) against different concentration of the GFP was generated (**Figure 2.19**). The amount of GFP from LysP-GFP fusion was done using the linear equation of **Figure 2.19** ($Y = 187523X + 918.59$, where Y, is the fluorescence and X, is the concentration of GFP) and thus the amount of LysP by correcting for the molecular weight ratio (2:1) of LysP:GFP respectively by multiplying the amount of GFP by 2, i.e., LysP has a molecular weight of 54 kDa and GFP, 28 kDa. In this case the molecular weight ratio in solution will be 2 LysP and 1 GFP (54 kDa/28 kDa). Thus, we need to multiply the GFP concentration by 2 to get the LysP concentration.

Table 2.III. Whole cell GFP fluorescence count of LysP-GFP. The GFP count was recorded from 100 μL of both uninduced and induced cells for three different culture media (GFP count 1, 2 and 3) using the SpectraMax microplate reader (excitation 485 nm and emission 512 nm).

	GFP count 1	GFP count 2	GFP count 3	Average count	Standard deviation	Coefficient of variation
Uninduced	45.3	40.8	42.9	43	2.25	0.05
Induced	3492.2	3473.1	3441.9	3435.7	81	0.02

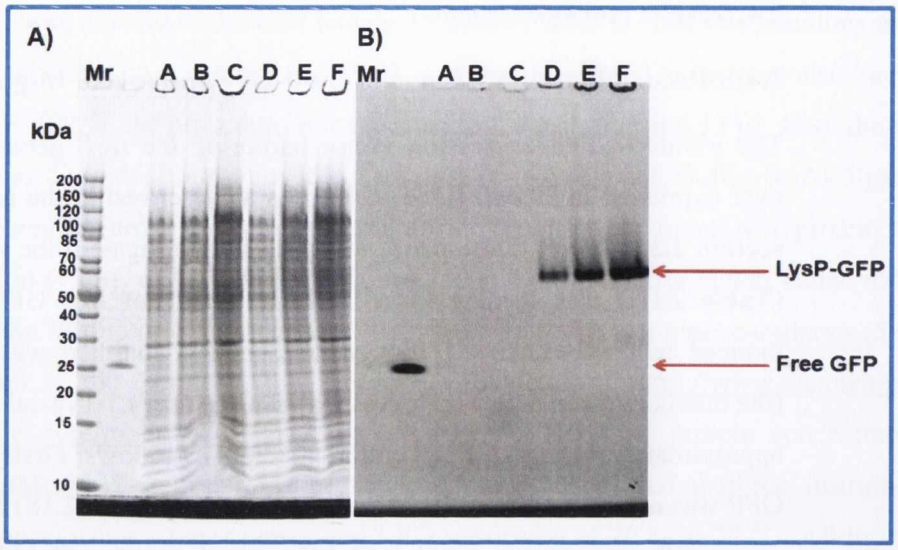


Figure 2.17. Accessing the integrity of overexpressed LysP-GFP by in-gel fluorescence and GelCode Blue Safe protein staining.

A) GelCode Blue Safe protein stained gel of LysP-GFP and B) In-gel fluorescence gel of LysP-GFP. Gel A and B are the same. After capturing the image of the gel (B) at 302 nm under UV light using a CCD camera after 5 s exposure, the gel was then stained afterwards with GelCode Blue Safe protein stain and the image was captured under visible light after 10 s exposure (Gel A). Lanes A, B and C correspond to 50, 100 and 200 μL of Uninduced whole cells resuspended by pipetting, in 50 μL of SDS sample buffer, 15 μL was loaded. D, E and F correspond to 50, 100 and 200 μL of induced whole cells resuspended in 50 μL of SDS sample buffer, 15 μL was loaded. Expected size of LysP-GFP = 54 kDa + 28 kDa = 82 kDa, free GFP runs at 28 kDa as expected but LysP-GFP runs at about 61 kDa higher than expected because it is highly folded compared to water soluble proteins.

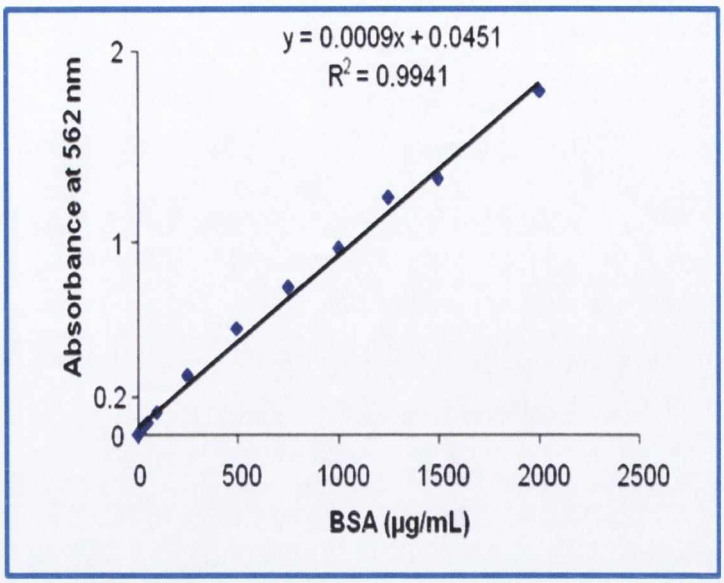


Figure 2.18. Bicinchoninic acid assay (BCA assay) for the determination of GFP concentration.

BSA was used to generate the standard curve (absorbance at 562 nm versus µg/mL concentration of BSA). From the standard curve the GFP concentration was determined.

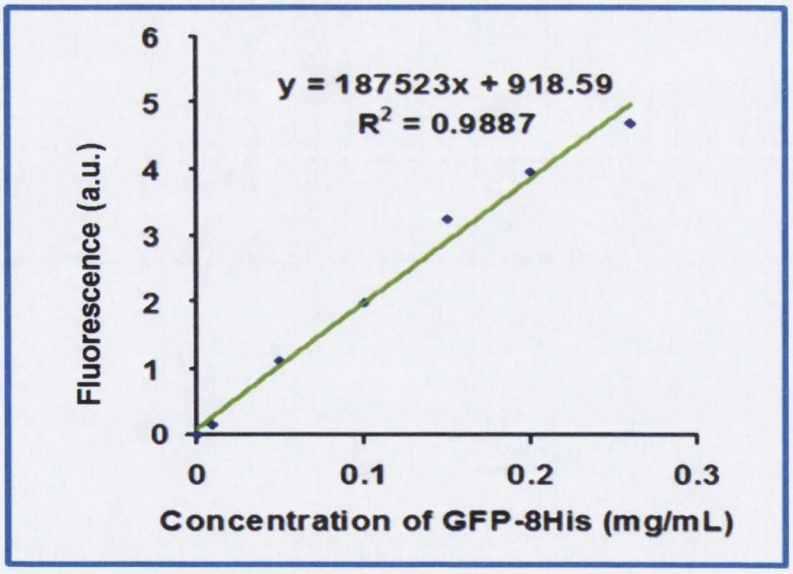


Figure 2.19. GFP-His$_8$ standard curve (fluorescence excitation at 485 nm and emission at 512 nm versus mg/mL concentration of GFP-His$_8$).

From the standard curve the concentration of GFP-His$_8$ from LysP-GFP-His$_8$ fusion was determined and thus the concentration of LysP.

2.4.2. Detergent screening of LysP-GFP

After the expression of LysP-GFP, membranes were isolated and detergent screening carried out. Detergent screening was carried out by solubilizing the membranes with 1 %(w/v) of the following detergent LMNG, DDM, DMNG, DM, OGNG, LDAO and with 2 %(w/v) NM and OG and the soluble membrane fraction analysed by FSEC. The top 4 detergents based on the detergent solubilization yield were DDM, DM, NM and LMNG (**Figure 2.20**) and monodisperse on SEC (**Figure 2.24**). For clarity, the top 4 detergents were shown separately in **Figure 2.22**. LDAO was eliminated based on lack of monodispersity on SEC. OGNG was eliminated based on solubilisation yield and OG based on both solubilisation yield and monodispersity on SEC. 2 %(w/v) were used for NM and OG because they have high CMC unlike the other detergents with low CMC compared to NM and OG (**Table 2.IV**).

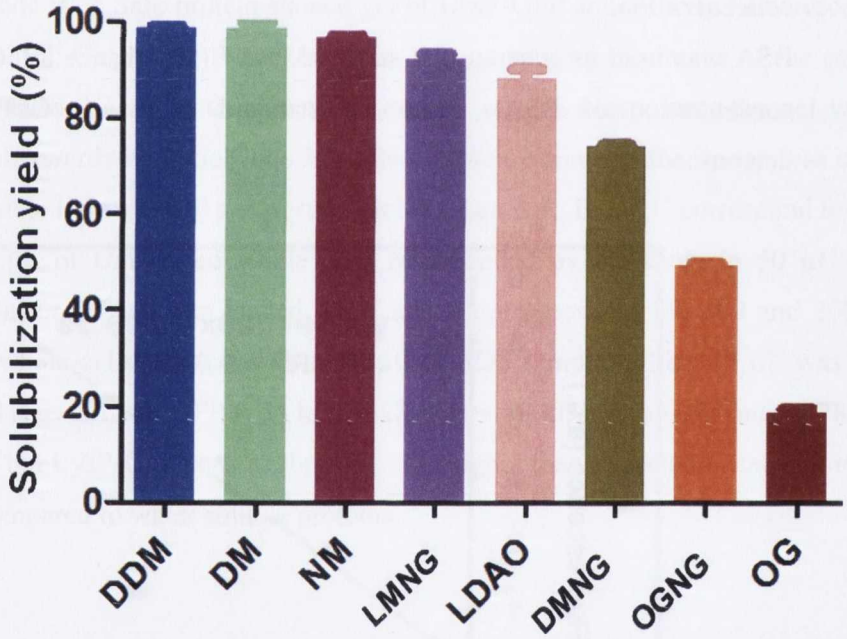


Figure 2.20. Detergent solubilization yield of LysP-GFP in the different detergents tested.

Membrane fractions containing LysP-GFP fusion were solubilized in 1 %(w/v) of all the detergents except NM and OG in which 2 %(w/v) were used. The relative percentage solubilization yield was calculated as follows: GFP count of the solubilized membrane supernatant after ultracentrifugation divided by the GFP count of the solubilized membrane before ultracentrifugation multiplied by 100. Error bars represent three different measurements.

Table 2.IV. List of detergents used, their CMC in water and their aggregation number in water.

Detergent	CMC in water (percent)	Aggregation number in water
DDM	0.0087	~ 78-149
DM	0.087	~ 69
NM	0.28	~ 25
NG	0.20	~ 133
OG	0.53	~ 27-100
LMNG	0.0019	N/A
DMNG	0.001	N/A
OGNG	0.001	N/A

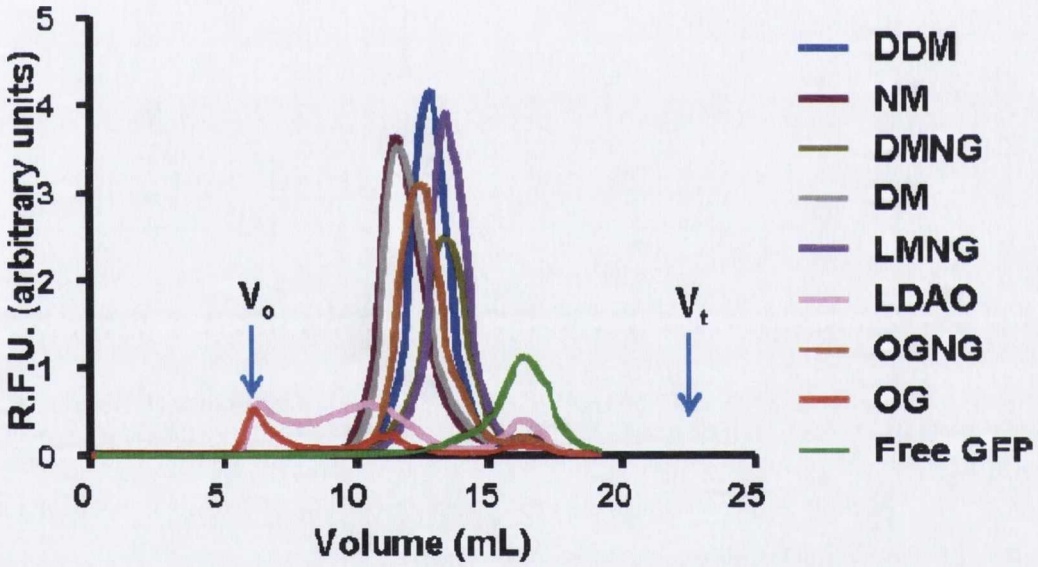


Figure 2.21. Fluorescence detection size-exclusion chromatography (FSEC) traces of LysP-GFP fusion in different detergents.

Membrane fractions containing LysP-GFP fusion were solubilised in 1 %(w/v) of all the detergents except NM and OG in which 2 %(w/v) were used. The GFP fluorescence of the collected fractions was monitored at 4 °C at a flow rate of 0.2 mL/min using Superose 6, 10/30 column connected to an Äkta Explorer. The column was pre-equilibrated and run with Buffer B. The GFP fluorescence (excitation wavelength of 485 nm and an emission wavelength of 512 nm) of 200 µL samples collected on a

Nunc black 96-well optical clear bottom plate. Free GFP from Section 2.3.5 was also analysed to detect the presence of free GFP in the samples. Vo is the excluded volume and Vt in the included volume.

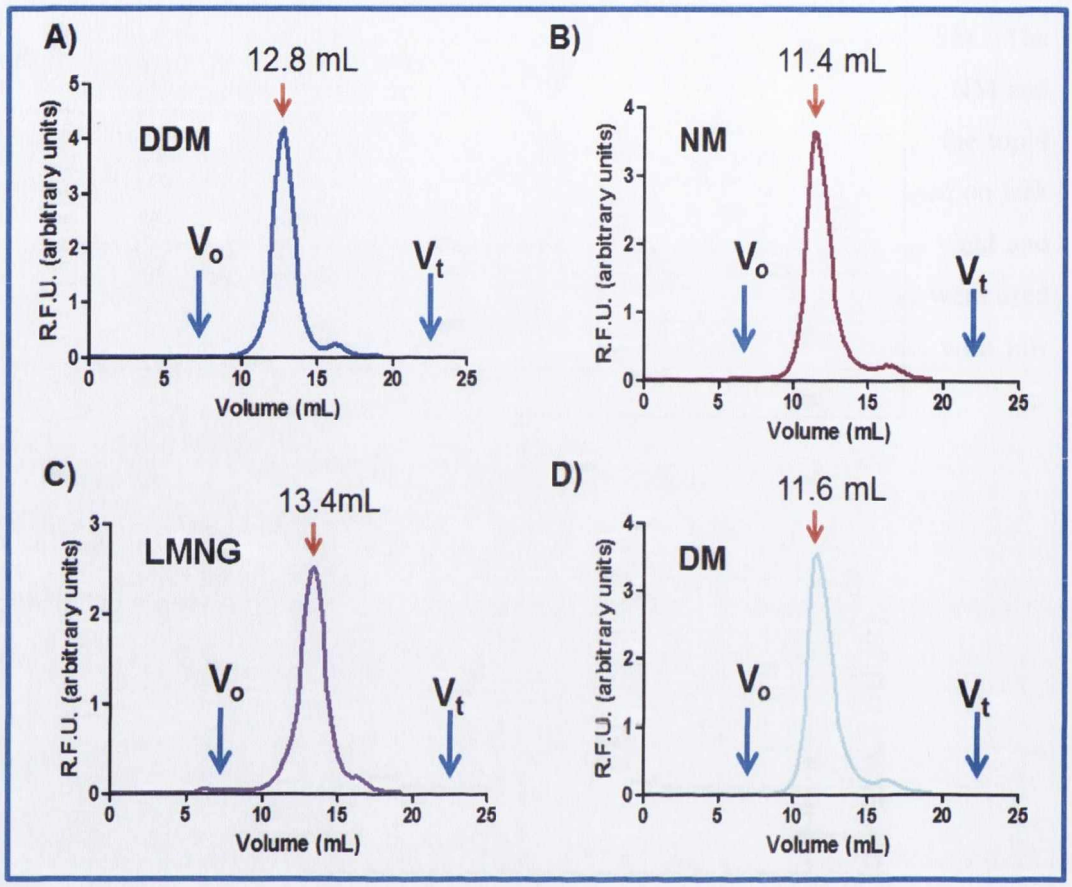


Figure 2.22. FSEC of LysP-GFP in the top 4 detergents (DDM, NM, LMNG and DM).

A) FSEC trace in DDM, **B)** FSEC trace in NM, **C)** FSEC trace in LMNG and **D)** FSEC trace in DM. Figures derived from **Figure 2.21**.

2.4.3. Purification of LysP from LysP-GFP Fusion

The top 4 detergents selected for further purification of LysP from LysP-GFP were LMNG, DDM, DM, NM as judged by their monodispersity (**Figure 2.21**) and solubilization yield (**Figure 2.20**). LysP was purified in all 4 detergents and all showed monodispersity on SEC except NM in which the SEC profile was not Gaussian (**Figure 2.24, this figure shows the SEC profiles for the TEV cleaved LysP only**). Despite the monodispersity of LysP on SEC in these detergents, only LMNG was the best in

keeping LysP stable on ice for at least 7 days after concentrated to 10 mg/mL in Buffer B (**Figure 2.30**). LysP precipitated in DDM on ice overnight at 10 mg/mL in Buffer A. In DM, LysP could not be concentrated to 10 mg/mL as the protein aggregates during concentration. Thus, LMNG was selected as the best detergent for LysP purification. The final yield of purified LysP after SEC was 0.5 mg/L of cell culture medium. The concentration of purified GFP free LysP after SEC was 0.4 mg/mL.

An analysis of the samples at different steps in the LysP purification from LysP-GFP using LMNG was carried out by Gelcode blue safe staining and in-gel fluorescence of the samples run on SDS PAGE gel (**Figure 2.23**). The overexpressed full length LysP-GFP remains intact throughout the purification process except during the cleavage with TEV. The TEV cleavage was not complete as some residual full length protein could be seen by in-gel fluorescence (**lane J**). This residual LysP-GFP was removed from the free LysP during reverse IMAC as seen in in-gel fluorescence lane M.

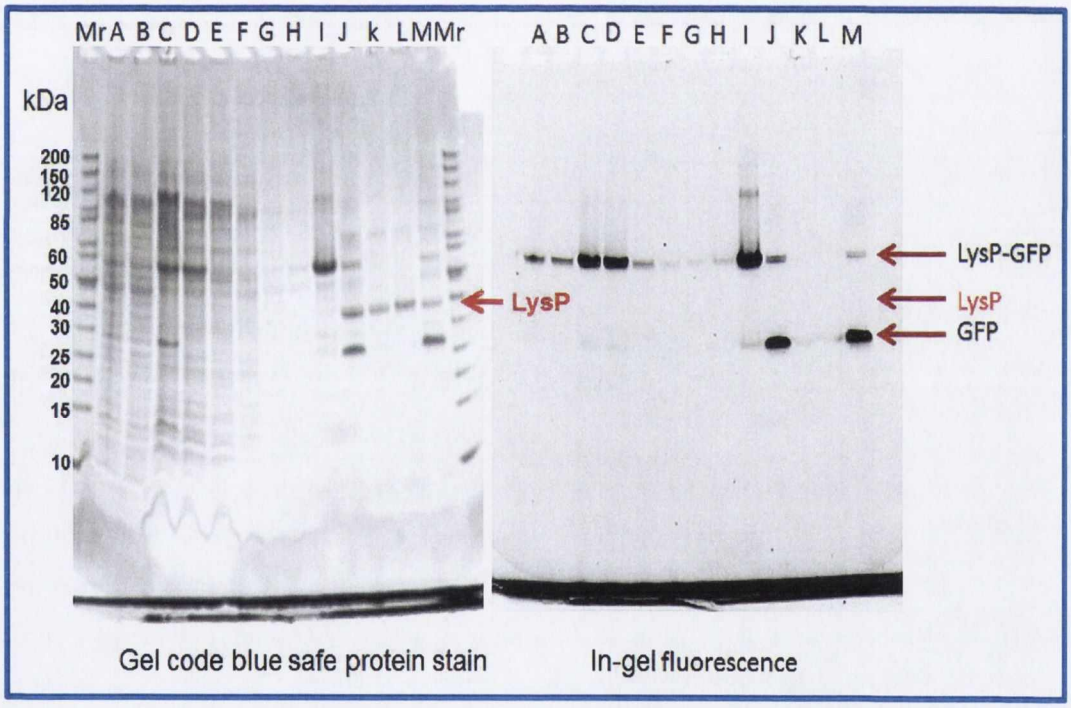


Figure 2.23. Purification of LysP. SDS PAGE of samples at every step in the purification process (left) and in-gel fluorescence of samples at every step in the purification process.

Mr = Marker, A= Cell homogenate, B= low speed supernatant, C= Membrane, D= solubilized membrane supernatant, E= Flow through from IMAC, F= 10 mM imidazole wash, G= 25 mM imidazole wash, H= 35 mM imidazole wash, I= Ni-NTA elution, J=

TEV cleavage mix, K= Flow through from reverse IMAC, L= Sample after SEC, M= 250 mM imidazole elution from reverse IMAC. Samples were run without heating.

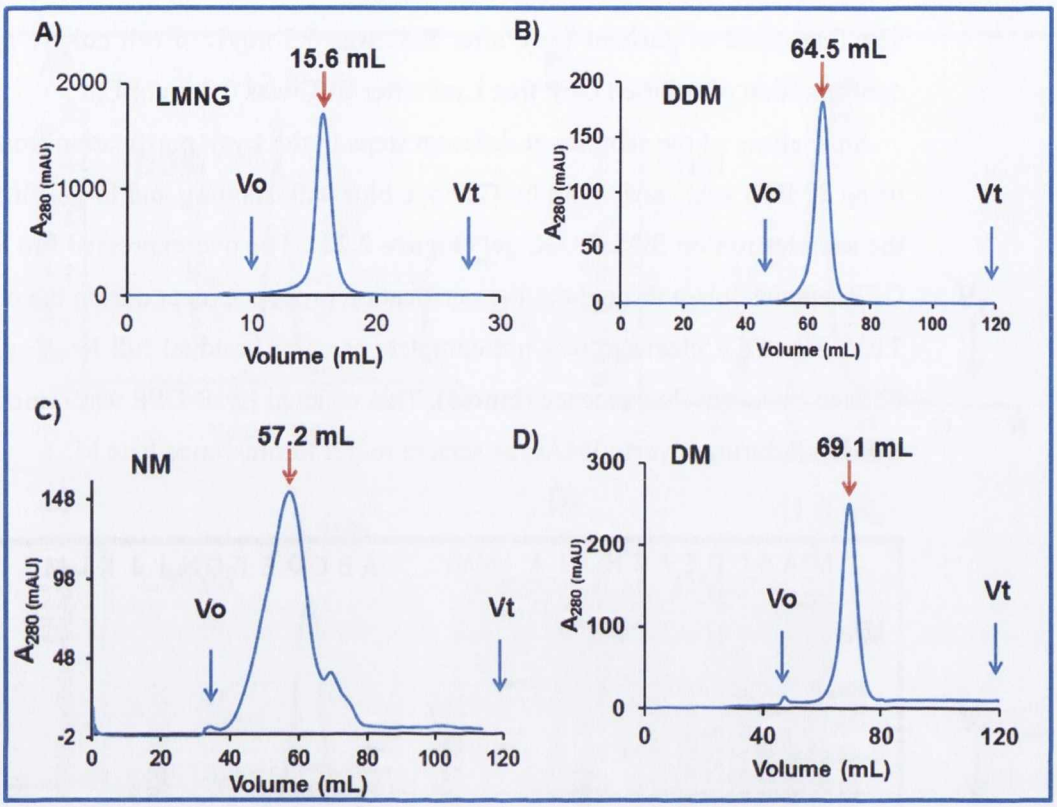


Figure 2.24. SEC chromatogram of TEV cleaved free LysP.

A) Protein was solublized and purified in LMNG, **B)** Protein was solubilized and purified in DDM, **C)** Protein was solubilized in DDM and exchanged to NM on SEC and **D)** Protein was solubilized in DDM and exchanged to DM on SEC. Samples (B, C and D) were loaded onto a HiLoad 16/60 Superdex 200 column (GE Healthcare) connected to an Äkta Explorer at a flow rate of 1 mL/min in Buffer A at 4 °C and the absorbance at 280 nm was monitored. Prior to loading the sample, column was pre-equilibrated with Buffer A (B) or Buffer A in which the DDM was replaced with NM (C) or DM (D) at 4 °C. For A) sample was loaded onto a Superose 6, 10/300 column (GE Healthcare) pre-equilibrated in Buffer B at a flow rate of 0.2 mL/min.

Free LysP is monodisperse in LMNG **(Figure 2.24 A),** DDM **(Figure 2.24 B),** and DM **(Figure 2.24 D),** which is consistent with FSEC (**Figure 2.22**). In NM, free LysP was not monodisperse as on SEC as there was a minor peak observed at 70 mL **(Figure 2.24 C)**, thus cleavage of GFP may have destabilizing the protein.

Molecular weight standards were run on the Superose 6, 10/300 column under same condition that was used to purify LMNG (Buffer B) to estimate its apparent molecular weight. **Figure 2.25** shows the gel filtration chromatogram of LysP in LMNG and inset is the calibration curve of the different standards run. 8 mg of the different standards (chymotrypsin, ovalbumin, albumin, conalbumin and aldolase) were dissolved in 1 mL of callibration buffer containing 150 mM NaCl and 20 mM Tris pH 7.2. 500 μL of each were run on the Superose 6, 10/300 at a flow rate of 0.2 mL/min using calibration buffer at 4 °C. 500 μL of GFP from Section 2.3.5 was also included. The peak elution volumes were used to generate the calibration curve. The size estimated as 66 kDa (15.6 mL) which is close to monomeric size of LysP (54 kDa), expected size (LysP + LMNG micelle, i.e., 54 + 70 respectively) is 114 kDa assuming micelle size of LMNG = DDM (That for LMNG is not yet available). This discrepancy in the determined size compared to the theoretical size could be as a result of the fact that the size of detergents varies on SEC. This data shows that LysP is probably a monomer in LMNG micelle. This results was confirmed by blue native page (**Figure 2.25**).

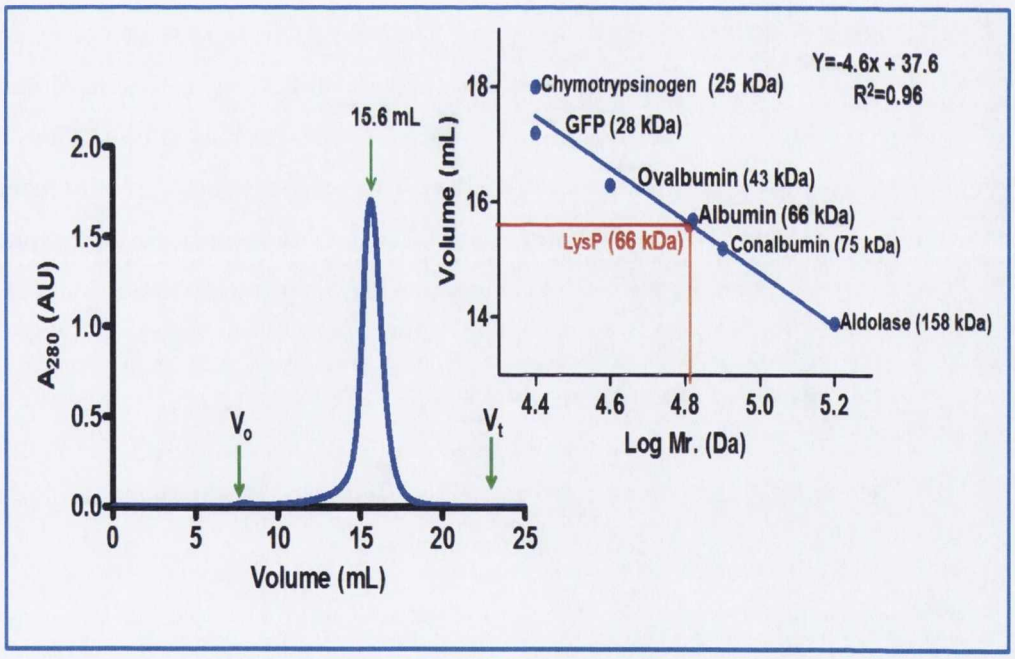


Figure 2.25. Size-exclusion chromatography (SEC) of TEV cleaved free LysP.
500 μL of 2 mg/mL LysP in Buffer B was loaded onto a Superose 6, 10/300 column (GE Healthcare) connected to an Äkta Explorer at a flow rate of 0.2 mL/min and the absorbance at 280 nm was monitored. Prior to loading the sample, column was pre-

equilibrated with Buffer A at 4 °C. Inset is a calibration curve of different molecular weight standards. From calibration curve Ve = 15.5 mL corresponding to 62.5 kDa, expected size = 114 kDa (LysP + LMNG = 54 + 70). This suggests that LysP purifies as a monomer in LMNG.

2.3.6. Purity Evaluation and Determination of Oligomerization State of GFP free LysP.

After successfully purifying LysP by SEC, we then estimated the purity by SDS PAGE and blue native page (BNP) which was greater than 95 % (**Figure 2.26**). The size and hence the oligomerization state of LysP could only be estimated from BNP and not SDS PAGE because LysP runs at a lower molecular weight than its expected size (36.6 kDa instead of 54 kDa) on SDS PAGE. This phenomenon has been seen for most membrane protein membrane proteins (Fang *et al.* 2006). Thus, we used BNP for size estimation. LysP runs at a molecular weight of 129 kDa on BNP. To account for the amount of Coomassie that was bound to LysP the mass on BNP was divided by a factor of 1.8 (Heuberger *et al.* 2002). Thus the mass obtained on BNP was estimated for LysP at 129/1.8 = 71 kDa. The expected size from amino acid sequence of LysP is 54 kDa which is 17 kDa less than what was obtained on BNP. The 17 kDa addition could be as a result of error in estimating how much Coomassie binds to LysP. Thus it is likely that LysP purifies as a monomer in LMNG. This results is comparable to that of monomeric β1AR6 and β1AR36 (Warne et al. 2009) whose expected sizes from amino acid sequence were 45 and 35.9 kDa respectively but all two proteins migrated close to 146 kDa on BNP: (146/1.8)= 81 kDa, with an extra mass addition of 36 and 45.1 kDa respectively. Also, we noticed that on SDS PAGE LysP runs always as a doublet (**Figure 2.26, A**) which could be the behavior of the protein in the denaturant SDS. This phenomenon was not seen in BNP and also protein elutes a single peak on SEC which is under non denaturing conditions.

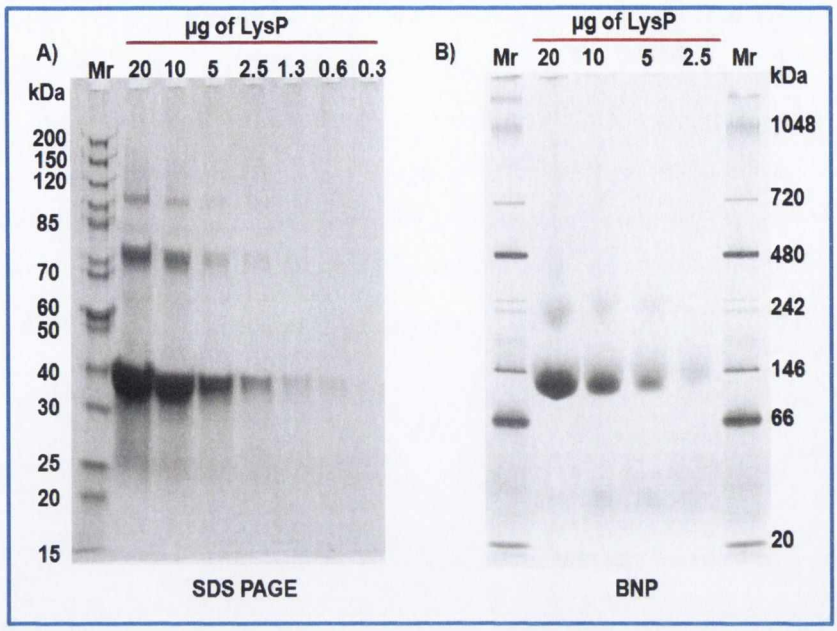


Figure 2.26. Purity evaluation and molecular weight estimation of GFP free LysP. **A)** SDS PAGE of LysP and **B)** BNP of LysP. LysP was purified in MNG-DDM and samples after SEC were analysed on 12 % SDS PAGE gel and 4-16 % Novex Bis-Tris Native PAGE gel.

2.4.4. Biochemical, Biophysical and Functional Characterization of LysP

2.4.4.1. Total Mass Spectrometry of LysP

In order to identify if purified protein was indeed LysP, one of the procedures was to precipitate the purified LysP and then determine the total weight by MALDI-TOF MS. The mass was determined by Dr. James Ault (Faculty of Biological Sciences, University of Leeds) (www.astbury.ac.uk). The weight determined was 54, 044.19 ± 5.65 Da. This is very close to the theoretical molecular weight of the LysP + *Bam*HI (GS) + TEV cleavage site (ENLYF) construct (54037.2 Da) we are working with. Thus, one can conclude that the purified protein is LysP.

2.4.4.2. In-gel Tryptic Digestion of LysP

Purified GFP free LysP in Buffer A from Section 2.3.5 was further identified by in-gel tryptic digestion followed by MALDI-TOF MS. For in-gel tryptic digestion, purified LysP was run on a 12 % SDS PAGE gel, the primary protein bands were cut out from the gel and shipped to Protein Analytics, Germany. **Figure 2.27,** shows the resulting spectrum analyses. From the spectrum analysis report the intensity coverage (which is the intensity signal accounted for by LysP) for this experiment was 53.0 % while the

sequence coverage (percentage of the database protein sequence covered by the matching peptides) was 8.8 %. The low sequence coverage is usually the case for very hydrophobic membrane proteins that makes it difficult for its positively charged residues to be exposed for digestion by trypsin (Wu & Yates 2003). **Figure 2.28, A)** shows the amino acid sequence of LysP and the peptide fragments that were identified and **Figure 2.28, B)** shows the cleavage position of trypsin in the topology map of LysP. Thus we could conclude that the purified protein was indeed LysP from *P. aeruginosa* strain PAO1.

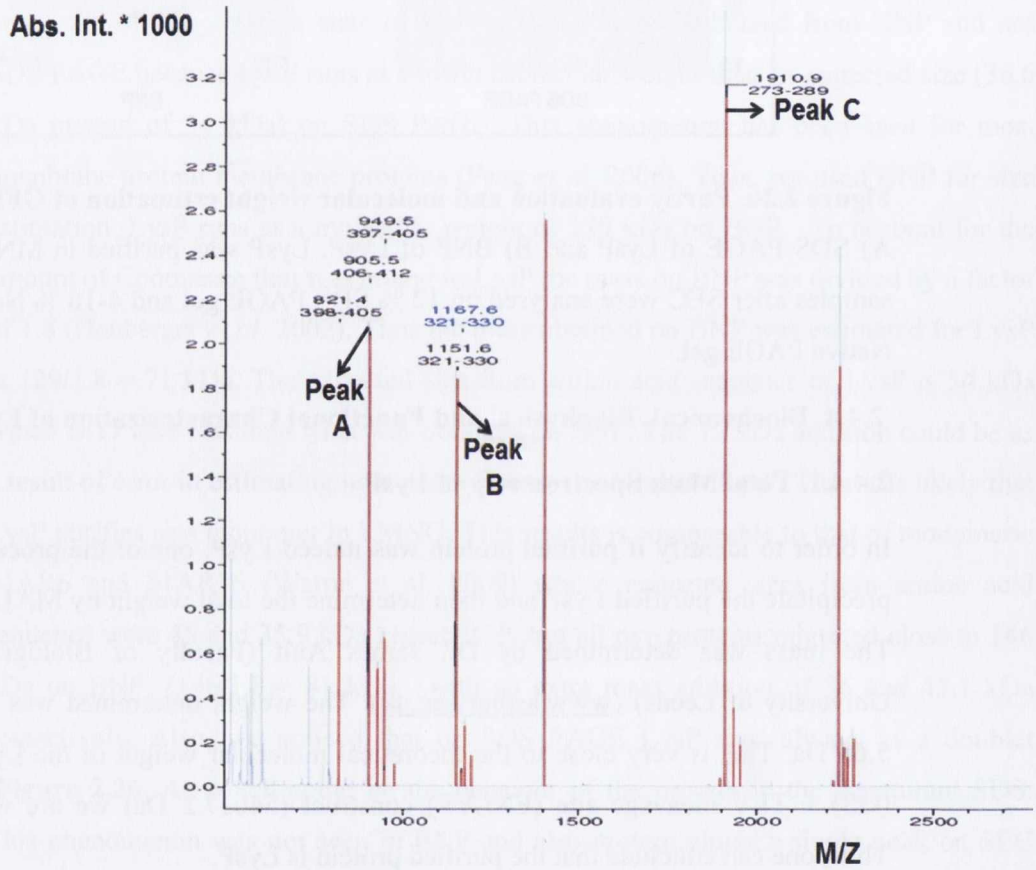


Figure 2.27. Spectrum analyses of GFP free LysP from MALDI-TOF MS after in-gel tryptic digesting.

The red lines correspond to the in-gel tryptic digested fragments of LysP. M/Z on the X-axis is the mass to charge ratio of the ionized peptide fragments. Absorbance intensity on the Y-axis, is the intensity of the UV laser light absorbance by the matrix solution mixed with the protein. Peaks A, B and C are the masses determined by MALDI-TOF MS.

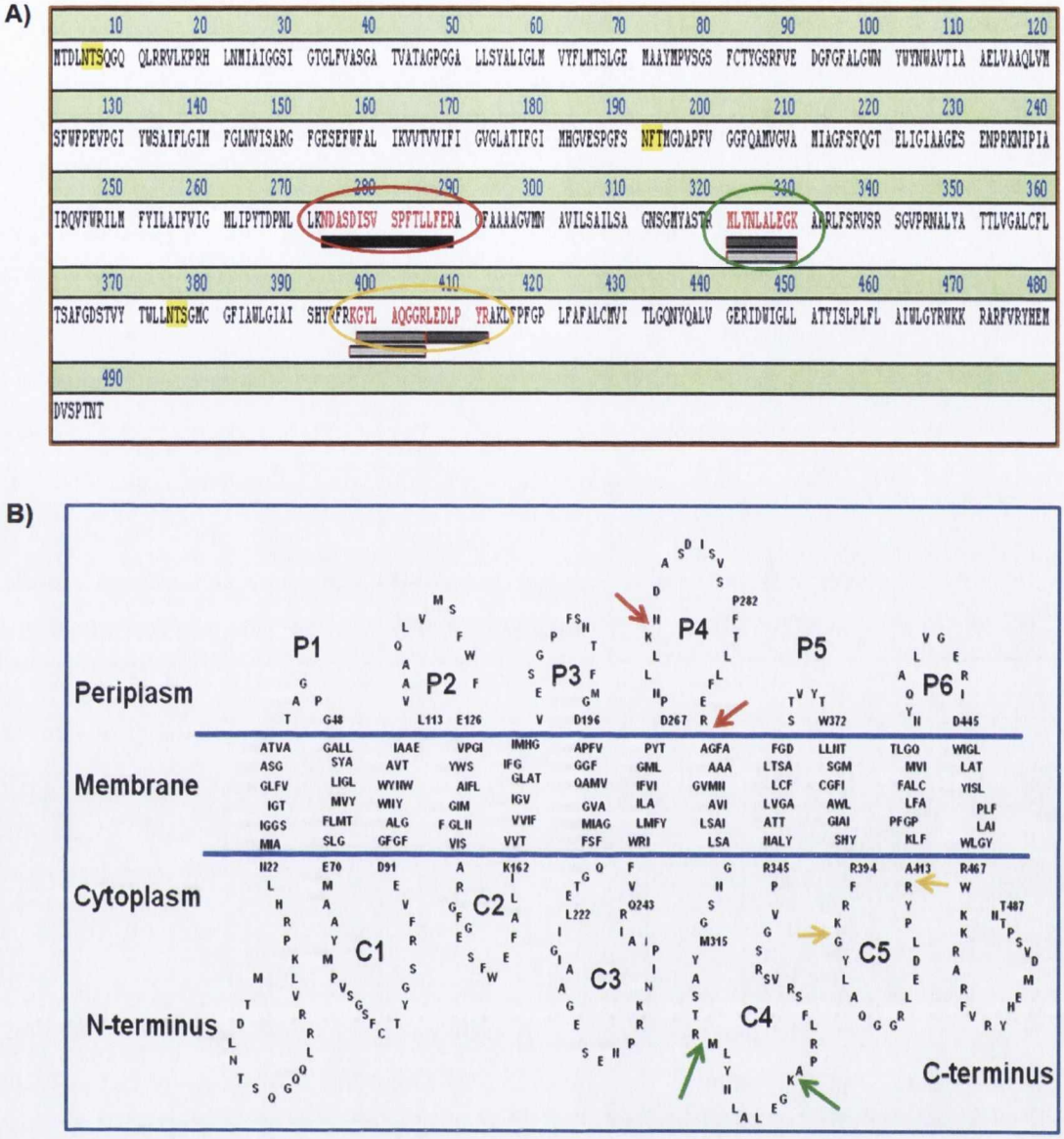


Figure 2.28. In-gel tryptic digestion of LysP followed by MALDI-TOF MS.

A) Amino acid sequence of LysP showing the identified peptide fragments in red from Figure 2.27. B) Topology map of LysP showing regions that were cleaved by trypsin. The sequence that was cut by trypsin in the periplasm (loop P4, Figure 2.28 B, red arrows) corresponds to the red circle in Figure 2.28 A and Peak C in Figure 2.27. The sequence that was cut by trypsin in the cytoplasm (loop C4, Figure 2.28 B, green arrows), correspond to the green circle in Figure 2.28 A and peak B in Figure 2.27 and also, the sequence that cut by trypsin in the cytoplasm (loop C5, Figure 2.28 B, yellow arrows), correspond to the yellow circle in Figure 2.28 A and peak A in Figure 2.27.

2.4.4.3. Glutaraldehyde (GA) Crosslinking to Investigate the Oligomerization States of GFP Free LysP

The oligomerization state of TEV cleaved free LysP was further investigated by GA crosslinking after being shown to be a monomer on BNP. GA crosslinks the primary amines found in lysine from one protein molecule to another. LysP has 9 lysines, 7 are located in the cytoplasm, 1 in the membrane and 1 in the periplasm (Figure 2.8). The cytoplasmic lysines likely are solvent exposed and are potential sites for crosslinking of LysP molecules (if it is a multimer) by GA. What is expected if there is crosslinking was a clear reduction in the primary bands to a higher band and should be dependent on GA concentration. A range of GA concentration up to 200 mM was used that has been successful for other transporters, e.g., the glutamate transporter from *Bacillus caldotenax and Bacillus stearothermophilus* was able to crosslinked at 0.25 mM GA (Yernool & Boudka *et al.* 2003) and also the arginine-agmantine antiporter from *E. coli* was able to crosslinked between 50-100 mM GA (Fang *et at.* 2006). Since there was no crosslinking for LysP up to 200 mM GA (**Figure 2.29, A**), DgkA was used as a positive control, it is a trimer and should crosslink in the presence of 16 mM GA (**Figure 2.29, B**). Since there was crosslinking by DgkA, it means LysP should also crosslink and that the experimental conditions were favourable for any protein that is a multimer to crosslink. Crosslinking in the presence of 0.5 %(w/v) SDS was used as a control to show crosslinking under denaturation condition (there should be no crosslinking under this condition). There was no clear crosslinking seen for LysP but was seen for DgkA. The shift in the presence of GA is probably binding of GA to the primary amines of lysine in LysP monomer. Thus, could conclude that LysP is a monomer and this result is supported by the SEC data (Figure 2.25) and BNP data (Figure 2.26).

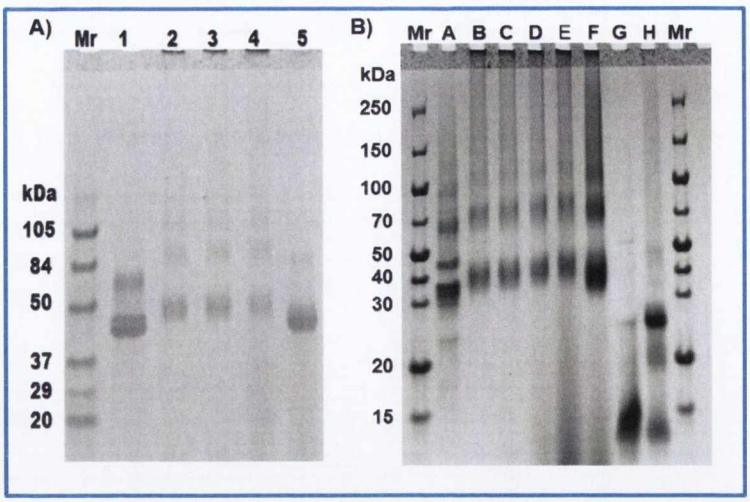


Figure 2.29. Glutaraldehyde crosslinking of TEV cleaved free LysP.

A) Mr is the marker and lane 1 is LysP with no GA added. Final glutaraldehyde (GA) concentrations were 50 mM (lane 2), 100 mM (lane3), 200 mM (lane 4). 0.5 % (w/v) SDS was added to the control sample shown in lane 5 in the presence of 200 mM GA to show crosslinking in under denaturing condition. B) Mr is the marker and lane 1 is LysP with no GA added. Final glutaraldehyde (GA) concentrations were 8 mM (lane B), 16 mM (lane C), 32 mM (lane D), 64 mM (lane E). 0.5 % (w/v) SDS was added to the control sample shown in lane F in the presence of 64 mM GA to show crosslinking under denaturing condition and there was no change (crosslinking). Lane H was DgkA with 16 mM GA and lane G is with no GA.

2.4.4.4. Stability studies of TEV Cleaved Free LysP

After identifying the purified protein after SEC to be LysP, by determination of the its total mass of LysP using ESI MS and by in-gel tryptic digestion followed by MALDI-TOF MS, its stability in the SEC Buffer B (150 mM NaCl, 20 mM Tris HCl pH 7.5, 5 % (v/v) glycerol and 0.003 % LMNG) was studied by concentrating the protein to 10 mg/mL and let it sits on ice for up to 7 days. If the sample is stable after 7 days, it will elute in the included volumn as a Gausian peak but if unstable, it should elute in the void volume (7.5 mL). Samples were analysed by SEC on days 0, 1, 3, 5 and 7. All SEC profiles (**Figure 2.30**) showed that LysP was still monodisperse on day 7 when purified in LMNG because it elutes in the included volume as a Gausian peak. At the same condition LysP precipitates completely when purified with DDM on day 1. This stability studies gives us a clear idea the quality of the protein prior to crystallization trials. Also this info can direct us how long we can keep the protein in case we need to

perform functional assays and also crystallization trials (7 days at 10 mg/mL in Buffer B).

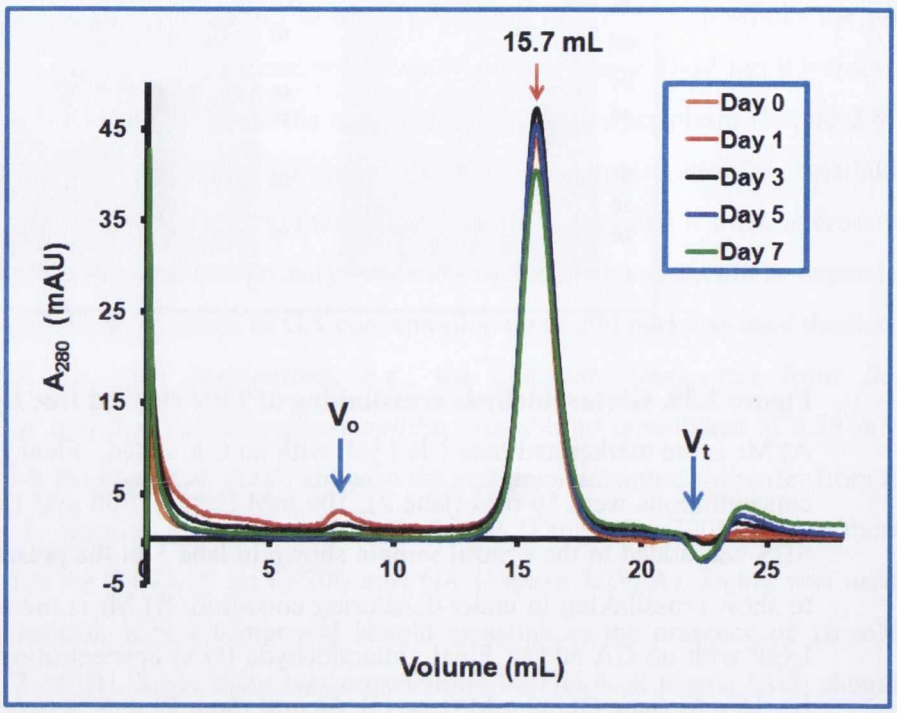


Figure 2.30. Stability studies of TEV cleaved free LysP.

LysP was concentrated to 10 mg/mL in buffer A. Aliquots were analysed on days 0, 1, 3, 5 and 7 using Superose 6, 10/30 column connected to an Äkta Explorer at a flow rate of 0.2 mL/min at 4 °C. Prior to analysis column was pre-equilibrated with SEC Buffer B.

2.4.4.5. CD Spectroscopy for Secondary Structural Determination of GFP free LysP

CD spectrum was used to investigate the secondary structure of GFP free LysP. CD was also used to find out if there is a change in the secondary structure upon addition if L-lysine to GFP free LysP. CD spectrum showed that the secondary structure of GFP free LysP is predominantly alpha helix. CD spectra were recorded using Jasco J-815 CD Spectrophotometer at 20 °C and the raw data were processed using the K2d analysis program in the online CD analysis webpage (DIOCROWEB) (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml). CD spectroscopy revealed 79 % alpha helical content and showed a right-handed coiled or spiral conformation (α-helix) (**Figure 2.31, A red spectrum**) thus one can conclude that LysP is an alpha helical

integral membrane protein. Also, to investigate the effect in the secondary structure of GFP free LysP, GFP free LysP in Buffer B from Section 2.3.5, was incubated with 5 mM L-Lysine (a mole ratio of 1 GFP free LysP and 26 L-lysine) and the spectrum was also recorded an analysed (**Figure 2.31, A blue spectrum**). The spectrum of LysP was subtracted from that of LysP in the presence of 5 mM L-lysine (**Figure 2.31, B**). The addition of 5 mM L-lysine caused a maximum change of 7.5 % at 205 nm (**Figure 2.31, B**), which suggest that the presence of L-lysine has an effect on the secondary structure of GFP free LysP. This conformational change could be exploited to screen different concentrations of L-lysine to determine the K_d of binding between L-lysine and GFP free LysP and also other ligands that binds to LysP could be screened with the goal of using them for co-crystallization trials. However, this was not carried out as MST analyse was used as an alternative.

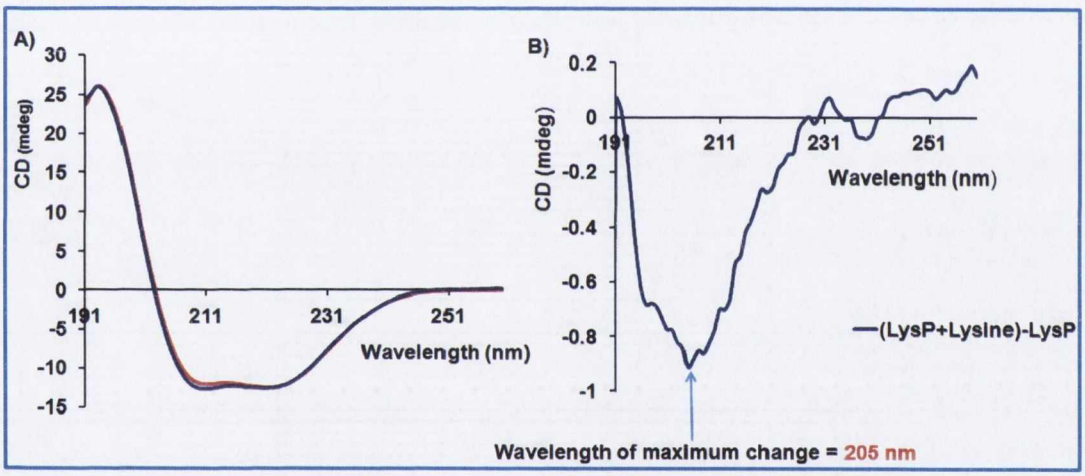


Figure 2.31. CD spectroscopy of GFP free LysP.

A) CD spectrum of LysP (red) and CD spectrum of LysP in the presence of 5 mM L-lysine (green) (a mole ratio of 1 GFP free LysP and 26 L-lysine). B) Subtraction of LysP spectrum from that of LysP plus 5 mM L-lysine to determine the wavelength of maximum change.

2.4.4.6. [^{3}H] L-lysine transport in whole cell

To investigate that the overexpressed LysP-GFP fusion is functionally active and transports [3H] L-lysine, we used radioactivity whole cell transport assay. In this experiment, the amount of [^{3}H] L-lysine transported by the induced cells were

compared with the uninduced cells (**Figure 2.32**). There was an increase in the amount of [^{3}H] L-lysine in the induced compare to the uninduced. However, the amount of [^{3}H] L-lysine being transported by the uninduced cells was significantly high compared to the induced. Uninduced uptake may be due to: 1) leaky expression of LysP-GFP, 2) *E. coli* LysP, since the *E. coli* expression host used for the overexpression of *P. aeruginosa* LysP as LysP-GFP has its own lysine transport system, it could be responsible for the transport in the uninduced cells, 3) promiscuous permeases in *E. coli* that allow hot L-lysine to cross into the cell since there are many proteins in *E. coli* that could be transporting L-lysine, and 4) non-specific binding of L-lysine to negatively charge biomolecules in the cell wall and periplasm of *E. coli.* Thus, we cannot conclude that the overexpressed LysP-GFP transports L-lysine.

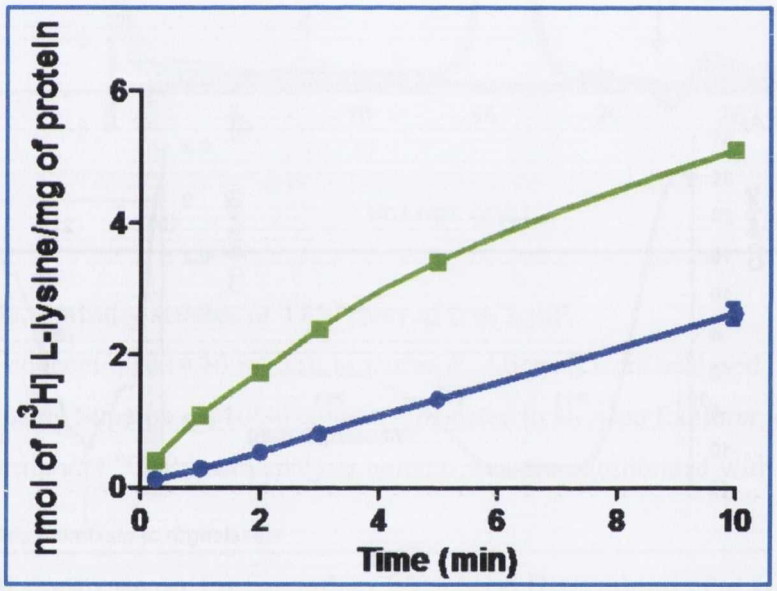


Figure 2.32. Whole cell transport of [^{3}H] L-lysine into *E. coli* C43 (DE3) by LysP-GFP.

Green represents transport in induced cells and blue represents transport in uninduced cells. The final concentration of [^{3}H] L-lysine (mixture of hot and cold L-lysine) was 50 μM. Error bars represent three different replicates.

To investigate the specificity of the transport of [^{3}H] L-lysine, different amino acids were screened to see if they will inhibit the transport of [^{3}H] L-lysine and also, a known inhibitor, L-4-thialysine was tested to show that it inhibits the transport [^{3}H] L-lysine. Among the inhibitors tested, only D-lysine, L-4-thialysine and cold-L-Lysine inhibited the transport of [^{3}H] L-lysine (**Figure 2.33**).

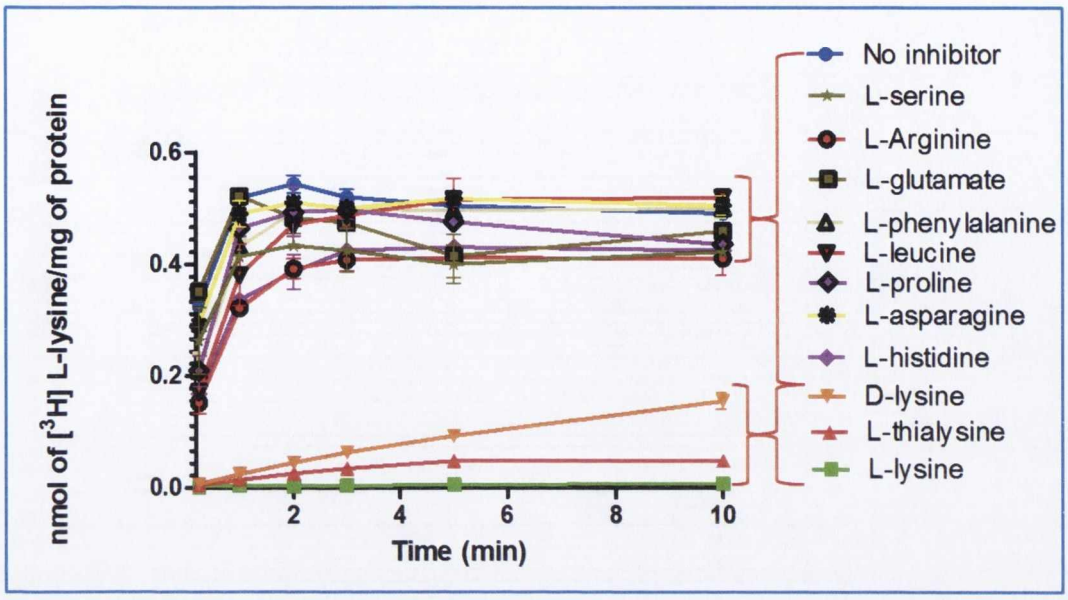


Figure 2.33. Whole cell transport inhibition of [^{3}H] L-lysine into *E. coli* C43 (DE3) by LysP-GFP.

Transport was performed after the addition of 3.3 mM of the different amino acids to the reaction mixture having the whole cell and glycerol (**Section 2.3.6.5**), prior to the addition of ^{3}[H] L-lysine. The final concentration of [^{3}H] L-lysine in this experiment was 1.2 μM, thus the reason for low amount of [^{3}H] L-lysine transported to when 50 μM final concentration of [^{3}H] L-lysine was used (**Figure 2.32**). Thus, **Figure 2.32** and **2.33** cannot be compared. Error bars represent three different replicates.

Since the whole cell inhibition studies (Figure 2.33) was carried at a concentration of hot L-lysine (1.2 μM) different from the transport assay carried out without inhibitors (Figure 2.32) (50 μM hot L-lysine was used), we decided to repeat the transport assay to compare the induced and uninduced at the same concentration that the inhibition studies in Figure 2.34 was done so as to generate a data that could be compared. In this repeat study we also included 3.3 mM each of cold L-lysine, L-4-thialysine and D-lysine as an inhibitor to both the induced and uninduced cells (Figure 2.34).

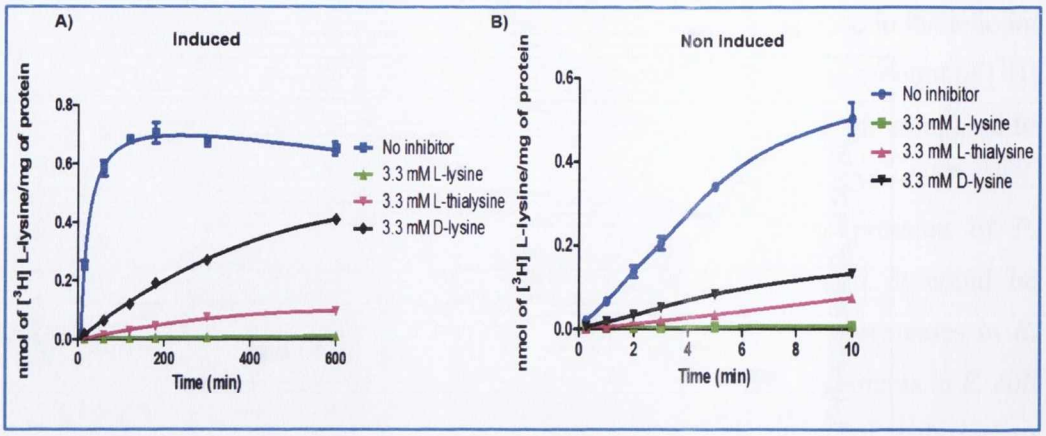


Figure 2.34. Whole cell transport assay of LysP-GFP.

A) Whole cell transport assay of [^{3}H] L-lysine into induced *E. coli* C43 (DE3) by LysP-GFP in the presence of 3.3 mM each of cold L-lysine, L-4-thialysine and D-lysine. There was a control with no inhibitor. **B)** Whole cell transport assay of [^{3}H] L-lysine into non induced *E. coli* C43 (DE3) by LysP-GFP in the presence of 3.3 mM each of cold L-lysine, L-4-thialysine and D-lysine. There was a control with no inhibitor. Error bars represent three different replicates.

Also, pH screening was carried out to establish the best pH range in which LysP-GFP was active (**Figure 2.35**). Best activity was obtained at pH 6.06, 6.6 and 7.17 after 0.25 and 2 min. No activity was obtained at pH 4.01 and pH 9 and 10 after 2 min. Similar study has been carried out in *E. coli* LysP were they screened the transport of [^{3}H] L-lysine pH 5.5, 6.8 and 8.0 (Steffes *et al.* 1992). Best activity was obtained at pH 5.5 followed by 6.8 while the transport of [^{3}H] L-lysine at pH 8.0 was very low.

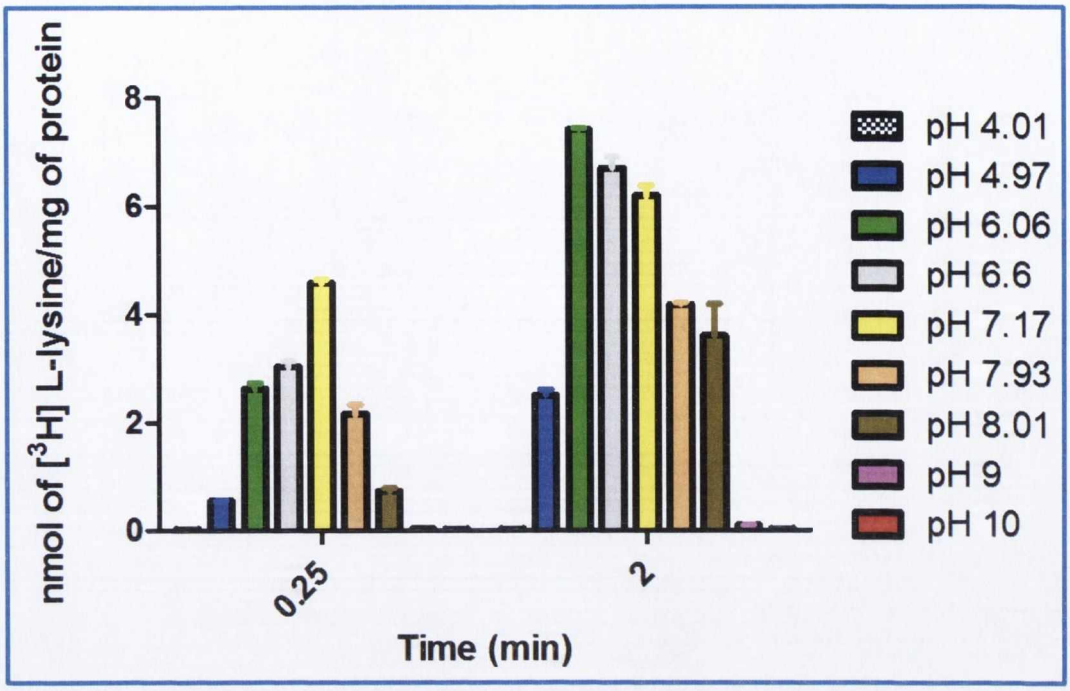


Figure 2.35. Whole cell assay pH screening of [^{3}H] L-lysine transport into induced *E. coli* C43 (DE3) harboring LysP-GFP.

Screening was carried out for the induced cells only. Error bars represents three different replicates.

2.4.4.7. Reconstitution and Counterflow Assays.

To verify that the purified GFP free LysP in Buffer A (Section 2.3.5) transports L-lysine, we carried out counterflow assay according to previously described protocol (Ward *et al.* 2000). LysP was reconstituted into proteoliposomes at a protein to lipid ratio of 1:100 (w/w) respectively. **Figure 2.36** shows a time course of [^{3}H] L-lysine accumulation in proteoliposomes. There was a high background of the liposome with no protein that could have resulted from unspecific binding of the hot L-lysine to the *E. coli* phospholipid used for the reconstitution. This could be because the *E. coli* phospholipid used for preparing the liposomes is composed of predominantly negatively charged phosphatidylethanolamine and phosphatidylglycerol that could be binding to positively charges [^{3}H] L-lysine.

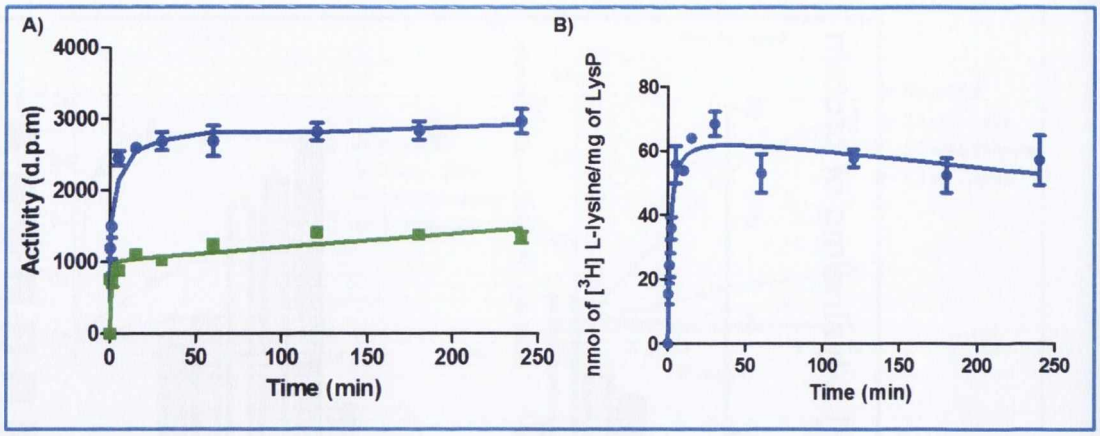


Figure 2.36. [^{3}H] L-lysine counterflow uptake by reconstituted GFP free LysP.

A) Blue shows transport into proteoliposomes containing LysP and green shows transport into liposomes with no LysP (control). **B)** nmol of [^{3}H] L-lysine transported/mg of GFP free LysP. Background transports (from liposomes, Figure 2.36 A, green) was subtracted from transport into proteolipisomes (Figure 2.36, blue).

We showed that the transport of hot L-lysine was highly specific by inhibiting the transport process with 10 mM cold of the 20 standard amino acids, L-4-thialysine and D-lysine (**Figure 2.37**). The inhibition studies were carried out after 3 min and after 20 min. The background activity (liposomes) was subtracted from the proteoliposomes with no inhibitor and also from the proteoliposomes with 10 mM cold inhibitors added. Thus, one can conclude that the transport activity of the proteoliposome was partly due LysP transporting L-lysine because if the transport was highly specific for lysine and was inhibited by the known inhibitor L-4-thialysine.

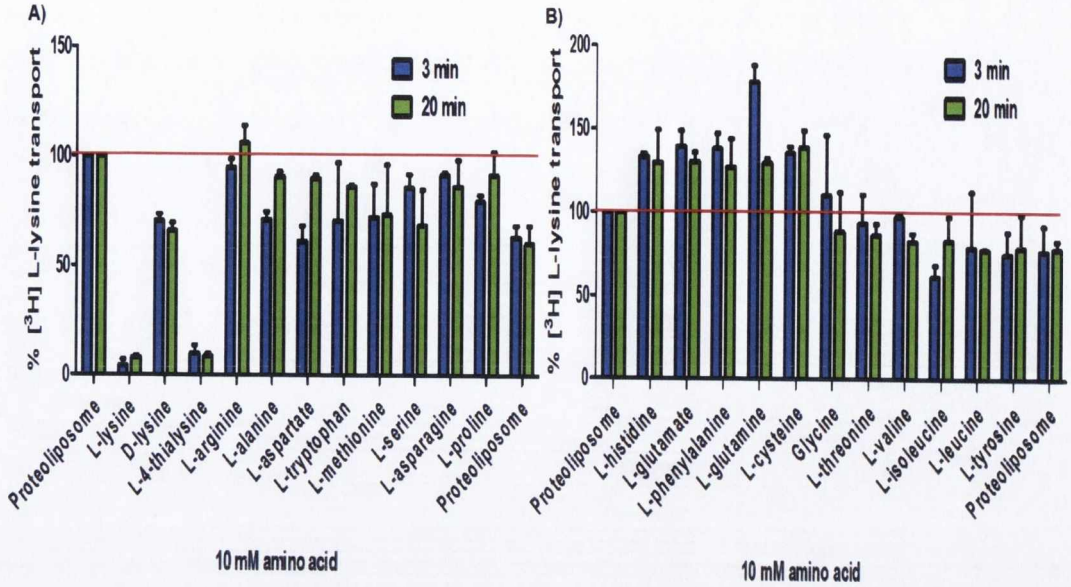


Figure 2.37. Transport inhibition of LysP reconstituted into proteoliposomes. Transport inhibition part 1, the inhibition was carried out in the presence of 10 mM of the different cold inhibitors at internal pH 7.0 and external pH 4.0. B) Transport inhibition part 2, the inhibition was carried out in the presence of 10 mM of the different cold inhibitors at internal pH 7.0 and external pH 4.0. The control, which is uptake without any inhibitor was carried out at the beginning and at the end of each set of experiments to ensure the proteoliposome was still active at the end of the experiment. Samples were collected after 3 min and 20 min. Error bars represents duplicate measurements. This experiment was done by Simon Patching in Prof. Peter Henderson's laboratory, University of Leeds.

Lastly, we showed that the transport of [^{3}H] L-lysine was Na^+ independent by comparing the transport with and without 50 mM NaCl for 30 min and 1 h. This experiment was carried out because ApcT which is a broad range amino acid transporter that belongs to the same APC family of transporters as LysP was shown to be Na^+ independent (Shaffer *et al.* 2009). Thus, we sought to investigate if the transport of L-lysine by LysP was Na^+ independent. There was no difference in the amount of [^{3}H] L-lysine transported after 30 min and 1 h. **Figure 2.38 A,** shows the transport into liposomes and proteoliposomes (with and without NaCl), in **Figure 2.38 B**, the liposome (background) was subtracted from proteoliposome (with and without NaCl) and was converted to nmol of [^{3}H] L-lysine transported/mg of LysP (**Figure 2.38 B**).

Clearly, there was no difference in the amount of [^{3}H] L-lysine transported in the presence of 50 mM NaCl compared to the proteoliposmes with no NaCl. Thus, one can conclude that the transport of [^{3}H] L-lysine by LysP is Na^+ independent because there should have been an increase in the transport activity in the presence of 50 mM NaCl compared to when no NaCl was added. ApcT has been shown to be a proton dependent transporter (Shaffer *et al.* 2009), thus it is likely that LysP will be a proton dependent transporter as well. However, this has to be investigated. It could be done following method of Shaffer *et al.* 2009, by keeping the pH inside proteoliposomes at 7.0 and 4.0 outside and see if there is a robust transport of [^{3}H] L-lysine compared to the case where the pH both in and out of the proteoliposomes in 4.0.

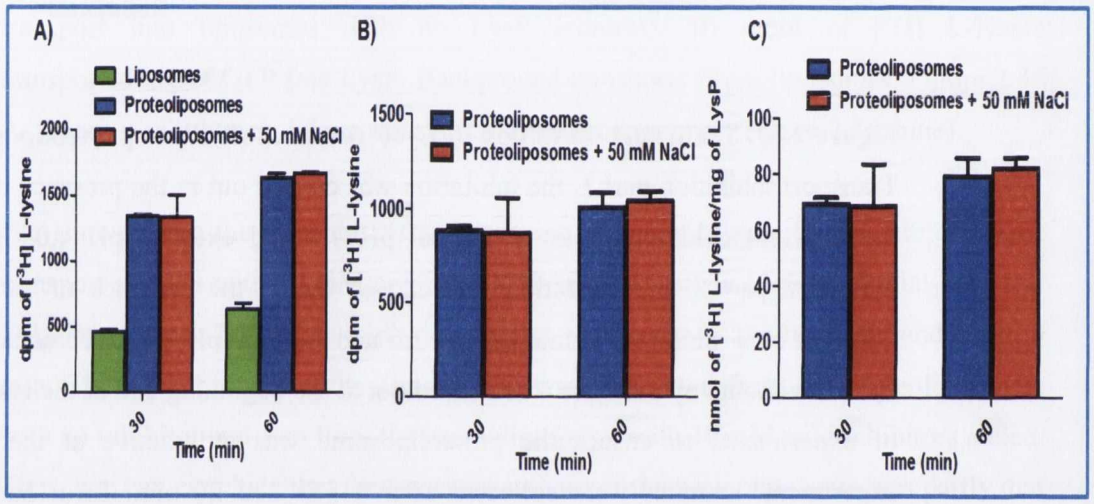


Figure 2.38. Sodium gradient test for GFP free LysP.

A) dpm values of both liposomes and proteoliposomes (with and without 50 mM NaCl) were shown after 30 min and 1 h. B) The dpm values of the liposomes (background) was subtracted from that of the proteoliposomes (with and without 20 mM NaCl) and C) dpm in B was converted to nmol of [^{3}H] L-lysine transporter/mg of LysP.

2.4.5.3. Microscale thermophoresis (MST) analysis of GFP free LysP.

To verify that LysP binds to lysine and the inhibitor L-4-thialysine and to screen for potential binders, MTS analysis was carried out on purified LysP in Buffer B from Section 2.3.5 as described in Section 2.3.6.7. MTS was the preferred choice because it does not require a lot of protein to perform full analyses (3.5 µg of protein only could be used for a set of experiment, i.e., 16 different concentrations of ligands can be

screened). **Figure 2.39** shows the binding profile of LysP to all ligands (L-lysine, L-4-thialysine, D-lysine, L-histidine, L-arginine, L-leucine, L-isoleucine, L-threonine, L-methionine, glycine, L-valine, L-alanine, L-glutamate, L-glutamine and L-proline) tested. The was binding seen for L-lysine, D-lysine, L-4-thialysine, L-histidine and L-arginine with K_d of 15 μM, 124 μM, 31 μM, 4.5 mM and 21 mM respectively (**Table 2.IV**). No binding was seen for the other amino acids tested. The binding for L-lysine was 8 times stronger than that of D-lysine, which suggests that L-lysine is the preferred substrate. However, the binding for L-histidine was 309 times weaker than that of L-lysine while that of L-arginine was 1446 times weaker and since no binding was seen for the other amino acids it suggests that LysP is a lysine specific permease.

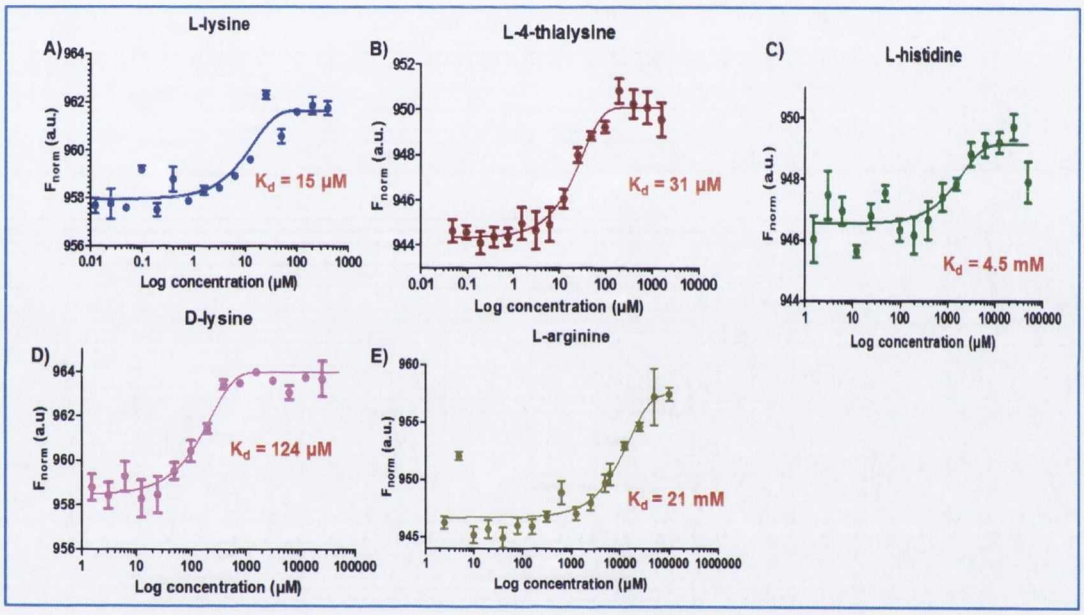


Figure 2.39. MST analysis of free LysP in Buffer B from Section 2.3.5.

A) MST binding profile of GFP free LysP in Buffer B from Section 2.3.5 to L-lysine with a K_d of 15 μM B) MST binding profile of LysP to L-4-thialysine with a K_d of 31 μM C) MST binding profile of LysP to L-histidine with a Kd of 4.5 mM and D) MST binding profile of LysP to D-lysine with a K_d of 124 μM E) MST binding profile of LysP to L-arginine, with a K_d of 21 mM , F) MST binding profile of LysP to L-alanine, no binding was seen, G) MST binding profile of LysP to glycine, no binding was seen, H) MST binding profile of LysP to L-isoleucine, no binding was seen, I) MST binding profile of LysP to L-methionine, no binding was seen, J) MST binding profile of LysP to L-proline, no binding was seen, K) MST binding profile of LysP to L-valine no binding was seen, L) MST binding profile of LysP to L-glutamine, no binding was

seen, M) MST binding profile of LysP to L-glutamate, no binding was seen, N) MST binding profile of LysP to L-threonine, no binding was seen and O) MST binding profile of LysP to L-Leucine, no binding was seen. Error bars represents three different replicates. The equations used to fit data was Y=Bmax*X/(Kd+X) + NS*X + Background. Where Bmax: is the maximum specific binding in the same units as Y. K_d: is the equilibrium binding constant, in the same units as X. It is the ligand concentration needed to achieve a half-maximum binding at equilibrium. NS: is the slope of nonspecific binding in Y units divided by X units. Background: is the amount of nonspecific binding with no added ligand. This represents instruments background.

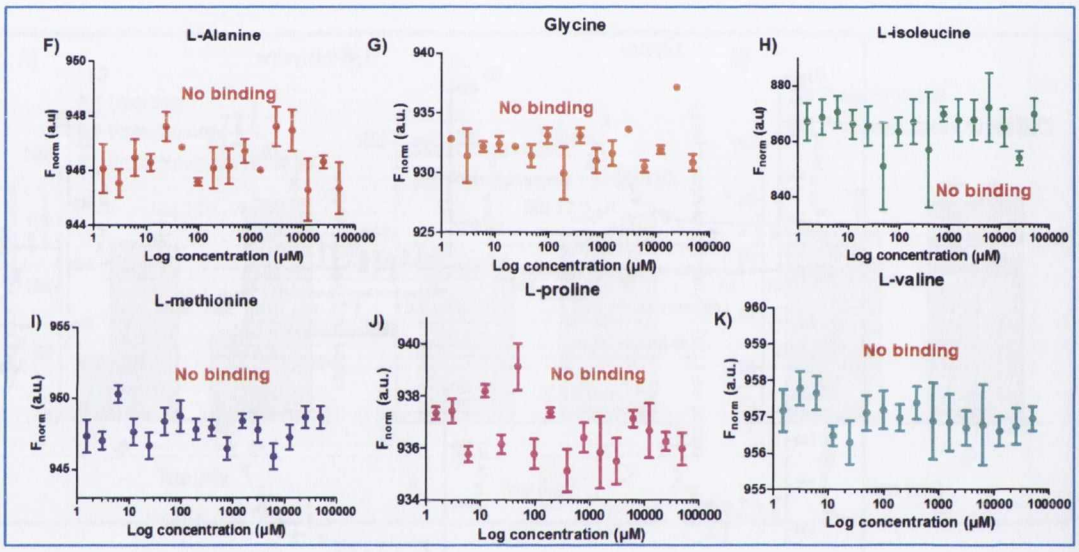


Figure 2.39. continued

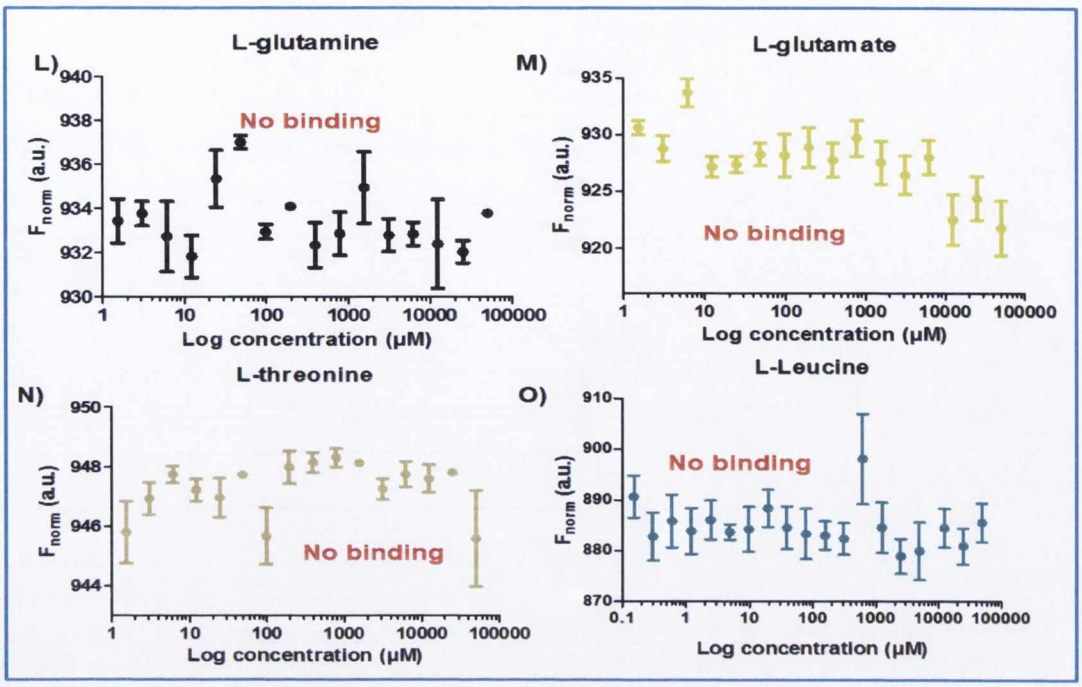


Figure 2.39. continued

Table 2.V. Summary of dissociation constant (K_d) of the different amino acids tested for GFP free LysP by MST analysis.

Amino acid	Dissociation constant (K_d)
L-lysine	15 μM
D-Lysine	124 μM
L-4-thialysine	31 μM
L-histidine	4.5 mM
L-arginine	21 mM
L-leucine	No binding
L-isoleucine	No binding
L-threonine	No binding
L-methionine	No binding
Glycine	No binding
L-valine	No binding
L-alanine	No binding
L-glutamate	No binding
L-glutamine	No binding
L-proline	No binding

2.4.5.4. Tryptophan Fluorescence Change

GFP free LysP in Buffer A from Section 2.3.5 (Buffer A was used throughout the experiment) to a final concentration of 0.26 μM into a 5 mL quartz cuvette and allow to thermally equilibrate for 5 min at 20 °C. The emission spectrum was then taking from 300 nm to 470 nm with excitation wavelength set at 280 nm. The integration time was 0.1 sec and the increment was 1 nm (**Blue spectrum, Figure 2.42**). The same procedure was repeated with the addition of 10 mM L-lysine (Fluka, Lot 1370953) (**Pink spectrum, Figure 2.42**). The mole ratio of LysP to L-lysine was 1:26 respectively. As controls, 10 mM L-lysine was added into Buffer A with no LysP, the spectrum was recorded as described above (**Yellow spectrum, Figure 2.42**) and also the spectrum was recorded for Buffer A alone (**Light green spectrum, Figure 2.42**). There was a fluorescence intensity change of 10 % upon addition of 10 mM L-lysine to LysP in the mole ratio of 1 LysP to 26 L-lysine. There was no significant intensity reading for buffer + 10 mM L-lysine and for the buffer alone. Thus it could be conclude that the addition of L-lysine caused a decrease in tryptophan fluorescence of LysP. Possible explanation could be the presence of tryptophan in or close to the binding pocket of LysP and upon binding of L-lysine to LysP quenches its fluorescence.

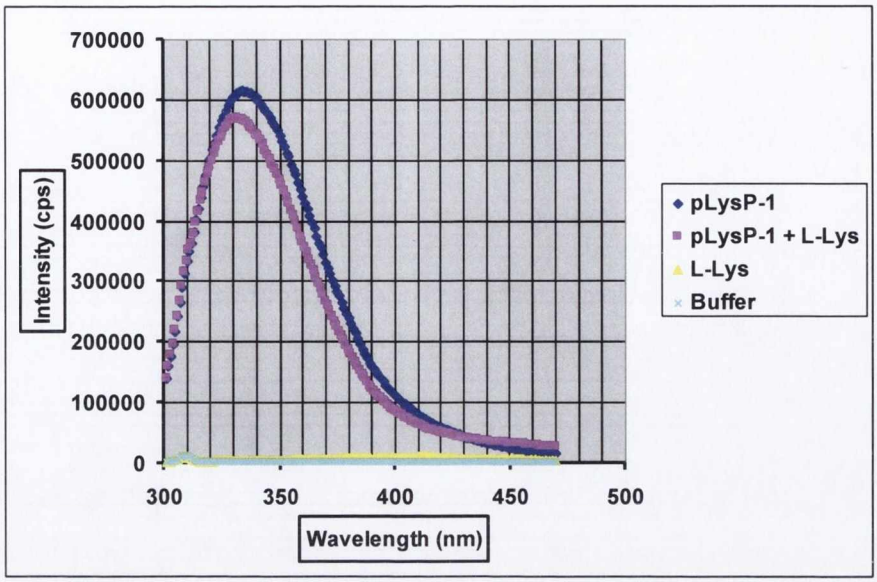


Figure 2.42. Emission spectrum for tryptophan fluorescence change of LysP in detergent solution.

Blue (Buffer A with LysP), Pink (Buffer A with LysP and 10 mM L-lysine) (mole ratio of 1 LysP to 26 L-lysine), Yellow (Buffer A with 10 mM L-lysine) and Light green (Buffer A). The excitation wavelength was set at 280 nm.

The 10 % change was not pursued further for the determination of K_d because the L-lysine used was absorbing light at 280 nm, thus it is likely that the decrease in fluorescence was not due to LysP.

2.5. Concluding Remarks

This chapter describes the cloning, overexpression, detergent selection, purification, biochemically/biophysical and functionally characterisation of LysP from LysP-GFP fusion. The *lysP* coding sequence was cloned and overexpressed in *E. coli* C43 (*DE3*) cells as an octa-histidine-tagged GFP fusion protein. In-gel fluorescence was used to show that full length protein was overexpressed. The fusion protein was shown to transport [^{3}H] L-lysine by a radioactivity transport assay with the transport being inhibited by both D-lysine and L-L-4-thialysine. Fluorescence-detection size exclusion chromatography (FSEC) showed that LysP-GFP is monodispersed in LMNG, DDM, DM, NM and DMNG. LMNG was the best in keeping TEV cleaved free LysP stable on ice for at least 7 days as judged by monodispersity on size exclusion chromatography. The purity of TEV cleaved LysP (0.5 mg/L of cell culture medium final yield) was estimated at greater than 95 % on both blue native PAGE (BNP) and SDS-PAGE. The identity of LysP was determined by mass spectrometry and by in-gel tryptic digestion. BNP suggests that LysP purifies as a monomer in LMNG micelle. CD spectroscopy revealed 79 % alpha helical content, 21 % random coil and no β strand. CD also revealed that LysP is α-helical protein. Microscale thermophoresis analysis (MTS) showed that LysP binds L-lysine, D-lysine, L-4-thialysine, L-histidine and L-arginine with a K_d of 15 μM, 124 μM, 31 μM, 4.5 mM and 21 mM respectively. No binding was seen for L-leucine, L-isoleucine, L-threonine, L-methionine, glycine, L-valine, L-alanine, L-glutamate, L-glutamine and L-proline. The binding for L-lysine was 8 times stronger than that of D-lysine, which suggests that L-lysine is the preferred substrate for the transporter. However, the binding for L-histidine was 309 times weaker than that of L-lysine while that of L-arginine was 1446 times weaker and since no binding was seen for the other amino acids it suggests that LysP is a lysine specific permease. Purified LysP reconstituted into proteoliposomes was shown to transports [^{3}H] L-lysine with the transport shown to be inhibited by L-4-thialysine and D-lysine. The bioinformatical analysis of LysP is also highlighted in this chapter. This study will aid the pursuit of a high resolution crystal structure and a detailed functional analysis of LysP. The fact that LysP was stable in LMNG on ice for at least 7 days coupled with its high expression yield makes it a very attractive target for detailed structure-function investigations.

Chapter 3: Crystallization and preliminary X-ray diffraction data of a lysine specific permease (LysP) from *Pseudomonas aeruginosa* Strain PAO1

3.1. Introduction

This chapter explains how LysP (purified in DDM) was set up into crystallization trials by both *in meso* and *in surfo* methods. Initial hits were obtained on commercial MemGold screen kit by the *in surfo* method that diffracts to 7 Å with reflections to 6 Å after exposure to synchrotron X-ray radiation. It also explained how the crystals were optimized by screening PEG concentration, temperature, precipitant/protein drop size, detergent additives, organic additives and the addition of phospholipids. The optimization unfortunately could only improve the diffraction 7 Å with reflections to 5 Å. This chapter also shows how nanobodies (CA5755 and CA5764) that were raised in a collaborative work were used for co-crystallizing LysP purified in LMNG. LMNG is the best detergent that keeps TEV cleaved GFP free LysP happy on ice for 7 days as judged by monodispersity on SEC. Crystallization trials both *in meso* and *in surfo* of the LysP/nanobody complex in the presence of both L-lysine and L-4-thialysine were carried out. An initial hit was obtained by the *in surfo* method that diffracts to 8 Å after several optimization methods. These methods include screening of PEG concentration, pH, salt concentration, precipitant/protein complex drop size screening, dehydration in PEG 400 and seeding. No crystal was obtained by the *in meso* method in my hands.

3.2. Experimental Procedures

The main buffers used in this section were:

Buffer A: 150 mM NaCl, 20 mM Tris pH 7.5, 0.03 %(w/v) DDM and 5 %(v/v) glycerol

Buffer B: 150 mM NaCl, 20 mM Tris pH 7.5, 0.003 %(w/v) LMNG and 5 %(v/v) glycerol.

The pH 7.5 is the pH is of Tris buffer before the other components were added. The pH of Tris buffer was adjusted to 7.5 using HCl and NaOH.

3.2.1. *In surfo* Crystallization Trials of LysP

Two protein samples were used; one incubated with 5 mM L-lysine (Sigma: Wicklow, Ireland, Lot 1370953) (dissolved in Buffer A) and the other with 5 mM L-4-thialysine (Sigma: Wicklow, Ireland, Lot A2636), also dissolved in buffer A) for 2 h on ice while the other part of the protein was not. The final protein concentration used for the experiment was 10 mg/mL and the mole ratio of LysP to both ligands was 1:26 respectively. Sitting drop crystallization trials were set up using 96-well MRC crystallization plates (Molecular Dimensions) using the TTP LabTech's mosquito® robot (TTP LABTECH) and using the MemGold and MemSys & MemStart screen kits (Molecular Dimensions). Few screens were used because MemGold screen kit already was designed for alpha helical membrane proteins (Newstead *et al.* 2008). The set up was done once in Dr. Amir Khan's Laboratory. 100 nL of both protein and precipitant solutions were dispensed (protein solution was dispensed first). The plates were stored at 18 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.2. *In Surfo* Crystallization Optimization of LysP

3.2.2.1. Precipitant Concentration Screening

The initial hit was obtained from reservoir solution containing 0.1 M NaCl, 0.1 M $MgCl_2$, 0.1 M Tris HCl pH 8.6 and 33 %(v/v) PEG 400. Purified LysP from SEC (0.25 mg/mL in 8 mL of Buffer A) was concentrated at 2, 500 g using rotor C0650 in the Allegra™ X-22R centrifuge (BECKMANN COULTER, Clare, Ireland) for 5 min at 4 °C using Centricon 50 kDa molecular weight cut off (MWCO) concentrator (Millipore: MA, USA, Lot R0KA67379). After the 5 min spin, the sample was resuspended by pipetting up and down. The 5 min spin and resuspension of the sample was carried out 6 times to get a final LysP concentration of 12 mg/mL in Buffer A (160 μL final

volume). The 12 mg/mL in Buffer A LysP clear solution was incubated with either 5 mM L-lysine (Sigma: Wicklow, Ireland, Lot 1370953) or L-4-thialysine (Sigma: Wicklow, Ireland, Lot A2636) and used to set up crystallization optimization trials by varying the precipitant solution concentration from 20-50 %(v/v) PEG 400 (steps 2 %). The other components of the reservoir solution were 0.1 M NaCl, 0.1 M $MgCl_2$ and 0.1 M Tris HCl pH 8.6. They were kept constant. In all the *in surfo* crystallization optimization experiments, the precipitant solution was dispensed first.

3.2.2.2. Precipitant/Protein Drop Size Screening

The variables used were as follows (2 μL + 2 μL, 1.5 μL + 2 μL, 1 μL + 2 μL and 2 μL + 1 μL) precipitant solution (38 %(v/v) PEG 400, 0.1 M NaCl, 0.1 M $MgCl_2$ and 0.1 M Tris pH 8.6) and protein samples (LysP+L-lysine) and (LysP+L-4-thialysine). By 2 μL + 2 μL, it means 2 μL of precipitant solution was mixed with 2 μL of protein solution). Sitting drop plates were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.2.3. Organic and Detergent Additive Screening

Sample for crystallization was prepared as in section 2.1.2.1 and used to set up crystallization optimization trials using organic and detergent additive screen I and II (Hampton Research). Purified protein (10 mg/mL in Buffer A) was incubated with 5 mM L-lysine and L-4-thialysine for 2 hours on ice prior to crystallization screening. The final protein concentration used for the experiment was ~12 mg/mL and the molar ratio of LysP to L-lysine and L-4-thialysine was 1:22. 1 μL of both complexes (LysP + lysine or L-4-thialysine) and precipitant solution (38 %(v/v) PEG 400, 0.1 M NaCl, 0.1 M $MgCl_2$, 0.1 M Tris pH 8.6) containing various organic and detergent additives were dispensed manually for sitting drops. The additives to be used for crystallization were prepared by adding 1 μL of the different additive stocks into 9 μL of the reservoir solution and the samples mixed by pipetting. 1 μL of the mixture was then used for the additive screening. 500 μL of reservoir solution was dispensed in each well. 24 well sitting drop plates (Hampton Research) were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.2.4. Addition of Polar *E. coli* Phospholipids

Polar *E. coli* phospholipid (PL) powder from Avanti consisting of 67 %(w/w) of phosphatidylethanolamine, 23.2 %(w/w) phosphatidylglycerol and 9.8 %(w/w) cardiolipin, was added into 0.5 %(w/v) DDM in 20 mM Tris HCl pH 7.5 to a final

concentration of 18.8 mM. The cloudy suspended PL was votex at 22 °C and sonicated for 30 min at room temperature in a BRANSONIC 2510 tabletop ultrasonic cleaner, aliquot into 1.5 mL Eppendorf tubes, snap frozen in liquid nitrogen and stored at -80 °C. PL was pipetted into LysP solution (10 mg/mL in Buffer A) just before crystallization trials and incubated for 1 hour on ice. The molar ratio of protein to PL used was 1:10 and 1:20, respectively. The precipitant solution used is made up of 0.1 M $MgCl_2$, 0.1 M NaCl, 0.1 M Tris HCl pH 8.6 and 28 -38 %(v/v) PEG 400 (steps 2 %(v/v)). Plates were set up in duplicate and stored at both 20 ^{0}C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions) and 4 ^{0}C in the cold room.

3.2.2.5. Optimization of LysP Using Similar Published Condition for Wild Type LacY.

This was a sitting drop experiment done manually using the standard condition (0.1 M NaCl, 0.1 M $MgCl_2$, 0.1 M Tris HCl pH 8.6 and 33 %(v/v) PEG 400) that has produced crystals of LysP. 1 μL of the LysP solution (30 mg/mL containing 1 mM polar *E. coli* lipid from Avanti with either 5 mM L-lysine or L-4-thialysine) was mixed with 1 μL of a freshly prepared solution consisting of 8 μL of reservoir [0.1 M Tris (pH 8.6)/0.1 M NaCl/0.1 M $MgCl_2$/30–36% PEG 400], 1 μL of 80 mM CHAPS and 1 μL of 30% 1,6-hexanediol. The 2 μL drops were equilibrated against 300 μL of reservoir solution at 20°C. The difference between wild type LysP and LacY is that the reservoir solution for LacY had 0.1 M ammonium sulfate, 100 mM HEPES pH 7.5 and 27-30 %(v/v) PEG 400. The mole ratio of LysP : lipid was 1:10 respectively.

3.2.3. *In Meso* Crystallization Trials of LysP

Protein sample concentrated to 10 mg/mL in Buffer A was used to prepare the cubic phase with 7.7 monoacylglycerol (MAG) (Lot; JL-4-92) and 7.8 MAG (Lot; tasA 21), each in a volume ratio of 1:1 (LysP:Lipid) (this was used because from the phase diagram at 50 %(w/w) composition of water, these lipids form the bicontinuous cubic phase Pn3m) (Caffrey *et al.* 2012) and with 9.9 MAG (Lot; M-239-M27-U) (NU-CHEK PREP INC) in a ratio of 2:3 (w/w) (LysP: 9.9 MAG respectively) was used because at 60 %(w/w) composition of water, 9.9 MAG forms the bicontinuous cubic phase Pn3m (Caffrey *et al.* 2012). 7.7 MAG has a total of 14 carbons with 7 at the head and 7 at the tail region, 7.8 MAG has 15 carbons (7 at the head and 8 at the neck) and 9.9 MAG has 18 carbons (9 at the head and 9 at the tail). The head region is separated from the tail region by the neck region. Prior to preparing the bicontinuous cubic phase (Pn3m) (Caffrey *et al.* 2012), LysP was incubated with 200 mM L-Lysine for 2 hours

on ice. The mole ratio of LysP to L-lysine was 1:880. 50 nL of the inverse bicontinuous cubic phase and 800 nL of the precipitant solution were dispensed using the *in meso* robot Xantus (Sias AG, Switzerland), equipped with a motor driven positive displacement Hamilton syringe (UMP II with a Micro4 controller world precision instrument Sarasota, FL, USA) (Cherezov *et al.* 2004). The final concentration of L-lysine is 5 mM (1:23 ratio of LysP:L-lysine) in the crystallization drop. The following commercial screens were set up: PACT premier (Molecular Dimensions), Memsys & Memstart (Molecular Dimensions), MemGold (Molecular Dimensions), MbClass (Qiagen), MemFac (Hampton Research), SaltRx (Hampton Research) and Cubic Screen (Emerald BioSystems). Plates were stored at 20 °C in the Formulatrix Rock Imager 1500 (RI1500) (Waltham, MA).

3.2.4. Bicelle Crystallization Trials of LysP/CA5755/L-lysine

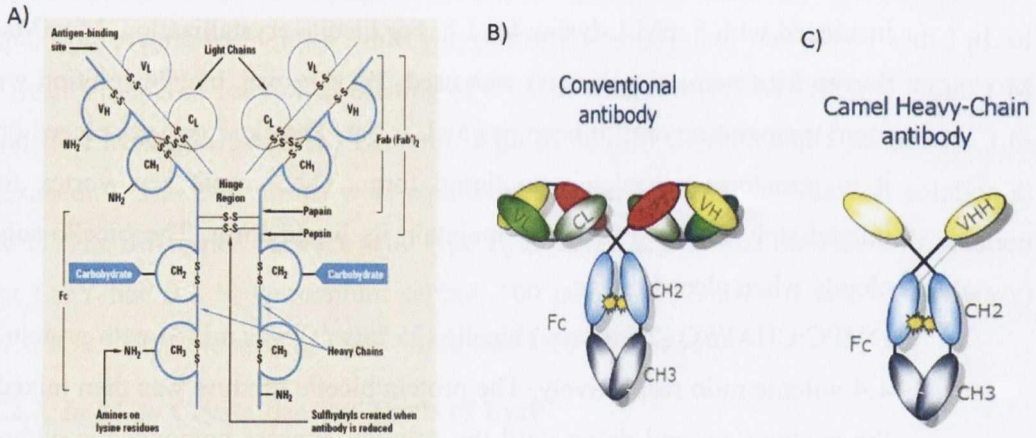
TEV cleaved free LysP from SEC (0.25 mg/mL in 8 mL of Buffer B) was concentrated as in Section 2.1.2.1, to 23 mg/mL in Buffer B (87 µL final volume) was incubated with 5 mM L-lysine for 1 h. For bicelle crystallization, MemMagic™ Bicelle Screen kit (memxbiosciences) was used. Prior to use, bicelle solution was allowed at room temperature until it becomes a clear gel. The clear gel was placed on ice to allow it to transform completely to liquid form. The sample was vortex for 3 sec and immediately returned to ice to maintain its liquid form. The bicelle solution became cloudy when placed on ice.

DMPC:CHAPSO (2.8:1 w/w) bicelle (35 %(w/v) was mixed with protein complex at a 1:4 volume ratio respectively. The protein/bicelle mixture was then mixed by pipetting the contents up and down until the solution appears homogenous without vortexing. The protein/bicelle mixture was then incubated on ice for 1 h before setting up crystallization trials. The final bicelle concentration was 7 %(w/v). The following screens (MemGold, MembFac and MemSys) were set up using the TTP LabTech's mosquito® robot (TTP LABTECH) (200 nL + 200 nL of protein complex/bicelle solution and precipitant respectively, was set up) and plates were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.5. Nanobody Production and Screening for Tight Binders against LysP.

An antibody is a large Y-shaped protein that is being used by the immune system to identify and neutralize foreign substances such as bacteria and viruses (**Figure 3.1 A & B**) (Woof & Burton 2004). Antibodies or immunoglobulin are composed of 2 heavy chain and 2 light chain linked together by disulfide bonds (Woof & Burton 2004 and

De Genst *et al.* 2006). The heavy chain is made up of the Fc fragment and the Fab fragment. The Fc fragment of the heavy chain has the second and third constant domain (CH2 and CH3) while the Fab fragment has the first constant domain (CH1) and variable domain (VH) (Woof & Burton 2004 and De Genst *et al.* 2006). The Fab fragment is separated from the Fc fragment by the hinge region (Woof & Burton 2004). The light is has only the Fab fragment which consists of the variable domain (VL) and the constant domain (CL): from the N-terminus (Woof & Burton 2004). "The antigen-binding sites are located at the tip of the Fab regions and are formed from the variable domains of both the heavy and light chains" (**Figure 3.1 A**) (Woof & Burton 2004). The differences between an antibody and a nanobody is that, a nanobody lacks the light chain and their heavy chain lacks the Cγ1 domain (**Figure 3.1 B & C**) (Muyldermans 2001 and De Genst *et al.* 2006).



Fugure 3.1. Structure of an antibody and a nanobody.

A & B) Conventional antibody has two heavy and two light chains. The heavy chains are made up of the Fc region and the Fab region while the light chains lack the Fc region. The Fab region of the heavy chains is made up of the CH1 domain and the V_H domain while that of the light chains is made up of the C_L and the V_L domains. The Fab and Fc regions of the the heavy chains are linked together by the hinge region. The Fc domain is made up two constant (in immunoglobulin G and A) or three constant domains (in immunoglobulin E) "The antigen-binding sites are located at the tip of the Fab regions and are formed from the variable domains of both the heavy and light chains" C) Structure of a nanobody. It lacks the light chain and their heavy chain lacks the Cγ1 domain (Figure 3.1 A was copied directly from

http://www.piercenet.com/files/1601675_AvBi_HB_INTL.pdf) while Figure 3.1 B & C were copied directlt from http://www.structuralbiology.be/chaperones.

Purified free LysP (obtained from LysP-GFP after cleavage of the GFP with TEV protease) for immunization and biotinylated LysP for screening for binders (1 mg/mL in Buffer B) were express mailed to Jan Steyaert's group at Vrije University Brussels where the immunization and screening for binders were carried out according to already described protocol (Muyldermans 2001). The biotinylation of GFP free LysP was carried out by adding it into 20 fold molar excess of EZ-link Sulfo NHS-LC-LC-Biotin (Thermo Scientific, Rockford, Illinois, USA, catalog number: 21338), a water soluble biotin analogue. EZ-link Sulfo NHS-LC-LC-Biotin (Figure 3.2. B) is an N-hydroxysuccinimide ester (NHS Ester) of biotin (Figure 3.2. C) which reacts with primary amines of LysP and form a stable amide bond. This biotinylation of LysP was done by adding 240 µL of 3 mg/mL LysP in Buffer B (55 µM final LysP concentration) to 26.4 µL of 6.7 mg/mL EZ-link Sulfo NHS-LC-LC-Biotin in Buffer B (1.1 mM final concentration of EZ-link Sulfo NHS-LC-LC-Biotin) and mixed by pipetting and incubated on ice for 2 h. Figure 3.2.A, shows a flow chart of the biotinylation reaction (http://www.piercenet.com/files/1601675_AvBi_HB_INTL.pdf). There are 9 lysine in LysP which are potential sites for biotinlyation, the biotin from the EZ-link Sulfo NHS-LC-LC-Biotin binds to primary amines of the lysines in LysP and EZ-link Sulfo NHS was displaced from EZ-link Sulfo NHS-LC-LC-Biotin (Figure 3.2. A). Excess biotin and the displaced EZ-link Sulfo NHS in the mixture were removed on a Superose 6, 10/30 column connected to an Akta prime FPLC at a flow rate of 0.2 mL/min using Buffer B at 4 ^{o}C. The column was pre-equilibrated with Buffer B prior to injection. The purpose of biotinylation of LysP was to use it for screening a biotin binds to streptavidin at a very high affinity and the streptavidin/biotin/LysP complex can then be immobilized on an ELISA plate for screening of nanobodies that will bind to LysP.

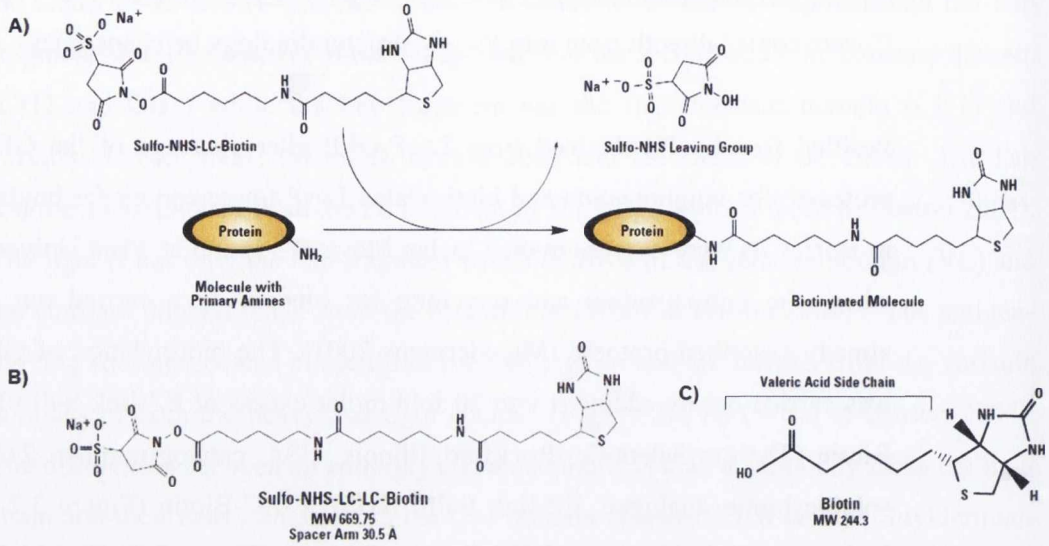


Figure 3.2. Mechanism of biotinylation of GFP free LysP with EZ-link Sulfo NHS-LC-LC-Biotin.

A) Reaction mechanism for the biotinylation of LysP. The biotin from the EZ-link Sulfo NHS-LC-LC-Biotin binds to primary amines of the lysines in LysP and EZ-link Sulfo NHS was displaced from EZ-link Sulfo NHS-LC-LC-Biotin. Excess biotin and the displaced EZ-link Sulfo NHS in the mixture were removed by SEC. **B)** Structure of Biotin. Biotin has a molecular weight of 244.3 Da **C)** EZ-link Sulfo NHS-LC-LC-Biotin (a soluble biotin analogue) is an N-hydroxysuccinimide ester (NHS Ester) of biotin which reacts with primary amines of LysP and form a stable amide bond. EZ-link Sulfo NHS-LC-LC-Biotin has a molecular weight of 669.75 Da and a spacer arm of 30.5 Å long (Figures copied directly from http://www.piercenet.com/files/1601675_AvBi_HB_INTL.pdf).

Figure 3.3 shows a schematic overview of strategies to clone and select the variable domain of heavy chain of heavy-chain antibody (VHH) genes from an immunised dromedary or llama (Muyldermans 2001). This was carried out by the group of Jan Steyeart at Vrije Universitat, Brussels. Briefly, the nanobody screening starts from peripheral blood lymphocyte isolated from an immunized llama (50 µL or 50 µg of LysP from 1 mg/mL LysP stock in Buffer B was loaded every week for 5 to 6 weeks) to prepare cDNA (Muyldermans 2001). All VHH are encoded by a single exons and share homologous borders of sequences because they belong to one single gene family (Muyldermans 2001). Therefore the mature VHH repertoire can be amplified by a single set of primers (Muyldermans 2001). ''The heavy chain of the llama or

dromedary heavy chain antibody is composed of the variable domain (VHH) followed immediately by the hinge region, CH1 and the second constant domain (CH2) (**Figure 3.3**)'' (Muyldermans 2001). The sequence of the hinge region is very different from the other isotypes therefore aiding cloning of the VHH region (Muyldermans 2001). The VHH region are cloned at the 3' end of a minor coat of gene 3 protein of filamentous phage, phage particles can then be produced that contain VHH-g3p fusion at their tips and the corresponding VHH gene encoded in the phagemid genome (Muyldermans 2001). VHH can therefore be selected for several sequential rounds if needed through panning of phage particles on immobilized immunogen (Muyldermans 2001).

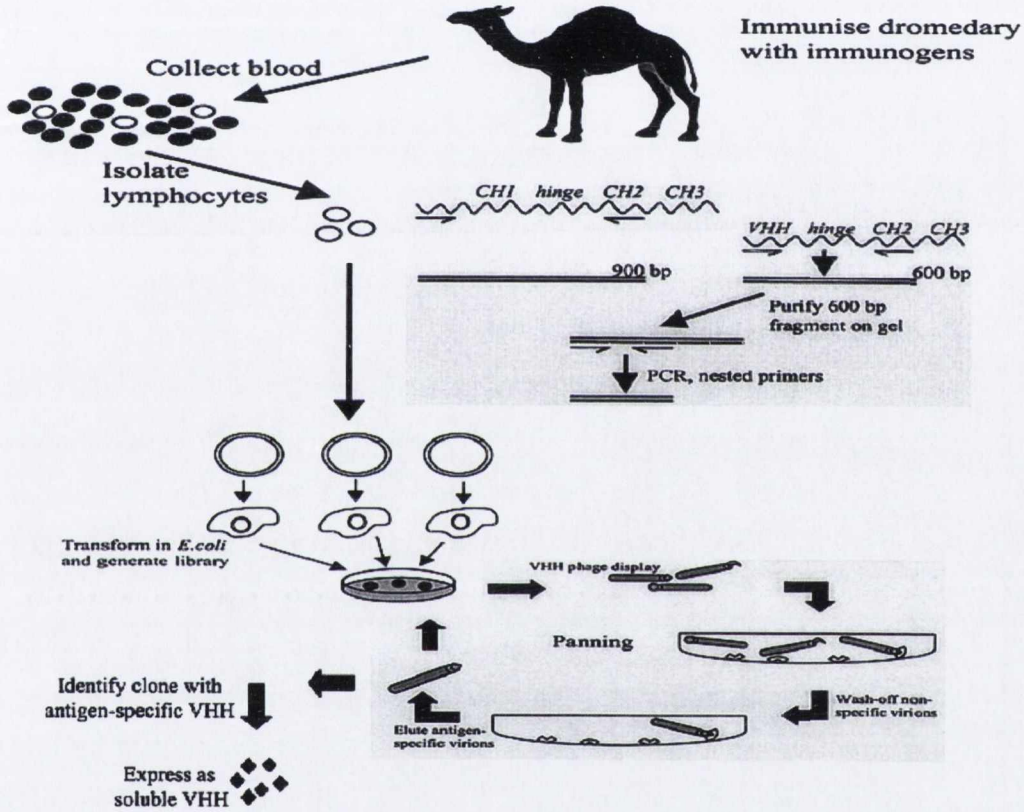


Figure 3.3. A Schematic overview of strategies to clone and select the VHH genes against LysP from an immunised dromedary or llama.

This was done by the group of Jan Steyeart at Vrije Universitat, Brussels. 1 mL of TEV cleaved GFP free LysP at 1 mg/mL in Buffer B was expressed mailed to them. 50 µL or 50 µg of LysP from 1 mg/mL LysP stock in Buffer B was loaded every week for 5 to 6 weeks) to prepare cDNA (figure was adopted form Muyldermans 2001).

A totally of 7 clones were identified and express mailed to us. They were CA5753, CA5755, CA5758, CA5760, CA5761, CA5764 and CA5765 (**Figure 3.4**).

```
>CA5764
QVQLQESGGGLVQAGGSLRLSCAASGRTFSTLAMGWFRQAPGKDRDFVARITTSGTNTDYAD
SVKGRFTISRDNAKNTLYLQMNSLKPEDTAVYYCAADRAAWVIKATIPSEYEYWGQGT
QVTVSSHHHHHHEPEA

>CA5755
QVQLQESGGGLVQTGGSLRLSCAASGRTFSIYGMAWVRQASGKEREFVAAMPRGGGTTNYA
DSVKGRFSISRDNAKNTVDLQMNSLKPEDTAVYYCVADRGFTLRLGIQHDYWGQGTQVT
VSSHHHHHHEPEA

>CA5765
QVQLQESGGGLVQVGGSLRLSCAASGFAFNDYAIGWFRQAPGKEREGISVISINDGSTWYADS
VKGRFTISSDNAKNTVYLQMNDLTPEDTAAYYCAADPTLRSWRIAANGVLSYEYDYS
GQGTQVTVSSHHHHHHEPEA

>CA5758
QVQLQESGGGLVQTGDSLRLSCAASGRTLSISAMGWFRQVPGKEREFVAGISRTVGRPYYVD
SVKGRFTISRDNAKNTVTLQMDSLEAEDTAAYYCAARVRGFVNFPSDRNEYEFWGQGT
QVTVSSHHHHHHEPEA

>CA5761
QVQLQESGGGLVQTGGSLRLSCAASGRTLSISAMGWFRQVPGKEREFVAGISRSVGRPNYVD
SVKGRFTISRDNAKNTMTLQMNSLEAEDTAAYYCAARVRGFVNFPSDRNEYDFWGQGT
QVTVSSHHHHHHEPEA

>CA5753
QVQLQESGGGLVQAGESLRLSCAASGKIFANNVMAWYRQAPGKQRELVASVDGGGRTNYAD
SVKGRFIISRDTKNMMYLQMHSLKPEDTGVYFCTVPAYDAWGQGTQVTVSSHHHHHHEPEA

>CA5760
QVQLQMNSLKPEDTAIYYCAANNGPYGVESHEYHYWGQGTQVTVSSHHHHHHEPEA
```

Figure 3.4. Amino acid sequence of the different nanobodies raised against TEV cleaved GFP free LysP.

The sequences in red are the CDR1, green are the CDR2, blue are the CDR3 and purple are the affinity tags (EPEA at the end of the hexahistidine tag is the stop codon). CDR is the antigen binding fragments in the VHH gene. CDR3 has longer amino acid sequence than CDR1 and CDR2 and it is the antigen binding fragment that gives the specificity of the VHH gene. The CDR regions were estimated by aligning the nanobody sequences with a nanobody whose structure is already solved (Kabula *et al.* 2010).

The clones were overexpressed in WK6 SU$^-$ cells and purified by IMAC and SEC. Single colony of the transformed WK6 SU$^-$ cells harboring the gene of the respective nanobody clone was inoculated into 20 mL of Luria Bertani (LB) media supplemented with 100 µg/mL ampicillin, 2 %(w/v) glucose and 1 mM $MgCl_2$. The inoculated media was grown for 16 h at 37 °C with shaken using the Sanyo Orbital incubator at 190 r.p.m. The starter culture was diluted 1 in 100 into 330 mL of terrific broth (TB) media in a 1 L baffled flask supplemented with 100 µg/mL ampicillin, 0.1 %(w/v) glucose and 1 mM $MgCl_2$. The cells were then grown for 3 h at 37 °C with shaking using the INFORS HT multitron standard shaker at 190 r.p.m to an OD_{600} of 0.7. The cells were then induced with 1 mM IPTG and grown further for 16 h with the temperature dropped to 28 °C. The cells were harvested by centrifugation using Beckman Coulter

Avanti J-20 XPI centrifuge with Beckman JLA-9.1000 rotor at 5,000 x *g* for 10 minutes at 4 °C the supernatants were discarded. The cells were then resuspended using a spatula in 15 mL of TES buffer (0.2 M Tris pH 8.0, 0.5 mM EDTA and 0.5 M Sucrose) and shaken for 1 h in the cold using 30 mL of TES buffer (0.05 M Tris pH 8, 0.125 mM EDTA and 0.125 M Sucrose) was then added and shook further for 45 min in the cold. The resuspended cells were then centrifuge at 7,500 x *g* for 10 min at 4°C suing rotor JLA-16,250 in the Beckman Coulter Avanti J-20 XPI centrifuge. The supernatant containing the overexpressed nanobodies in the perisplasmic fraction was kept and used directly for IMAC purification. 500 uL of Ni-NTA beads was added to the supernatant and 10 mM imidazole pH 7.5 was added to prevent unspecific binding of contaminating protein. The beads were packed into a 10 mL glass Econo column (BioRad) and washed with 10 column volumes of buffer containing 150 mM NaCl, 20 mM Tris pH 7.2 and 20 mM imidazole pH 7.5 at 22 °C. Bound protein was eluted with 2 mL of buffer containing 150 mM NaCl, 20 mM Tris pH 7.2 and 500 mM imidazole pH 7.5. Excess imidazole was then removed by SEC on a hiLoad 16/60 Superdex 200 column connected to an Äkta prime FPLC (GE Healthcare), at a flow rate of 1 mL/min using Buffer B at 4 °C. Prior to injecting the samples, the column was pre-equilibrated with buffer containing 150 mM NaCl and 20 mM Tris pH 7.2. The purified nanobodies were analyzed by SDS PAGE.

Purified nanobodies were mixed with purified LysP by pipetting up and down inside a 1.5 mL Eppendorf tube, at a molar ratio of 1 protein : 3 nanobody and incubated on ice for 1 h. Samples were then subjected to SEC to remove free nanobody and to detect co-elution of the complex using a Superose 6, 10/300 column connected to an Äkta prime FPLC (GE Healthcare), at a flow rate of 0.2 mL/min at 4 °C. The column was pre-equilibrated in Buffer B before injecting the samples. As a control LysP without any nanobody add to it was also subjected to SEC. All samples were run in Buffer B. The LysP peak fraction as well as the free nanobody fractions was analyzed on SDS PAGE.

3.2.6. *In Surfo* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764.

Concentrated free LysP (10 mg/mL in Buffer B was incubated with either nanobody CA5755 or CA5764 at a molar ratio of 1 protein:1.5 nanobodies for 1 h on ice in the presence of 5 mM L-lysine. Sitting drop crystallization trials were set up using 96-well MRC crystallization plates (Molecular Dimensions) using the TTP LabTech's mosquito® robot (TTP LABTECH) and using the MemGold and MemSys & MemStart screen kits (Molecular Dimensions). 100 nL of both protein/nanobody/L-lysine complex and precipitant solutions were dispensed. The plates were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.7. *In Surfo* Crystallization Optimization of LysP with Nanodody Clone CA5755

In the initial crystallization trials in the presence of nanobody clones CA5755, an initial hit (**Figure 3.13**) appeared after 3 weeks on the MemGold screen kit (0.1 M $CaCl_2$ + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME). Crystals were tiny rods with the longest of about 20 µm. The goal of this optimization was to grow bigger crystals using this initial hit.

3.2.7.1. PEG 2000 MME Screening.

The PEG 2000 MME concentration was varied from 9 to 19 % (w/v) with a step of 2 %(w/v). The $CaCl_2$ and Tris pH 6.5 concentrations were kept at 0.1 M. 1 µL of both complexes (LysP + CA5755 + L-lysine) and precipitant solution were dispensed manually for sitting drops. 500 µL of reservoir solution was dispensed in each well. 24 well sitting drop plates (Hampton Research) were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.7.2. PEG 2000 MME and pH Screening.

The PEG 2000 MME concentration was varied as in panel 2.1.7.1 but four different pH were used (0.1 M Tris pH 6.5, 7.5 and 8.6 and also 0.1 M sodium citrate pH 5.5 was used). The CaCl2 concentration maintained at 0.1 M. 1 µL of both complexes (LysP + CA5755 + L-lysine) and precipitant solution were dispensed manually for sitting drops. The was precipitation in the drops having 0.1 M sodium citrate when CaCl2 was added as result of the formation of calcium citrate complex formation. This condition was not used for further optimization. 500 µL of reservoir solution was dispensed in each well.

24 well sitting drop plates (Hampton Research) were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.7.3. $CaCl_2$ Concentration Screening.

Three different concentrations of $CaCl_2$ were used (0.05 M, 0.1 M and 0.2 M). The PEG 2000 MME concentration was 13 %(w/v) and the buffer used was 0.1 M Tris pH 6.5 as in the initial hit. 1 µL of both complexes (LysP + CA5755 + L-lysine) and precipitant solution were dispensed manually for sitting drops. 500 µL of reservoir solution was dispensed in each well. 24 well sitting drop plates (Hampton Research) were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions). Crystals were obtained only at 0.1 M $CaCl_2$, thus we further rescreening from 0.08 to 0.13 M $CaCl_2$ (steps 0.01 M).

3.2.7.4. Precipitant/Protein Drop Size Screening.

Concentrated free LysP (10 mg/mL in buffer containing 150 mM NaCl, 20 mM Tris pH 7.5, 0.003 %(w/v) LMNG and 5 %(v/v) glycerol) was incubated with nanobody CA5755 at a molar ratio of 1:1.5 respectively for 1 h on ice in the presence of either 5 mM L-lysine or 5 mM L-L-4-thialysine. The volume ratio of precipitant:protein complexes used were, 1:1, 2:2, 2:1, 1:2 and 1.5:1 respectively. 500 µL of reservoir solution was dispensed in each well in a 24 well sitting drop plates (Hampton Research) and plates were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.7.5. Dehydration of LysP/CA5755/L-lysine Crystals in PEG 400.

The goal was to dehydrate the crystals at high concentration of PEG 400 to improve the diffraction quality (Kuo *et al.* 2003) and also, PEG 400 at high concentration will serve as cryo protectant for the crystals since the initial hit had 13 %(w/v) PEG 2000 MME which cannot cryo protect the crystals during freezing in liquid nitrogen. In a sequential manner, crystals from the crystallization drop (**Figure 3.17**) were harvested using MiteGen microloops of size 20 µm and transferred to 2 µL sitting drop of reservoir solution + 5 mM L-lysine + 34-52 %(v/v) PEG 400 (2 % steps). Plates were sealed and stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions). Also, 2 µL of 100 %(v/v) PEG 400 was added to drops having crystals (**Figure 3.16**). Crystals were incubated in the PEG 400 overnight prior to harvesting.

3.2.7.6. Seeding of LysP/CA5755/L-lysine Crystals

Seeding makes use of the phase diagram in which the protein and precipitant concentrations are related by; zone of undersaturation, metastable zone, labile or nucleation zone and the precipitation zone. In the undersaturation zone the precipitant concentration and protein concentration are very low such that nucleation cannot occur for crystal growth, in the metastable zone (clear zone) there is no nucleation unlike the labile zone where there is nucleation and crystal growth and in the precipitation zone the protein and/precipitant concentration is very high such that the protein comes out of solution. In seeding seed from the labile zone are transferred to the metastable zone to trigger the protein in it to crystallize (Friedmann *et al.* 2011). The labile zones (Figure 1.14) were assumed to be those conditions that gave crystals while the metastable zones were assumed to be those conditions that did not give crystals.

Initial crystals from (0.1 M Tris pH 6.5, 0.1 M $CaCl_2$ and 13 %(w/v) PEG 2000 MME) were optimized by setting up a fresh trial with the addition of PEG 400 (28-54 %(v/v) steps 2 %. After 3 weeks crystals were not obtained unlike the initial condition that gave crystals after 3 weeks. Seeds from the condition shown in **Figure 3.16** were transfer to the conditions that had PEG 400 added. Horse hair was used to transfer the seeds by streaking into the drops that had crystals and then to the drops that did not give crystals (Friedmann *et al.* 2011) and plates were stored at 20 °C in the RUMED® Rubarth Apparate GmbH incubator (Molecular Dimentions).

3.2.8. *In Meso* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764.

LysP concentrated to 10 mg/mL in buffer B was used to prepare the cubic phase with 7.7 MAG (Lot; JL-4-92) and 7.8 MAG (Lot; tasA 21), each in a volume ratio of 1:1 (LysP:Lipid) and with 9.9 MAG (Lot; M-239-M27-U) (NU-CHEK PREP INC) in a ratio of 2:3 (w/w) (LysP: 9.9 MAG). These ratios were used because they have been shown to form the bicontinuous cubic phase (Pn3m) (Caffrey *et al.* 2012). Prior to preparing the bicontinuous cubic phase (Pn3m), LysP was incubated with either nanobody clone CA5755 or CA5764 at a molar ratio of 1:1.5 respectively on ice for 1 h in the presence of 200 mM L-Lysine. 50 nL of cubic phase and 800 nL of the precipitant solution were dispensed using the *in meso* robot Xantus (Sias AG, Switzerland), equipped with a motor driven positive displacement Hamilton syringe (UMP II with a Micro4 controller world precision instrument Sarasota, FL, USA) (Cherezov *et al.* 2004). The final concentration of L-lysine was 5 mM (200 mM added

to the protein stock prior to set up and was diluted to 5 mM upon addition of 800 nL of precipitant solution) (1:23 ratio of LysP:L-lysine) in the crystallization drop. The following commercial screens were set up: PACT premier (Molecular Dimensions), Memsys & Memstart (Molecular Dimensions), MemGold (Molecular Dimensions), MbClass (Qiagen), MemFac (Hampton Research), SaltRx (Hampton Research) and Cubic Screen (Emerald BioSystems). Plates were stored at 20 °C in the Formulatrix Rock Imager 1500 (RI1500) (Waltham, MA).

3.2.9. X-ray Diffraction Studies

Crystals of LysP purified in DDM were harvested by myself using microloops (MiTiGen), cryo-cooled in liquid nitrogen without the addition of any cryo-protectant, because the crystallization condition had 38 %(v/v) PEG 400 which is itself a cryo-protectant at that concentration. Crystals were stored and transported in liquid nitrogen Dewar to the Diamond Light Source, UK (DLS) and the Advanced Photon Source, USA (APS). Crystals were exposed to synchrotron X-radiation and 180° data were collected at both APS and DLS on beamlines 24-ID-E (40 % transmission) and I24 (100 % transmission) respectively. The APS data for LysP was collected by Dr. Amir Khan and Humberto Fernandes.

For LysP/CA5755/L-lysine or L-4-thialysine crystals, they were harvested using 20, 30 and 50 µm microloops (MiTiGen), cryo-cooled in solution containing the respective reservoir solution + 30-50 %(v/v) PEG 400 [steps 2 % (v/v)] or 6-24 %(v/v) glycerol [steps 2 %(v/v)]. Crystals were stored and transported in liquid nitrogen dewar to the Diamond Light Source, UK (DLS) Crystals were shot and data were collected at the DLS on beam line I24 (100 % transmission, detector distance = 600 mm, exposure time = 0.2 sec, oscillation = 0.2 deg, energy = 12.8 keV, 0 and 90 degree images were collected by myself with the guidance of Dr Declan Doyle and Dr. Valerie Pye).

3.3. Results

A flow chart summarizing the progress through the crystallogenesis process is shown in **Figure 3.5.**

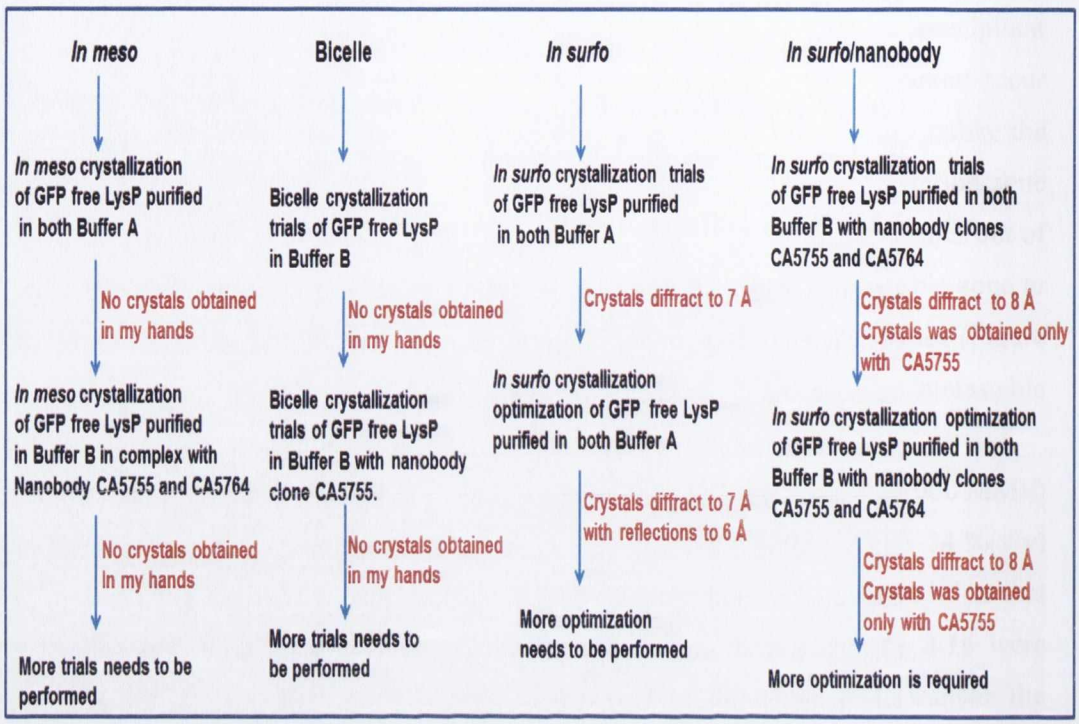


Figure 3.5. A flow chart summarizing the progress through the crystallogenesis process of GFP free LysP.

3.3.1. *In surfo* Crystallization Trials of LysP

Concentrated TEV-cleaved LysP (12 mg/mL) was set up in sitting drop vapour diffusion crystallisation trials using a Mosquito robot (TTP LABTECH) and commercially available crystal screen kits (MemGold and MemSys & MemStart) and the plates were stored at 18 °C. Examples of crystals that grew from the MemGold screen kit (**Table 3.I**) are shown in **Figure 3.6**.The Korima UV microscope (excitation set at 280 nm) was used to confirm that the crystals were protein crystals. **Table 3.II,** shows the initial condition obtained from MemGold screen kit in the presence of L-4-thialysine.

Table 3.I. Initial sitting drop crystallization hits of LysP using the MemGold screen (Molecular Dimensions).

5 mM L-Lysine was added to protein sample prior to crystallisation. Crystals grew at 20 °C.

Condition	Salt	Buffer	pH	Precipitant
D5	0.05 M Magnesium acetate	0.1 M MES	5.4	26 %(v/v) PEG 350 MME
D6	0.2 M calcium chloride	0.1 M Tris HCl	6.5	39 %(v/v) PEG 400
G3	0.1 M Magnesium chloride/Sodium chloride	0.1 M Tris HCl	8.5	33 %(v/v) PEG 400

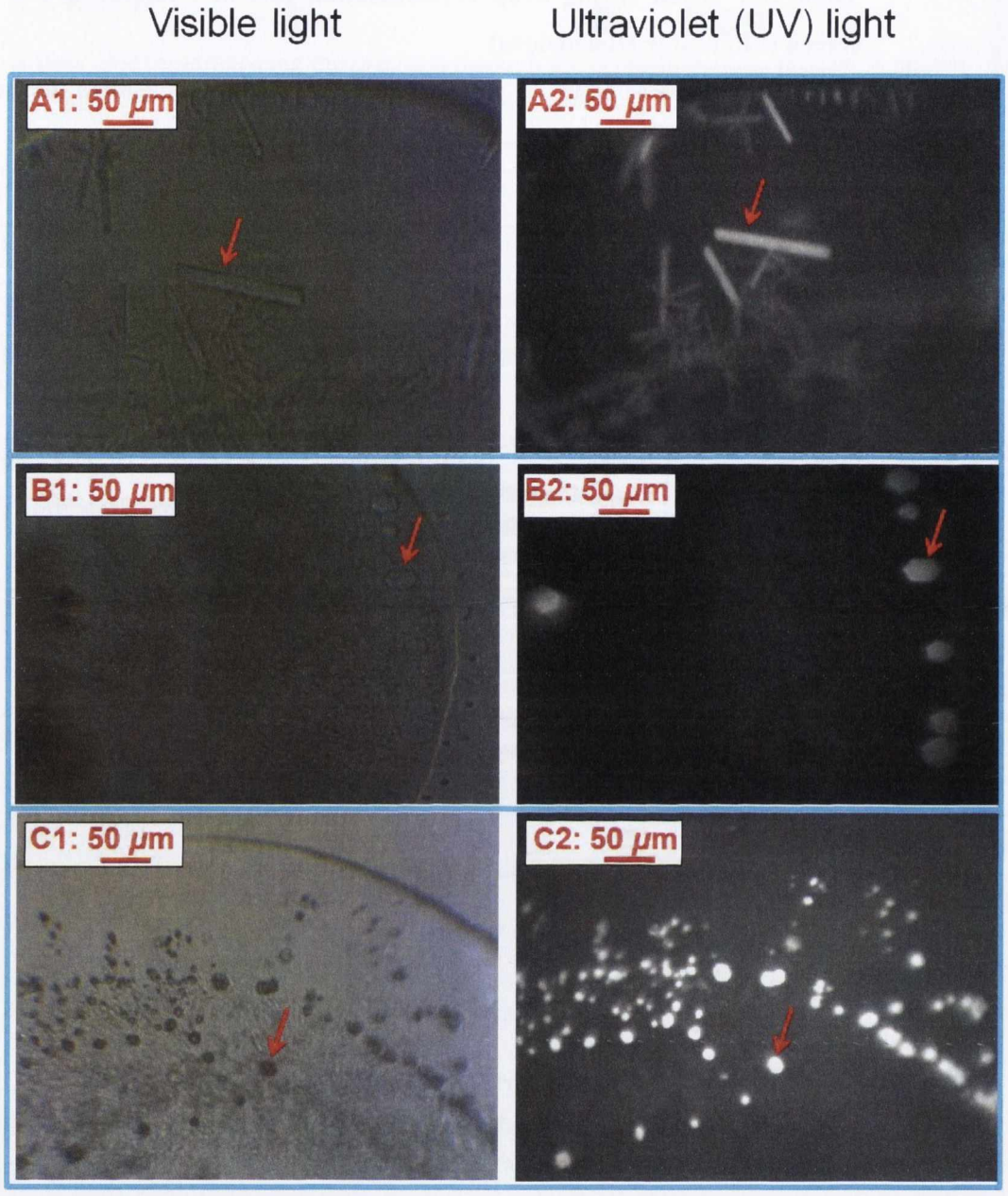


Figure 3.6. Initial *in surfo* crystallisation hits of LysP obtained by vapour diffusion using the MemGold screen kit.

Left column shows images obtained using visible light and right column shows same wells viewed under UV light. All images were taken using a Korima UV microscope (excitation set at 280 nm). Crystals in A, B and C from **Figure 3.6.** correspond to condition D6, G3 and D5 respectively in **Table 3.I**. Crystals grew at 18 °C.

Table 3.II. Initial crystallization hits of LysP obtained by sitting drop on a MemGold crystallization screen kit.

5 mM L-4-thialysine was added to protein sample prior to crystallisation. Crystals grew at 20 °C.

Condition	Salt	pH	Buffer	Precipitant
B3	0.04 M NaCl	8.0	0.04 M Tris HCl	27 %(v/v) PEG 350 MME
D1	0.2 M CaCl2	7.5	0.1 M HEPES	53 %(v/v) PEG 400
D3	-	7.5	0.05 M HEPES	22 %(w/v) PEG 4000
D7	0.1 M KCl	8.5	0.1 M Tris HCl	53 %(v/v) PEG 400
H7	0.2 M CaCl2	6.5	0.1 M MES	33 %(v/v) PEG 400

The crystals in **Figure 3.6 B)** were further shown to be LysP crystals by harvesting them and washed several times in reservoir solution, dissolved in SDS sample buffer, run on SDS PAGE and silver stained (**Figure 3.7 A**). The doublet bands around 36.6 kDa was seen for both mother liquor and the crystals. Also, there were higher molecular weight aggregates seen in both LysP solution, mother liquor and the crystals. This behavior has been seen for LysP (**Figure 3.7 B)** in which the sample was heated at 45 °C and analysed on SDS PAGE which saw a disappearance of the monomeric doublet bands at 36.6 kDa to higher molecular weight bands. Thus, it is likely that the crystallization condition (0.1 M NaCl, 0.1 M $MgCl_2$, 0.1 M Tris pH 8.6 and 36 %(v/v) PEG 400) or the incubation of the plates for 2 weeks at 20 °C may have caused the aggregation of the protein in the mother liquor which was also seen in the crystals. There was a minor contaminant below the doublet band in both LysP solution and mother liquor which was not seen in the crystals suggestion the crystals were formed from a single species which was LysP. The contaminant could not be seen in gel B because it was stained by a less sensitive GelCode Blue Safe protein stain and not silver stain as gel A.

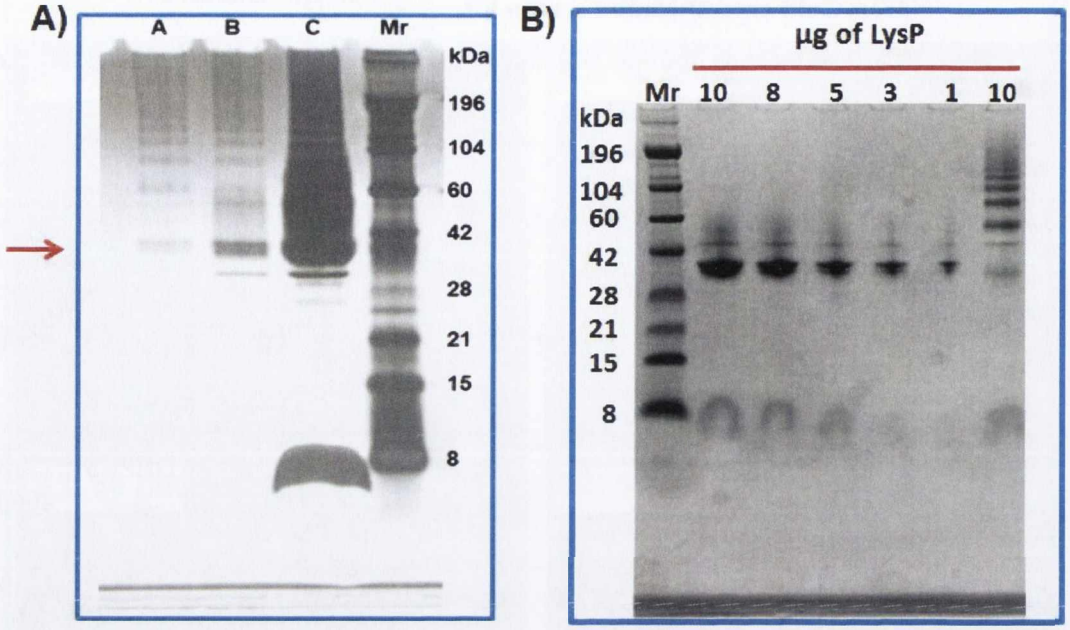


Figure 3.7. Silver staining of *in surfo* crystals of LysP and SDS PAGE analysis of the effect of heating on purified GFP free LysP.

A) Silver stained gel of LysP crystal. Lane A (LysP crystals), Lane B (mother liquor), lane C (purified LysP solution) and Mr (molecular weight standard). Crystals of LysP shown in **Figure 3.7 B,** were harvested, dissolved in SDS sample buffer, run on SDS PAGE and silver stained. **B)** SDS PAGE of LysP purified in Buffer A. The 10 µg sample at the end was boiled at 45 °C prior to loading. Gel B was stained by GelCode Blue Safe protein stain.

3.3.2. *In surfo* Crystallization Optimization LysP

3.3.2.1. Precipitant Concentration Screening

The initial hit was obtained from reservoir solution containing 0.1 M NaCl, 0.1 M $MgCl_2$, 0.1 M Tris HCl pH 8.6 and 33 % (v/v) PEG 400. In attempt to improve on the crystal quality and/or size, the PEG 400 concentration was varied from 20 to 50 %(v/v) (steps 2%(v/v). In this experiment, crystals were obtained between 30 and 45 %(v/v) PEG 400 but only crystals that grew at concentrations higher than the initial PEG concentration were bigger (~ 200 µm) than that obtained in the initial crystallization trials (~ 100 µm). Unfortunately, the diffraction quality of the optimized crystals did not improve (7 Å).

3.3.2.2. Precipitant/Protein Drop Size Screening

Precipitant and protein drop size were varied to see if it could improve on the diffraction quality of the crystals. The variables used were as follows (2 μL + 2 μL, 1.5 μL + 2 μL, 1 μL + 2 μL, 2 μL + 1 μL) precipitant solution (38 %(v/v) PEG 400, 0.1 M NaCl, 0.1 M $MgCl_2$ and 0.1 M Tris pH 8.6) and protein samples (LysP+L-lysine) and (LysP+L-4-thialysine). However, crystals could only be obtained when the precipitant solution drop volume was either greater or equal to the LysP/ligand solution drop volume. There was no difference in the crystal size or quality obtained. There was clear protein precipitation with no crystals obtained when the LysP/ligand solution drop volume was greater than the precipitant solution drop volume.

3.3.2.3. Organic and Detergent Additive Screening

Another optimization strategy used was the addition of 0.5 %(v/v) each of various organic solvents and different detergents as additives from organic additives screen and detergent additives screen I and II from Hampton Research to the initial hit obtained (38 %(v/v) PEG 400, 0.1 M NaCl, 0.1 M $MgCl_2$ and 0.1 M Tris (HCl) pH 8.6. Surprisingly, crystals grow continuously to ~ 100 μm in 6 weeks. Despite the large sizes obtained crystals diffract very poorly (~12 Å). **Figure 3.8** shows an example of crystals that grew in the presence of 0.5 %(v/v) 1-propane diol as an organic additive.

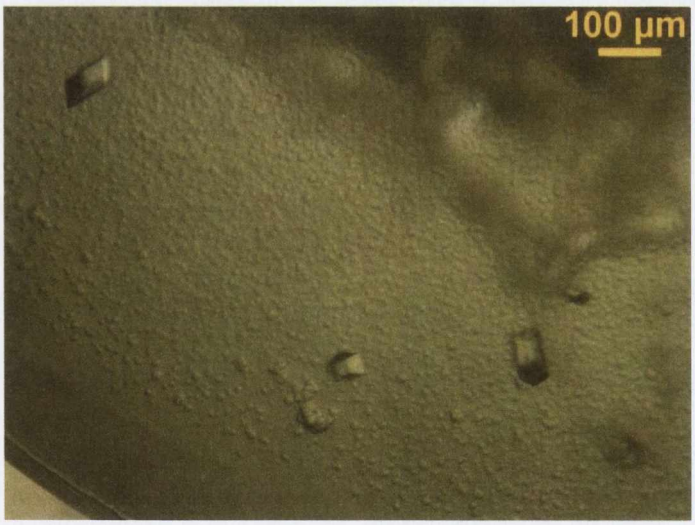


Figure 3.8. *In surfo* srystals of LysP in the presence of 0.5 %(v/v) 1-propane diol as an organic additive.

Initial crystals grew in the presence of 0.1 M NaCl, 0.1 M MgCl2, 0.1 M Tris pH 8.6, 38 %(v/v) PEG 400. Crystals took 6 weeks to grow to max size at 20 °C.

3.3.2.4. Addition of Polar *E. coli* Phospholipid

Phospholipid (PL) concentration was used between 0 and 2 mM with the goal of improving the diffraction quality of the crystals. The precipitant solution that was used was made up of 0.1 M $MgCl_2$, 0.1 M NaCl, 0.1 M Tris HCl pH 8.6 and 28 -38 %(v/v) PEG 400 (steps 2 %(v/v)). Plates were set up in duplicate and stored at both 20 ^{0}C and 4 ^{0}C. Although crystals were obtained at both temperatures, those at 4 ^{o}C were tiny (~ 10 μm) and while those at 20 ^{o}C were about 100 μm. However, addition of PL alone could not improve the diffraction quality of LysP.

3.3.2.5. Optimization of LysP Using Similar Published Condition for Wild Type LacY.

Lastly, we were able to improve the diffraction with reflections seen to 5 Å using similar condition that was used for LacY. 1 μL of the protein solution (30 mg/mL containing 1 mM polar *E. coli* lipid from Avanti with either 5 mM L-Lysine or L-4-thialysine) was mixed with 1 μL of a freshly prepared solution consisting of 8 μL of reservoir (0.1 M Tris pH 8.6, 0.1 M NaCl/0.1 M $MgCl_2$/30–36% PEG 400), 1 μL of 80 mM CHAPS and 1 μL of 30 %(w/v) 1,6-hexanediol. The 2 μL drops were equilibrated against 300 μL of reservoir solution at 20°C. The difference between the crystallization condition of wild type LysP and LacY is that the reservoir solution for LacY had 0.1 M ammonium sulfate, 100 mM HEPES pH 7.5 and 27-30 %(v/v) PEG 400. Crystals of LysP appeared after 4 days and only in the presence of L-4-thialysine and in 36 %(v/v) PEG 400. These are rods of average size ~ 100 μm (**Figure 3.9**) and diffracted to 7 Å with reflections to 5 Å. The slight improve in the diffraction may have worked because the addition of PL may have improved the stability of LysP. Guan *et al.* 2006, showed that by adding different amounts of PL led to the formation of different crystal forms of *E. coli* LacY and also they saw a correlation between the amounts of PL added with the diffraction quality (Guan *et al.* 2006). Excessive removal of PL may have occurred during the wash steps in the purification process. However, more screening needs to be carried out to determine the right amount of PL to be added.

Figure 3.9. *In surfo* crystals of LysP obtained using similar published condition for wild type LacY.

LysP crystals grew in [0.1 M Tris HCl (pH 8.6)/0.1 M NaCl/0.1 M $MgCl_2$/30–36% PEG 400, 8 mM CHAPS and 3 %(w/v) 1,6-hexanediol].

3.3.3. *In Meso* Crystallization Trials of LysP

Concentrated TEV-cleaved LysP (15 mg/mL in 150 mM NaCl, 20 mM Tris pH 7.5 and 0.03 %(w/v) DDM) was set up into crystallization trials using the method *in meso* method. This involved reconstituting the protein into the LCP using commercially available lipid (9.9 MAG) and novel lipids synthesized in the Caffrey Laboratory (7.8 MAG and 7.7 MAG). The following commercial screens were set up: PACT premier (Molecular Dimensions), Memsys & Memstart (Molecular Dimensions), MemGold (Molecular Dimensions), MbClass (Qiagen), MemFac (Hampton Research), SaltRx (Hampton Research) and Cubic Screen (Emerald BioSystems). Plates were stored at 20 ^{o}C. No crystal was obtained but however several conditions (**Table 3.III**) gave protein precipitates with all three lipids tested that could be re-screened to see if crystals can be obtained. **Figure 3.10** shows an example of a precipitate obtained on a MemGold screen kit in 0.1 M HEPES pH 7.5 and 11 %(w/v) PEG 4000 after 30 days. 7.8 MAG was used to prepare the cubic phase. Crystals may not have been obtained because only 3 lipids (9.9 MAG, 7.8 MAG and 7,7 MAG) were tried. Thus, more lipids need to be tested to find out the one LysP can call home in order to crystallize. Also, those conditions that gave precipitate needs to be rescreened (**Table 3.III**)

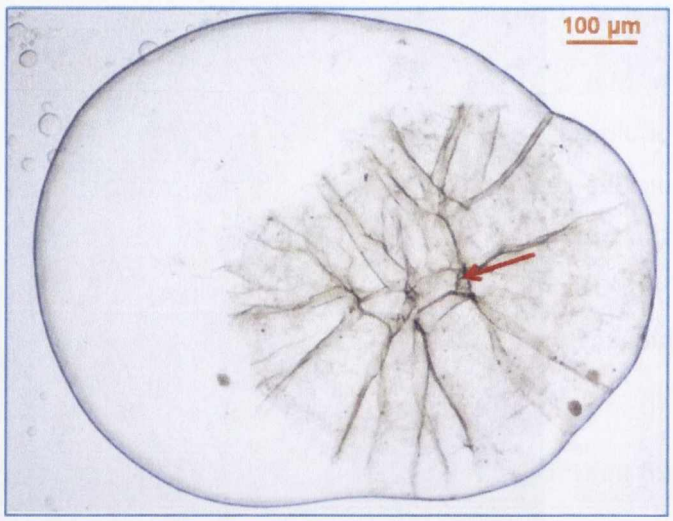


Figure 3.10. *In meso* crystallization trials of GFP free LysP showing precipitated protein in the bolus of cubic phase.

This is an example of a precipitate that was obtained after 30 days at 20 °C on the Memgold screen kit (0.1 M HEPES pH 7.5 and 11 %(w/v) PEG 4000). 7.8 MAG was used to prepare the cublic phase.

Table 3.III. Summary of some *in meso* crystallization conditions which resulted in precipitation of LysP.

MemGold	Salt	buffer	pH	Precipitant
B2	0.1 M Sodium chloride	0.1 M Sodium citrate	5.6	5.5 %(w/v) PEG 3350
B3	0.05 M Calcium chloride/0.05 M Barium chloride	0.1 M Tris HCl	8.2	32 %(v/v) PEG 400
B12	None	0.1 M Tris HCl	8.0	5.5 %(w/v) PEG 4000
C1	0.1 M Sodium chloride	0.02 M Tris HCl	7.0	7.7 %(w/v) PEG 4000
C3	0.04 M Sodium cholride	0.04 M Tris HCl	8.0	27 %(v/v) PEG 350 MME
C4	0.05 M sodium chloride/0.05 M Magnesium chloride	0.1 M Sodium citrate	6.0	22 %(v/v) PEG 400
C11	0.02 M nickel sulfate	0.01 M HEPES	7.0	33 %(v/v) Jeffamine-M600

Table 3.III. Continued.

D1	0.2 M calcium chloride	0.01 M HEPES	7.5	53 %(v/v) PEG 400
D2	0.05 M magnesium acetate	0.05 M sodium acetate	5.0	28 %(v/v) PEG 400
D4	0.2 M calcium chloride	0.1 M Tris HCl	8.0	44 %(v/v) PEG 400
D5	0.05 M magnesium acetate	0.05 M sodium acetate	5.4	24 %(v/v) PEG 400
D7	0.1 M potassium chloride	0.1 M Tris HCl	8.5	39 %(v/v) PEG 400
E1	0.22 M sodium citrate	0.1 M Tris HCl	8.0	35 %(v/v) PEG 400
E4	None	0.02 M Tris HCl	7.5	22 %(v/v) PEG 550 MME
F2	0.1 M sodium acetate	0.02 M Tris/0.1 M sodium citrate	7.5/ 5.0	
F7	0.02 M potassium nitrate	0.03 M potassium citrate	6.5	7.7 %(w/v) PEG 4000
G1	0.2 M ammonium sulfate	0.1 M sodium acetate	4.6	28 %(v/v) PEG 550 MME
G2	None	0.05 M glycine	9.0	55 %(v/v) PEG 400
G5	0.2 M magnesium chloride/ 0.1 M potassium chloride	0.025 M sodium citrate	4.0	33 %(v/v) PEG 400
G8	0.1 M sodium chloride/4% v/v ethylene glycol	0.1 M MES	6.5	33 %(v/v) PEG 400
H1	0.1 M sodium chloride	0.1 M sodium phosphate	7.0	33 %(v/v) PEG 300
H7	0.2 M calcium chloride	0.1 M MES	6.5	33 %(v/v) PEG 400

Table 3.III. Continued.

H11	0.001 M cadmium chloride/0.03 M magnesium chloride	0.1 M MES	6.5	30 %(v/v) PEG 400
MemSys & Memstart	**Salt**	**buffer**	**pH**	**Precipitant**
B9	0.1 M sodium chloride	0.1 M sodium citrate	5.6	30 %(v/v) PEG 400
C1	0.1 M Ammonium sulphate	0.1 M Sodium HEPES	7.5	30 %(v/v) PEG 400
D8	0.2 M Zinc acetate	0.1 M MES	6.5	10 %(w/v) PEG 8000
F3	None	0.1 M Sodium HEPES	7.5	30 %(v/v) PEG 400
F4	0.1 M sodium chloride	0.1 M Sodium HEPES	7.5	30 %(v/v) PEG 400
G12	None	0.1 M MOPS	7.0	12 %(w/v) PEG 4000
H3	None	0.1 M Sodium HEPES	7.5	12 %(w/v) PEG 4000
Index HT	**Salt**	**buffer**	**pH**	**Precipitant**
D2	None	0.1 M HEPES	7.0	30 %(v/v) Jeffamine M-600 pH 7.0
D3	None	0.1 M HEPES	7.0	30 %(v/v) Jeffamine ED-2001 pH 7.0
E6	0.05 M Calcium chloride dihydrate	0.1 M BIS-TRIS	6.5	30 %(v/v) PEG 550 MME
F1	0.2 M L-Proline	0.1 M HEPES	7.5	10 %(w/v) PEG 3350
PACT Premier	**Buffer/salt**	**Salt/buffer**	**pH**	**Precipitant**
A2	0.1 M Succinic Acid, Phosphate, Glycine system	None	5.0	25 %(w/v) PEG 1500
C1	0.1 M Propionic acid, Cacodylate, Bis-tris propane system	None	4.0	25 %(w/v) PEG 1500
C2	0.1 M Propionic acid, Cacodylate, Bis-tris propane system	None	5.0	25 %(w/v) PEG 1500

3.3.4. Bicelle Crystallization Trials of LysP/CA5755/L-lysine

Crystallization trial of LysP/CA5755/L-lysine was carried out in the presence of DMPC:CHAPSO bicelle as described in **section 2.1.4**. Two conditions gave salt crystals by the bicelle method. Upon exposure to synchrotron X-radiation they did not show any protein diffraction, suggesting that crystals were likely salt crystals. **Figure 3.11** shows an example of the salt crystals obtained by the bicelle method. Bicelle crystallization may not have worked because only one bicelle concentration was tested (7 %(w/v)). Therefore, more concentrations need to be tested to see if crystals will be obtained.

Figure 3.11. Example of *in surfo* salt crystals obtained in the bicelle crystallization trials at 20 °C.

The crystallization condition was 0.1 M Magnesium acetate, 0.1 M Sodium citrate pH 5.8 and 14 %(w/v) PEG 5000 MME. The protein concentration was 23 mg/mL in Buffer B and crystals appeared after 7 days.

3.3.5. Nanobody Production and Screening for Tight Binders to LysP.

Nanobodies that will be used as co-crystallants for structural investigations were raised and screened according to already described protocol (Muyldermans 2001, Vincke & Muyldermans 2012). Nanobodies were over expressed in WK6 SU^- cells and purified by a combination of IMAC and SEC. **Figure 3.12 A)** shows the SEC chromatogram of the different nanobodies and **B)** SDS PAGE gel of the purified nanobodies. The final yield for the nanobodies was 10 mg/mL of cell culture medium.

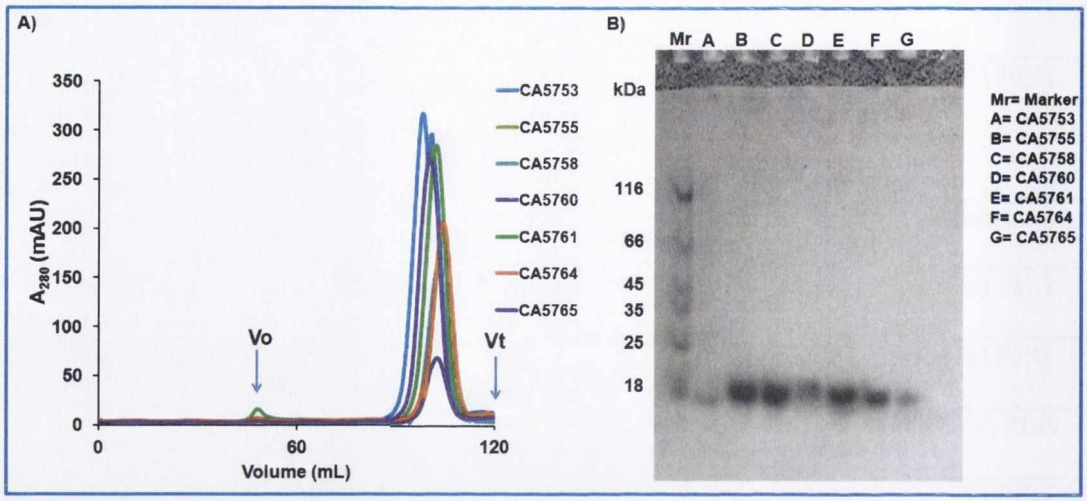


Figure 3.12. Purification of nanobodies against GFP free LysP.

A) Gel filtration chromatogram of the different clones. B) SDS PAGE analysis of the purified nanobodies against LysP. The gel was stained with GelCode Blue Safe protein stain. Samples after IMAC were run on a hiLoad 16/60 Superdex 200 column connected to an Äkta prime at 4 °C at a flow rate of 1 mL/min using buffer containing 150 mM NaCl and 20 mM Tris pH 7.2. The column was pre-equilibrated with the same buffer prior to injecting the samples.

We further rescreened these nanobodies to select those that will bind to LysP in buffer A. LysP was incubated with all nanobodies at a molar ratio of 1:3 respectively for 1 h on ice and the samples were analyzed on SEC for co-elution. By co-elution we were looking for peak shift on SEC. Two clones (CA5755 and CA5764) out of 7 showed a shift in the maximum elution volume from 15.7 mL to 15.5 mL (**Figure 3.13**). This suggests that there is binding of both clones to LysP. No peak shift was seen for the other clones and for the control which was free LysP with no nanobody. This result was further validated on SDS PAGE in which for the two suggested positive clones the doublet bands at about 36.6 kDa was seen for LysP and a lower band corresponding to the nanobodies were seen (**Figure 3.14, lanes D and L**).

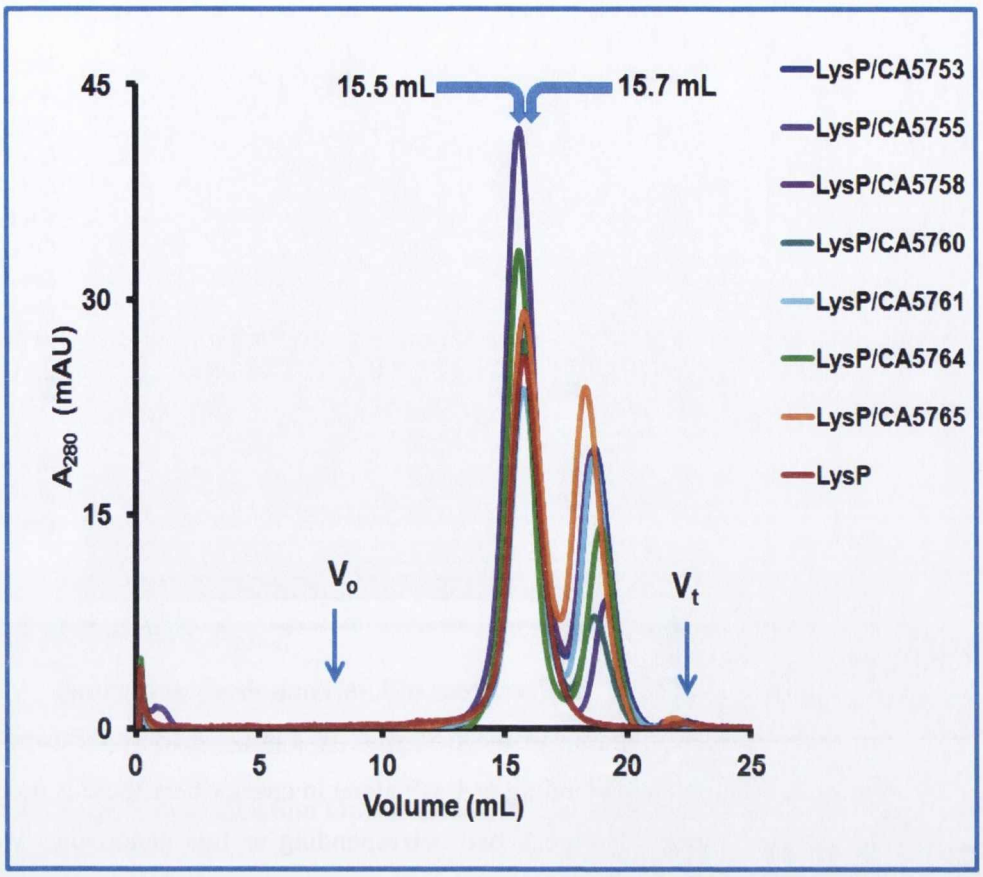


Figure 3.13. SEC of LysP/nanobody complex screening.

LysP was incubated with purified nanobodies at a molar ration of 1:3 respectively and incubated on ice for 1 h. Samples were then analysed by injecting onto a Superose 6, 10/30 connected to an Äkta Explorer at a flow rate of 0.2 mL/min using Buffer B at 4 °C. Prior to injecting the sample column was pre-equilibrated in Buffer B.

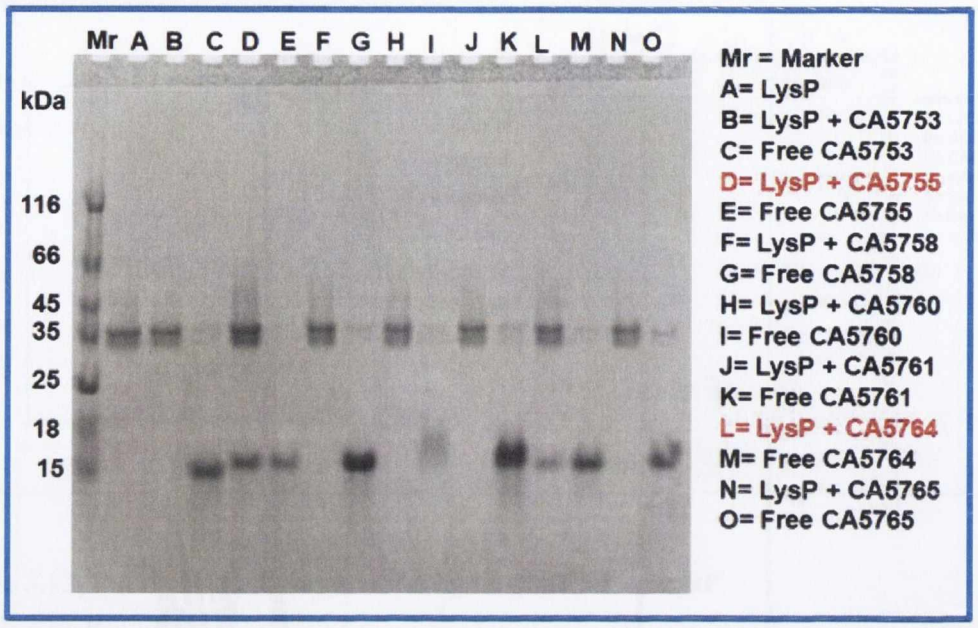


Figure 3.14. SDS PAGE analysis of LysP/nanobody screening.

Samples corresponding to the first peak in **Figure 3.13** (LysP/nanobody complex in cases where there is binding or LysP alone in cases where there is no binding) were run alongside samples in peak two corresponding to free nanobodies were run on 12 % SDS PAGE gels (120 V for 2 h at 22 °C) and stained with GelCode Blue Safe protein stain. Positive clones can be seen in lanes D and L where bands corresponding to LysP and the nanobodies can be seen running together.

3.3.6. *In Surfo* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764.

LysP purified in MNG DDM was incubated with either nanobody clones CA5755 or CA5764 in the presence of L-lysine as described in section **2.1.6. Figure 3.15**, shows an initial hit of LysP crystals obtained from MemGold (0.1 M $CaCl_2$ + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) in the presence of CA5755 and L-lysine. No crystal was obtained in the presence of nanobody clone CA5764.

Figure 3.15. Initial *in surfo* crystallization hit of LysP in the presence of nanodody clone CA5755 and L-lysine.

Crystals grew in the presence of 0.1 M $CaCl_2$ + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME. The crystals appeared after 3 weeks at 20 °C.

3.3.7. *In Surfo* Crystallization Optimization of LysP with Nanodody Clone CA5755

In the initial crystallization trials *in surfo* of LysP in the presence of CA5755, crystals were obtained in the presence of (0.1 M $CaCl_2$ + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME (**Figure 3.15**). Crystals were very tiny (15-20 µm) in size. The goal of this optimization was to increase the size of these crystals in this condition. The following strategies were employed.

3.3.7.1. PEG 2000 MME Screening.

The first strategy employed was to vary the PEG 2000 MME concentration from 9 to 19 % (w/v) with a step of 2 %(w/v). The $CaCl_2$ and Tris pH 6.5 concentrations were kept at 0.1 M. Crystals were obtained only at 13 %(w/v) PEG 2000 MME (**Figure 3.16**). There was a slight improvement of the sizes (from 20-30 µm to 50-60 µm).

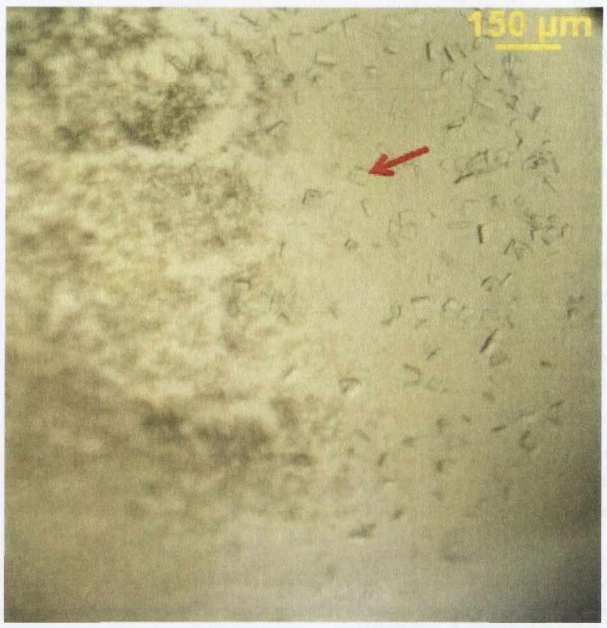


Figure 3.16. *In surfo* crystallization optimization of LysP/CA5755/L-lysine by screening PEG 2000 MME concentration.

The PEG 2000 MME concentration was screened and crystals grew only in 13 %(w/v) PEG 2000 MME. Square plate crystals appeared after 3 weeks at 20 °C and were 50-60 µm in size.

3.3.7.2. PEG 2000 MME and pH Screening

The second strategy used was the screening of pH. The PEG 2000 MME concentration was varied as in **section 3.1.8.1** but four different pH were used (0.1 M Tris pH 6.5, 7.5 and 8.6 and also 0.1 M sodium citrate pH 5.5 was used). Crystals were obtained only at pH 6.5 and 13 %(w/v) PEG 2000 MME similar to **Figure 3.16**.

3.3.7.3. $CaCl_2$ Concentration Screening.

Also, another strategy tested was the screening of $CaCl_2$ concentration. Three different concentrations of $CaCl_2$ were used (0.05 M, 0.1 M and 0.2 M). The PEG 2000 MME concentration was 13 %(w/v) and the buffer used was 0.1 M Tris pH 6.5 as in the initial hit. Only 0.1 M $CaCl_2$ produced crystals as in the initial hit (**Figure 3.17**). No crystal was seen at 0.05 M and 0.2 M $CaCl_2$ but precipitate (**Figure 3.18 and 3.19**) respectively.

Figure 3.17. ***In surfo*** **$CaCl_2$ screening of LysP/CA5755/L-lysine.**

Three concentrations were tested 0.05, 0.1 and 0.2 M but crystals grew only in 0.1 M. Square plate crystals appear after 3 weeks at 20 °C and were 50-60 micron in size.

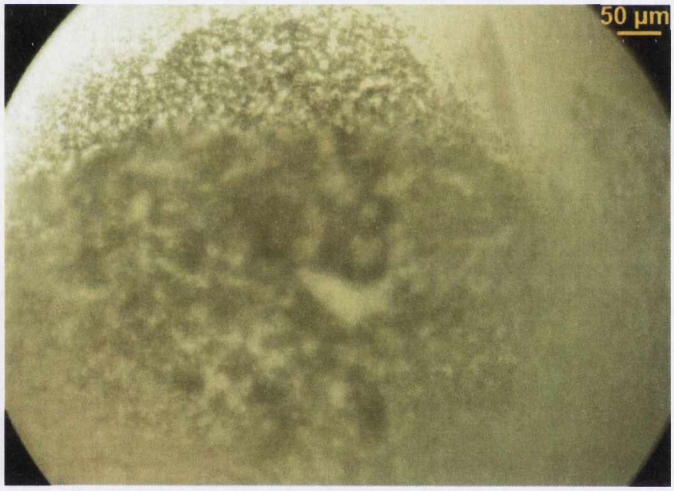


Figure 3.18. ***In surfo*** **$CaCl_2$ screening of LysP/CA5755/L-lysine.**

Three concentrations were tested 0.05, 0.1, 0.2 M. This condition had 0.05 M $CaCl_2$ and only precipitates could be seen.

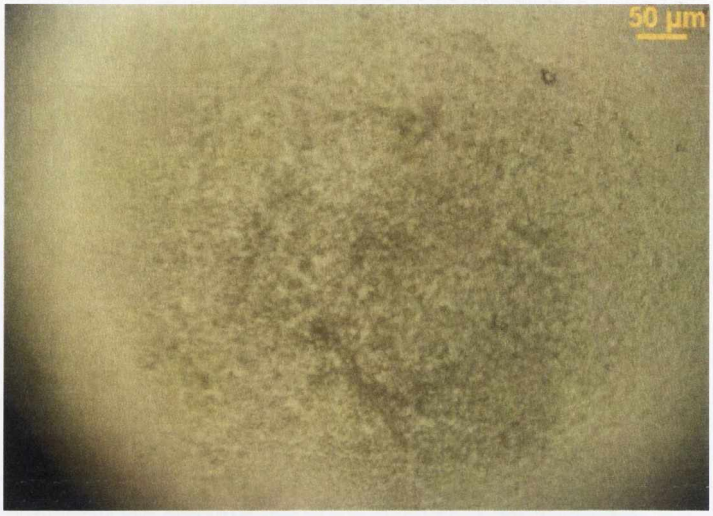


Figure 3.19. *In surfo* $CaCl_2$ screening of LysP/CA5755/L-lysine.
Three concentrations were tested 0.05, 0.1, 0.2 M. This condition had 0.2 M $CaCl_2$ and only precipitates could be seen.

Since crystals were obtained only at 0.1 M $CaCl_2$, further screening was carried out above and below 0.1 M $CaCl_2$. The screening done was from 0.08 to 0.13 M $CaCl_2$ (steps 0.01 M), 13 %(w/v) PEG 2000 MME and 0.1 M Tris pH 6.5 were maintained. Similar square plate crystals were obtained at 0.08 M and 0.09 M $CaCl_2$ concentration as in **Figure 3.17**. However a different crystal form showed up at 0.1 M $CaCl_2$ concentration which was not obtained previously (**Figure 3.20**), the square plate crystals were also obtained for this concentration. No crystal was obtained after 4 weeks at concentrations above 0.1 M $CaCl_2$.

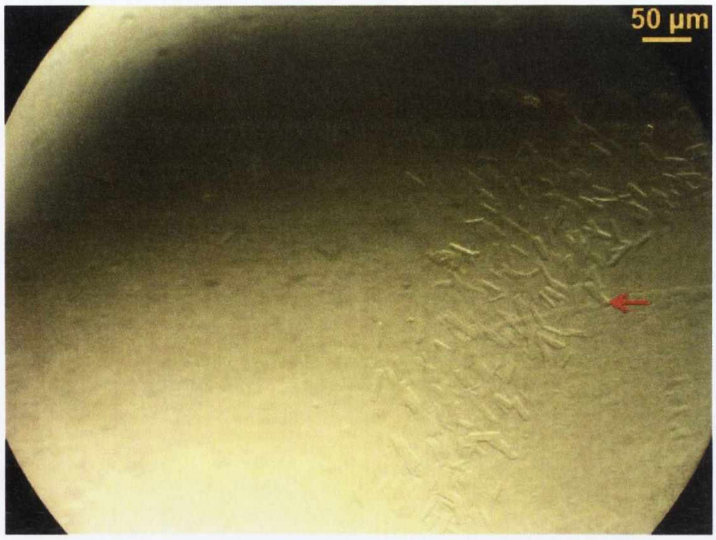


Figure 3.20. *In surfo* $CaCl_2$ screening of LysP/CA5755/L-lysine.

The concentrations tested were from 0.08 to 0.13 M CaCl2 (steps 0.01 M), Rectangular plated of size 50-60 µm were obtained at 0.1 M CaCl2 after 3 weeks at 20 oC. This was a new crystal form unlike the square plates obtained at 0.1 M CaCl2.

3.3.7.4. Precipitant/Protein Drop Size Screening

Here, crystals were optimized by screening different ratios of precipitant and protein drop sizes. LysP incubated with CA5755 in the presence of either L-lysine or L-4-thialysine as described in **section 2.1.1.** The ratio of precipitant:protein complexes used were, 1:1, 2:2, 2:1, 1:2 and 1.5:1 respectively. In the presence of L-lysine similar square plate crystals were obtained in 1:1 (**Figure 3.21**) and 2:2 (**Figure 3.22**) µL ratio of precipitant:protein respectively. Heavy precipitate was seen with the 2:1 µL ratio of precipitant:protein respectively (**Figure 3.23**) which was similar to that obtained in the 1.5:1 µL ratio (**data not shown**). Light precipitate was seen with the 1:2 µL ratio of precipitant:protein respectively (**Figure 3.24**). In the presence of L-4-thialysine no crystal was seen at the 1:1 µL ratio of precipitant:protein respectively (**Figure 3.25**). While crystals of about 50-60 micron were seen after 3 weeks in the 2:2 (**Figure 3.26**) and 1.5:1.5 (**Figure 3.27**) µL ratio of precipitant:protein respectively. Heavy precipitate was seen at the 2:1 (**Figure 3.28**) while light precipitate was seen at the 1:2 (**Figure 3.29**) µL ratio of precipitant:protein respectively.

Figure 3.21. ***In surfo*** **optimized crystals of LysP/nanobody CA5755 complex.** Crystals were obtained in 1:1 µL drop size of precipitant (0.1 M $CaCl_2$ + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-lysine after 3 weeks at 20 °C.

Figure 3.22. ***In surfo*** **optimized crystals of LysP/nanobody CA5755 complex.** Crystals were obtained in 2:2 µL drop size of precipitant (0.1 M $CaCl_2$ + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-lysine after. 3 weeks at 20 °C.

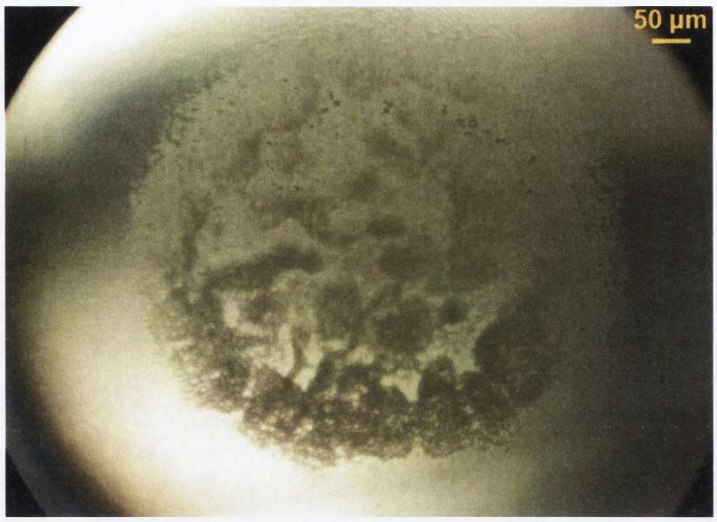


Figure 3.23. ***In surfo*** **optimized condition of LysP/nanobody CA5755 complex.** Precipitate were obtained in 2:1 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-lysine.

Figure 3.24. ***In surfo*** **optimized condition of LysP/nanobody CA5755 complex.** Precipitate were obtained in 1:2 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-lysine.

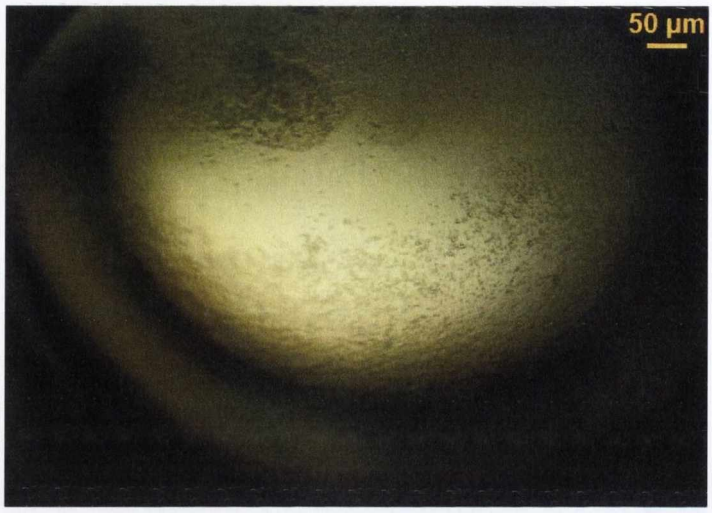


Figure 3.25. *In surfo* optimized condition of LysP/nanobody CA5755 complex. Precipitate were obtained in 1:1 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-L-4-thialysine.

Figure 3.26. *In surfo* optimized crystals of LysP/nanobody CA5755 complex. Crystals were obtained in 2:2 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-4-thialysine after 3 weeks at 20 °C.

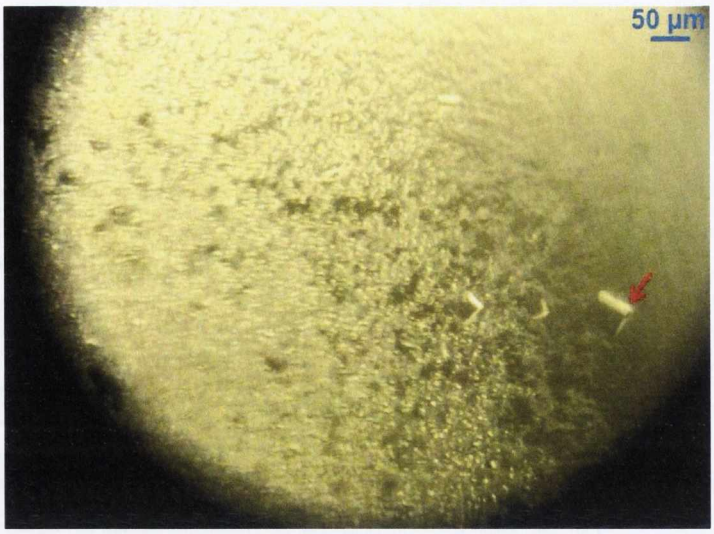


Figure 3.27. *In surfo* optimized crystals of LysP /nanobody CA5755 complex. Crystals were obtained in 1.5:1.5 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-4-thialysine after 3 weeks at 20 °C.

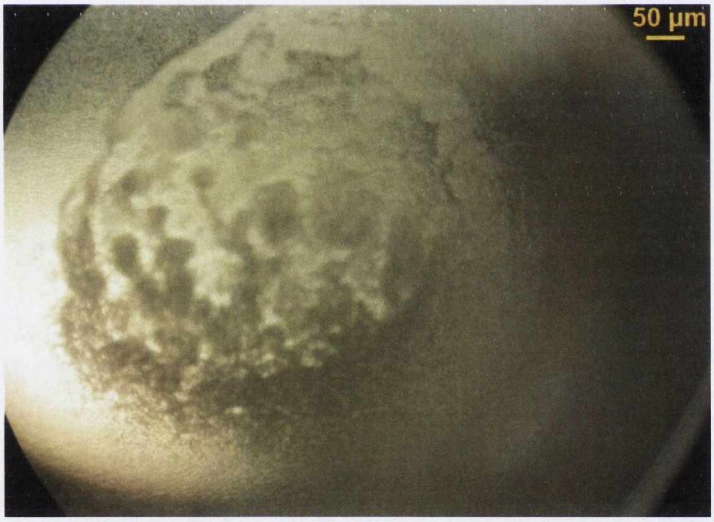


Figure 3.28. *In surfo* optimized condition of LysP/nanobody CA5755 complex. Precipitate were obtained in 2:1 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-L-4-thialysine.

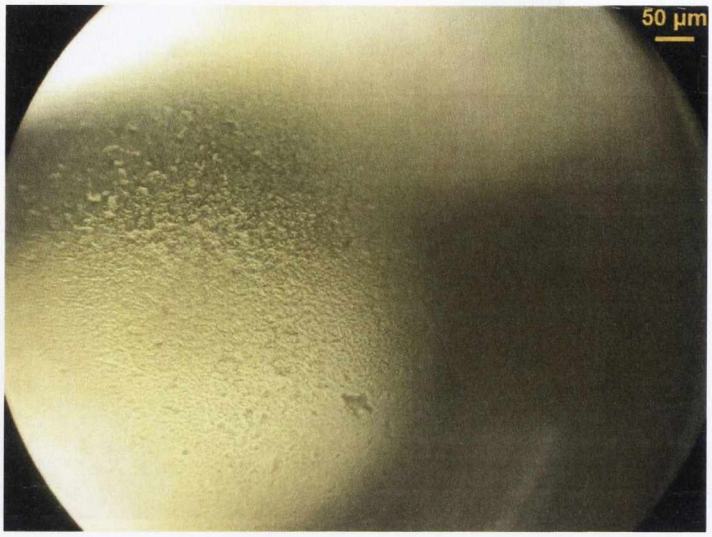


Figure 3.29. *In surfo* optimized condition of LysP/nanobody CA5755 complex. Precipitate were obtained in 1:2 µL drop size of precipitant (0.1 M CaCl2 + 0.1 M Tris pH 6.5 + 13 % PEG 2000 MME) and protein respectively in the presence of 5 mM L-L-4-thialysine.

3.3.7.5. Dehydration of LysP/CA5755/L-lysine Crystals in PEG 400

In this experiment, crystals from the initial condition (0.1 M Tris pH 6.5 + 0.1 M $CaCl_2$ + 13 %(w/v) PEG 2000 MME) were harvested and transferred to 2 µL of initial condition + 5 mM L-4-thialysine or L-lysine + 34-52 %(v/v) PEG 400 (2 % steps). After overnight incubation crystals stayed undissolved, prior to harvesting and cryocooling. Unfortunately, this did not improve the diffraction.

3.3.7.6. Seeding of LysP/CA5755/L-lysine Crystals

The logic of seeding has been described in **section 2.1.7.6** (Friedmann *et al.* 2011). Initial crystals from (0.1 M Tris pH 6.5 + 0.1 M $CaCl_2$ + 13 %(w/v) PEG 2000 MME) were optimized by setting up a fresh trial with the addition of PEG 400 [28-54 %(v/v) (steps 2 %(v/v)]. After 3 weeks crystals were not obtained unlike the initial condition were crystals were obtained in 3 weeks. Seeds from the initial condition (**Figure 3.16**) were transferred to the conditions that had PEG 400 added. Horse hair was used to transfer the seeds and plates were stored at 20 °C. No crystal was obtained.

3.3.8. *In Meso* Crystallization Trials of LysP with Nanobody Clones CA5755 and CA5764.

Purified LysP was incubated with either nanobody CA5755 or CA5764 in the the presence of L-lysine and L-4-thialysine and crystallization trials were set up as described in section **2.1.3**. Unfortunately, no crystal was obtained in the presence of either nanobody. However, precipitates were obtained.

3.3.9. X-ray Diffraction Studies of LysP with and without Nanobodies

Crystals from the initial crystallization trials as well as the optimized crystals (without nanobodies) were tested for diffraction at the Advanced Photon Source (APS) in Chicago on beamline NE-CAT and at Diamond Light Source in UK on beamline I24. **Figure 3.6, B1&2** shows the crystal morphology that produced the best crystal diffraction limits of 7 Å (**Data not shown**) from the initial crystallization trials. Also, crystals from **Figure 3.9** using similar published condition to wild type LacY diffracted to 7 with reflections to 6 Å (**Figure 3.30**), the data was indexed tentatively as $P2_1$ a = 175 Å, b = 291 Å, c = 176 Å, $\beta = 119^o$.

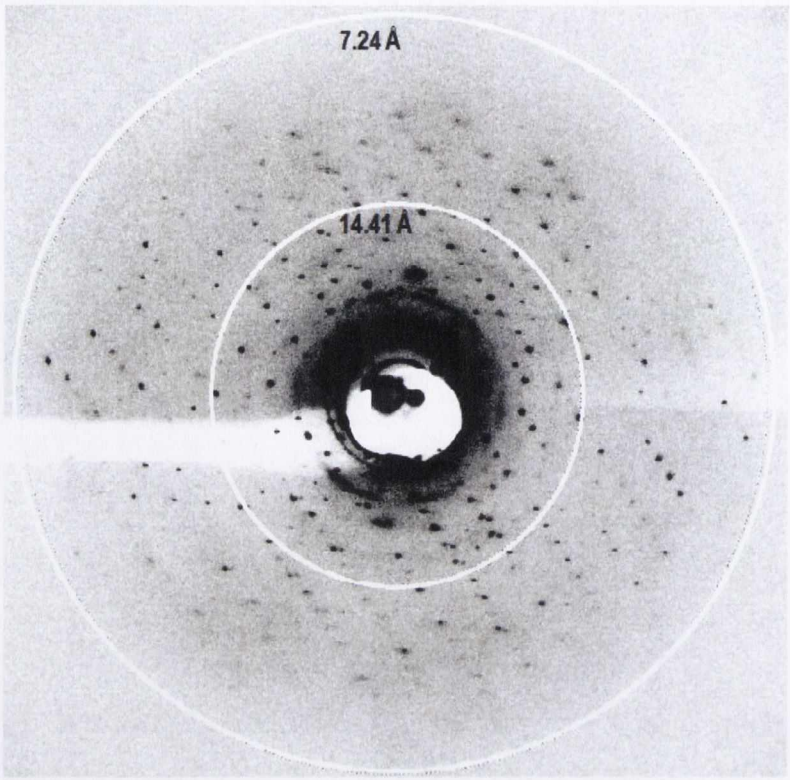


Figure 3.30. X-ray diffraction pattern of LysP at 7 Å with reflections to 6 Å. Crystals of the type shown in **Figure 3.9** was shot at the DLS on beam line I24 (100 % transmission, detector distance = 450 mm, exposure time = 1 sec, oscillation = 0.2 deg, energy = 12.8 keV, 180 degree images were collected on a 8 x 8 μm beam with a flux of 10^{12} photons/s at -173 °C). 22 images only were collected because the crystals died as a result of radiation damage.

Crystals of LysP/CA5755/L-lysine or L-4-thialysine were shipped to DLS and were shot and 0 and 90° data were collected on beam line I24 (100 % transmission on a 8 x 8 μm beam with a flux of 10^{12} photons/s at -173 °C) (**Figure 3.31**).

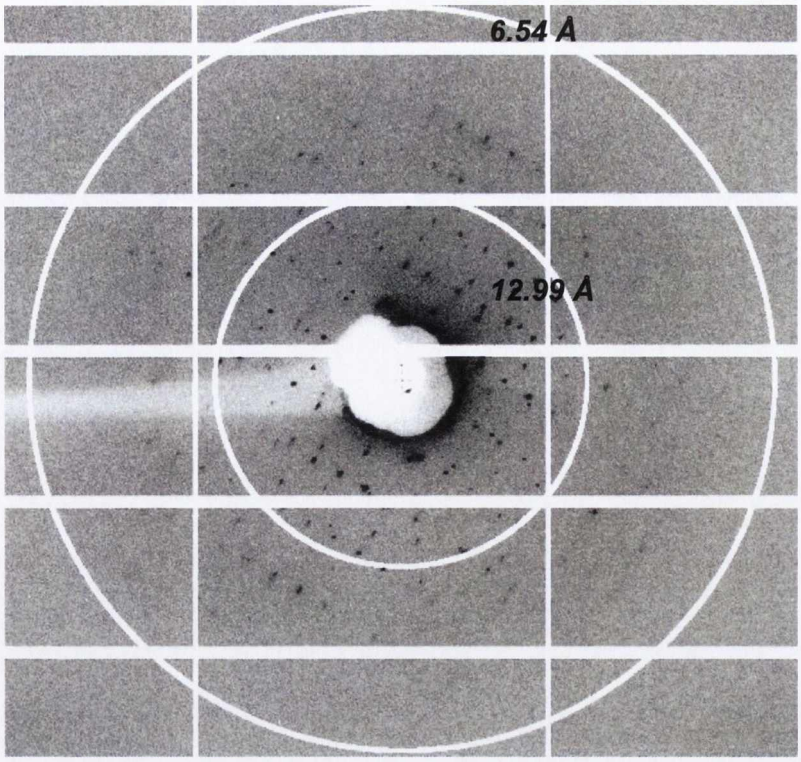


Figure 3.31. Diffraction pattern of LysP/CA5755/L-4-thialysine crystals.
Crystals diffract to 8 Å at the DLS on beam line I24 (100 % transmission, detector distance = 600 mm, exposure time = 0.2 sec, oscillation = 0.2 deg, energy = 12.8 keV, 0 and 90 degree images were collected on a 8 x 8 μm beam with a flux of 10^{12} photons/s at -173 °C). The data could not be index because of the poor diffraction quality.

3.4. Concluding Remarks

LysP from *Pseudomonas aeruginosa* PAO1 was purified from LysP-GFP fusion using DDM and the purified TEV cleaved free LysP was concentrated and crystallization trials both *in meso* and *in surfo* were set up. Crystals were obtained by the *in surfo* method both in the presence of 5 mM L-lysine and L-4-thialysine. The crystals were shown to be protein crystals by exposing them to UV light (**Figure 3.7**). The initial crystals were tested for diffraction on the synchrotron X-ray radiation and they diffracted the beam to 7 Å. Several optimization strategies were employed to improve the diffraction quality of the crystals with the goal of solving the structure of LysP. They include: screening of PEG concentration, screening of precipitant/protein drop size, screening of different detergent and organic additives, screening of temperature and the addition of *E. coli* polar lipid. Despite all the optimization strategies employed, only the *E. coli* polar lipids could improve the diffraction of LysP in the presence of L-4-thialysine with reflections seen to 6 Å. Despite setting up crystallization trials with LysP purified in DDM, it was obvious that the protein was unstable in DDM as it precipitates overnight at 4 °C when concentrated to 10 mg/mL. Thus we have to either set up crystallization trials immediately after the purification or preserve the protein at -80 °C for later use. Despite the instability of LysP in DDM, it was easily crystallizable as crystals appeared only after 16 h. LMNG was the best detergent to stabilize LysP as judged by monodispersity of the protein on SEC after concentrating the protein to 10 mg/mL and kept on ice for 7 days. Although LysP was stable in LMNG, it did not crystallize when purified in LMNG with either L-lysine or L-4-thialysine present. This drives us to go into collaboration with the goal of raising nanobodies to be used as cocrystallants with LysP. This collaboration was with Jan Steyeard's group at Vrije University in Brussels. Basically we expressed mailed both free and biotinylated LysP (for immunization and screening for binders respectively) to them and they performed the immunization and screening and identified 7 clones. The 7 clones were express mailed back to us and we further rescreened to identify those that bind tightly to LysP purified in LMNG. Two clones (CA5755 and CA5764) showed binding by coelution on SEC after incubating the nanobodies to LysP at a molar ratio of 3:1 respectively. The binding was further confirmed by running the coeluting peak on SDS PAGE and bands corresponding to LysP and the two nanobodies were seen. Both nanobodies were incubated with LysP at a molar ratio of 1.5:1 respectively in the presence of both L-lysine and L-4-thialysine. Crystallization trials were set up both *in meso* and *in surfo.*

Initial hit of the LysP/CA5755/L-lysine was obtained on the MemGold screen kit by the *in surfo* methods. Crystals could be harvested and washed several times in reservoir solution and the crystals will be dissolved in SDS sample buffer, run on SDS PAGE and silver stained to detect the presence of the CA5755 in the crystals. This result hopefully should show that the crystals were indeed from at least LysP/CA5755 complex. No crystal was obtained in the presence of CA5764. Unfortunately after several optimizations of the crystals could only diffract to 8 Å upon exposure to synchrotron X-ray radiation, which was LysP/CA5755 in the presence of 5 mM L-4-thialysine. More optimization is required here and more crystals need to be tested.

All in all, LysP was purified in DDM and crystals were obtained by *in surfo* method that diffracts to 7 Å with spot to 5 Å after exposure to synchrotron X-ray radiation after optimization. LysP was also cocrystallized with nanobody CA5755 and the cocrystals diffract to 8 Å after optimization. No crystal was obtained by the *in meso* method.

The reason why the diffraction could not be improved in DDM and LMNG could be the large micelle size of these detergents. As a result of the large micelle size protein-protein contact needed to form well ordered crystals was not available as the protein was surrounded by the detergents. For the case of LysP/CA5755 crystallization more optimization needs to be carried out and more crystals need to be tested.

Future work for this project could be to screen shorter chain detergents of the LysP/nanobody complex. This is because the nanobody could stabilize the protein in the shorter chain detergents. This can be done by purifying the protein in LMNG and then aliquots of say 50 µg of the purified protein will be diluted say 1 in 20 with buffer containing 1 %(w/v) of say β-octylglucoside (OG), nonylglucoside (NG) and n-lauryl-n,n-dimethylamineoxide (LDAO). The sample will be left at 20 °C overnight and analyse on SEC for monodispersity. Crystallization trials will follow with the shorter chain detergents identified in which the LysP/nanobody complex was stable with the goal of improving the diffraction quality. In this case the will be more protein-protein contact needed to form well ordered crystals (Sonoda *et al.* 2011). Also, since only one bicelle concentration (7 %(w/v)) was tested during the bicelle crystallization, more concentrations could be tried to see if crystals will be obtained. Lastly, only three lipids have been tried by the *in meso* method (9.9 MAG, 7.8 MAG and 7.7 MAG), other lipids could be tried to see if crystals will be obtained.

Also, those conditions that gave precipitate *in meso* could be rescreened to see if they will form crystals.

Chapter 4: General Conclusions and Discussions and Future Plans

4.1. General Conclusions and Discussions

LysP belongs to the APC transporter family (Shaffer *et al.* 2009). It is an alpha helical integral membrane lysine transporter from *Pseudomonas aeruginosa* strain PAO1 which has been predicted to have 12 transmembrane helices (http://www.cbs.dtu.dk/services/TMHMM). *Pseudomonas* uses lysine as a carbon and nitrogen source (Weill-Thevenet *et al.* 1979). Carbon and/or nitrogen are building blocks of biomolecules. To understand the mechanism of lysine selectivity and transport by LysP we sought to solve the high resolution crystal structure and to understand its function using biochemical and biophysical means.

In this project, *lysP* from strain PAO1 was cloned into the expression vector pWaldo-GFPd and over-expressed in *E. coli* C43 (DE3) cells as a C-terminal His-tagged Green Fluorescent Protein (GFP) fusion protein. *E. coli* BL21 (C43) DE3 strain is a special strain for toxic membrane protein over-expression (Shaw & Miroux 2003). The best yield based on whole cell GFP fluorescence (44 mg/L of cell culture medium) was obtained by lowering the temperature from 37 °C to 25 °C after induction with 0.4 mM IPTG. Such is the case with most membrane proteins were best yields are obtained by lowering the temperature after induction (Drew *et al.* 2006). The GFP addition proved particularly useful as its fluorescence could be easily monitored to quantify protein yield at every step in the expression and purification process and in the screening for membrane solubilizing detergents in which the transporter was monodispersed to enable purification of a homogenous target protein. Using FSEC, we screened 9 different detergents that have been either successful in solving the structures of membrane proteins or the production of a stable and homogenous membrane protein. They were; DDM, MNG DDM, DM, DMNG, NM, OG, OGNG, NG and LDAO. The top 4 detergents selected further for the purification of GFP free LysP from LysP-GFP based on solubilization yield and monodispersity on FSEC were: DDM, LMNG, DM and NM. LysP-GFP fusion was solubilized in DDM or LMNG, purified by Ni^{2+}-NTA affinity and size-exclusion chromatography. In the case of NM and DM, all purification steps were done using DDM except that the final SEC was done in the respective detergents. Prior to SEC, LysP was cleaved from the GFP fusion tag using TEV protease during dialysis and the resulting protein was monodispersed by SEC in DDM,

LMNG and DM which was consistent with the result from FSEC trace with a purity estimated at >95 % on SDS PAGE. When LysP was exchanged from DDM to DM on SEC, it could not be concentrated to above 2 mg/mL as the protein precipitated. In the case of NM, the SEC profile was not a Gausian peak compared to the FSEC trace. The reason could be that after the GFP was cleaved from LysP by TEV protease and the GFP free LysP exchanged from DDM to NM, it becomes unstable. Again, LysP purified in DDM precipitated overnight when concentrated at 10 mg/mL. Unlike the LysP purified in DDM, LysP purified in LMNG was still monodisperse on SEC for 7 days after concentrated to 10 mg/mL at 4 °C. The reason why LysP could be concentrated in DDM and LMNG easily more than DM and NM, could be because of its long acyl chains having 12 carbons than the other shorter chain detergents (10 carbons for DM and 8 carbons for NM). The long acyl chain produces larger micelles that surrounds the protein and keeps it stable. Also, the reason why LysP was stable in LMNG more than the other detergents could be because, LMNG is made up of two conventional DDM molecules linked together by a quaternary carbon (**Figure 4.1**). The central quaternary carbon and the bulky nature of the head group may reduce conformational flexibility in LysP making it stable, thus the reason why LMNG was the best in keeping LysP stable (Chae *et al.* 2010).

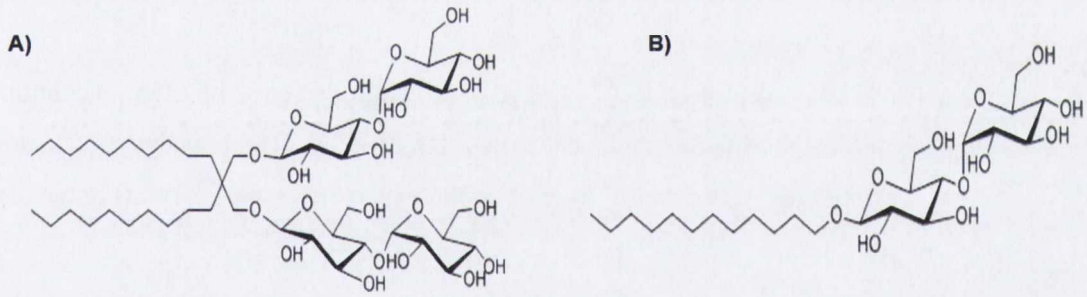


Figure 4.1. Chemical structure of the maltose neopentyl glycol detergent and the conventional detergent counterpart

A) Chemical structure of lauryl maltose neopentyl glycol. **B)** Chemical structure of n-dodecyl-β-D-maloside (figures copied directly from http://www.affymetrix.com) (Chae *et al.* 2010).

On SDS PAGE, there were always three bands seen for the purified GFP free LysP with the primary LysP band running as a doublet (Figure 2.29 A). The primary LysP doublet bands were not seen on blue native PAGE with also very minor aggregation

bands seen at higher molecular weight compared to SDS PAGE. Also, under another non denaturing condition (SEC) GFP free LysP runs as a single Gausian peak, suggesting the higher molecular weight bands seen on SDS PAGE could be aggregated forms of the protein upon addition of the denaturant, SDS (Haga *et al.* 2012). Similar behaviour was observed with the human M_2 muscarinic acetylcholine receptor (**Figure 4.2**) (Haga *et al.* 2012).

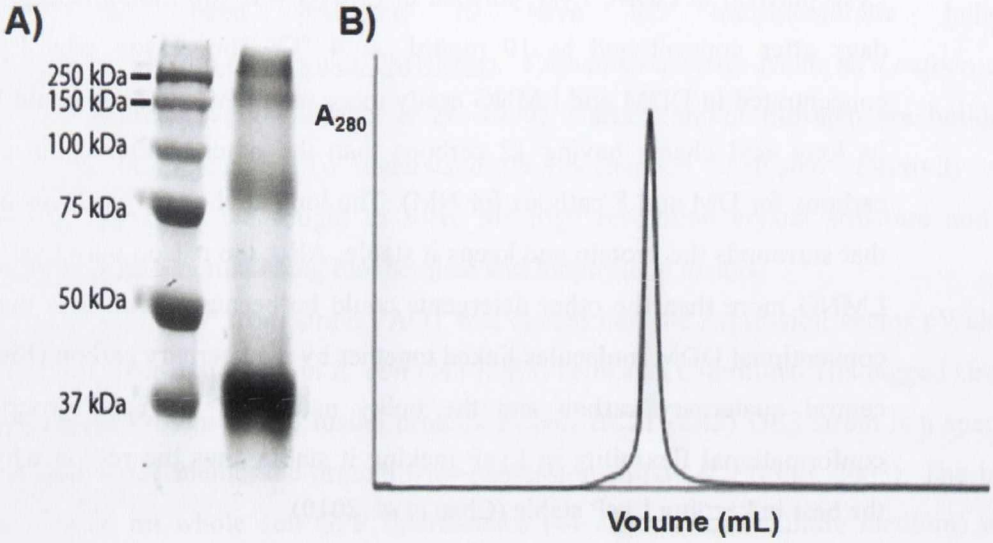


Figure 4.2. "SDS-PAGE and SEC analysis of M_2-T4L (human M_2 muscarinic acetylcholine receptor.

A) The receptor appears to aggregate in the presence of SDS, accounting for higher molecular weight bands on a gel. B) Size exclusion analysis is a non-denaturing technique, and receptor appears monodisperse by this assay" (Figure copied directly from Haga *et al.* 2012).

The major monomeric bands of LysP on SDS PAGE runs at ~36 kDa instead of 54 kDa (Figure 2.26) which is typical for most integral membrane proteins (Gaillard, Slotboom *et al.* 1996, Fang *et al.* 2007). The high molecular weight bands of LysP were proved not to be contaminants before proceeding into crystallization trials by total mass spectrometry. For total mass spectrometry, purified LysP in DDM micelle was precipitated and sent for total mass determination. The size obtained was 54044.19 ± 5.65 Da. which is very close to the theoretical molecular weight of the LysP + *Bam*HI (GS) + TEV cleavage site (ENLYF) construct (54037.2 Da) being purified. Although there was an additional mass ~7 Da, one could still conclude that the purified protein

was LysP. The mass increase may be due to the fact that LysP was isolated from the lipid bilayer with detergents and this makes it, as the case with membrane proteins, difficult to be isolated with buffers that are suitable for mass spectrometry. Also, membrane proteins bind the detergents in which they were purified and also bind lipid from the cell that produces them. This makes it difficult to produce positively charged peptide ions for electrospray ionization (Mirza *et al.* 2007). We further identified LysP using in-gel tryptic digestion followed by mass spectrometry. From the mass spectrum analysis report of the in-gel tryptic digestion fragments, the intensity coverage was 53.0 % while the sequence coverage was 8.8 %. This is usually the case for very hydrophobic membrane proteins that makes them difficult for its positively charged residues to be exposed for digestion by trypsin (the reason for the low sequence coverage). Combining all the results with the goal of identifying LysP, it was concluded that the purified protein was indeed LysP and was used further for crystallization trials and for functional assays. The oligomerization state of LysP was investigated by glutaraldehyde cross linking and blue native PAGE. All results suggested that LysP is a monomer in detergent solution.

Although LysP purified using DDM was pure enough to be used for crystallization trials, it was very unstable at concentrations above 5 mg/mL. For example it precipitated within 3 hours on ice at 10 mg/mL in Buffer A. Thus, after purification and concentration of LysP, it has to be set up directly into crystallization trials or frozen at -80 °C. This problem is similar to the case of wild type LacY in which phospholipids (PL) were added after purification to restore its stability before the structure could be solved (Guan *et al.* 2006). A possible explanation for this instability could be that during purification large amount of PL that help to keep the protein stable and functional were washed away. This explanation could be supported by the fact that the group of Prof. Ronald Kabback, University of California, Los Angeles, has modified the protocol for LacY purification by reducing the steps used for purification (there is no SEC in the protocol) prior to crystallogenesis (**Unpublished work, personal communication**). This modification has contributed to structural determination of LacY (Abramson *et al.* 2003).

Purified LysP was set up into crystallization trials using *in surfo* as well as the *in meso* method in the presence of both L-lysine and L-4-thialysine. Similar trials were also set without the above mentioned substrates. Several conditions gave crystals by the *in surfo* method in both substrates. No crystal was obtained without substrate, suggesting that the presence of substrate may help to stabilize the protein. The crystals from the

initial crystallization trials appeared after one day and grew to maximum size (~100 μm) within one week. Several screening around the initial conditions that gave crystals *in surfo* were carried out that improves the size of the crystals to 200 μm. The screening methods used included; varying the concentration of precipitant solution, varying the precipitant/protein drop size, the use of organic and detergent additives, and the addition of *E. coli* polar phospholipid. The LysP concentration used for the optimizations was 10 mg/mL in Buffer A. Unfortunately, even though the sizes of the crystals were improved during screening they could only diffract to 7 Å with reflections to 6 Å (the crystals from initial crystallization trials diffracted to 7 Å only). The reasons for this poor diffraction could be due to the fact that DDM, used for crystallization, has a large micelle size and this reduces protein to protein contact necessary for the formation of crystal contacts in well packed crystals. This problem could have been overcome by using shorter chain detergents but because of the dynamics of membrane transporters, they are difficult to be stabilized and to crystallize in shorter chain detergents (Sonoda *et al.* 2010). Purified free LysP in DDM easily gave crystals *in surfo* using MemGold screen kit more than any other detergents. This result is supported by the fact that the majority of membrane transporters in the MPDB (Raman *et al.* 2006; www.mpdb.tcd.ie) have been solved in DDM (Newstead *et al.* 2008 & Sonoda *et al.* 2010). Although most transporters have been crystallized in DDM, the median resolution is the poorest compared to other classes of membrane proteins (Figure 4.3. A) (Sonoda *et al.* 2010). In 2010, Sonoda *et al.* explored the relationship between the diffraction of different membrane protein families whose structures have been solved and the crystallization detergents (Figure 4.3. B) and found that 'no native X-ray structure of a secondary-active transporter has been solved to <2.9–3.0 Å in DDM, but has in either of the short chain detergents DM (sodium galactose transporter vSGLT, 2.70 Å (Faham *et al.* 2008), NM (benzyl-hydantoin transporter MhpI, 2.85 Å (Weyand *et al.* 2008), OG (leucine transporter LeuT, 1.6 Å (Yamashita *et al.* 2005), or OTG (Na+-independent amino acid transporter ApcT, 2.35 Å) (Shaffer *et al.* 2009)'.

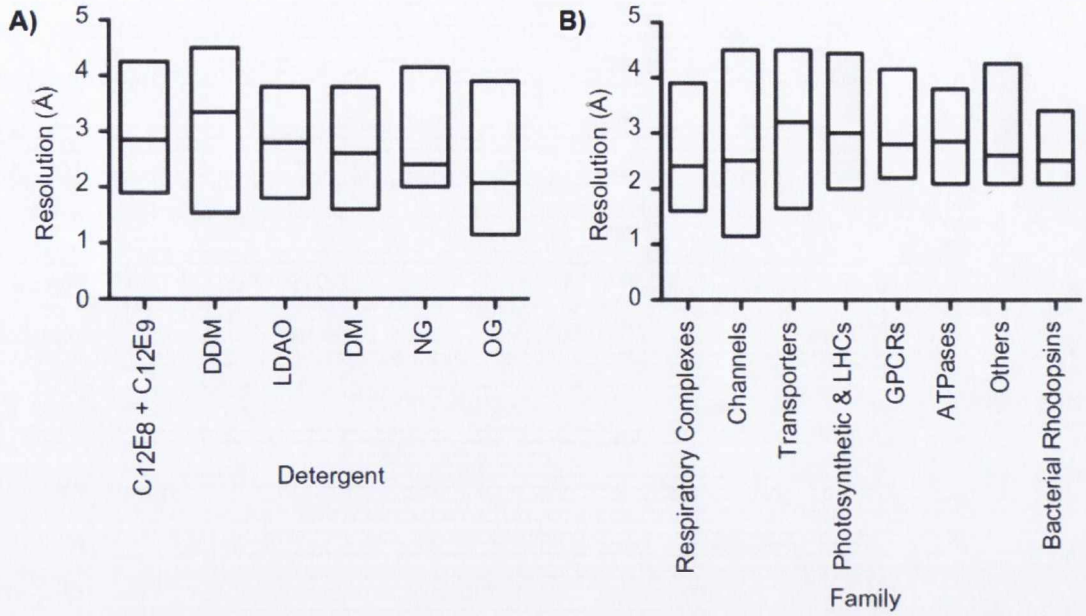


Figure 4.3. "Correlation between structural resolution and the type of crystallization detergents" (Sonoda *et al.* 2010).

A) Relationship between median resolution and detergents used for solving the structures of membrane transporters across all species, DDM has the poorest median resolution (~ 3.3 Å) and **B**) relationship between resolution and different membrane protein family across all species whose structures have been solved (taken from Sonoda *et al.* 2010).

In attempts to improve that diffraction quality of LysP, nanobodies were raised against TEV cleaved GFP free LysP and used for co-crystallization both *in meso* and *in surfo* (Van Bockstaele *et at.* 2009). This was a collaborative project with the group of Prof. Jan Steyaert at Vrije Universitat, Brussels. The purpose of using nanobodies was that they might bind to the protein, makes it more stable and increase the protein-protein contact need for well-ordered crystals to be formed. In this collaboration we supplied them with biotinylated and non biotinylated LysP and they carried out the immunization and screening for clones that binds to LysP. A total of 7 clones were identified by panning as described in **Figure 3.3** (CA5753, CA5755, CA5758, CA5760, CA5761, CA5764 and CA5765) of which only two (CA5755 and CA5764) got bound to LysP purified in LMNG micelle after screening for tight binders using SEC and SDS PAGE. Using TEV cleaved free LysP purified in LMNG in complex with nanobody clone CA5755, crystals were obtained *in surfo* in the presence of 5 mM L-4-thialysine and 5 mM L-lysine, and after several optimizations they could only

diffract to 8 Å. The reason for this poor diffraction could be that the right crystallization condition that will give crystals suitable for structural determination has not yet been identified (Parker & Newstead 2012). Thus, more screening and optimization needs to be carried out. No crystal was obtained with nanobody clone CA5764.

In meso crystallization was carried out with LysP purified in DDM and MNG DDM with or without nanobodies. In this method the target protein was reconstituted in the bilayer of the bicontinuous cubic phase using commercially available lipid (9.9 MAG) and novel lipids synthesized in Caffrey's Lab (7.8 MAG and 7.7 MAG). Crystallization trials were carried using commercially available 96-well screen kits [PACT premier (Molecular Dimensions), Memsys & Memstart (Molecular Dimensions), MemGold (Molecular Dimensions), MbClass (Qiagen), MemFac (Hampton Research), SaltRx (Hampton Research) and Cubic Screen (Emerald BioSystems)]. No crystal was obtained so far in my hands, but however several conditions gave protein precipitates in all three lipids used. These precipitates could be micro crystals, thus, these conditions could be re-screened to see if crystals can be obtained. In the Lab of Prof. Martin Caffrey several membrane proteins have been crystallized by this method of re-screening conditions that gave protein precipitations (Lyons *et al.* 2012). This method mainly varies the PEG concentration above and below the initial PEG concentration and the addition of different concentrations of salt and/or buffers at different pH. Dr. Joseph Lyons a member of Prof. Caffrey's group has designed a 96 conditions screen in which he mixed 20 and 40 %(v/v) PEG 400 with 100 mM and 400 mM salts with the addition of different buffers from pH 6 to 7.5. This screen was recently used in collaboration work with the group of Dr. Simon Newstead at Oxford University for the crystallization of PepT St (a peptide transporter). Several hits have been obtained with this screen (**unpublished observation**). However, on 5th of October 2012, when this thesis is being written, these crystals have not yet been tested for protein diffraction. It is in Dr. Lyons plans to test this screen with other alpha helical membrane proteins for further optimization of the screen. This screen will be used for LysP, to see if crystals can be obtained *in meso*. This was the screen designed by the group of Dr. Brian Kobilka as advised by Dr. Lyons (**personal communication with Dr. Lyons)** and they solve the structure of the β(2)-adrenergic receptor in complex with an agonist and its cognate G protein (Rosenbaum *et al.* 2011). Also, Dr. Lyons used the same screen and solved the structure of the caa3-type cytochrome oxidase (Lyons *et al.* 2012)

After cloning, expression and purification we showed that purified TEV cleaved LysP binds to L-lysine, D-Lysine, L-4-thialysine, L-histidine and L-arginine. This was carried out using MST analysis. MTS analysis revealed that LysP binds L-lysine, D-lysine, L-4-thialysine, L-histidine and L-arginine with K_d of 15 μM, 124 μM, 31 μM, 4.5 mM and 21 mM respectively. No binding was seen for L-leucine, L-isoleucine, L-threonine, L-methionine, glycine, L-valine, L-alanine, L-glutamate, L-glutamine and L-proline. The binding for L-lysine was 8 times stronger than that of D-lysine, which suggests that L-lysine is the preferred substrate for the transporter. However, the binding for L-histidine was 309 times weaker than that of L-lysine while that of L-arginine was 1446 times weaker and since no binding was seen for the other amino acids it suggests that LysP is a lysine specific permease. CD spectroscopy confirmed that LysP is an alpha helical integral membrane protein which is consistent with the bioinformatics analysis. CD revealed 79 % alpha helical content, 21 % random coil and no β strand. In the presence of 5 mM L-lysine there was a change in the CD spectrum of 7.5 % at 205 nm, suggesting a change in the secondary structure of LysP when L-lysine is added to it. Tryptophan fluorescence change revealed a decrease in fluorescence intensity of 10 % upon addition of L-lysine to LysP with respect to LysP without any addition of L-lysine. Possible explanation for this could be that when L-lysine is added to LysP, it quenches the fluorescence of a tryptophan closed to the binding pocket of LysP. From the tryptophan fluorescence study, it is likely that the purified LysP changes conformation upon addition of L-lysine. However, the tryptophan fluorescence change experiment was not pursued further for the determination because the L-lysine used was absorbing light at 280 nm.

MST analysis, Trp fluorescence change and CD, however, do not reveal whether LysP actually transports L-lysine. Thus radioactivity assay of LysP using radiolabelled L-lysine was carried out. From the radioactivity assay, it shows that LysP reconstituted into proteoliposomes transports L-lysine with the transport being inhabited by D-lysine and L-4-thialysine which is consistent with the result obtained from the MST analysis. The transport of L-lysine in LysP reconstituted proteoliposomes was much higher when compared with the liposomes with no LysP in it. Also, the transport of [^{3}H] L-lysine was shown to be Na^+ independent as there was no difference in the transport of [^{3}H] L-lysine when 50 mM NaCl was added to the buffer used for the counterflow assay compared to when no NaCl was added to the counterflow buffer. This experiment tells us that LysP does not use a Na^+ gradient to transports L-lysine. Thus, LysP likely uses

a proton gradient to transports L-lysine as has been shown for ApcT which is a member of the APC transporter family which LysP belong (Shaffer *et al.* 2007).

Finally we carried out whole cell assay to investigate if the cloned LysP-GFP was active and transporting [^{3}H] L-lysine and also to test other possible amino acid substrate for LysP-GFP and thus LysP by whole cell inhibition of [^{3}H] L-lysine. In this study we compared the transport of [^{3}H] L-lysine in induced and uninduced cells. There was an increase in transport of the induced cells compared to the uninduced, suggesting there was transport of [^{3}H] L-lysine. However, we cannot conclude that the transport was due to LysP-GFP since I did not use LysP knock-out *E. coli* strain for the assay. This is because there is also a LysP in *E. coli* that could be transporting [^{3}H]L-Lysine that we are measuring and also with the presence of broad spectrum amino acid transporters and other promiscuous amino acid transporters in *E. coli*, it was unreasonable to conclude that the transport was due to LysP-GFP. The whole cell transport assay was inhibited by D-lysine, L-4-thialysine and cold L-lysine but not with L-histidine, L-arginine, L-glutamate, L-phenylalanine, L-leucine, L-proline, L-asparagine and L-serine. Weaknesses in this assay were: 1. the fact that we used a LysP-GFP construct as the GFP may be affecting the transport of [^{3}H] L-lysine by LysP. 2. A control having a similar protein that does not transport [^{3}H] L-lysine such as LacY-GFP, such that we can have an idea of the magnitude of the background transport and 3. We did not measure the concentration of LysP-GFP being overexpressed to be certain the transport of [^{3}H] L-lysine could be atleast coming from LysP-GFP.

4.2. Future Plans

Future plans will focus on improving the resolution of LysP crystals in order to solve the crystal structure. To achieve this, focus has to be on obtaining a stable protein. This can be achieved by proteolytic truncation of the N and C terminus of LysP. Previous work has been shown on full length GlpG in which the full length protein was very active and difficult to work with. Proteolytic truncation was then carried out to remove the first 100 amino acid to obtain a stable core domain and the structure was solved at 2.1Å (Wang *et al.* 2006). Also, in situ proteolysis could be carried out in order to remove floppy and disordered regions in the protein by adding trace amount of proteases into the crystallization drop (Dong *et al.* 2007 and Wernimont & Edwards 2009). Wernimont & Edwards 2009, carried out in situ proteolysis on 69 membrane proteins that resisted structural determination and were able to crystallize 10 of them and their structures were solved. The success rate for this case was 14.5 % (Wernimont & Edwards 2009).

Also, to reduce conformational flexibility of LysP, mutagenesis studies will be carried out. Nie *et al.* 2008 showed that a mutation (cysteine 154 to glycine) in LacY makes it more stable. Since LacY is related transporter to LysP in having 12 transmembrane segments with a sequence identity of 28.6 % (http://www.ch.embnet.org/software/LALIGN_form.html), it is worth mutating these cysteines to see if a stable construct can be obtained. There are four cysteines (C) in LysP (C82, C358, C380, and C427). C82 is found in the cytoplasm while the rest are embedded in the membrane. C358 and C427 are conserved among homologues of LysP while C82 is replaced by Alanine (A) and C380 replaced by either Threonine (T) or Serine (S). A homology model would be useful to determine whether there is C in LysP conserved in an equivalent region to that of C154 in LacY, which would provide a promising mutagenesis target.

Also, shorter chain detergents will be screened to identify those in which the LysP in complex with the nanobodies will be stable as shorter chain detergents may increase the likelihood of obtaining well ordered structure grade crystals (Sonoda *et al.* 2011). This will be carried out by incubating the LysP/nanobody complex into excess of the respective shorter chain detergents OG, NG, OTG, LDAO and NM and incubate at room temperature overnight and the samples will be analysed on SEC for monodispersity.

LysP from different homologues will be cloned, produced and crystallization trials will be set up. There might be a likelihood of obtaining a stable protein from a different homologue that will produce structure grade crystals. This kind of work was carried out with the glutamate transporter in which three homologues from *Bacillus caldotenax, B. stearothermophilus* and *P. horikoshii* were screened and the protein from *P. horikoshii* was the most stable and the structure was solved (Yernool & Boudker *et al.* 2003 & 2004).

Also, different bicelle concentrations will be screened in bicelle crystallization experiments (Faham & Bowie 2002) to find out if structure grade crystals will be obtained. The bicelle crystal structure of LeuT has been reported (Wang *et al.* 2012), thus the method works for secondary active transporters.

Furthermore, the *in meso* conditions that gave precipitates will be re-screened to see if they can produce crystals and also more crystallization trials will be carried out using the new screen designed by Dr. Lyons described in **Section 4.1**.

Lastly, new crystallization screen kit will be used such as the modified Memgold screen kit (Parker & Newstead 2012) with further 133 conditions added to it to identify new conditions that would be suitable for structure determination (Parker & Newstead 2012). "This screen has crystallization and additive screens that should prove useful for the initial crystallization trials and for subsequent crystal optimization" (Parker & Newstead 2012).

References

Abramson, J., I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, and S. Iwata. (2003). Structure and mechanism of the lactose permease of Escherichia coli. Science **301**:610-615.

Albritton, L. M., L. Tseng, D. Scadden, and J. M. Cunningham. (1989). A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. Cell **57**:659-666.

Asherie, N. (2004). Protein crystallization and phase diagrams. Methods **34**:266-272.

Bickler, S. W., M. C. Heinrich, and G. C. Bagby. (1992). Magnesium-dependent thermostability of DNase I. Biotechniques **13**:64-66.

Blundell, T. L., H. Jhoti, and C. Abell. (2002). High-throughput crystallography for lead discovery in drug design. Nat. Rev. Drug Discov. **1**:45-54.

Boudker, O., R. M. Ryan, D. Yernool, K. Shimamoto, and E. Gouaux. (2007). Coupling substrate and ion binding to extracellular gate of a sodium-dependent aspartate transporter. Nature **445**:387-393.

Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghagen, and J. C. Venter. (1996). Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science **273**:1058-1073.

Caffrey, M. (2000). A lipid's eye view of membrane protein crystallization in mesophases. Curr. Opin. Struct. Biol.**10**:486-497.

Caffrey, M. (2003). Membrane protein crystallization. J. Struct. Biol. **142**:108-132.

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

Caffrey, M., D. Li, and A. Dukkipati. (2012). Membrane protein structure determination using crystallography and lipidic mesophases: recent advances and successes. Biochemistry **51**:6266-6288.

Chae, P. S., S. G. Rasmussen, R. R. Rana, K. Gotfryd, R. Chandra, M. A. Goren, A. C. Kruse, S. Nurva, C. J. Loland, Y. Pierre, D. Drew, J. L. Popot, D. Picot, B. G. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka, and S. H. Gellman. (2010). Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. Nat. Methods **7**:1003-1008.

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 Biol. Crystallogr. **60**:1795-1807.

Cherezov, V., D. M. Rosenbaum, M. A. Hanson, S. G. Rasmussen, F. S. Thian, T. S. Kobilka, H. J. Choi, P. Kuhn, W. I. Weis, B. K. Kobilka, and R. C. Stevens. (2007). High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor. Science **318**:1258-1265.

Cherezov, V., E. Yamashita, W. Liu, M. Zhalnina, W. A. Cramer, and M. Caffrey. (2006). In meso structure of the cobalamin transporter, BtuB, at 1.95 A resolution. J. Mol. Biol. **364**:716-734.

Claros, M. G. and G. von Heijne. (1994). TopPred II: an improved software for membrane protein structure predictions. Comput. Appl. Biosci. **10**:685-686.

Costerton, J. W., P. S. Stewart, and E. P. Greenberg. (1999). Bacterial biofilms: a common cause of persistent infections. Science **284**:1318-1322.

Daniels, D. L., G. Plunkett, 3rd, V. Burland, and F. R. Blattner. (1992). Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes. Science **257**:771-778.

Davies, J. and D. Davies. (2010). Origins and evolution of antibiotic resistance. Microbiol Mol. Biol. Rev. **74**:417-433.

De Genst, E., K. Silence, K. Decanniere, K. Conrath, R. Loris, J. Kinne, S. Muyldermans, and L. Wyns. (2006). Molecular basis for the preferential cleft recognition by dromedary heavy-chain antibodies. Proc. Natl. Acad. Sci. U S A **103**:4586-4591.

Dong, A., X. Xu, A. M. Edwards, C. Chang, M. Chruszcz, M. Cuff, M. Cymborowski, R. Di Leo, O. Egorova, E. Evdokimova, E. Filippova, J. Gu, J. Guthrie, A. Ignatchenko, A. Joachimiak, N. Klostermann, Y. Kim, Y. Korniyenko, W. Minor, Q. Que, A. Savchenko, T. Skarina, K. Tan, A. Yakunin, A. Yee, V. Yim, R. Zhang, H. Zheng, M. Akutsu, C. Arrowsmith, G. V. Avvakumov, A. Bochkarev, L. G. Dahlgren, S. Dhe-Paganon, S. Dimov, L. Dombrovski, P. Finerty, Jr., S. Flodin, A. Flores, S. Graslund, M. Hammerstrom, M. D. Herman, B. S. Hong, R. Hui, I. Johansson, Y. Liu, M. Nilsson, L. Nedyalkova, P. Nordlund, T. Nyman, J. Min, H. Ouyang, H. W. Park, C. Qi, W. Rabeh, L. Shen, Y. Shen, D. Sukumard, W. Tempel, Y. Tong, L. Tresagues, M. Vedadi, J. R. Walker, J. Weigelt, M. Welin, H. Wu, T. Xiao, H. Zeng, and H. Zhu. (2007).

In situ proteolysis for protein crystallization and structure determination. Nat. Methods **4**:1019-1021.

Drew, D., M. Lerch, E. Kunji, D. J. Slotboom, and J. W. de Gier. (2006). Optimization of membrane protein overexpression and purification using GFP fusions. Nat. Methods **3**:303-313.

Drew, D., S. Newstead, Y. Sonoda, H. Kim, G. von Heijne, and S. Iwata. (2008). GFP-based optimization scheme for the overexpression and purification of eukaryotic membrane proteins in Saccharomyces cerevisiae. Nat. Protoc. **3**:784-798.

Drew, D., D. J. Slotboom, G. Friso, T. Reda, P. Genevaux, M. Rapp, N. M. Meindl-Beinker, W. Lambert, M. Lerch, D. O. Daley, K. J. Van Wijk, J. Hirst, E. Kunji, and J. W. De Gier. (2005). A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. Protein Sci. **14**:2011-2017.

Drew, D. E., G. von Heijne, P. Nordlund, and J. W. de Gier. (2001). Green fluorescent protein as an indicator to monitor membrane protein overexpression in Escherichia coli. FEBS Lett. **507**:220-224.

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.

Fang, Y., L. Kolmakova-Partensky, and C. Miller. (2007). A bacterial arginine-agmatine exchange transporter involved in extreme acid resistance. J. Biol. Chem. **282**:176-182.

Forrest, L. R., R. Kramer, and C. Ziegler. (2011). The structural basis of secondary active transport mechanisms. Biochim. Biophys. Acta. **1807**:167-188.

Friedmann, D., T. Messick, and R. Marmorstein. (2011). Crystallization of macromolecules. Curr. Protoc. Protein Sci. **Chapter 17**:Unit17 14.

Fucile, S., E. Palma, A. Martinez-Torres, R. Miledi, and F. Eusebi. (2002). The single-channel properties of human acetylcholine alpha 7 receptors are altered by fusing alpha 7 to the green fluorescent protein. Proc. Natl. Acad. Sci. U S A **99**:3956-3961.

Gao, X., F. Lu, L. Zhou, S. Dang, L. Sun, X. Li, J. Wang, and Y. Shi. (2009). Structure and mechanism of an amino acid antiporter. Science **324**:1565-1568.

Giacomini, K. M., S. M. Huang, D. J. Tweedie, L. Z. Benet, K. L. Brouwer, X. Chu, A. Dahlin, R. Evers, V. Fischer, K. M. Hillgren, K. A. Hoffmaster, T. Ishikawa, D. Keppler, R. B. Kim, C. A. Lee, M. Niemi, J. W. Polli, Y. Sugiyama, P. W.

Swaan, J. A. Ware, S. H. Wright, S. W. Yee, M. J. Zamek-Gliszczynski, and L. Zhang. (2010). Membrane transporters in drug development. Nat. Rev. Drug. Discov. **9**:215-236.

Gordeliy, V. I., J. Labahn, R. Moukhametzianov, R. Efremov, J. Granzin, R. Schlesinger, G. Buldt, T. Savopol, A. J. Scheidig, J. P. Klare, and M. Engelhard. (2002). Molecular basis of transmembrane signalling by sensory rhodopsin II-transducer complex. Nature **419**:484-487.

Guan, L., I. N. Smirnova, G. Verner, S. Nagamori, and H. R. Kaback. (2006). Manipulating phospholipids for crystallization of a membrane transport protein. Proc. Natl. Acad. Sci. U S A **103**:1723-1726.

Haga, K., A. C. Kruse, H. Asada, T. Yurugi-Kobayashi, M. Shiroishi, C. Zhang, W. I. Weis, T. Okada, B. K. Kobilka, T. Haga, and T. Kobayashi. (2012). Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. Nature **482**:547-551.

Henderson, P. J. and A. J. Macpherson. (1986). Assay, genetics, proteins, and reconstitution of proton-linked galactose, arabinose, and xylose transport systems of Escherichia coli. Methods Enzymol. **125**:387-429.

Heuberger, E. H., L. M. Veenhoff, R. H. Duurkens, R. H. Friesen, and B. Poolman. (2002). Oligomeric state of membrane transport proteins analyzed with blue native electrophoresis and analytical ultracentrifugation. J Mol Biol **317**:591-600.

Hofer, N., D. Aragao, and M. Caffrey. (2010). Crystallizing transmembrane peptides in lipidic mesophases. Biophys. J. **99**:L23-25.

Hoffmann, W. (1985). Molecular characterization of the CAN1 locus in Saccharomyces cerevisiae. A transmembrane protein without N-terminal hydrophobic signal sequence. J. Biol. Chem. **260**:11831-11837.

Hoiby, N., T. Bjarnsholt, M. Givskov, S. Molin, and O. Ciofu. (2010). Antibiotic resistance of bacterial biofilms. Int. J. Antimicrob. Agents **35**:322-332.

Honore, N. and S. T. Cole. (1990). Nucleotide sequence of the aroP gene encoding the general aromatic amino acid transport protein of Escherichia coli K-12: homology with yeast transport proteins. Nucleic Acids Res. **18**:653.

Hsieh, J. M., G. M. Besserer, M. G. Madej, H. Q. Bui, S. Kwon, and J. Abramson. (2010). Bridging the gap: a GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. Protein Sci **19**:868-880.

Hunte, C., E. Screpanti, M. Venturi, A. Rimon, E. Padan, and H. Michel. (2005). Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. Nature **435**:1197-1202.

Jauniaux, J. C. and M. Grenson. (1990). GAP1, the general amino acid permease gene of Saccharomyces cerevisiae. Nucleotide sequence, protein similarity with the other bakers yeast amino acid permeases, and nitrogen catabolite repression. Eur. J. Biochem. **190**:39-44.

Jeanthon, C., S. L'Haridon, A. L. Reysenbach, M. Vernet, P. Messner, U. B. Sleytr, and D. Prieur. (1998). Methanococcus infernus sp. nov., a novel hyperthermophilic lithotrophic methanogen isolated from a deep-sea hydrothermal vent. Int. J. Syst. Bacteriol. **48 Pt. 3**:913-919.

Jerabek-Willemsen, M., C. J. Wienken, D. Braun, P. Baaske, and S. Duhr. (2011). Molecular interaction studies using microscale thermophoresis. Assay Drug Dev. Technol. **9**:342-353.

Kawate, T. and E. Gouaux. (2006). Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. Structure **14**:673-681.

Khoury, G. A., R. C. Baliban, and C. A. Floudas. (2011). Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. Sci. Rep. **1**.

Krogh, A., B. Larsson, G. von Heijne, and E. L. Sonnhammer. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305**:567-580.

Kubala, M. H., O. Kovtun, K. Alexandrov, and B. M. Collins. (2010). Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. Protein Sci. **19**:2389-2401.

Kuhn, P., K. Wilson, M. G. Patch, and R. C. Stevens. (2002). The genesis of high-throughput structure-based drug discovery using protein crystallography. Curr. Opin. Chem. Biol. **6**:704-710.

Kunji, E. R., M. Harding, P. J. Butler, and P. Akamine. (2008). Determination of the molecular mass and dimensions of membrane proteins by size exclusion chromatography. Methods **46**:62-72.

Kuo, A., M. W. Bowler, J. Zimmer, J. F. Antcliff, and D. A. Doyle. (2003). Increasing the diffraction limit and internal order of a membrane protein crystal by dehydration. J Struct. Biol. **141**:97-102.

Lewenza, S., J. L. Gardy, F. S. Brinkman, and R. E. Hancock. (2005). Genome-wide identification of Pseudomonas aeruginosa exported proteins using a consensus computational strategy combined with a laboratory-based PhoA fusion screen. Genome Res. **15**:321-329.

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. J. Vis. Exp.

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

Luecke, H., B. Schobert, H. T. Richter, J. P. Cartailler, and J. K. Lanyi. (1999). Structure of bacteriorhodopsin at 1.55 A resolution. J. Mol. Biol. **291**:899-911.

Lyons, J. A., D. Aragao, O. Slattery, A. V. Pisliakov, T. Soulimane, and M. Caffrey. (2012). Structural insights into electron transfer in caa3-type cytochrome oxidase. Nature **487**:514-518.

MacLeod, C. L., K. Finley, D. Kakuda, C. A. Kozak, and M. F. Wilkinson. (1990). Activated T cells express a novel gene on chromosome 8 that is closely related to the murine ecotropic retroviral receptor. Mol. Cell. Biol. **10**:3663-3674.

Mao, L., Y. Tang, S. T. Vaiphei, T. Shimazu, S. G. Kim, R. Mani, E. Fakhoury, E. White, G. T. Montelione, and M. Inouye. (2009). Production of membrane proteins for NMR studies using the condensed single protein (cSPP) production system. J. Struct. Funct. Genomics **10**:281-289.

Mirza, S. P., B. D. Halligan, A. S. Greene, and M. Olivier. (2007). Improved method for the analysis of membrane proteins by mass spectrometry. Physiol. Genomics **30**:89-94.

Morris, M. C., J. Depollier, J. Mery, F. Heitz, and G. Divita. (2001). A peptide carrier for the delivery of biologically active proteins into mammalian cells. Nat. Biotechnol. **19**:1173-1176.

Mueckler, M., C. Caruso, S. A. Baldwin, M. Panico, I. Blench, H. R. Morris, W. J. Allard, G. E. Lienhard, and H. F. Lodish. (1985). Sequence and structure of a human glucose transporter. Science **229**:941-945.

Muyldermans, S. (2001). Single domain camel antibodies: current status. J. Biotechnol. **74**:277-302.

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

Nie, Y., F. E. Sabetfard, and H. R. Kaback. (2008). The Cys154-->Gly mutation in LacY causes constitutive opening of the hydrophilic periplasmic pathway. J. Mol. Biol. **379**:695-703.

Niegemann, E., A. Schulz, and K. Bartsch. (1993). Molecular organization of the Escherichia coli gab cluster: nucleotide sequence of the structural genes gabD and gabP and expression of the GABA permease gene. Arch. Microbiol. **160**:454-460.

Parker, J. L. and S. Newstead. (2012). Current trends in alpha-helical membrane protein crystallization: An update. Protein Sci. **21**:1358-1365.

Pi, J., P. J. Wookey, and A. J. Pittard. (1991). Cloning and sequencing of the pheP gene, which encodes the phenylalanine-specific transport system of Escherichia coli. J. Bacteriol. **173**:3622-3629.

Pi, J., P. J. Wookey, and A. J. Pittard. (1993). Site-directed mutagenesis reveals the importance of conserved charged residues for the transport activity of the PheP permease of Escherichia coli. J. Bacteriol. **175**:7500-7504.

Prilusky, J., C. E. Felder, T. Zeev-Ben-Mordehai, E. H. Rydberg, O. Man, J. S. Beckmann, I. Silman, and J. L. Sussman. (2005). FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. Bioinformatics **21**:3435-3438.

Racher, K. I., R. T. Voegele, E. V. Marshall, D. E. Culham, J. M. Wood, H. Jung, M. Bacon, M. T. Cairns, S. M. Ferguson, W. J. Liang, P. J. Henderson, G. White, and F. R. Hallett. (1999). Purification and reconstitution of an osmosensor: transporter ProP of Escherichia coli senses and responds to osmotic shifts. Biochemistry **38**:1676-1684.

Raman, P., V. Cherezov, and M. Caffrey. (2006). The Membrane Protein Data Bank. Cell Mol. Life Sci. **63**:36-51.

Reizer, J., K. Finley, D. Kakuda, C. L. MacLeod, A. Reizer, and M. H. Saier, Jr. (1993). Mammalian integral membrane receptors are homologous to facilitators and antiporters of yeast, fungi, and eubacteria. Protein Sci. **2**:20-30.

Rosen, B. P. (1971). Basic amino acid transport in Escherichia coli. J. Biol. Chem. **246**:3653-3662.

Rosen, B. P. (1973). Basic amino acid transport in Escherichia coli: properties of canavanine-resistant mutants. J. Bacteriol. **116**:627-635.

Rosenbaum, D. M., C. Zhang, J. A. Lyons, R. Holl, D. Aragao, D. H. Arlow, S. G. Rasmussen, H. J. Choi, B. T. Devree, R. K. Sunahara, P. S. Chae, S. H.

Gellman, R. O. Dror, D. E. Shaw, W. I. Weis, M. Caffrey, P. Gmeiner, and B. K. Kobilka. (2011). Structure and function of an irreversible agonist-beta(2) adrenoceptor complex. Nature **469**:236-240.

Shaffer, P. L., A. Goehring, A. Shankaranarayanan, and E. Gouaux. (2009). Structure and mechanism of a Na+-independent amino acid transporter. Science **325**:1010-1014.

Shaw, A. Z. and B. Miroux. (2003). A general approach for heterologous membrane protein expression in Escherichia coli: the uncoupling protein, UCP1, as an example. Methods Mol. Biol. **228**:23-35.

Sobczak, I. and J. S. Lolkema. (2005). Structural and mechanistic diversity of secondary transporters. Curr. Opin. Microbiol. **8**:161-167.

Solcan, N., J. Kwok, P. W. Fowler, A. D. Cameron, D. Drew, S. Iwata, and S. Newstead. (2012). Alternating access mechanism in the POT family of oligopeptide transporters. EMBO J. **31**:3411-3421.

Sonnhammer, E. L., G. von Heijne, and A. Krogh. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6**:175-182.

Sonoda, Y., A. Cameron, S. Newstead, H. Omote, Y. Moriyama, M. Kasahara, S. Iwata, and D. Drew. (2010). Tricks of the trade used to accelerate high-resolution structure determination of membrane proteins. FEBS Lett. **584**:2539-2547.

Sonoda, Y., S. Newstead, N. J. Hu, Y. Alguel, E. Nji, K. Beis, S. Yashiro, C. Lee, J. Leung, A. D. Cameron, B. Byrne, S. Iwata, and D. Drew. (2011). Benchmarking membrane protein detergent stability for improving throughput of high-resolution X-ray structures. Structure **19**:17-25.

Steffes, C., J. Ellis, J. Wu, and B. P. Rosen. (1992). The lysP gene encodes the lysine-specific permease. J. Bacteriol. **174**:3242-3249.

Stover, C. K., X. Q. Pham, A. L. Erwin, S. D. Mizoguchi, P. Warrener, M. J. Hickey, F. S. Brinkman, W. O. Hufnagle, D. J. Kowalik, M. Lagrou, R. L. Garber, L. Goltry, E. Tolentino, S. Westbrock-Wadman, Y. Yuan, L. L. Brody, S. N. Coulter, K. R. Folger, A. Kas, K. Larbig, R. Lim, K. Smith, D. Spencer, G. K. Wong, Z. Wu, I. T. Paulsen, J. Reizer, M. H. Saier, R. E. Hancock, S. Lory, and M. V. Olson. (2000). Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. Nature **406**:959-964.

Takanishi, Y., M. Toshimitsu, M. Nakata, N. Takada, T. Izumi, K. Ishikawa, H. Takezoe, J. Watanabe, Y. Takahashi, and A. Iida. (2006). Frustrated smectic layer structures in bent-shaped dimer liquid crystals studied by x-ray microbeam diffraction. Phys. Rev. E. Stat. Nonlin. Soft. Matter. Phys. **74**:051703.

Tanaka, H. and K. Soda. (1974). Enzymatic N-acetylation of lysine analogs. Purification and properties of acetyl coenzyme A:S-(beta-aminoethyl)-L-cysteine omega-N-acetyltransferase. J. Biol. Chem. **249**:5285-5289.

Tanaka, J. and G. R. Fink. (1985). The histidine permease gene (HIP1) of Saccharomyces cerevisiae. Gene **38**:205-214.

Thomson, R., T. C. Hodgman, Z. R. Yang, and A. K. Doyle. (2003). Characterizing proteolytic cleavage site activity using bio-basis function neural networks. Bioinformatics **19**:1741-1747.

Vandenbol, M., J. C. Jauniaux, and M. Grenson. (1989). Nucleotide sequence of the Saccharomyces cerevisiae PUT4 proline-permease-encoding gene: similarities between CAN1, HIP1 and PUT4 permeases. Gene **83**:153-159.

Vincke, C. and S. Muyldermans. (2012). Introduction to heavy chain antibodies and derived nanobodies. Methods Mol. Biol. **911**:15-26.

Wagner, S., M. M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K. J. van Wijk, D. J. Slotboom, J. O. Persson, and J. W. de Gier. (2008). Tuning Escherichia coli for membrane protein overexpression. Proc. Natl. Acad. Sci. U S A **105**:14371-14376.

Wallin, E. and G. von Heijne. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci. **7**:1029-1038.

Wang, H., J. Elferich, and E. Gouaux. (2012). Structures of LeuT in bicelles define conformation and substrate binding in a membrane-like context. Nat. Struct. Mol. Biol. **19**:212-219.

Wang, Y., Y. Zhang, and Y. Ha. (2006). Crystal structure of a rhomboid family intramembrane protease. Nature **444**:179-180.

Warne, T., M. J. Serrano-Vega, C. G. Tate, and G. F. Schertler. (2009). Development and crystallization of a minimal thermostabilised G protein-coupled receptor. Protein Expr. Purif. **65**:204-213.

Wernimont, A. and A. Edwards. (2009). In situ proteolysis to generate crystals for structure determination: an update. PLoS. One **4**:e5094.

West, I. C. (1970). Lactose transport coupled to proton movements in Escherichia coli. Biochem. Biophys. Res. Commun. **41**:655-661.

Wright, E. M., D. D. Loo, and B. A. Hirayama. (2011). Biology of human sodium glucose transporters. Physiol. Rev. **91**:733-794.

Winsor, G. L., D. K. Lam, L. Fleming, R. Lo, M. D. Whiteside, N. Y. Yu, R. E. Hancock, and F. S. Brinkman. (2011). Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes. Nucleic Acids Res. **39**:D596-600.

Woof, J. M. and D. R. Burton. (2004). Human antibody-Fc receptor interactions illuminated by crystal structures. Nat. Rev. Immunol. **4**:89-99.

Wienken, C. J., P. Baaske, U. Rothbauer, D. Braun, and S. Duhr. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. Nat. Commun. **1**:100.

Wu, C. C. and J. R. Yates, 3rd. (2003). The application of mass spectrometry to membrane proteomics. Nat. Biotechnol. **21**:262-267.

Yamashita, A., S. K. Singh, T. Kawate, Y. Jin, and E. Gouaux. (2005). Crystal structure of a bacterial homologue of Na+/Cl--dependent neurotransmitter transporters. Nature **437**:215-223.

Yang, Z. R., R. Thomson, P. McNeil, and R. M. Esnouf. (2005). RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. Bioinformatics **21**:3369-3376.

Yernool, D., O. Boudker, E. Folta-Stogniew, and E. Gouaux. (2003). Trimeric subunit stoichiometry of the glutamate transporters from Bacillus caldotenax and Bacillus stearothermophilus. Biochemistry **42**:12981-12988.

Yernool, D., O. Boudker, Y. Jin, and E. Gouaux. (2004). Structure of a glutamate transporter homologue from Pyrococcus horikoshii. Nature **431**:811-818.

Yoshimoto, T., E. Yoshimoto, and D. Meruelo. (1991). Molecular cloning and characterization of a novel human gene homologous to the murine ecotropic retroviral receptor. Virology **185**:10-17.