



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Neves, Angela T F

Title:

On the Structure and Function of LicB, a Secondary Transporter from Human Pathogens

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode> This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

On the Structure and Function of LicB, a Secondary Transporter from Human Pathogens



University of
BRISTOL

Angela Neves

A Dissertation Submitted to the University of Bristol in Accordance with the Requirements for
Award of the degree of Doctor of Philosophy in the Faculty of Life Sciences

School of Biochemistry

June 2021

Word Count: 61,499

Abstract

The transport of solutes across biological membranes is a fundamental process in nature. This thesis concerns a novel membrane transport protein known as LicB that is essential for the persistence and virulence of pathogenic bacteria in the human respiratory tract. LicB catalyses the uptake of choline from the host, which is an essential first step in a short metabolic pathway that attaches choline to the lipopolysaccharide. LicB is a putative high-affinity choline transporter with no significant sequence homology or functional similarity to other choline permeases and is a potential drug target. However, the structure and function of LicB remain unknown.

Here, LicB was successfully overexpressed in *Escherichia coli*, solubilised in dodecyl-maltoside detergent and purified by affinity chromatography. Size exclusion chromatography, as well as blue native PAGE, showed that LicB could be purified as a monomer, although some nonspecific oligomerisation occurred. Purified LicB was experimentally determined to contain ten transmembrane α -helices, in agreement with bioinformatic predictions. A novel fluorescence assay for ligand binding *in vitro* was introduced that allows the first studies of the ligand binding affinity and specificity of LicB. LicB was selective in binding to choline and acetylcholine over related small molecules, confirming previous reports on substrate specificity. Finally, choline uptake assays were established after LicB was successfully reconstituted in synthetic proteoliposomes. This work thus provides the first experimental characterisation of the Choline Uptake Transporter (LicB-T) protein family.

A parallel project was focused on further optimisation of the expression and purification of a class of peripheral membrane-bound proteins, the acyl-CoA:alcohol *O*-acyltransferases, here exemplified by the *Saccharomyces cerevisiae* Atf1 and specific enzymes from papaya, melon and strawberry fruit. Efforts to express, purify and characterise these proteins are described in the final chapter of the thesis.

Acknowledgements

Firstly, a huge thank you to my supervisor, Dr Paul Curnow, for his support during my PhD and all his guidance and encouraging words. Similarly, I would like to thank Professor Paul Race for his support during the rotation project. I also wish to thank to my progression panel, Professor Ian Collinson and Professor Mark Dillingham, for their discussions and valuable insights into the project.

I must say a huge thank you to Dr Virginie Dufour, for training me in the lab and for regular troubleshooting. Thanks also to current lab members, Ben Hardy and Sam Williams, and the undergraduate and MSc students who helped along the way. Thank you to all the members of A100 and C101 who have, over the years, become amazing friends and made my PhD the fun and enjoyable experience that it was! To Alice, Carl and Laurence for the advice in the lab!

Lastly, I would also like to thank my family for all of their support, interest and patience over these last four years. Particularly a special thank you to Tiago for his endless support and always believing in me.

Declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's *Regulations and Code of Practice for Research Degree Programmes* and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED:..... DATE:.....

Table of contents

List of abbreviations	xv
List of figures	xix
List of tables	xxv
Thesis outline	xxix
Chapter 1. Introduction to membrane proteins	1
1.1. Biological membranes.....	1
1.1.1. General properties of biological membranes	1
1.1.2. Main characteristics of biological lipids.....	3
1.2. Membrane proteins.....	4
1.2.1. Peripheral, integral or amphitropic membrane proteins	5
1.2.2. Transporters as a vital class of membrane proteins	6
1.3. Structure of transmembrane proteins	9
1.3.1. Main characteristics of α -helical proteins	10
1.3.2. Tertiary structure of transmembrane proteins.....	12
1.4. From transmembrane features to prediction models	13
1.5. Experimental limitations on the study of membrane proteins.....	15
1.6. Membrane mimetics.....	16
1.6.1. Detergents	18
1.6.2. Protein-based reconstitution systems	28
1.6.3. Synthetic polymer-based reconstitution systems.....	30
1.6.4. Peptide-based reconstitution systems.....	32
PART I - Studies of the secondary transporter LicB	34
1.7. Colonisation of the human respiratory tract.....	34
1.8. Types of respiratory tract infections	35
1.8.1. <i>H. influenzae</i> strains in RTIs – colonisation and disease.....	37
1.9. The outer cell wall of Gram-negative bacteria	38
1.10. Virulence factors in colonising bacteria.....	39
1.10.1. Structural differences between LPS and LOS.....	40
1.10.2. Main LOS modifications.....	41

1.11.	Phosphorylcholine is an important and widespread virulence factor	42
1.11.1.	Phosphorylcholine supports cell adhesion and affects host recognition	43
1.11.2.	Host defenses against phosphorylcholine expressing bacteria	45
1.11.3.	Current and future perspectives on ChoP based vaccines	46
1.12.	The biosynthesis of phosphorylcholine modified structures	47
1.12.1.	Phosphodiesterases involved in choline uptake	49
1.12.2.	Enzymes of the <i>lic1</i> operon	49
1.13.	LicB and its relationship with other choline transporters in bacteria	51
1.13.1.	The BCCT family	51
1.13.2.	The uptake of choline by LicB in <i>H. influenzae</i>	53
1.13.3.	The LicB transporter	54
1.14.	Project aims	55
Chapter 2.	Materials and methods	57
2.1.	Specific laboratory equipment	57
2.2.	Chemicals, reagents and solutions	58
2.3.	Enzymes	60
2.4.	Preparation and constituents of buffers	60
2.5.	Bioinformatic analysis	61
2.5.1.	Computer software	61
2.5.2.	Secondary structure and transmembrane topology prediction	62
2.5.3.	Homology modelling	62
2.5.4.	Identification of LicB sequence homologues and conserved motifs	64
2.6.	General microbiology techniques	64
2.6.1.	<i>E. coli</i> strains and vectors	64
2.6.2.	Media and plates	65
2.6.3.	Antibiotics	65
2.7.	General DNA techniques	66
2.7.1.	Cloning of <i>licB</i> gene into pET28 vector	66
2.7.2.	Site-directed mutagenesis	67
2.7.3.	PCR product transformation	68
2.7.4.	Plasmid miniprep	68
2.7.5.	DNA sequencing	69
2.7.6.	DNA transformation into chemically competent cells	69
2.7.7.	Agarose gel electrophoresis	70
2.8.	Protein expression in <i>E. coli</i> and protein purification	70
2.8.1.	Large scale <i>E. coli</i> expression	70
2.8.2.	Isolation of membranes	71
2.8.3.	Membrane fraction solubilisation	71
2.8.4.	Purification of recombinant protein	72
2.9.	Total protein concentration determination	72

2.10.	Biophysical techniques	73
2.10.1.	Polyacrylamide gel electrophoresis (PAGE)	73
2.10.2.	Direct immunoblot ('Dot Blot')	74
2.10.3.	MALDI mass spectrometry of LicB protein	75
2.10.4.	Size exclusion chromatography	75
2.10.5.	Size exclusion chromatography coupled to multi angle light scattering	76
2.10.6.	Circular dichroism spectroscopy	78
2.11.	Ligand binding assays	79
2.11.1.	Isothermal titration calorimetry	79
2.11.2.	Fluorescence resonance energy transfer	79
2.12.	Transport assay	81
2.12.1.	Liposome formation	81
2.12.2.	DDM-mediated reconstitution	81
2.12.3.	'Flotation' assay	82
2.12.4.	Pyranine transport assay	82
Chapter 3. Recombinant expression and initial biophysical characterisation of the choline permease LicB		83
3.1.	Introduction and aim	83
3.1.1.	Special considerations for the recombinant expression of membrane proteins	84
3.1.2.	Special considerations in applying biophysical techniques to membrane proteins	85
3.2.	Results	87
3.2.1.	Bioinformatic analysis of LicB	87
3.2.2.	Expression and purification of LicB in <i>E. coli</i>	95
3.2.3.	Biophysical characterisation of purified recombinant LicB	105
3.3.	Discussion	110
3.3.1.	Expression and purification results	110
3.3.2.	Bioinformatic and biophysical characterisation	111
3.4.	Conclusion	112
Chapter 4. Functional characterisation of LicB		113
4.1.	Introduction and aim	113
4.1.1.	Protein-ligand interactions	113
4.1.2.	Fluorescence Resonance Energy Transfer	115
4.1.3.	Isothermal Titration Calorimetry	117
4.1.4.	Transport energetics	119
4.2.	Results	120
4.2.1.	LicB ligand binding assays	120
4.2.2.	Studying ligand binding through ITC	129
4.2.3.	Proteoliposome transport assay	131
4.3.	Discussion	137

4.4.	Conclusion.....	142
Chapter 5.	Probing the function of LicB by site-directed mutagenesis	143
5.1.	Using mutagenesis to characterise LicB	143
5.1.1.	Site-directed mutagenesis.....	146
5.2.	Results.....	147
5.2.1.	Identifying conserved residues through multiple sequence alignment	147
5.2.2.	Expression and purification of LicB variants	153
5.2.3.	Structural characterisation of LicB variants	156
5.2.4.	Competition assay	159
5.3.	Discussion	161
5.4.	Conclusion	164
Chapter 6.	Summary and future work	165
PART II - The study of acyl-CoA:alcohol O-acyltransferases from yeast and fruit		167
Chapter 7.	Literature review.....	167
7.1.	Volatile esters in the aroma of food and drink	167
7.2.	General considerations on AATases	170
7.3.	Ester production in yeast <i>S. cerevisiae</i>	170
7.3.1.	Ester biosynthesis pathways.....	172
7.3.2.	AATs genes - <i>atf1</i> and <i>atf2</i>	173
7.3.3.	Atf1, Atf2 and Lg Atf1 proteins – homology and cellular localisation.....	176
7.3.4.	<i>In vitro</i> studies on Atf1 and Atf2 enzymes	178
7.4.	Ester production in fruits	180
7.4.1.	Ester biosynthesis pathways in fruits	180
7.4.2.	Main characteristics of fruit AATases	181
7.4.3.	Cellular localisation and conserved motifs in fruit AATases.....	182
7.4.4.	<i>In vivo</i> studies using fruit crude extract	185
7.4.5.	<i>In vitro</i> experiments with plant AATases	185
7.4.6.	<i>In silico</i> studies on plant AATases.....	186
7.5.	Perspectives on the industrial use of AATases	187
7.6.	Aims of this chapter	187
Chapter 8.	Materials and methods	189
8.1.	Chemicals, reagents and solutions	189
8.2.	Enzymes.....	190

8.3.	Prepared buffers	190
8.4.	General microbiology techniques	191
8.4.1.	Strains and expression vectors	191
8.4.2.	Media and agar plates	192
8.4.3.	Antibiotics	192
8.5.	Bioinformatic analysis	193
8.6.	General DNA techniques	193
8.6.1.	Polymerase chain reaction.....	193
8.6.2.	Cloning of modified <i>atf1</i> gene constructs into pYES2 vector	194
8.6.3.	Cloning of <i>atf1</i> gene into pET28 vector	195
8.6.4.	Synthetic genes of AATases.....	196
8.6.5.	Cloning of AATase genes in pYES2 vector	197
8.6.6.	Cloning of AATase genes in pOPIN vectors.....	197
8.6.7.	PCR product transformation.....	199
8.6.8.	Plasmid miniprep and DNA sequencing.....	199
8.6.9.	DNA transformation into chemically competent cells.....	200
8.7.	General protein expression and purification protocols.....	200
8.7.1.	Expression in <i>E. coli</i>	201
8.7.2.	Protein expression in <i>S. cerevisiae</i>	202
8.7.3.	Isolation of membranes.....	203
8.7.4.	Detergent screening assay	203
8.7.5.	Membrane fraction solubilisation.....	204
8.7.6.	Purification of recombinant protein.....	204
8.8.	Total protein concentration determination	204
8.9.	Biophysical techniques	205
8.9.1.	Western blotting.....	205
8.9.2.	Size exclusion chromatography	206
8.9.3.	SEC-MALS and circular dichroism	206
8.10.	Enzymatic activity assays.....	206
8.10.1.	Acyltransferase activity.....	206
8.10.2.	Thioesterase assay	206
Chapter 9.	Expression and characterisation of AATases	207
9.1.	Engineering yeast alcohol acetyltransferase Atf1 for structural biology	207
9.1.1.	Introduction and aim	207
9.1.2.	Results.....	209
9.1.3.	Discussion	218
9.2.	Detergent screen of Atf1 expressed in <i>S. cerevisiae</i>	219
9.2.1.	Introduction and aim	219
9.2.2.	Results.....	219
9.2.3.	Discussion	225
9.3.	Heterologous expression of fruit AATases in <i>S. cerevisiae</i>	227
9.3.1.	Introduction and aim	227
9.3.2.	Results.....	227

9.3.3. Discussion	233
9.4. Using pOPIN vectors to optimise protein expression.....	234
9.4.1. Introduction and aim	234
9.4.2. Results.....	234
9.4.3. Discussion	250
Chapter 10. Conclusions on alcohol acyltransferase work	253
Chapter 11. Overall conclusions	255
Chapter 12. References	257

List of abbreviations

Abbreviation	Definition
AAT	Alcohol acetyltransferase
AATase	Acyl-CoA:alcohol <i>O</i> -acyltransferase
ADP	Adenosine diphosphate
AEAT	Acyl-CoA:ethanol <i>O</i> -acyltransferase
AIM	Auto-induction media
ANN	Artificial neural network
ATP	Adenosine triphosphate
BCCT	Betaine-carnitine-choline transporter
BN	Blue native
CB	Column buffer
CD	Circular dichroism
ChoP	Phosphorylcholine
CMC	Critical micelle concentration
CMT	Critical micellar temperature
CoA	Coenzyme A
COPD	Chronic obstructive pulmonary disease
CRP	C-reactive protein
DBN	Dynamic bayesian networks
DC	Dansyl choline
DDM	<i>n</i> -dodecyl- β -D-maltoside
DM	<i>n</i> -decyl- β -D-maltoside
DME	Drug/metabolite exporter family
DMT	Drug/metabolite transporter superfamily
ER	Endoplasmic reticulum
FC-14	Fos-choline 14
FRET	Fluorescence resonance energy transfer
GC-MS	Gas chromatography-mass spectrometry
GFP	Green fluorescent protein
GPC	Glycerophosphorylcholine

GPCR	G-protein coupled receptor
HMM	Hidden markov model
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β -D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
KDO	Phosphorylated 2-keto-3-deoxyoctulosonic acid
LB	Lysogeny broth
LOS	Lipooligosaccharide
LPN	Lipid-protein nanodisc
LPS	Lipopolysaccharide
LRTI	Lower respiratory tract infection
LTA	Lipoteichoic acid
MALS	Multi-angle light scattering
MCFA	Medium-chain fatty acid
MP	Membrane protein
MRE	Mean residue ellipticity
MSP	Membrane scaffold protein
MW	Molecular weight
NMR	Nuclear magnetic resonance
OD ₆₀₀	Optical density at 600 nm
OM	Otitis media
PA	Phosphatidic acid
PAF	Platelet-activating factor
PAGE	Polyacrylamide gel electrophoresis
PC	Phosphatidylcholine
PCR	Polymerase chain reaction
PDB	Protein data bank
PDC	Protein-detergent complex
PE	Phosphatidylethanolamine
PGN	Peptidoglycan
PS	Phosphatidylserine
rPAF	Receptor for platelet-activating factor
RT	Respiratory tract
RTI	Respiratory tract infection
SB	Super broth
SDS	Sodium dodecyl sulphate

SEC	Size exclusion chromatography
SMA	Styrene-maleic acid
SMALP	Styrene-maleic acid copolymer-lipid particle
SP	Signal peptide
SVM	Support vector machine
TA	Teichoic acids
TCDB	Transporter classification database
T _m	Melting temperature
TM	Transmembrane
UFA	Unsaturated fatty acids
URTI	Upper respiratory tract infection
WB	Western blot
WT	Wild-type

Microbial species abbreviations

<i>E. coli</i>	<i>Escherichia coli</i>
<i>H. influenzae</i>	<i>Haemophilus influenzae</i>
<i>M. catarrhalis</i>	<i>Moraxella catarrhalis</i>
NTHi	Non-typeable <i>Haemophilus influenzae</i>
<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
<i>S. aureus</i>	<i>Staphylococcus aureus</i>
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
<i>S. pneumoniae</i>	<i>Streptococcus pneumoniae</i>

List of figures

Figure 1.1. Representation of different membrane models.	2
Figure 1.2. Examples of common glycerophospholipids.	3
Figure 1.3. Common lipidic assembly models in water.	4
Figure 1.4. Types of membrane proteins.	5
Figure 1.5. Overview of the different types of active transport found in nature.	7
Figure 1.6. Representation of the main types of membrane proteins.	10
Figure 1.7. Representation of two types of transmembrane protein.	11
Figure 1.8. Distribution of protein structure and amino acid residues.	12
Figure 1.9. Exemplification of a protein hydrophathy index, plotted against residue number.	13
Figure 1.10. Representation of the membrane proteins in different membrane mimetics.	17
Figure 1.11. Representation of the different phases encountered during solubilisation of membrane protein as a function of the unbound detergent.	20
Figure 1.12. Schematic representation of the Krafft point.	21
Figure 1.13. Percentage (%) of studies in which the detergent is used for solubilisation and crystallisation.	24
Figure 1.14. New classes of detergents.	26
Figure 1.15. Overview of the different approaches used for reassembly of membrane proteins into liposomes.	28
Figure 1.16. Schematic representation of membrane proteins in MSP or saposin-A forming nanodiscs or lipid nanoparticles, respectively.	29
Figure 1.17. Chemical structure of SMA polymer at neutral pH.	30
Figure 1.18. Chemical structure of the A8-35 amphipol.	31
Figure 1.19. Different molecular structures of peptergents at neutral pH.	33
Figure 1.20. Main stages of colonisation in the human respiratory tract.	35
Figure 1.21. Bacteria cell wall structures.	39
Figure 1.22. General structures of LPS.	40
Figure 1.23. Loci responsible for LOS phase variation in NTHi <i>H. influenzae</i> strain.	41
Figure 1.24. Schematic representation of the role of ChoP in <i>H. influenzae</i>	44
Figure 1.25. Schematic representation of the recognition of ChoP by CRP and the complement activation molecule C1q.	46
Figure 1.26. The biosynthetic pathway for ChoP.	48

Figure 1.27. Betaine binding site in BetP transporter.....	52
Figure 1.28. The uptake of [¹⁴ C]-choline into <i>H. influenzae</i> strain was dominated by LicB... 53	53
Figure 2.1. General workflow of homology modelling.	63
Figure 2.2. Size exclusion chromatography calibration curve.	75
Figure 3.1. Schematic representation of the critical steps from expression to characterisation of LicB.....	84
Figure 3.2. Representation of (A) IPTG induction and (B) auto-induction (by using a combination of glucose, lactose and glycerol in the media) for expression of LicB protein. .	85
Figure 3.3. Differences in the spectra of proteins with different secondary structures.....	87
Figure 3.4. Full protein and DNA sequences of LicB expressed in <i>E. coli</i>	88
Figure 3.5. Topology diagram of LicB drawn using TOPO2 based on TOPCONS prediction.	91
Figure 3.6. Hydrophobicity analysis of LicB sequence.....	91
Figure 3.7. Quality parameters of the LicB homology model.	93
Figure 3.8. Comparison between YddG and LicB transporters.....	94
Figure 3.9. Schematic representation of purification protocol.	96
Figure 3.10. Coomassie-stained SDS-PAGE gels demonstrating the successful expression and purification of recombinant LicB after IMAC and, subsequent SEC chromatograms.....	98
Figure 3.11. Expression of LicB protein in T7 host strain (SB _{AIM}) and purification after solubilisation with DDM.	101
Figure 3.12. Expression of LicB protein in different host <i>E. coli</i> strains and culture media.	102
Figure 3.13. SEC profile of LicB protein purified from diverse expression conditions (A-D). (E) % of Recovery of LicB in the different conditions mentioned.	104
Figure 3.14. Circular dichroism to assess the secondary structure and long-term stability of LicB.....	106
Figure 3.15. Full melting scans of LicB in different detergent micelle between temperatures of 25°C and 95°C.....	107
Figure 3.16. SEC-MALS analysis of protein-detergent complex (PDC) of LicB in DDM.....	108
Figure 3.17. Blue Native-PAGE gel of purified LicB samples confirming the presence of aggregates.....	109
Figure 4.1. Schematic representation of the binding of the fluorescent derivative dansyl choline to LicB.....	117
Figure 4.2. Main parts of the ITC instrument.	118
Figure 4.3. Fluorescence emission plotted against wavelength.....	122
Figure 4.4. Experimental fluorimetry demonstrating the FRET phenomenon.....	123
Figure 4.5. Control experiments using increasing concentrations of either choline or the corresponding volume of buffer.	126

Figure 4.6. Competition experiments with different small molecules.....	127
Figure 4.7. Isothermal titration calorimetry of LicB.....	130
Figure 4.8. Normalised absorbance at 540 nm (A ₅₄₀) of either the lipid mixture before extruding or after extrusion (liposomes) with increasing percentage of DDM.	132
Figure 4.9. Reconstitution of LicB into proteoliposomes.....	133
Figure 4.10. Confirmation of pyranine inside liposomes and proteoliposomes.....	133
Figure 4.11. Schematic representation of the experimental setup.	135
Figure 4.12. Representative pyranine fluorescence traces produced from the proteoliposome transport assays.....	136
Figure 4.13. Representative pyranine fluorescence traces produced when no chemical gradient (Δ pH) was applied.	137
Figure 4.14. Part of the multiple sequence alignment between BCCTs and LicB protein...	141
Figure 5.1. Multiple sequence alignment between Choline Uptake Transporter Family members.....	144
Figure 5.2. Multiple sequence alignment of LicB in selected opportunistic pathogens of the human nasopharynx and the corresponding phylogenetic tree.....	146
Figure 5.3. Multiple sequence alignment of LicB from <i>H. influenzae</i> against sequences selected from BLAST search.....	150
Figure 5.4. Figure highlighting conserved residues between LicB homologues.....	151
Figure 5.5. Phylogenetic tree based on the multiple sequence alignment.	152
Figure 5.6. LicB homology model generated in Chapter 3.....	152
Figure 5.7. Purification of LicB variants.	154
Figure 5.8. Size exclusion profiles of LicB variants solubilised in 1% DDM.	155
Figure 5.9. Circular dichroism spectroscopy of LicB variants in 1% DDM.....	157
Figure 5.10. Thermal melts of LicB variants in 1% DDM.	158
Figure 5.11. Summary of biophysical properties of LicB variants.	159
Figure 5.12. Competition assays of each LicB variant.....	160
Figure 5.13. Part of the multiple sequence alignment between LicB homologues coloured by a degree of conservation.....	161
Figure 5.14. Illustration of the possible polar interactions performed by mutated residues.	163
Figure 7.1. List of the main esters and alcohol compounds in (A) fruits and (B) beers.....	168
Figure 7.2. List of some esters produced by apple, strawberry and melon fruits.....	169
Figure 7.3. General overview of the synthesis of esters by alcohol acyltransferases.....	170
Figure 7.4. Representation of the two major enzyme families in <i>S. cerevisiae</i> responsible for the synthesis of volatile esters.	171
Figure 7.5. Main pathways involved in the formation of esters.	173

Figure 7.6. Schematic pathway regulation of <i>atf1</i> and <i>ole1</i> genes in aerobic conditions and in presence of UFAs.	175
Figure 7.7. Cellular localisation of co-expressing Atf1-GFP (in green) and the lipid particle marker Erg6-DsRed (in red).	177
Figure 7.8. Cellular localisation of Atf2 protein.	177
Figure 7.9. Biochemical assays of Atf1 protein performed by Nancolas <i>et al.</i> (2017) ³⁶⁸	179
Figure 7.10. Phylogenic tree analysis of fruit AATases.	182
Figure 7.11. Vinorine synthase crystal structure (PDB entry 2BGH).	183
Figure 8.1. Representation of a part of the galactose-inducible pYES2 yeast expression plasmid cloned with the <i>atf1</i> gene.	195
Figure 8.2. Representation of the strategy employed to clone <i>atf1</i> , <i>VpAAT1</i> , <i>FaAAT1</i> and <i>MAAT</i> genes into pOPIN vectors.	198
Figure 9.1. Secondary structures of the N- and C-terminal hydrophobic domains of Atf.	208
Figure 9.2. Representation of the project workflow and methodology.	209
Figure 9.3. Multiple sequence alignment of WT and truncated variants of Atf1.	210
Figure 9.4. Expression and purification of Atf1 in Thesit.	212
Figure 9.5. Truncated Atf1 proteins were still localised in membrane fraction.	213
Figure 9.6. Size exclusion chromatogram of Atf1 variants.	215
Figure 9.7. Coupled reaction used for acyltransferase activity.	215
Figure 9.8. Coupled enzyme assay for Atf1 WT and Atf1 Δ N/ Δ C activities.	216
Figure 9.9. <i>E. coli</i> does not express Atf1 truncations.	217
Figure 9.10. Detergent screening assay of Atf1 protein in 15 different detergents.	219
Figure 9.11. Large scale expression of Atf1 in Cymal-5 and FC-14.	221
Figure 9.12. SEC-MALS analysis of the Atf1-FC-14 protein-detergent complex.	222
Figure 9.13. Circular dichroism spectrum of purified Atf1.	223
Figure 9.14. Thioesterase and biochemical acyltransferase activity assay of Atf1 in FC-14.	224
Figure 9.15. Protein sequences of FaAAT1, VpAAT1 and MAAT used in this work.	229
Figure 9.16. Expression of MAAT, VpAAT1 and FaAAT1 in <i>S. cerevisiae</i>	230
Figure 9.17. IMAC purification of cellular fractions of FaAAT1 and MAAT proteins using 2% Thesit.	232
Figure 9.18. Protein sequence of non-fusion Atf1 and fused with MBP and NusA protein tags.	235
Figure 9.19. Protein sequence of non-fusion FaAAT1 and fused with MBP and NusA protein tags.	236
Figure 9.20. Protein sequence of non-fusion MAAT and fused with MBP and NusA protein tags.	237

Figure 9.21. Protein sequence of non-fusion VpAAT1 and fused with MBP and NusA protein tags.....	238
Figure 9.22. Small expression of non-fusion Atf1 or fused with protein tags.....	240
Figure 9.23. Large scale expression of Atf1 and MBP-Atf1.	241
Figure 9.24. Detergent screening of Atf1 expressed in <i>E. coli</i> BL21(DE3) cells cloned in pOPINF and growth in LB _{AIM}	242
Figure 9.25. Coomassie-stained SDS-PAGE gels showing non-fusion FaAAT1 and MAAT expressed in the insoluble fraction.	243
Figure 9.26. Small scale expression of non-fusion VpAAT1 or fused with protein tags.....	244
Figure 9.27. Small scale expression of non-fusion MAAT or fused with protein tags.	245
Figure 9.28. Small expression of FaAAT1 fused with protein tags in LB cultures induced with IPTG.	247
Figure 9.29. Small scale expression trials of FaAAT1 fused with protein tags.	248
Figure 9.30. MBP-FaAAT1 was localised to sedimenting membranes.	248
Figure 9.31. Large scale expression and purification of FaAAT1 fused with MBP solubilised with 1% DDM.	249

List of tables

Table 1.1. List of some superfamilies present in TCDB.	8
Table 1.2. List of popular methods for the bioinformatic prediction of membrane protein topology and their main characteristics.	14
Table 1.3. Different membrane mimetic systems used for membrane protein solubilisation.	17
Table 1.4. Types of detergents.	23
Table 1.5. Rankings of the top 15 leading causes of death in 2002 and the predicted causes for 2030.	36
Table 1.6. Main bacterial species present in otitis media, tonsillitis, sinusitis and pneumoniae.	37
Table 1.7. Bacteria that express ChoP modified structures.	42
Table 1.8. Examples of BCCT family proteins.	51
Table 1.9. Kinetic parameters determined from transport assays in cell strains.	54
Table 2.1. Laboratory equipment used.	57
Table 2.2. List of chemicals, reagents and commercial kits.	58
Table 2.3. List of enzymes used.	60
Table 2.4. Buffers used in the thesis, ordered alphabetically by colloquial name.	60
Table 2.5. List of servers and the correspondent links used for TM topology prediction.	62
Table 2.6. List of online resources used for template-based homology.	63
Table 2.7. List of <i>E. coli</i> strains and vectors used in LicB project.	64
Table 2.8. Media recipes and general preparation.	65
Table 2.9. List of antibiotics used.	65
Table 2.10. Primer sequences for the cloning of <i>licB</i> gene in pET28 vector.	67
Table 2.11. List of the forward (F) and reverse (R) primers used for site-directed mutagenesis of LicB, and the correspondent annealing temperature set in PCR reactions.	68
Table 2.12. Heat shock conditions used for each cell line transformation.	69
Table 2.13. Conditions used for LicB expression.	71
Table 2.14. List of chemicals used for membrane solubilisation and subsequent protein purification.	72
Table 2.15. Theoretical molecular weight (Da) and extinction coefficient (ϵ) for each protein studied based on Expasy ProtParam online tool.	73

Table 2.16. Values of the molecular weight (kDa) and Stokes radius (nm) for each calibrant and the correspondent elution volume.	76
Table 3.1. Output of bioinformatic analyses of the primary structure of LicB.	90
Table 3.2. Best scored templates from SWISS-Model and HHpred.....	92
Table 3.3. Yields of recombinant LicB obtained using DDM, DM, Cymal-5 and SMA2000 as solubilising chemicals.....	97
Table 3.4. Calculated MW_{app} PDC and LicB based on K_{av} values.....	98
Table 3.5. LicB expression conditions and final purification yields after affinity and size exclusion chromatography.	100
Table 3.6. Top 5 results from the BLAST search on the peptide sequence -SLSGLQVYFLR-	100
Table 4.1. Experimental determined binding constants for LicB.	123
Table 4.2. Results of competition assays.	128
Table 4.3. Summary of choline concentrations tested and the experiment outcome.	129
Table 4.4. Gradients associated with the outside-to-inside direction for the movement of a proton.	134
Table 5.1. Members of the Choline Uptake Transporter Family (TCDB 2.A.7.18).....	143
Table 5.2. List of the 30 sequences used for the multiple sequence alignment.	148
Table 7.1. Molecular characteristics of known AATs genes.....	174
Table 7.2. Characterisation of BAHD acyltransferases from different commercial fruits - papaya, melon, strawberry, apple and kiwi.....	184
Table 8.1. List of reagents and commercial kits used.	189
Table 8.2. List of enzymes used in the AATases project.	190
Table 8.3. List of prepared buffers.....	190
Table 8.4. <i>E. coli</i> strains used in AATase project.	191
Table 8.5. <i>S. cerevisiae</i> strain used in AATase project.....	191
Table 8.6. Expression vectors used in AATase project.....	191
Table 8.7. Media and general preparation.....	192
Table 8.8. List of antibiotics used.	192
Table 8.9. Typical PCR reaction mix composition.	193
Table 8.10. Typical PCR amplification conditions.....	194
Table 8.11. DNA template, name and sequence of reverse and forward primers used for PCR amplification of the entire plasmid containing the <i>atf1</i> gene.....	195
Table 8.12. Sequence of the forward (F) and reverse (R) primers used to amplify <i>atf1</i> gene.	196
Table 8.13. Sequence of forward (F) and reverse (R) primers used for alcohol acyltransferase gene amplification by PCR.....	197

Table 8.14. Sequence of forward (F) and reverse (R) primers used for gene cloning into pOPIN vectors.	198
Table 8.15. Features of the pOPIN vectors.	199
Table 8.16. List of primers used for sequencing service.	200
Table 8.17. List of detergents used for detergent screening assay and the correspondent critical micelle concentration (CMC) ^{105,130,134}	203
Table 8.18. List of detergents used to solubilise <i>E. coli</i> and <i>S. cerevisiae</i> membrane fractions solubilisation.	204
Table 8.19. Theoretical molecular weight and extinction coefficient (ϵ) for each alcohol acyltransferase protein calculated based on ExPASy ProtParam online tool.	205
Table 9.1 Purification yield for the Atf1 truncated proteins after IMAC.	214
Table 9.2. Yields obtained from Atf1 purification using Cymal-5 and FC-14 as solubilising agents.	220
Table 9.3. Crystallisation group of AATase from papaya (VpAAT1), cultivated strawberry (FaAAT1), kiwi (AcAT16 and AcAT9), apple (MpAAT1) and melon (MAAT).	228
Table 9.4. Results of AATase expression in <i>S. cerevisiae</i>	231
Table 9.5. Screening expression constructs against cell media and expression conditions for each protein of interest (FaAAT1, MAAT, VpAAT1 and Atf1).	239

Thesis outline

The PhD project aims to develop reliable and robust methodologies to study two specific membrane proteins, one peripheral and one integral, and to use these methods to generate new knowledge on their structure and function. Because this naturally carries a strong element of method development, the experimental limitations on the study of membrane proteins and the use of detergents as their main solubilising agents will be introduced in the first chapter.

Part I.

The central part of the thesis will be focused on a previously unstudied membrane transporter, the high-affinity choline permease LicB. This transporter plays a crucial role in the virulence and persistence of several 'commensal pathogens' in the human respiratory tract. Part I comprises chapters 1 to 6:

- Chapter 1 gives a general introduction to membrane protein studies, then provides a comprehensive review of phosphorylcholine (ChoP) role during respiratory infection and the relevance of LicB in ChoP expression.
- Chapter 2 describes the materials and methods used in this work.
- Chapter 3 presents results from the expression of LicB in different conditions, such as cell strains and culture media, and subsequent solubilisation and purification. Moreover, the biophysical characterisation of LicB is also included in this chapter.
- Chapter 4 comprises different functional analyses of LicB. A novel FRET-based method is introduced for direct measurements of ligand and competition assays. This chapter also describes efforts to establish substrate transport assays in reconstituted proteoliposomes.
- Chapter 5 analysis how mutations in the *licB* gene can affect recombinant protein expression, purification, secondary structure conformation and ligand binding.
- Chapter 6 presents a summary of the findings and future directions of this work.

Part II.

The second part of this thesis is focused on a family of peripheral membrane proteins found in yeast and plants. These are alcohol acetyltransferases typified by the *S. cerevisiae* enzyme Atf1. These proteins are involved in the formation of volatile esters, contributing to the aroma of fruits and the taste of yeast-fermented beverages.

Part II comprises chapters 7 to 9. A suitable overview of these proteins' contribution to volatile ester synthesis and a critical review of the prior literature on these enzymes will be provided in Chapter 7. Materials and methods specific to their study are described in Chapter 8. Results corresponding to the study of alcohol acetyltransferases are described in Chapter 9.

A conclusion on the study of membrane proteins is provided in the final chapter.

Chapter 1. Introduction to membrane proteins

1.1. Biological membranes

Biological membranes are fundamental to life. As well as defining the external boundary of every living cell, lipid membranes are essential for the compartmentalisation of subcellular organelles. Membranes allow the “generation of difference” embodied by electrochemical and solute gradients, organise and regulate enzyme activities, and even supply substrates for biosynthesis and signalling molecules ¹. Intracellular membranes can be highly diverse, with specialised functions such as insulation in the myelin sheath surrounding nerve axons, nutrient absorption in the intestines, or protein degradation in the lysosome ².

1.1.1. General properties of biological membranes

All biomembranes consist of a rich mixture of lipid molecules with which proteins and carbohydrates may be associated or covalently linked. This structure is generally referred to as a phospholipid bilayer since phospholipids are the major constituents; however other components, such as glycolipids and sterols, can also be present ^{3,4}. The protein to lipid ratio, as well as the percentage of each component, differ depending on the membrane, indicating the variety of biological functions ^{5,6}. Furthermore, many biological membranes are asymmetric, with the inner and outer leaflets differing in their lipid and protein composition ⁷.

Membranes are highly dynamic structures, and the concept of biological membranes as a ‘fluid mosaic’ remains a valuable model ⁴. Lateral diffusion of membrane components in the bilayer plane allows for short-lived and long-lived protein-protein and protein-lipid interactions important to membrane function ¹. However, it is increasingly evident that many membranes have *lipid rafts*, in which the concentrations of cholesterol and sphingolipids are elevated ^{8,9}. Rafts float effortlessly through the membrane and are more closely packed and organised than the adjacent bilayer. Particular signalling and trafficking events are allowed by these specialised membrane microdomains ^{3,10,11}. Hence membranes are more like a ‘patchwork quilt’ with specialised microdomains and are generally more crowded with protein than is suggested by standard textbook figures (e.g., Figure 1.1) ¹².

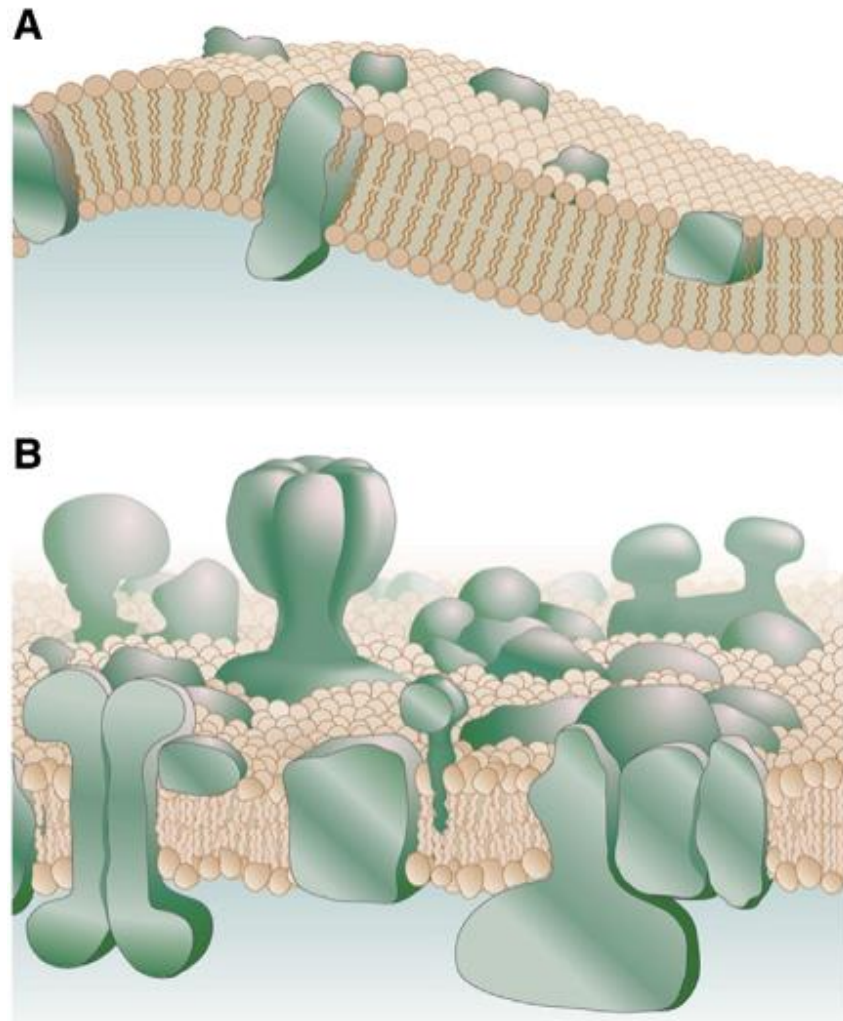


Figure 1.1. Representation of different membrane models.

(A) An interpretation of the original Fluid-Mosaic model, as proposed by Singer–Nicolson ⁴. **(B)** An updated version that shows more significant protein crowding, the widespread formation of protein complexes, and local membrane distortions that can accommodate proteins. Adapted from Engelman (2005) ¹².

One key responsibility of biological membranes is their selective permeability, which prevents the passage of charged and polar biomolecules. This allows the maintenance of electrochemical gradients across the membrane. Membrane permeability is defined by both the lipid and protein components, with the selective transport of hydrophilic compounds and ions enabled by protein channels or transporters ⁵. These latter solute transporters are a central focus of this thesis.

1.1.2. Main characteristics of biological lipids

Biological lipids can show considerable variation, influencing interactions inside the lipid bilayer and physical properties such as fluidity, permeability and elasticity ⁶. For example, the most common lipids – the glycerophospholipids – can vary the length and saturation of the fatty acid ‘tails’. Derivatives of the phosphorylcholine headgroup, which determines the surface interactions of the membrane, include choline, serine or ethanolamine, leading to lipids denoted phosphatidylcholine (PC), phosphatidylserine (PS) and phosphatidylethanolamine (PE), respectively ^{4,13}. These headgroups cause the lipids to be zwitterionic (such as PC or PE) or negatively charged (PS or phosphatidic acid (PA)) and can potentially engage in electrostatic interactions (Figure 1.2).

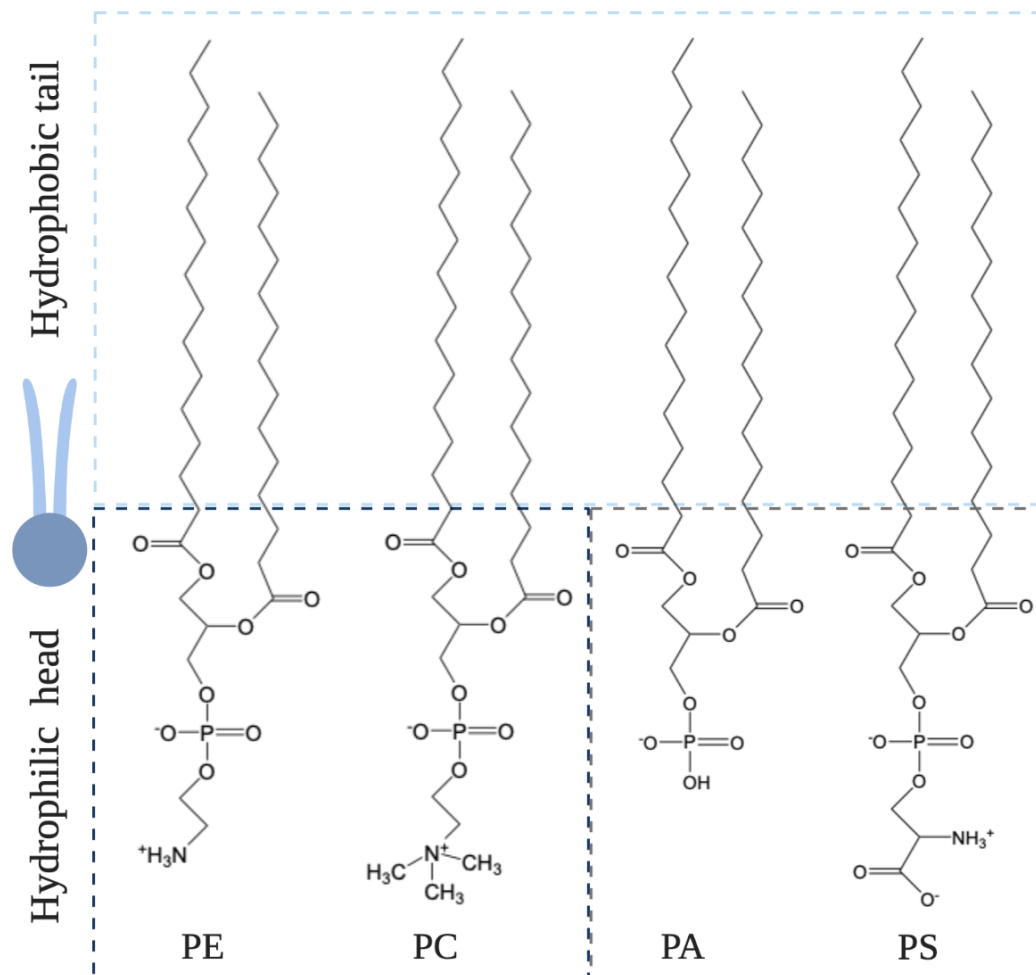


Figure 1.2. Examples of common glycerophospholipids.

While sharing the same basic skeleton, substitutions to the headgroup can generate zwitterionic (PE and PC) and negatively charged (PA and PS) phospholipids. Lipid chain length and saturation can also be varied (not shown). Adapted from Whited and Johs (2015) ¹³.

All biological lipids are amphipathic, with the general structure of a hydrophilic 'head' and two hydrophobic fatty acid 'tails'. This amphipathicity, and the broadly cylindrical shape of the lipid molecule, mean that lipids aggregate and spontaneously self-organise into bilayers and vesicles in water (Figure 1.3B and C). Within the biological cell, lipid bilayers can adopt a number of structures, depending on the lipid/water ratio and external conditions ¹⁴. Also relevant to this thesis is the organisation of lipid-like surfactants into micelles (Figure 1.3A), which is preferred because these surfactants have only a single hydrophobic 'tail' (alkyl chain). This gives detergents a conical shape that favours micelle assembly.

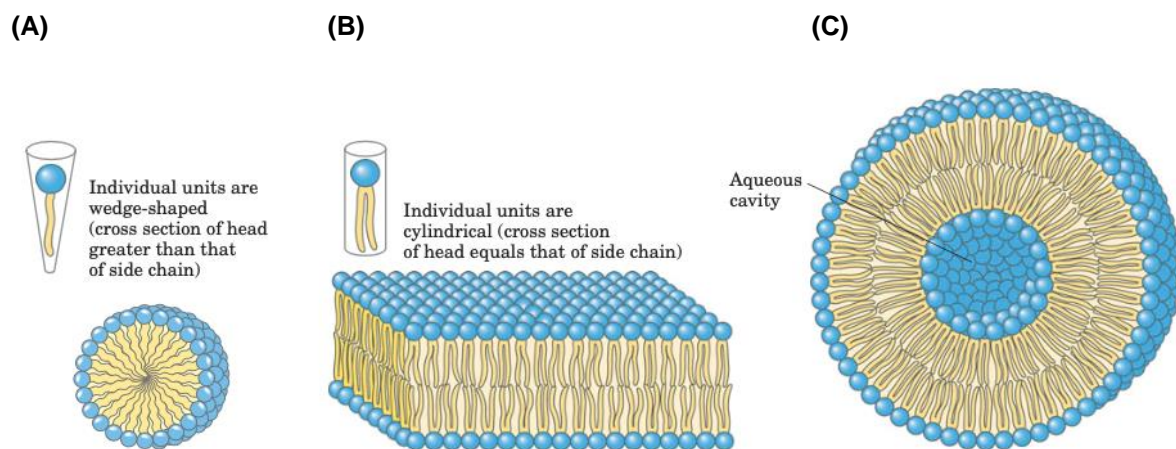


Figure 1.3. Common lipidic assembly models in water.

(A) Micelle, (B) bilayer and (C) vesicle are shown. Hydrophilic head groups of lipids are shown as blue spheres, while hydrophobic tails are shown in yellow. Adapted from Lehninger Principles of Biochemistry (2008) ⁵.

1.2. Membrane proteins

Many biological membranes are densely packed with protein, making up around 50% of the membrane by weight ¹. Based on hydrophobicity analysis, between 20% to 30% of all open reading frames in sequenced genomes are suggested to be membrane proteins (MPs) ^{15,16}. This abundance represents the ubiquitous role of membrane proteins in vital biological functions, including solute and ion transfer (transporters and channels), energy generation (ATP synthases), osmotic regulation (porins, water channels), signal transduction (receptors), enzymatic reactions, and more ¹⁷. This means that malfunctional integral membrane proteins are implicated in the development of serious illnesses, including neurological and cardiac diseases ^{18,19}, cystic fibrosis ²⁰, or inherited disorders ²¹. Because of these essential biological

functions, membrane proteins account for the vast majority of current drug targets ²². Additionally, there is an increasing interest in exploiting membrane proteins in the synthetic biology field ^{23–26}, specifically in the design of ‘*de novo*’ membrane proteins that could be included in living biosystems ^{27–29}.

Membrane proteins are broadly classified into four types based on how they are attached to the membrane: peripheral, amphitropic, lipid-linked or integral (Figure 1.4) ⁵.

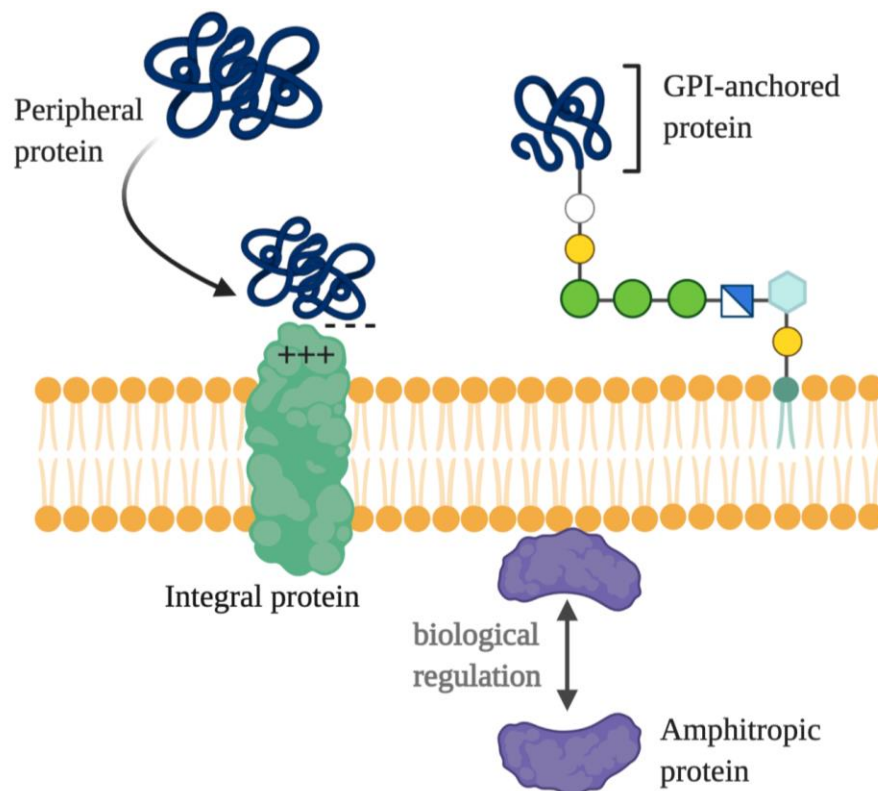


Figure 1.4. Types of membrane proteins.

Integral, peripheral, amphitropic and lipid-linked proteins are shown. Adapted from Lehninger Principles of Biochemistry (2008) ⁵.

1.2.1. Peripheral, integral or amphitropic membrane proteins

Although there are exceptions, in general *peripheral* (extrinsic) membrane proteins do not interact with the membrane's hydrophobic centre. In contrast, they bind with the hydrophilic portions of integral membrane proteins or the polar domains of membrane lipids ¹³. These associations are mediated through electrostatic interactions or hydrogen bonding and can

easily be disrupted by the addition of urea or carbonate, removal of Ca^{2+} by a chelating agent and variations in pH ⁵. As a result, extracting peripheral membrane proteins from the membrane surface is a relatively simple process. Examples of peripheral membrane proteins include cytoskeleton proteins ³⁰, cationic toxins like charybdotoxin ³¹, or electron carriers, for instance, cytochrome C, which binds to integral membrane proteins ³².

In contrast, *integral* (intrinsic) membrane proteins have at least one or more domains that pass entirely through the membrane. These transmembrane domains are intimately associated with the bilayer and cannot be extracted without disrupting the membrane (Figure 1.4) ^{5,17}. They interact closely with nearby lipids as well as other proteins through extensive non-polar interactions ³³. Since integral membrane proteins are deeply embedded within the membrane bilayer's hydrophobic core, the transmembrane segments are strongly biased towards amino acids with hydrophobic side chains, with charged and polar groups disfavoured ³⁴.

The anchored proteins are covalently linked to lipids that retain the protein at the membrane level (Figure 1.4). Long-chain fatty acids, sterols, isoprenoids, or glycosylated derivatives of phosphatidylinositol (GPI) are examples of these linked lipids ¹³. In general, linked proteins anchor asymmetrically in the membrane; for instance, GPI anchors exclusively on the plasma membrane's outer face ⁵.

Lastly, amphitropic proteins can be found either in the cytosol or associated with the membrane (Figure 1.4). Amphitropic proteins' reversible interaction with the membrane surface is generally regulated; for instance, a conformational shift in the protein, due to ligand binding, may expose a previously inaccessible membrane-binding site ³⁵. Some examples of amphitropic membrane proteins are Src kinase ³⁶, protein kinase C or phospholipase C ³⁷.

The remainder of this review will focus on the integral membrane proteins.

1.2.2. Transporters as a vital class of membrane proteins

Transporters are a vital class of membrane proteins that are involved in the active or passive movement of compounds through the cell membrane. They perform crucial biological functions such as nutrient uptake, metabolites and toxic compounds efflux, and neurotransmitters elimination from the synaptic cleft ³⁸.

1.2.2.1. Transporters – classifications and main characteristics

The movement of materials across membranes can be classified into either passive or active transport ⁵. Passive transport primarily concerns integral membrane proteins that act as pores and channels. For this reason, passive transport is also known as ‘facilitated diffusion’, with the pore or channel providing a conduit for molecules to move across the bilayer down their concentration gradient - often referred to as ‘downhill’ transport. As implied by the name, passive transport does not consume energy ^{5,7}. Active transport also involves integral membrane proteins, and these carriers can move solutes against their concentration gradient - often referred to as ‘uphill’ transport. Proteins that carry out active transport can be subdivided into primary active transporters and secondary active transporters, depending on their energy source (Figure 1.5) ^{39,40}. Primary active transporters use chemical energy from the hydrolysis of adenosine triphosphate (ATP) or light ⁴¹. Conversely, secondary active transporters couple the energy of electrochemical ion gradients to the transport of another substrate. This movement can be in the same or opposite directions, typically termed as symport or antiport, respectively. The focus of this thesis is a secondary transporter, known as LicB, responsible for the ‘uphill’ active transport of the small molecule choline.

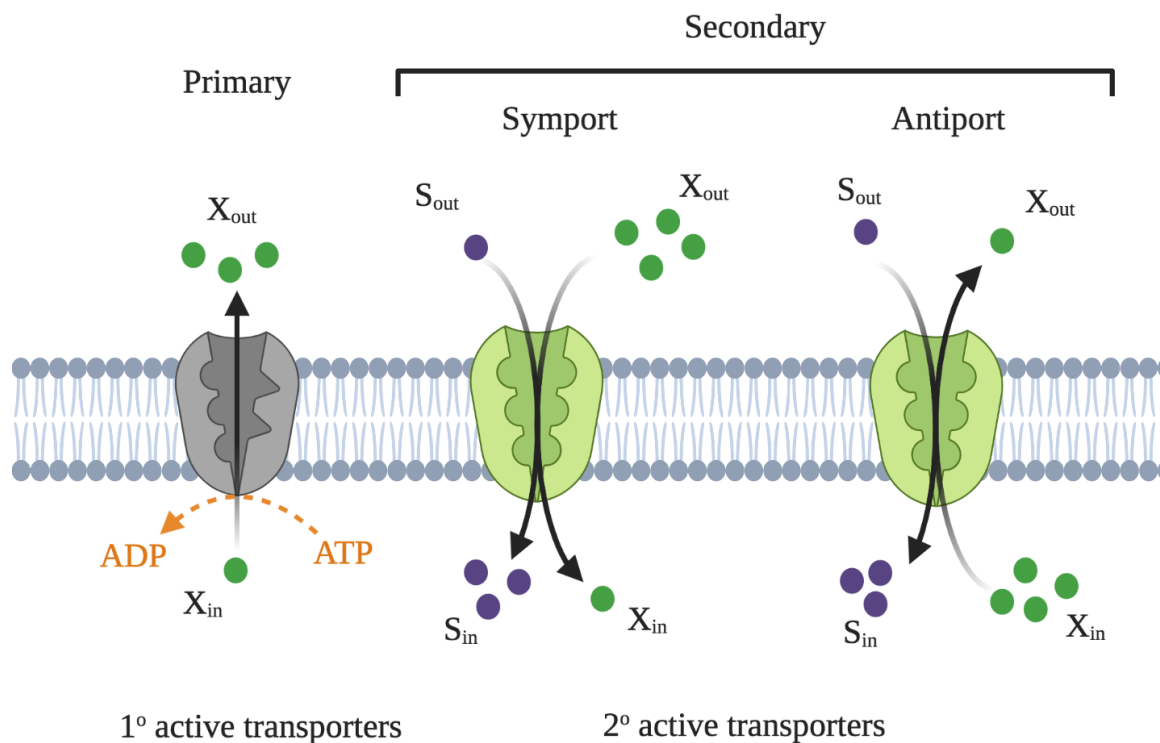


Figure 1.5. Overview of the different types of active transport found in nature.

Primary transporter (in grey) uses energy from ATP hydrolysis, whereas secondary transport (in green) uses an electrochemical gradient. Adapted from Watson (2015) ⁷.

1.2.2.2. Transporter Classification Database

Two different transporter classification schemes have now been established. Lolkema & Slotboom's MemGen method classifies membrane proteins into structural groups depending on their hydropathy profile, which is thought to illustrate a particular fold^{42,43}. Moreover, the well-known Transporter Classification Database (TCDB, <http://www.tcdb.org>) categorises and groups transporters by function and evolutionary relationships – namely into systems, clusters, families and superfamilies^{40,44,45}. Proteins within this database are classified by unique 'TC numbers', analogous to the familiar 'EC numbers' of soluble enzymes. These TC numbers are used in the following sections. At the time of writing, more than 60 superfamilies and 1000 transport protein families have been identified⁴⁴. Table 1.1 lists some of the most well-characterised superfamilies.

Table 1.1. List of some superfamilies present in TCDB.

The corresponding number of recognised families, examples and the equivalent TC numbers are shown. Adapted from Saier *et al.* (2016)⁴⁴.

Superfamily	Number of families	Examples of constituent families (Abbr.)	TC #s
Voltage-gated Ion Channel (VIC)*	37	VIC	1.A.1
		IRK-C	1.A.2
		PCC	1.A.5
Major Facilitator (MF)	100	MFS	2.A.1
		GPH	2.A.2
		POT	2.A.17
Amino acid/polyamine/organocation (APC)	27	APC	2.A.3
		BCCT	2.A.15
		AAAP	2.A.18
Drug/Metabolite Transporters (DMT)	35	CCC	2.A.30
		BAT	2.A.7.2
		DME	2.A.7.3
		GMA	2.A.7.13
ATP-Binding Cassette (ABC1, 2 & 3)	96	LicB-T	2.A.7.18
		P-RFT	2.A.87
		VUT	2.A.88
		ABC	3.A.1

(* Although the Voltage-gated Ion Channel (VIC) superfamily is considered in the TCDB website, the members of this family are not transporters.

Family abbreviations: VIC (Voltage-gated Ion Channel), IRK-C (Inward Rectifier K⁺ Channel), PCC (Polycystin Cation Channel), MFS (Major Facilitator Superfamily), GPH (Glycoside-Pentoside-Hexuronide), POT (Proton-dependent Oligopeptide Transporter), APC (Amino Acid-Polyamine-Organocation), BCCT (Betaine/Carnitine/Choline Transporter), AAAP (Amino Acid/Auxin Permease), CCC (Cation-Chloride Cotransporter), BAT (Bacterial/Archaeal Transporter), DME (Drug/Metabolite Exporter), GMA (GDP-Mannose:GMP Antiporter), LicB-T (Choline Uptake Transporter), P-RFT (Prokaryotic Riboflavin Transporter), VUT (Vitamin Uptake Transporter).

The Drug/Metabolite transporter (DMT) superfamily is of direct relevance to this thesis and will be further addressed in this review. The DMT superfamily started from a typical two transmembrane span (TMS) precursor, leading to four TMS proteins which then added one and duplicated to give ten TMS proteins. The evolutionary pathway has been established as follows: $2 \xrightarrow{\times 2} 4 \xrightarrow{+1} 5 \xrightarrow{\times 2} 10$ (TMS) ^{46,47}. Therefore, DMT superfamily consists of transporters with diverse structural characteristics, reflecting a range of functions, from nucleotide-sugar transporters to an isoprenoid flippase ⁴⁸.

Many proteins within the DMT remain relatively uncharacterised. A particular example studied in this thesis is the Choline Uptake Transporter (LicB-T) family, which was identified relatively recently in human pathogens. A detailed review of this transporter family can be found in Section 1.13.

1.3. Structure of transmembrane proteins

The transmembrane (TM) regions of integral membrane proteins adopt two types of secondary structure within the bilayer region. These are α -helices and β -strands (Figure 1.6). The α -helices can be thought of as individually-stable folded units, whereas β -strands are always arranged into a super-secondary structure, β -barrels ⁴⁹. The two types of secondary structures are exclusive within the membrane region; proteins are either all-alpha or all-beta in the membrane core. Although these structures differ in many respects, both can fully satisfy the hydrogen bonding potential of the carbonyl group and amide group of the protein backbone ⁴⁹. This is essential to membrane insertion since free hydrogen bonding groups are unfavourable in the hydrophobic membrane interior.

Of the two structural types, α -helical TM proteins are the most prevalent ^{50,51}, with β -barrels only occurring in bacterial outer membranes, mitochondria, and chloroplasts ^{52,53}. Only helical-type integral membrane proteins are relevant to this thesis, and β -barrels will not be considered further.

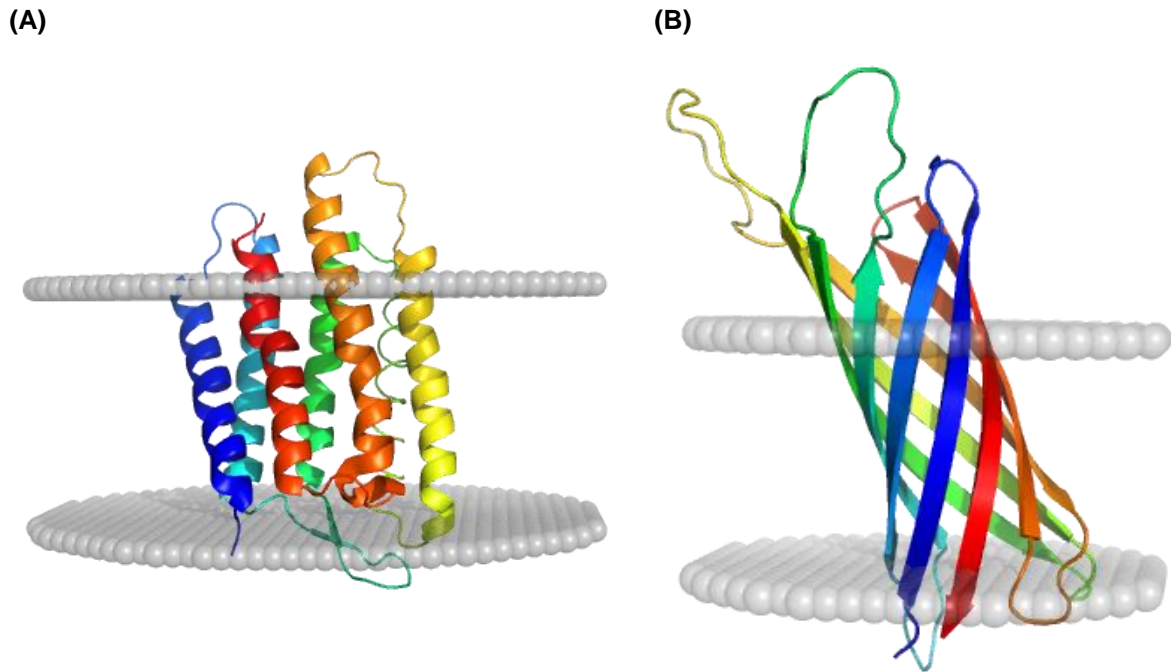


Figure 1.6. Representation of the main types of membrane proteins.

(A) Is an α -helical TM protein – Bacteriorhodopsin (PDB entry 2AT9). **(B)** Is a β -barrel TM protein - FepA (PDB entry 1FEP). Lipid bilayer regions (grey spheres) and are predicted by PPM server ⁵⁴, and figures were produced by Pymol.

1.3.1. Main characteristics of α -helical proteins

It is estimated that around 27% of most organisms' genes encode α -helical integral proteins ^{50,51,55}. These helices generally consist of a stretch of about 18-26 amino acid residues, which is sufficient to span across the lipid membrane (around 30 Å thickness) ⁵⁶. The helical integral membrane proteins can be further defined by the number of times the amino acid sequence passes through the membrane, often referred to as the transmembrane topology. Proteins containing a single TM helix are known as bitopic or single-pass (Figure 1.7A); proteins with more than one TM helix, connected by loop regions outside the membrane, are called polytopic or multi-pass (Figure 1.7B) ^{5,28}.

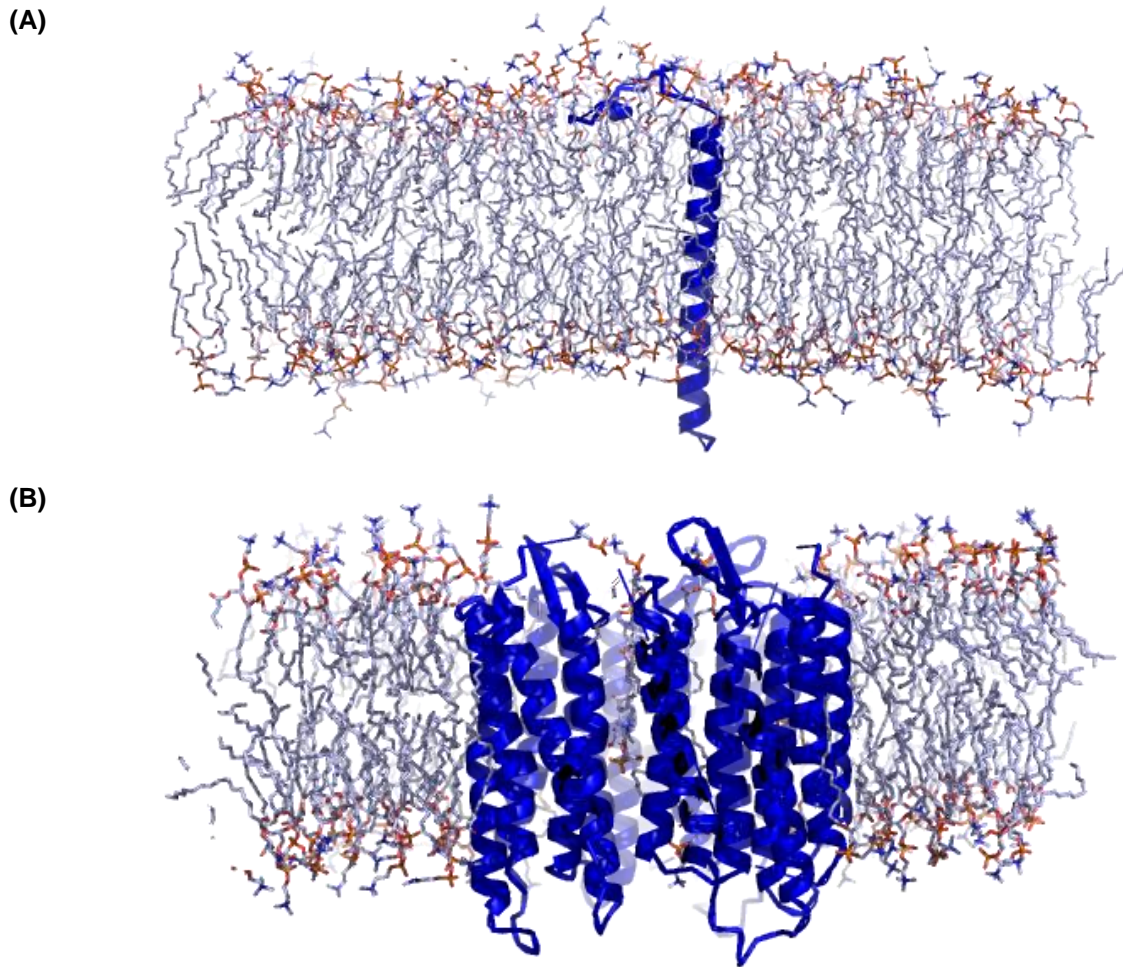


Figure 1.7. Representation of two types of transmembrane protein.

(A) Single-pass helical membrane protein (PDB entry 2MV6). **(B)** Multi-pass helical membrane protein (PDB entry 2AT9). Figures were produced by Pymol based on a molecular dynamics calculation by MemProtMD ⁵⁷.

The amino acid frequency in helical membrane proteins is different across three environments: inside the lipid membrane (hydrophobic environment), outside the lipid membrane (hydrophilic water environment) and the phospholipid headgroup-rich interface area. Consequently, as illustrated in Figure 1.8, different amino acid residues are favoured at different protein locations. Broadly, hydrophobic amino acids are biased towards the membrane core, whereas tryptophan and tyrosine residues enrich the ends of TM helices at the non-polar/polar interface due to amphipathic nature of their side chains ^{34,58}. Moreover, the orientation in the membrane, or topology, dictates the location of the termini and loops. The so-called 'positive inside rule' reflects the preference for loops towards the cell's inside to carry an overall positive charge ⁵⁹. The loops themselves can range from short structureless linkers to large, complex and highly structured soluble domains.

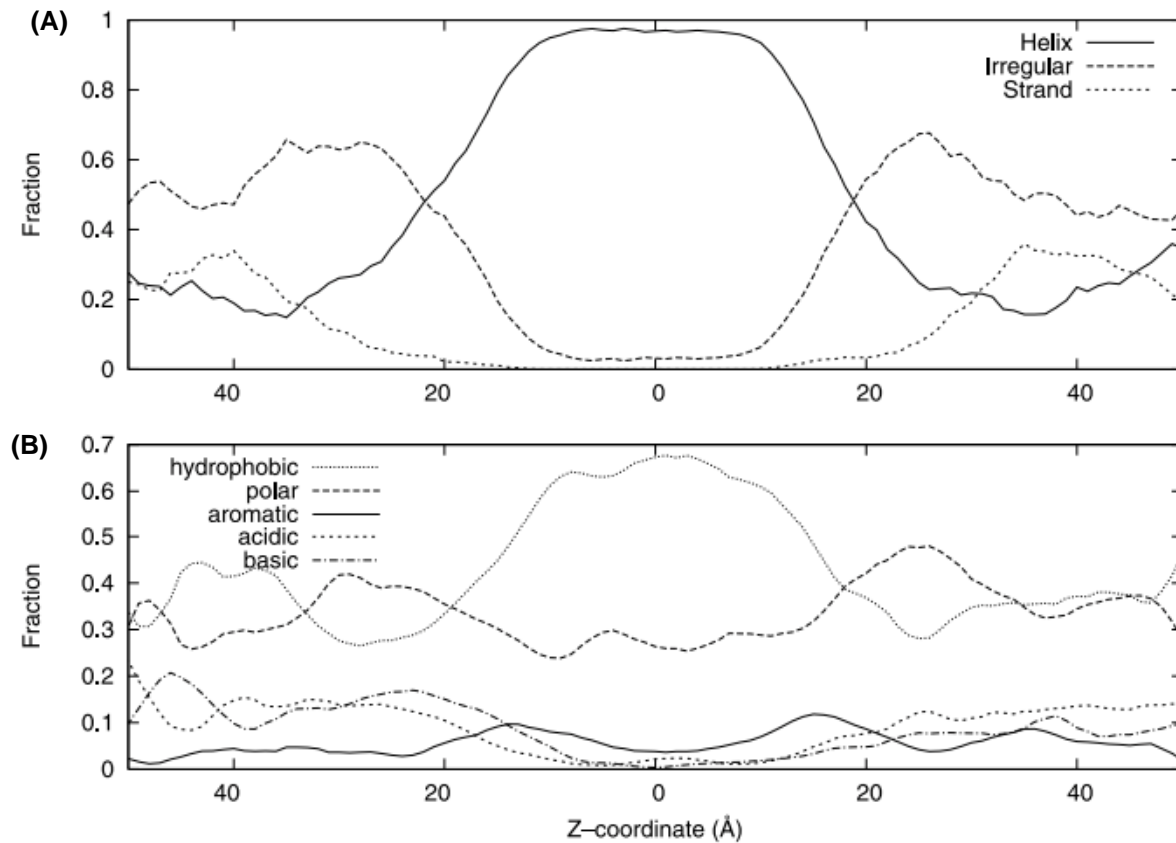


Figure 1.8. Distribution of protein structure and amino acid residues.

(A) Secondary structure between -50 to 50 Å from the centre of the membrane. **(B)** Distribution of the different classifications of amino acid residues in the same region. Adapted from Granseth *et al.* (2005)⁵⁸.

1.3.2. Tertiary structure of transmembrane proteins

Although the hydrophobic effect is thought to be the primary driving force for the proper folding of soluble proteins⁶⁰, these interactions are not responsible for maintaining membrane protein structure. In these particular case, Van der Waals interactions between unique side chains are thought to be required for the individual α -helices to pack together in a well-defined tertiary structure^{49,61}. This gives rise to particular membrane helix packing motifs featuring small residues which allow transmembrane helices to approach and pack together. Some examples of these motifs are the GAS_{LEFT}⁶² or the Glycine zipper⁶³, which can be found in multiple structural contexts. In receptors, transporters or channels, due to the structural rearrangements needed for functionality, the interfaces between helices are not as strongly packed. Therefore, hydrogen bonds are more commonly observed, specifically close to water-filled cavities⁶⁴.

1.4. From transmembrane features to prediction models

Adopting the correct TM topology is essential for membrane protein structure and function. Several bioinformatics tools exist for topology prediction, which includes not only the regions of the primary sequence that will form membrane segments but also the inside/outside orientation of these segments with respect to the membrane ⁶⁵. The first methods for topology prediction of TM segments were only based on the physicochemical properties, namely identifying probable hydrophobic segments on the protein sequence using the Kyte/Doolittle hydrophobicity scale ⁶⁶. Based on this hydropathy index, it is possible to predict where the TM segments will be (Figure 1.9). Later, with TopPred, the prediction models' accuracy improved considerably by including the 'positive inside rule', as discussed above ^{59,67}.

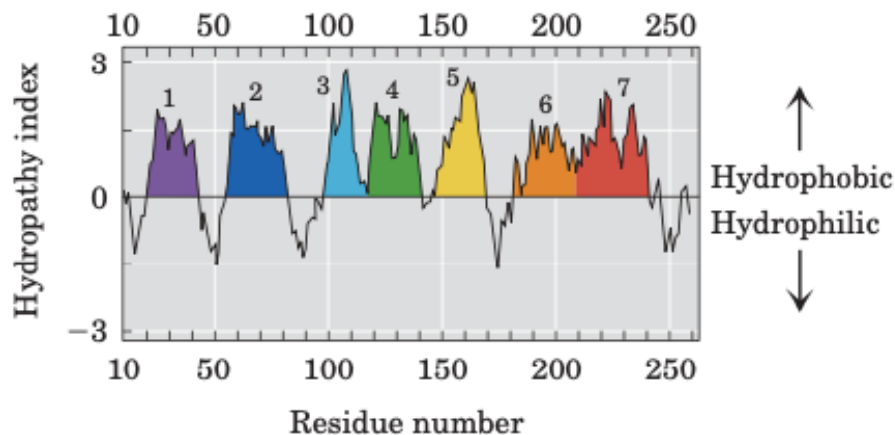


Figure 1.9. Exemplification of a protein hydropathy index, plotted against residue number.

Adapted from Lehninger Principles of Biochemistry (2008) ⁵.

Significant progress was observed with the introduction of multiple sequence alignments and the use of Hidden Markov Models (HMM) ⁶⁸⁻⁷⁰. The first HMM-based transmembrane predictors were TMHMM ⁷¹ and HMMTOP ⁷². Furthermore, the addition of a more precise biological hydrophobicity scale ⁷³, a mixture of machine learning approaches ^{74,75} – such as Artificial Neural Networks (ANN), Support Vector Machines (SVM) or Dynamic Bayesian Networks (DBN) ⁷⁶⁻⁷⁸ - and the use of consensus methods ⁷⁹ improved prediction accuracy even further. It is important to highlight that the identification of the hydrophobic TM helices and the 'positive inside rule' are still the essential basis of these computational methods. Table 1.2 lists some transmembrane topology prediction methods and their main properties.

Despite all the progress and the improving accuracy in predicting TM domains, there are still three significant challenges. First, it remains difficult to make the correct distinction between signal peptides (SP) sequences and N-terminal TM segments ⁷⁷. The first predictor that could do this with some accuracy level was Phobius ⁸⁰; however, some other approaches have been acknowledged ^{74,76,81}. The second major challenge is the prediction of more complex membrane protein structures such as re-entrant helices or the identification of membrane domains of peripheral proteins and the prediction of amphipathic helices ⁸². OCTOPUS is an example of a predictor that can identify TM regions and re-entrant regions, and interface helices by combining HMM with ANN ⁸³. Although this method has improved the topology prediction's accuracy, the ability to identify non-standard features is still restricted. Consequently, when different prediction methods were applied on the human proteome, results differed significantly ⁸⁴.

Table 1.2. List of popular methods for the bioinformatic prediction of membrane protein topology and their main characteristics.

Computational methodology, type of sequence alignment and identification of signal peptides are cited. Adapted from Venko *et al.* (2017) ⁸⁵.

Topology predictors	Methodology	Type of alignment	Identification of SP	Ref.
TMHMM	HMM	single	no	69
HMMTOP	HMM	single	no	70,72
SCAMPI2	HMM	single/multi	no	86
Philius	DBN	single	yes	76
PolyPhobius	HMM	multi	yes	87
SCOPTOPUS	HMM + ANN	multi	yes	74
MEMSAT3	ANN	multi	no	75
MEMSAT	SVM	multi	yes	81
TOPCONS (2015)	Consensus	multi	yes	88

1.5. Experimental limitations on the study of membrane proteins

Despite their abundance and biological importance, many membrane proteins remain poorly understood. The difficulty of studying MPs stems mainly from the high average hydrophobicity of TM regions, which are adapted to reside in the membrane and cause non-specific aggregation once removed from this environment. This presents challenges to recombinant expression and purification as well as to subsequent biophysical studies of structure and function. Although there have been significant advances over the past few years, mainly due to the development of new tools and reagents, a series of challenges remain to the study of any new membrane protein. These are set out below.

Many MPs are found at low levels in their native biological membranes. This means that recombinant expression is required to produce these proteins for biophysical studies. This has led to attempts to overexpress membrane proteins in several different recombinant hosts, with *E. coli* being the most widely used^{89,90}. However, this overexpression can frequently lead to problems such as reduced growth and hampered division, formation of inclusion bodies, or non-functional proteins⁹¹. Moreover, refolding of any aggregate or misfolded membrane proteins is often time-consuming and not successful^{89,92}. Specific *E. coli* strains have been developed to overcome cytotoxicity problems, namely the “Walker strains” C41(DE3) and C43(DE3), which were identified from a screen of BL21(DE3) derivative strains that best-expressed membrane proteins⁹³. Later it was discovered that these two strains contained three mutations in the *lacUV5* promoter, resulting in reduced rates of transcription and translation of recombinant proteins under the control of T7 promoter⁹⁴. This enabled the cell machinery to have a better response to their toxic effect, resulting in higher overexpression yields⁹⁴. Moreover, Wagner and colleagues hypothesised that due to a high transcription level of membrane proteins, the *Sec* translocon might be saturated, leading to unfolded proteins and cytoplasmic aggregation⁹⁴.

Alongside *E. coli*, the Gram-positive *Lactococcus lactis* has been proven to be a substitute to overexpress membrane proteins with a reduced tendency to produce inclusion bodies^{90,95,96}. The main advantages of *Lactococcus lactis* over *E. coli* are the single membrane and the relatively mild proteolytic activity, whereas the main drawback is the absence of chaperones and other cell machinery to fold more complex MP⁹⁵. Besides, *Pichia pastoris* or *S. cerevisiae* have been used for eukaryotic membrane protein expression due to their low cost, ability to grow at high density, and ease of genetic modification⁹⁷. Moreover, the highly regulated

quality control system of *S. cerevisiae* provides an accurate indication of correcting folding and localisation of membrane proteins⁹⁸. Furthermore, there is a growing interest in using the baculovirus-insect cell lines to overexpress membrane proteins⁹⁹.

If a high-level recombinant expression can be achieved, the following experimental barrier to overcome relates to protein extraction and purification. Conditions must be identified to maintain protein stability, oligomeric state and viability for further characterisation¹⁰⁰. One of the significant factors to consider during these processes is the choice of the extraction agent from the native membrane. The success of extraction can be evaluated mainly by three methods: size exclusion chromatography, negative stain electron microscopy or light scattering¹⁰¹. The selection of extraction agents is a recurring topic of this thesis and is considered in greater detail below.

1.6. Membrane mimetics

Seeking an appropriate mimic of the natural lipid bilayer is critical during *in vitro* studies of membrane proteins. The ideal membrane mimetic system should extract the maximum amount of the membrane protein from the biological membrane; while keeping it folded and stable for structural studies and preserve its function to interact with cofactors or ligands¹⁰².

Detergents, also called surfactants, are often the first choice for extraction because they are relatively straightforward and well-characterised systems. Moreover, for the past ten years, detergents have been the most popular mimetic system for structural studies (Cryo-EM, crystallography or NMR)¹⁰³.

More recently, alternative membrane mimetics have emerged, such as:

- 1) polymer-based reconstitution systems (amphipol and styrene-maleic acid),
- 2) protein-based reconstitution systems (membrane scaffold protein and saposin-A lipoprotein),
- 3) peptide-based reconstitution systems (peptergent, lipopeptide, nanostructured [beta]-sheet peptide and ApoA1 mimetic peptide).

Figure 1.10 demonstrates the organisation of some popular membrane mimetic complexes. Each methodology, which will be briefly discussed in Table 1.3, has specific strengths and limitations.

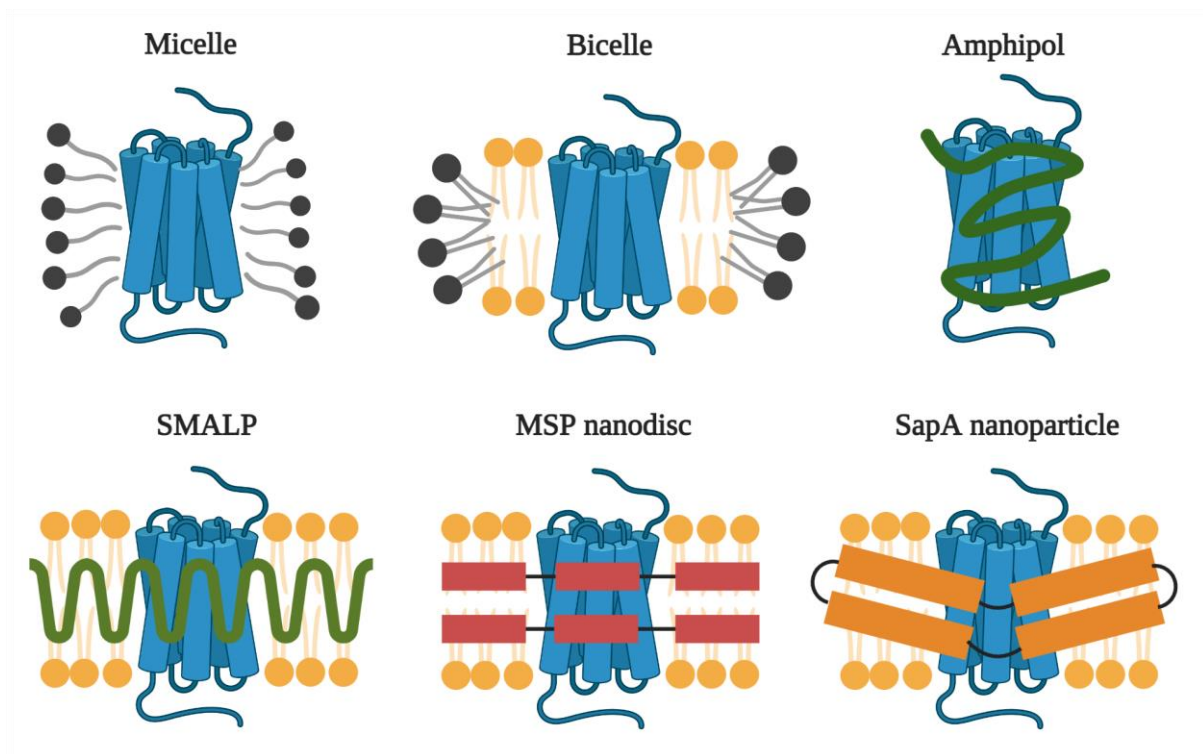


Figure 1.10. Representation of the membrane proteins in different membrane mimetics.

Detergent micelle, bicelle, amphipol, SMALP, MSP nanodisc and Saposin-A (SapA) nanoparticle are shown. Detergent molecules are in grey and lipid molecules in yellow. Adapted from Mineev and Nadezhdin (2017) ¹⁰⁴.

Table 1.3. Different membrane mimetic systems used for membrane protein solubilisation.

Type	Mimetic system	Unnatural functional groups	Requires prior detergent solubilisation?	Instability at low pH or divalent cations
Detergent	Micelle / Bicelle	No	-	No
Polymer based	Amphipol	Yes	No/Yes	Yes
	SMALP	Yes	No	Yes
Protein based	MSP	No	Yes	No
	Saposin-A	No	Yes	No
Peptide based	Peptergent	Yes	No	Unknown
	Lipopeptide	Yes	Yes	Unknown
	[Beta]-sheet peptide	Yes	Yes	Unknown
	ApoA1 mimetic peptide	Yes	Yes	Unknown

1.6.1. Detergents

Detergents are amphipathic molecules that self-associate to form spherical or spheroidal micelles in a concentration-dependent manner. This is introduced in Section 1.6.1.3. Micelles can dissolve cell membranes and extract integral membrane proteins, which are transferred from the membrane bilayer into the hydrophobic core of the micelle. This effectively renders hydrophobic MPs soluble outside of their native lipid environment ¹⁰⁵. Detergents are thus used at every step of membrane protein purification and analysis ^{106–108}.

1.6.1.1. Identifying appropriate detergents

Many different detergents are commercially available, with various different properties. There is no universal rule that can determine which detergent is appropriate to be used with any particular membrane protein, and so a screening process is typically required to find the most suitable. Kotov *et al.* demonstrated the potential use of nanoDSF technology (DSF) to screen protein-detergent complexes' stability based on their melting temperature ¹⁰⁹. Further protocols for high throughput detergent screening have been described recently, namely the BMSS (Biotinylated Membranes Solubilisation & Separation) ¹¹⁰ or fully-automated detergent screening ¹¹¹, which different detergents can be screened at once. Another established technique is the use of membrane proteins fused to Green Fluorescent Protein (GFP) to rapidly screen different solubilisation conditions by using fluorescence size exclusion chromatography ^{112,113}. The fusion complex is easily visualised and monitored during all stages of expression and purification by fluorescence light emission at ~512 nm ^{114,115}. Moreover, Drew and his colleagues demonstrated that membrane proteins fused with GFP can be overexpressed (higher than 1 mg per litre) in *S. cerevisiae* and are stable and monodisperse in both detergents FC-12 and DDM ¹¹³.

To add further complexity, the detergent that better extracts the protein from biological membranes may not be compatible with purification or subsequent biochemical studies. Therefore, a 'trial and error' process is commonly required to extract the protein of interest in its active biological form. Secondly, detergents may partly or entirely remove native lipid molecules, which might be critical for protein's functionality and stability ^{116–119}. Moreover, some detergents may disrupt enzymatic activity and ligand binding assays. They can also interfere with protein dynamics, unfold secondary structure or disrupt protein-protein interactions ^{120–124}. Lastly, detergent micelles might not be the ideal membrane mimetic due to their geometric and physicochemical properties. Micelles tend to have a spherical form with

higher curvature when compared to the native lipid membrane ¹⁴. Moreover, the rapid exchange of the detergent monomers between the protein-detergent complex and the reservoir of 'free' detergent molecules may result in unfavourable dynamics that can significantly differ from the native environment ¹²⁵.

1.6.1.2. Membrane solubilisation process

Membrane solubilisation by detergents includes three main steps (Figure 1.11). When detergents are added to biological membranes at low concentrations, the monomers are able to perturb the membrane structurally by partitioning into the lipid bilayer. However, when detergents are added at higher concentrations (equal or higher than their critical micelle concentration - Section 1.6.1.3), the bilayer becomes saturated with detergent monomers and breaks apart, producing lipid-protein-detergent mixed micelles. An additional increase in detergent concentration causes progressive delipidation to form lipid/detergent and protein/detergent mixed micelles. This last aggregate can also be called a protein-detergent complex (PDC) ^{105,125,126}. The total mass contribution of the detergent to the PDC can vary between 30 to 70% ^{126,127}, and the detergent molecules in the PDC are assumed to create a layer around the protein's hydrophobic exterior ^{125,126}.

Following solubilisation, the excess of free micelles should be removed since any uncomplexed micelles can interfere with further biophysical as well as structural characterisation of the membrane protein. For example, an excess of empty micelles can prohibit protein crystallisation ⁹⁷ and interfere with vitrification for Cryo-EM ¹²⁸. The exchange of detergents monomers can change the binding pattern and generating detergent artefacts in sensitive techniques such as isothermal titration calorimetry ¹²⁹. Four main methods are usually applied to remove detergent excess: dialysis, hydrophobic adsorption, gel chromatography or ion-exchange chromatography ¹³⁰.

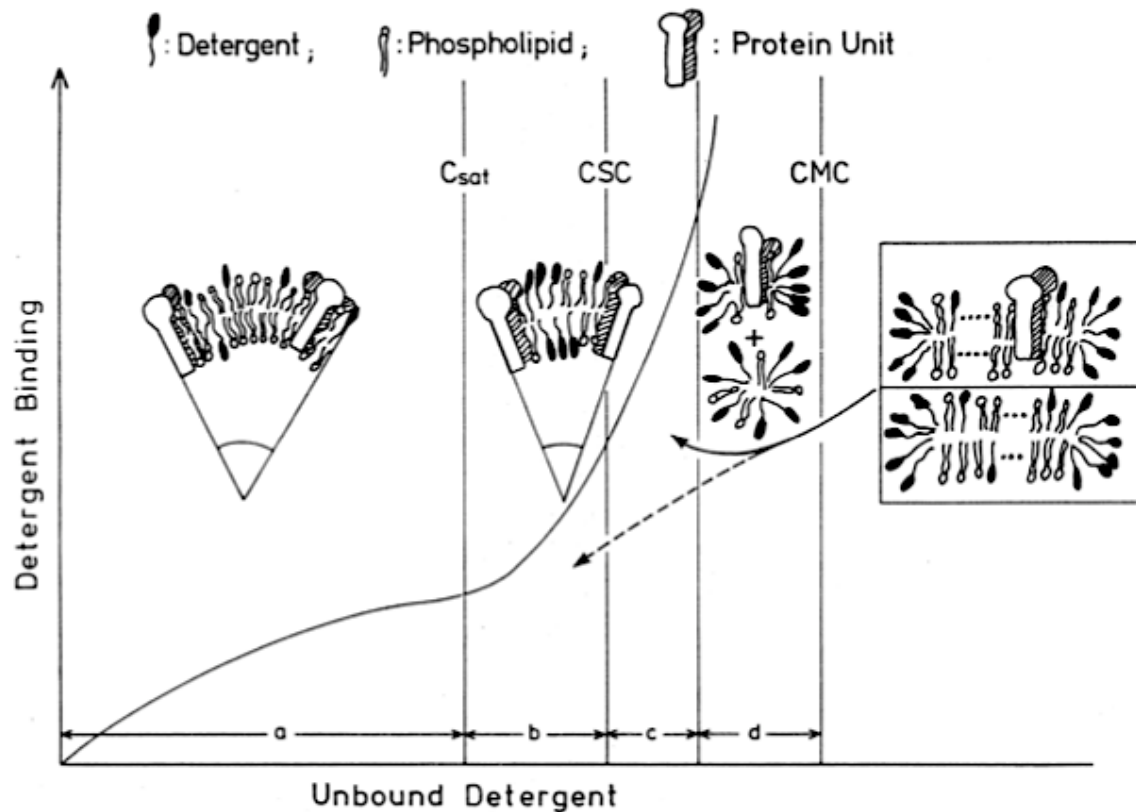


Figure 1.11. Representation of the different phases encountered during solubilisation of membrane protein as a function of the unbound detergent.

C_{sat} , onset of membrane solubilisation; CSC, critical solubilisation concentration, at which point both lipid-detergent and protein-lipid-detergent complexes begin to be released from the membrane; CMC, critical micelle concentration. Adapted from le Maire *et al.* (2000) ¹²⁵.

1.6.1.3. Critical micelle concentration

The critical micelle concentration (CMC) is described as the minimum concentration at which detergent monomers form aggregates spontaneously, termed micelles ¹³⁰. This is specific to each individual detergent, as is the number of monomers within the micelle, known as the aggregation number ¹³¹. Since micelle formation is essential for extracting proteins from membranes, during solubilisation steps, detergents are used at a concentration significantly greater than their CMC, typically 10 to 20 times higher.

The CMC can be determined experimentally by surface tension, light scattering, fluorescence or dye solubilisation measurements ¹³². Furthermore, CMC varies depending on a number of factors, including pH, temperature, ionic strength, and the presence of lipid, protein or other detergent molecules ¹²⁵. Concerning temperature, the critical micellar temperature (CMT) is the minimum temperature at which a particular detergent concentration can reach its CMC

value. In contrast, the Krafft Point is the temperature at which the crystalline phase, monomers and micelles occur in equilibrium ¹³¹. In most instances, this will be the same as the CMT (Figure 1.12).

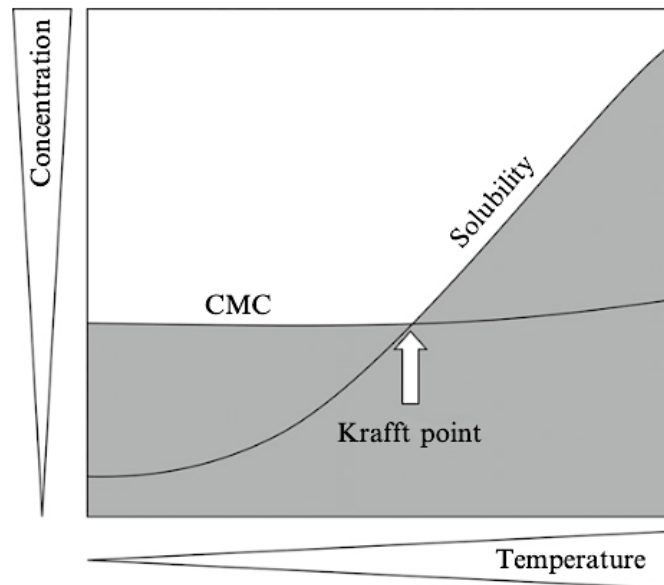


Figure 1.12. Schematic representation of the Krafft point.

Grey areas correspond to detergent monomers, whereas white areas correspond to detergent micelles. Adapted from Linke (2009) ¹³¹.

1.6.1.4. Types of detergents

There are numerous reviews on detergents which include detailed information regarding the physicochemical parameters of the monomer and the properties of the micelle ^{131,133}. The properties of some common detergents can be found in Table 1.4. Broadly, detergent compounds can be categorised according to their headgroup chemistry as either ionic, non-ionic or zwitterionic.

Ionic detergents (Table 1.4A) have either negative or positive charged headgroups, such as sodium dodecyl sulphate (SDS) and cetyl trimethylammonium bromide (CTAB). Although these detergents are effective in terms of protein extraction from the membrane, they are often considered harsh and tend to be denaturing since they disrupt protein-protein interactions. However, a particular sub-class of the ionic detergents is the bile acid salts (such as sodium

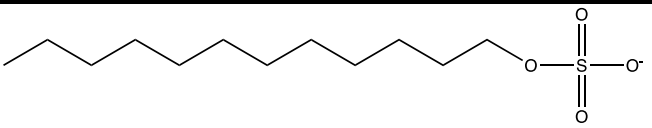
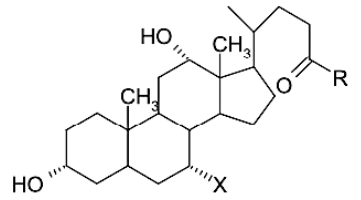
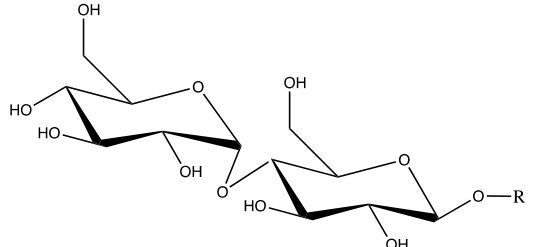
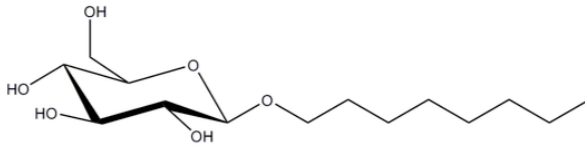
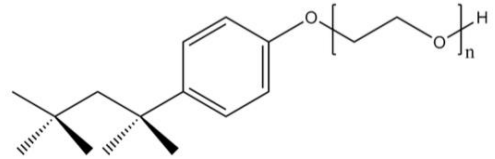
cholate), consisting of rigid steroidal groups in their backbone (Table 1.4B). In general, these are milder detergents that are less disruptive of protein structure.

Non-ionic detergents, including maltosides, glucosides or polyoxyethylene glycols, are characterised by having uncharged headgroups (Table 1.4C). These tend to be mild and non-denaturing; they act to replace lipid-protein interactions rather than protein-protein interactions. Therefore, the majority of the membrane protein structures in the PDB were solved in sugar-based non-ionic detergents, namely *n*-dodecyl- β -D-maltoside (DDM), *n*-decyl- β -D-maltoside (DM) and *n*-octyl- β -D-glucopyranoside (OG) (Figure 1.13)^{134,135}. Even though DDM is the most used for protein purification and crystallisation, it forms larger detergent micelles of about 65 - 70 kDa, masking the hydrophilic regions of membrane proteins. Consequently, the development of protein-protein interactions indispensable for crystal formation is often inhibited. In opposition, as short-chain detergents have smaller micelles, they can allow crystal contacts to form and so may be better suited to crystallisation.

Lastly, detergents with both positively and negatively charged headgroups fall into the zwitterionic class (Table 1.4D). This includes the fos-choline detergents, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate (CHAPSO), and lauryldimethylamine-*N*-oxide (LDAO). These have also been successfully used in membrane protein studies.

Table 1.4. Types of detergents.

(A) Ionic, (B) Bile acids, (C) Non-ionic and (D) Zwitterionic - and their main properties (structure, CMC, MW, micelle size). Adapted from le Maire *et al.* (2000)¹²⁵ and Seddon *et al.* (2004)¹³⁰. % corresponds to weight (g) per 100 mL.

Detergent	MW (g/mol)	CMC (% / mM)	Micelle size (kDa)	Chemical structural
(A) Ionic detergents				
SDS	288.4	0.17 – 0.23 / 6 - 8	18	
(B) Bile acid salts				
Sodium cholate X = OH, R = O ⁻ Na ⁺	430.6	0.41-0.60 / 9 - 15	0.9	
Sodium deoxycholate X = H, R = O ⁻ Na ⁺	414.6	4 - 8	1.7 - 4.2	
(C) Non-ionic detergents				
DDM, R = C ₁₂ H ₂₅	510.6	0.0087 / 0.17	65 - 70	
DM, R = C ₁₀ H ₂₁	482.6	0.087 / 1.8	40	
Cymal – 5, R = C ₁₁ H ₂₁ Cymal – 6, R = C ₁₂ H ₂₃	494.6 508.6	0.12 / 2.5 0.028 / 0.56	23 32	
OG	292.4	0.53 / 20	25	
Triton X-100 n = 9,10	624.8 (av.)	0.01 / 0.2	60 - 90	

(D) Zwitterionic detergents

CHAPS X = H	614.9	0.5 / 8 - 10	6
CHAPSO X = OH	630.9	0.5 / 8 - 10	7
LDAO	229.4	0.023 / 1 - 2	21.5

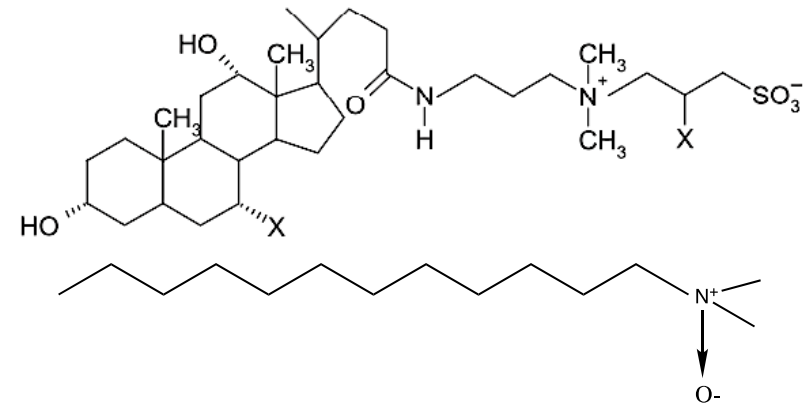
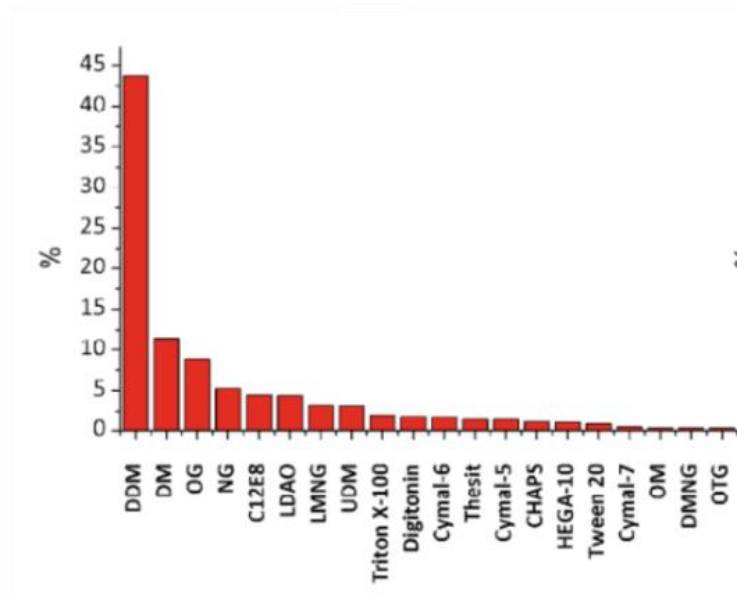
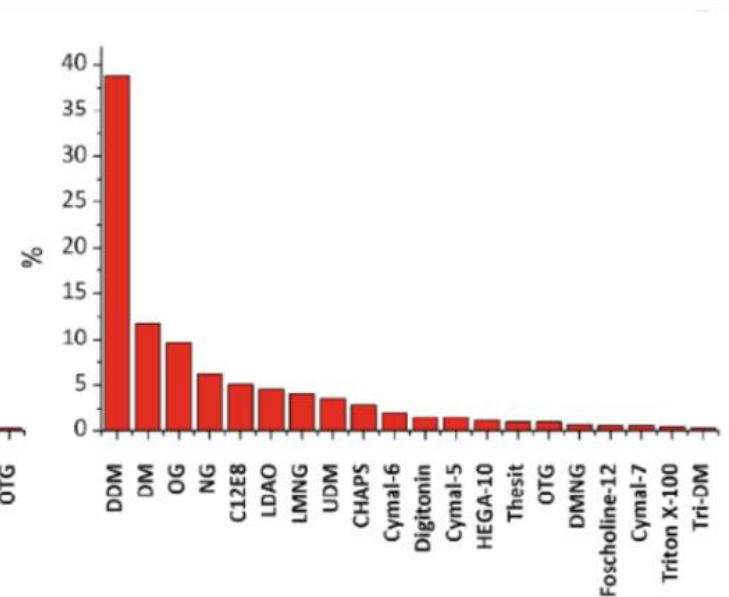
**(A)****(B)**

Figure 1.13. Percentage (%) of studies in which the detergent is used for solubilisation and crystallisation.

(A) Detergents used for the initial extraction and purification of a set of integral membrane proteins for which structures have been obtained. **(B)** Detergents enabling the crystallisation of the same set of proteins. Adapted from Stetsenko and Guskov (2017) ¹³⁴.

1.6.1.5. Bicelles

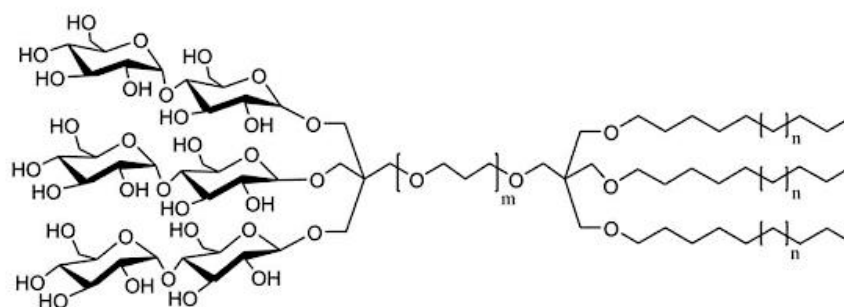
Bicelles are constituted of a planar lipid bilayer (typically constituted by DMPC molecules) and detergent molecules (*i.e.*, DHPC or CHAPSO) that solvate the bilayer edges (Figure 1.10)¹³⁶. Bicelles can be formed with diverse lipid to detergent ratios, thus varying their size¹⁰⁴. The amount of detergent in the bicelles is much lower than in mixed micelles¹³⁰. Therefore, bicelles provide a more native-like environment, which is helpful to investigate specific lipid-protein interactions. Moreover, bicelles are magnetically oriented, and thus, the effect of lipid composition on structural properties can be studied by NMR^{137,138}.

1.6.1.6. Alternative classes of detergents

New detergents continue to be developed. This effort has concentrated on identifying compounds that improve the purified membrane protein's stability by increasing membrane domain rigidity, compared to the traditional detergents, and close to that inflicted by lipids in the native environment.

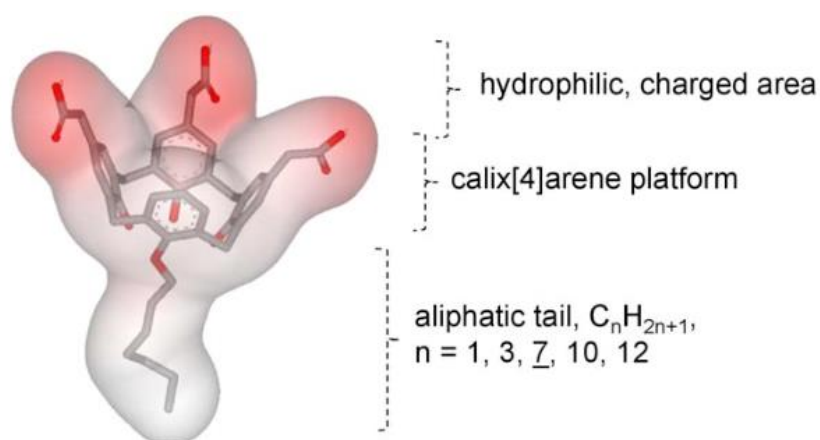
Recently, Bae *et al.* presented the tandem-neopentyl glycol maltosides (TNMs), which improved protein stability over time compared to DDM (Figure 1.14A)¹³⁹. Another example of novelty is the anionic calix[4]arene (C4Cn, n = 1-12) detergent, based on a rigid calixarene scaffold to which hydrophilic and charged headgroups and hydrophobic tail are attached (Figure 1.14B)¹⁴⁰. This structure stimulates salt-bridge interactions with the positively charged residues of the membrane proteins existing at the cytosol-membrane interface. Matar-Merheb *et al.* successfully demonstrated the extraction of BmrA, ABCG2, AcrB and the SR-Ca²⁺-transporting ATPase using C4Cn derivatives¹⁴⁰. Lastly, the steroid-based facial amphiphiles (FAs) have been used for solubilisation/stabilisation protocols. They are synthesised by attaching sugar headgroups to a cholate scaffold (Figure 1.14C)^{141,142}. Such design has proven not only to stabilise membrane proteins (due to their large hydrophobic surface), but also to improve crystallisation since more rigid protein-detergent complexes can mediate closer intermolecular contacts between molecules of membrane proteins¹⁴³.

(A)



TNM-C9L : m = 1, n = 1	TNM-C10L: m = 1, n = 2
TNM-C11L: m = 1, n = 3	TNM-C12L: m = 1, n = 4
TNM-C13L: m = 1, n = 5	TNM-C14L: m = 1, n = 6
TNM-C11S: m = 0, n = 3	TNM-C12S: m = 0, n = 4
TNM-C13S: m = 0, n = 5	TNM-C14S: m = 0, n = 6

(B)



(C)

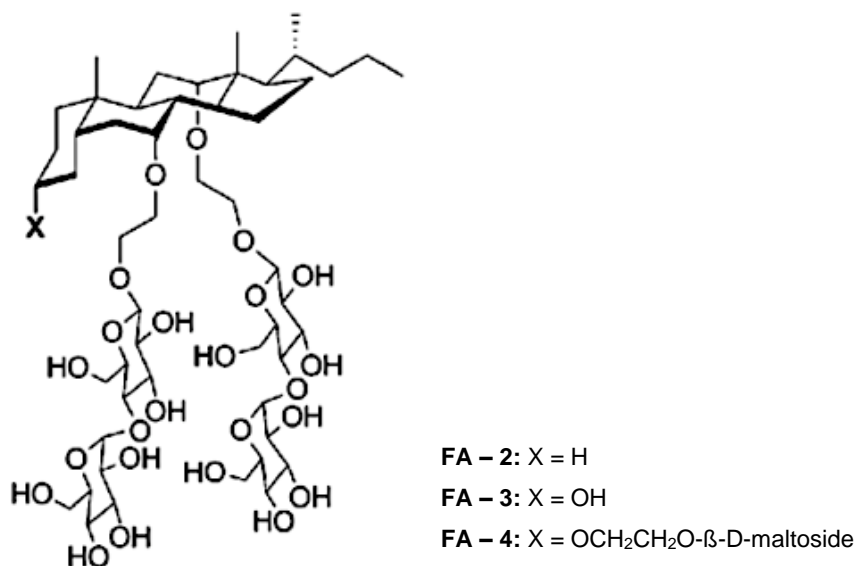


Figure 1.14. New classes of detergents.

(A) TNMs, with either long (TNM-Ls) or short (TNM-Ss) spacers between the two neopentyl glycol units. Adapted from Bae *et al.* (2016) ¹³⁹. (B) C4Cn, constituted by a rigid backbone, an aliphatic tail, and a hydrophilic area. Adapted from Matar-Merheb *et al.* (2011) ¹⁴⁰. (C) FAs – example of FA-2 to 4 with different groups at the position X. Adapted from Lee *et al.* (2013) ¹⁴².

1.6.1.7. Protein reconstitution in proteoliposomes

Liposomes are systems that can mimic phospholipid membranes to which a protein could be incorporated (trapped) or inserted into the bilayer; once proteins are introduced, these are termed proteoliposomes. Proteoliposomes are an essential tool for biophysical studies, especially in the study of transport proteins, as well as for nanobiotechnological applications ¹⁴⁴.

In comparison to natural membrane systems, proteoliposomes have major advantages since they can be obtained with a minimum number of components (lipids and proteins), simplifying experimentation and data interpretation. On the other hand, there are some important criteria to take into consideration for ideal reconstituted proteoliposomes, which will be briefly discussed. Proteoliposomes should be of a homogenous size, achieved by extrusion, with the protein being evenly distributed in the liposome ¹⁴⁵. The activity of the protein must be retained in the proteoliposome, which can be influenced by the number of proteins per liposomes and the lipid to protein ratio ¹⁴⁶. Moreover, and especially important for transport assays, the protein must be incorporated in the liposomes in the correct orientation - inside-out or outside-in - ensuring the correct transport direction of substrates ¹⁴⁷. However, many transport proteins seem to work equally well in either direction, and orientation is less critical ¹⁴⁸. Lastly, proteoliposomes should have low passive permeability of their membrane, ensuring the limited diffusion of solutes and counterions so that chemical and electrochemical gradients can be maintained ¹⁴⁶.

Figure 1.15 highlights three main technical approaches to insert membrane proteins into liposomes that have been described in the literature. The most successful method uses a detergent-mediated pathway since detergents are already present in the initial purification step. Reconstitution can also be mediated by organic solvent or by mechanical disruption of the liposome, such as with sonication or freeze-thaw cycles ¹⁴⁹. It is worth noting that the organic-mediated and mechanical reconstitution have been somewhat associated with protein denaturation and inactivity ¹⁵⁰.

There are two main approaches to detergent-mediated reconstitutions. The first of these might be termed “simple dilution”, whereby the protein-detergent complex is diluted into a liposome solution to a detergent concentration below the CMC. After micelles being destabilised, the membrane protein moves from this structure into the liposome. In the second approach, detergent is used first to saturate the liposome bilayer, resulting in disruption of lipid-lipid interactions. Consequently, the bilayer tends to be more permeable and receptive to

incorporating the protein once this is introduced. After proteoliposome formation, the excess of detergent is removed by size exclusion or adsorption to hydrophobic beads, as described previously. An excellent review on the formation of proteoliposomes is given by the classic review of Rigaud *et al.* ¹⁵⁰, to which the reader is directed for more information.

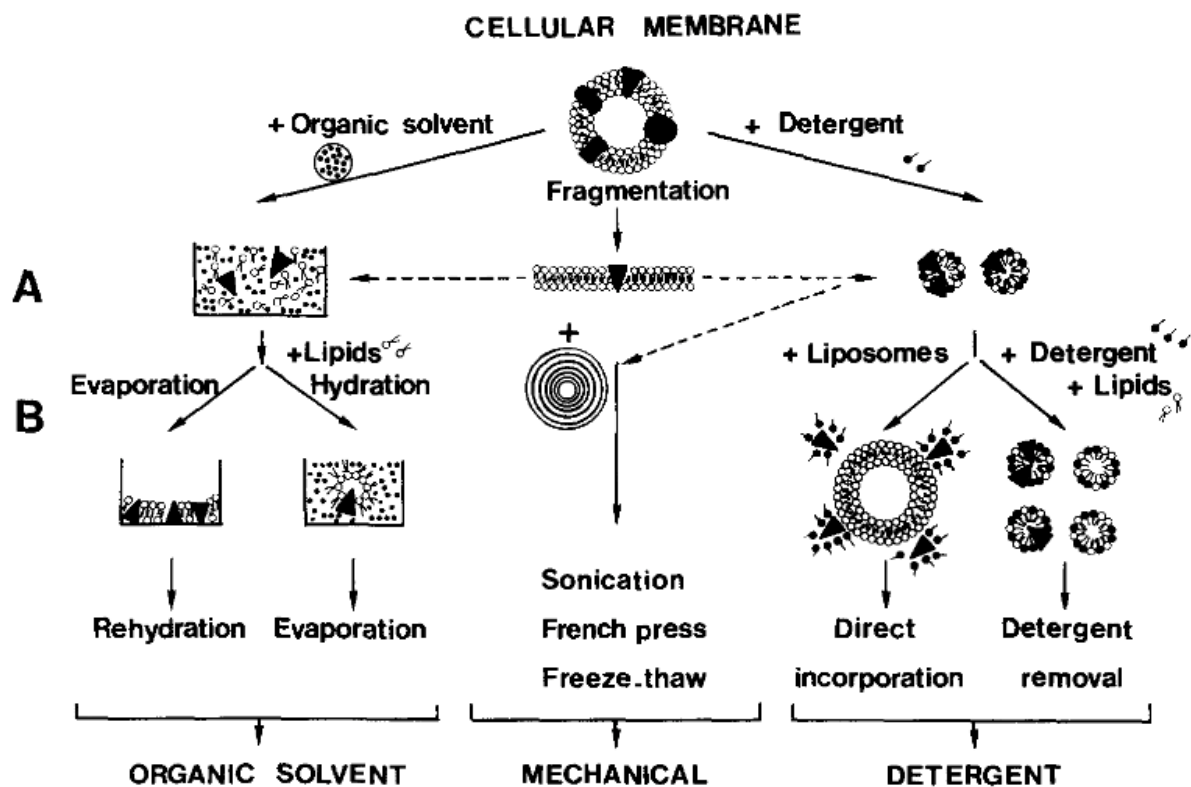


Figure 1.15. Overview of the different approaches used for reassembly of membrane proteins into liposomes.

(A) Membrane proteins can be purified and isolated in an organic solvent or as PDCs. (B) Three main strategies used for reconstitution into liposomes: organic solvent-mediated reconstitutions, mechanical means, and detergent-mediated reconstitutions. Adapted from Rigaud *et al.* (1995) ¹⁵⁰.

1.6.2. Protein-based reconstitution systems

Membrane protein reconstitution using protein scaffolds include two existing methods, the nanodisc system and the Salipro method. Both methods make use of amphipathic protein scaffolds, the membrane scaffold protein (MSP) ^{151,152} or the saposin-A ¹⁵³, respectively. In these reconstitution systems, the purified protein is surrounded by a small patch of phospholipids. This lipid-protein complex is wrapped by a protein 'belt' that forms and

maintains its structure. The lipid to protein ratio and the length of the protein scaffold will influence the nanoparticle size ¹⁵⁴. Figure 1.16 reports the process of nanodiscs or nanoparticles assembly. This method is not detergent free since membrane protein is assembled in detergent micelles before the MSP/saposin and lipids are applied. Nanodiscs and lipid nanoparticles are formed from this mixture once the detergent is removed.

Besides giving the membrane protein a more native-like environment, nanodiscs have several advantages over traditional detergent. They can improve long-term protein stability as well as enzymatic activity ¹⁵⁵. Moreover, they allow precise control of lipid composition and the oligomeric state of the protein of interest ¹⁵². Nanodiscs can be broadly applicable independently of the protein type, and they have been extensively used for structural studies (including Cryo-EM ^{156,157} and X-ray crystallography ^{158,159}), as well as functional assays, biotechnology and medical applications ¹⁵².

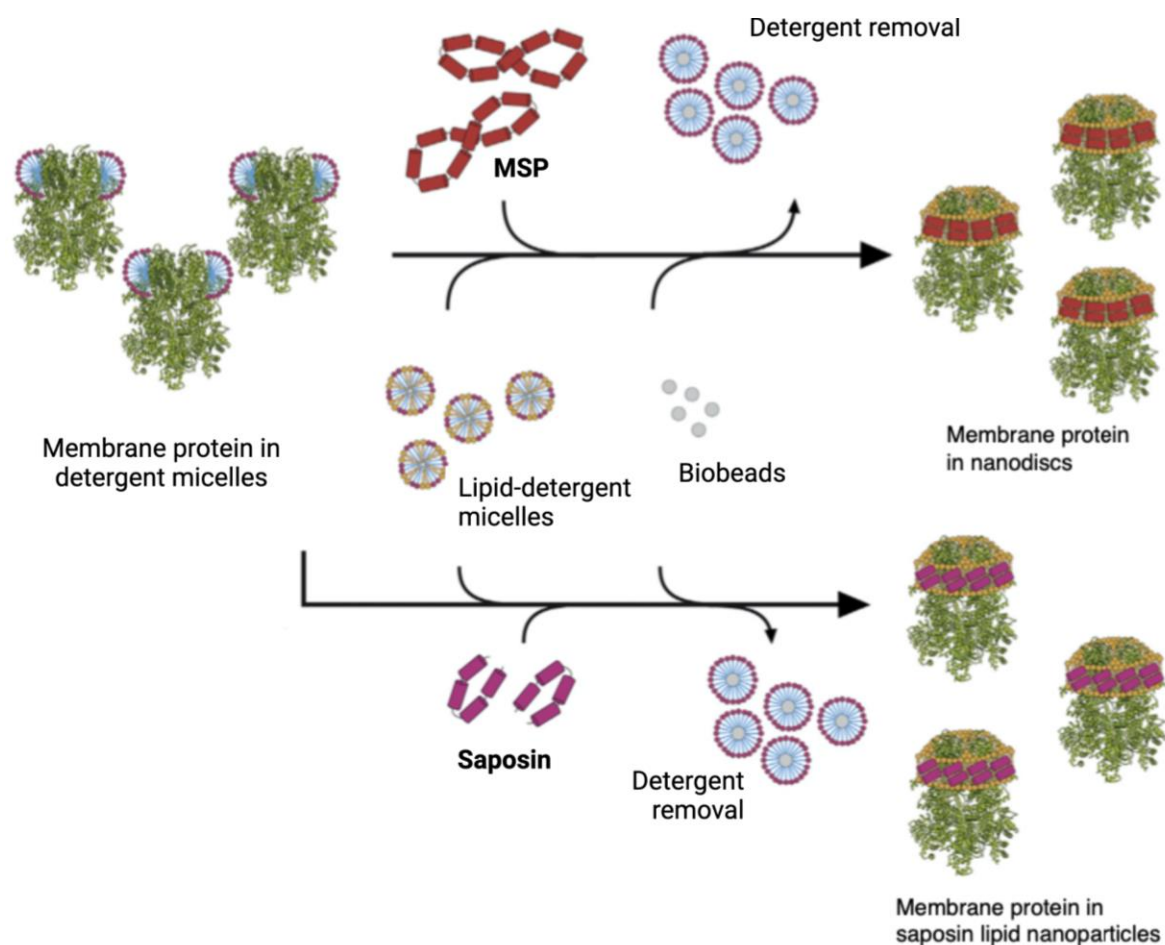


Figure 1.16. Schematic representation of membrane proteins in MSP or saposin-A forming nanodiscs or lipid nanoparticles, respectively.

Adapted from Autzen *et al.* (2019) ¹⁵⁴.

1.6.3. Synthetic polymer-based reconstitution systems

The protein scaffolds mentioned above have been produced in recombinant cell culture, but their scalability, purity and homogeneity are often compromised ¹⁶⁰. Furthermore, these membrane mimicking systems continue to rely on detergents to extract the protein of interest from the membrane ^{153,161}. Researchers have been developing a variety of synthetic polymers to address these constraints, which will be briefly discussed below.

One of these synthetic polymers is the styrene-maleic acid copolymer (SMA); And among the main properties of SMA is its amphipathicity, comprising styrene (S) and maleic (M) acid units, being hydrophobic and negative charged, respectively (Figure 1.17) ¹⁶². The 2:1 and 3:1 S:M polymer ratios have been the most used for protein solubilisation ¹⁶². The SMA copolymer spontaneously forms discoidal particles with biological lipids with a maximum diameter close to 15 nm, termed styrene-maleic acid copolymer-lipid particles (SMALPs) ^{163,164}. The mechanism for SMALPs formation has been proposed to have three main steps ¹⁶⁵, in which the release of membrane proteins from lipid membrane is envisaged as a 'cookie-cutter' - where the SMA polymer wraps around a lipid bilayer patch ^{166,167}.

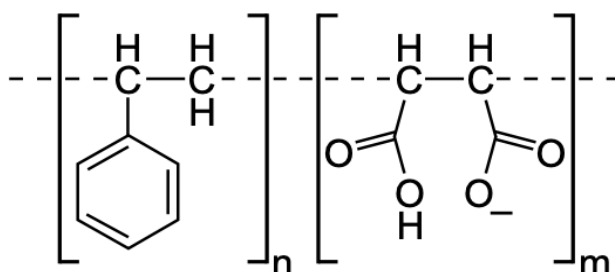


Figure 1.17. Chemical structure of SMA polymer at neutral pH.

Adapted from Dörr *et al.* (2014) ¹⁶⁶.

The main advantage of using SMA is that it extracts the membrane protein of interest with its related lipids, which means that the protein's immediate environment is still the native lipid membrane – lipid bilayer composition and asymmetry are conserved ¹⁶³. Membrane proteins incorporated in SMALPs are generally more stable compared to detergents, as demonstrated by their increased thermostability ^{164,166}. The proteins in the SMALP particle can be purified using standard chromatography without the addition of any other chemical, for instance, detergent. The use of SMALPs has been recognised to be consistent with solid-state NMR, X-ray crystallography and Cryo-EM ¹⁶⁸⁻¹⁷⁰.

Despite the successful application of SMA in the study of several membrane proteins – e.g., AcrB¹⁶⁹, Pgp¹⁷¹, KcsA¹⁶⁶ and PBP2A¹⁷² - intolerance to low pH or divalent metal ions limits its applications. Below pH ~ 6.5, SMA loses its hydrophilicity and thus precipitates in solution^{163,173}. The same occurs when divalent cations interact with the carboxyl groups through electrostatic interactions, leading to SMA precipitation¹⁶³. Therefore, membrane proteins that need cofactor ions such as Mg²⁺ or Ca²⁺ will not be suitable to be solubilised by SMA. Moreover, since solution NMR prefers low pH, using SMALPs in these experiments can be challenging¹⁷⁴. Another important aspect is the size of the protein to be solubilised since proteins or protein complexes with a molecular weight greater than 400 kDa will probably not fit into the SMALP particle¹⁶³. This limitation eliminates the solubilisation of many membrane complexes. Considering all these restrictions, researchers are working on alternative synthetic polymers with enhanced ion compatibility and pH^{175,176}.

Similar to SMA, amphipols are a class of short and amphipathic synthetic polymers consisting of a hydrophilic backbone (carboxylate groups) and hydrophobic side chains¹⁷⁷. The A8-35 variant has been the most studied amphipol to date (Figure 1.18). However, its sensitivity to acidic buffers (below pH 7) and divalent cations such as Ca²⁺ has limited its use. As a result, other variants, especially those with lower pH insensitivity, such as non-ionic amphipols, have been developed^{178–180}.

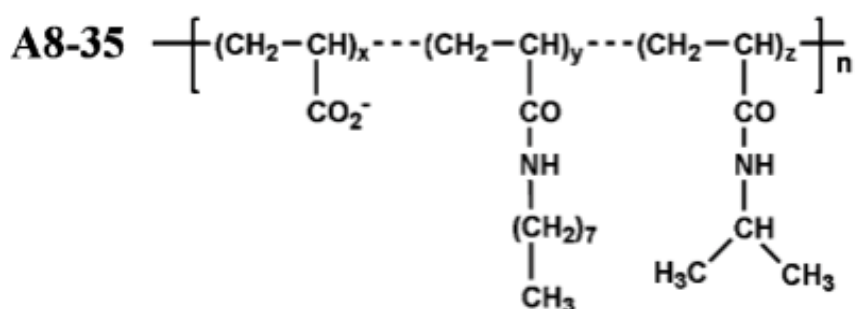


Figure 1.18. Chemical structure of the A8-35 amphipol.

The polymer displays a negative charge at neutral and basic pH. Adapted from Picard *et al.* (2006)¹⁷⁹.

Although amphipols can partition into biological membranes, there are weakly dissociating surfactants^{181,182}. Even though there are a few exceptions¹⁸³, typically, MPs require to be first solubilised in detergent and subsequently exchanged into amphipols¹⁸⁴. Amphipols have been helpful to improve the stability of MPs in amphipol-protein complexes^{183,185} or to refold denatured MPs^{186,187}.

1.6.4. Peptide-based reconstitution systems

Membrane mimicking systems based on peptides have been considered as possible substitutes for synthetic polymers and protein scaffolds. This is mainly because peptides can be synthesised to a high degree of homogeneity and controlled amino acid sequence. However, the scientific community has not widely adopted these systems for membrane protein characterisation as previous methods, mainly due to their high production cost and difficult accessibility¹⁸⁸. Although a series of peptide reconstitution systems have now been described^{126,189–191}, for the purpose of this review, only peptergents will be briefly discussed since they can be used directly as solubilising agents.

The first α -helix peptide detergent was first developed by Stroud and colleagues in 1993. It had a long hydrophobic surface, enough to solvate the hydrophobic membrane spanning region of an integral membrane protein¹⁹². Stroud *et al.* showed that this peptergent could maintain the solubility of bacteriorhodopsin and rhodopsin for two days¹⁹².

More generally, peptergents are amphipathic peptides composed of repeating units of hydrophobic residues (as alanine, valine or (iso)leucine) with ends coated by a positively (lysine) or negatively charged residue (aspartic acid)¹⁹³. To preserve ionic nature in the peptide, the N-terminus is acetylated and the C-terminus amidated, or both^{194,195}. As a result, peptides form spherical micelles that are able to extract and stabilise membrane proteins. Some examples are the olfactory receptors¹⁹⁵, the G-protein coupled receptor bovine rhodopsin¹⁹³ or the *E. coli* glycerol-3-phosphate dehydrogenase¹⁹⁶.

Different types of peptides have been produced by varying their length and their N or C-terminus charge (Figure 1.19), leading to the formation of micelles with different aggregation number and surface charge¹⁹⁴.

Peptergents can function similarly to conventional detergents and exhibit a range of benefits over new classes of solubilising chemicals. They can be systematically engineered and commercially manufactured at high purity, keeping their stability over long periods at room temperature. Moreover, they can be used in combination with other surfactants in a number of ways^{193,195}.

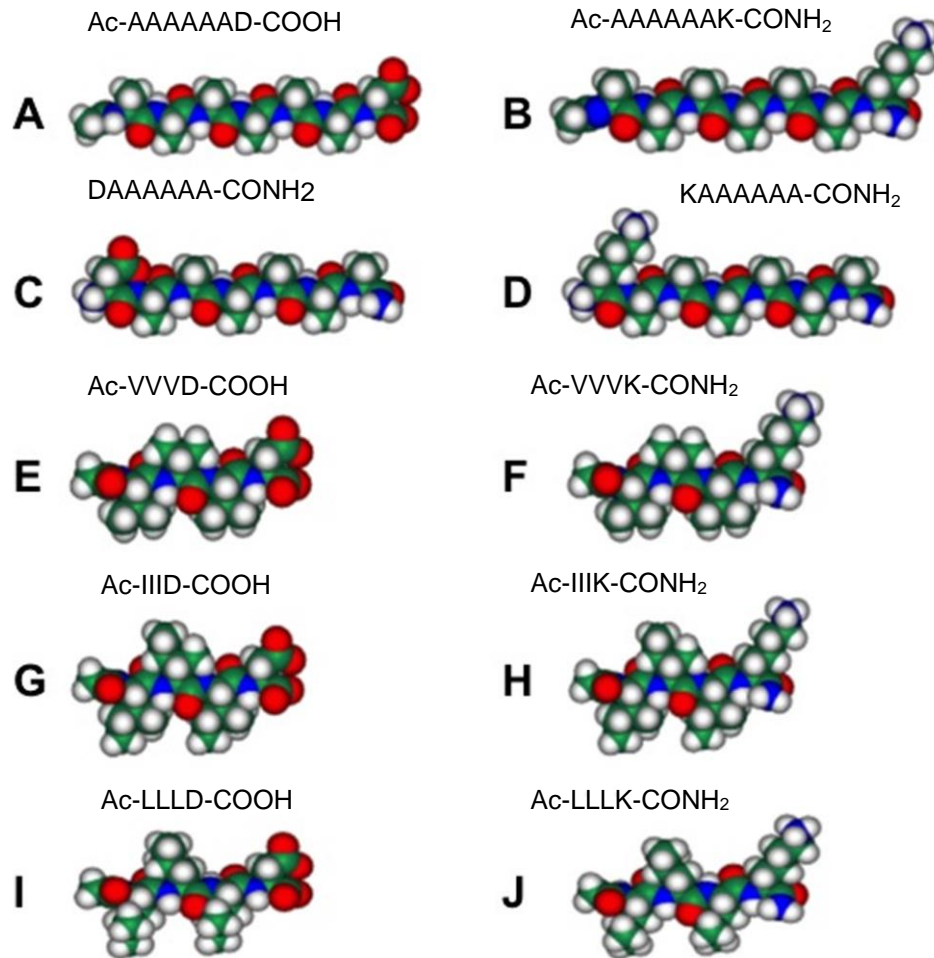


Figure 1.19. Different molecular structures of peptergents at neutral pH.

Amino acid residues are represented as a one-letter code. Aspartic acid and lysine are located at the peptide heads, whereas hydrophobic residues in the middle. Colour code: teal, carbon; red, oxygen; blue, nitrogen; and white, hydrogen. Adapted from Corin *et al.* (2011) ¹⁹⁵.

PART I - Studies of the secondary transporter

LicB

1.7. Colonisation of the human respiratory tract

The microbial flora of the human respiratory tract (RT) is usually settled during the first year of life, and its composition varies significantly with age ¹⁹⁷. It includes a diverse community of commensal or mutualistic bacteria as well as potential pathogens, for instance, *H. influenzae*, *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Moraxella catarrhalis* and *Pseudomonas*, *Actinobacteria*, and *Neisseria* species ^{198–202}. Most of the time, these microorganisms do not cause any clinical symptoms, being considered a normal part of the human microbiome. However, in some individuals, these ‘commensal pathogens’ can overgrow and invade to cause respiratory or invasive diseases, especially in adults with weak immune system or children with an immature immune system ^{200,203}.

In all cases, respiratory tract infections begin with the successful colonisation of the epithelial surface of the respiratory tract. Broadly speaking, the colonisation cycle involves the following phases: enter the lumen and pass through mucus, adherence of the pathogen to the mucosal surface, growth in the nutrient-limited environment, evasion of the host immune system and transmission to a new host (Figure 1.20) ²⁰⁴.

Because colonisation is a prerequisite for infection, colonising bacteria have evolved cell surface modifications that promote their adherence to the mucosal surface. At the same time, these modifications may also prevent antibody recognition and reduce host immune system responses. One central mechanism that underpins colonisation, which will be extensively discussed here, is modifying the bacterial surface with phosphorylcholine (ChoP). The structure of ChoP mimics host membrane lipids, such as phosphatidylcholine, supports binding to host epithelia, and can mask bacterial surface structures such as the lipopolysaccharides ²⁰⁵. Many of the commensal and pathogenic bacteria found in the respiratory tract can adapt their outer surface with ChoP. The rest of this introduction will focus on the biosynthesis and biological role of ChoP and describe previous studies on a membrane transport protein, known as LicB, that is essential in ChoP production and which is the primary concern of this thesis. This will be introduced with reference to the most well-characterised

example of LicB in *Haemophilus influenzae*. Therefore, special attention will be given to *H. influenzae* strains and the clinical manifestations associated with their colonisation in the human respiratory tract.

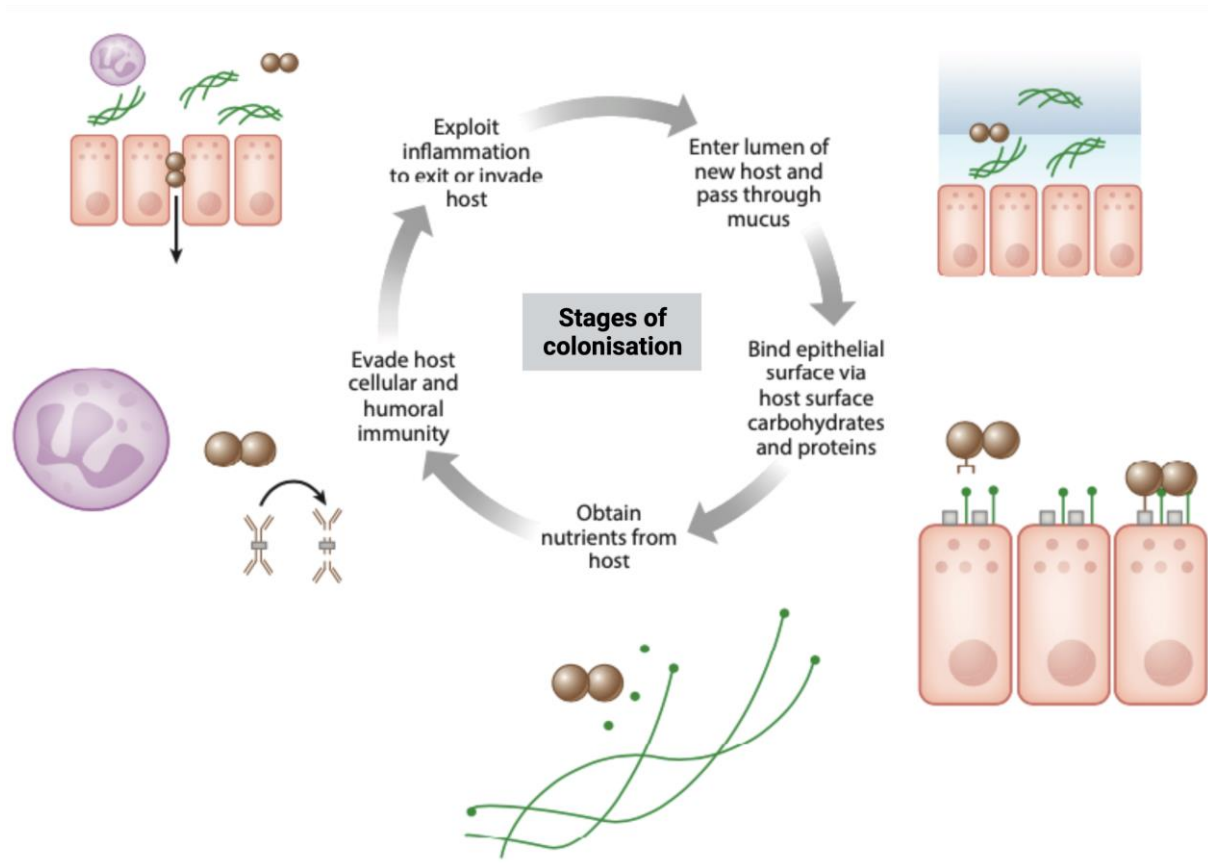


Figure 1.20. Main stages of colonisation in the human respiratory tract.

Adapted from Siegel and Weiser (2015) ²⁰⁴.

1.8. Types of respiratory tract infections

Respiratory tract infections (RTIs) are among the leading causes of mortality worldwide and are predicted to be in the top 5 leading causes of deaths in 2030 (Table 1.5) ²⁰⁶. Respiratory infections cause 100,000 hospital admissions and 5 million GP visits each year in the UK ²⁰⁷ and have a high socioeconomic cost associated ²⁰⁰. In 2014 conditions including asthma, chronic obstructive pulmonary disease (COPD) and lower RTIs cost the UK economy £6.6 billion ²⁰⁸.

Table 1.5. Rankings of the top 15 leading causes of death in 2002 and the predicted causes for 2030.Adapted from Mathers and Loncar (2006) ²⁰⁶.

Disease or injury	2002 rank	2030 rank
Ischaemic heart disease	1	1
Cerebrovascular disease	2	2
Lower respiratory infections	3	5
HIV/AIDS	4	3
COPD	5	4
Perinatal conditions	6	9
Diarrhoeal diseases	7	16
Tuberculosis	8	23
Trachea, bronchus, lung cancers	9	6
Road traffic accidents	10	8
Diabetes mellitus	11	7
Malaria	12	22
Hypertensive heart disease	13	11
Self-inflicted injuries	14	12
Stomach cancer	15	10

Respiratory infections are conventionally divided into upper respiratory tract infections (URTIs) and lower respiratory tract infections (LRTIs). URTIs affect the nose, sinuses, pharynx, larynx, and large airways, leading to sinusitis, rhinitis, pharyngitis, tonsillitis, and laryngitis ^{201,204}. In addition, URTIs in young children are often complicated by otitis media (OM) ²⁰⁹. On average, 700 million cases of OM have been reported every year worldwide ²¹⁰. On the other hand, LRTIs are infections in the lungs or below the voice box, causing pneumonia, bronchitis, tuberculosis, and exacerbations of chronic obstructive pulmonary disease (COPD). In worst cases, these can develop into invasive bloodstream infections and sepsis. Among all conditions, LRTIs have the highest impact on human well-being, with six and two-fold greater disability-adjusted life years (DALYs) than diabetes mellitus and ischaemic heart disease, respectively ²¹¹. In underdeveloped countries, pneumonia has been the leading cause of death in children under five years, higher than AIDS, malaria and measles combined ²¹².

In many cases, the molecular basis for these infections is still not well-understood. Nonetheless, it is known that other health conditions may induce a higher susceptibility to bacterial infections in the RT, as, for example, allergies, viral infections, smoking and airborne environment pollutions ^{197,201,203}. Both URTIs and LRTIs can be caused by the 'commensal

pathogens' of the nasopharynx. Generally, the nasopharyngeal flora serves as a source or reservoir from which bacteria may spread to the lower and upper respiratory tract ²⁰⁰. *S. pneumoniae*, *H. influenzae* (predominantly the non-typeable strains – discussed below), and *M. catarrhalis* have been recognised as the leading cause of both URTIs and LRTIs ^{200,213}. Table 1.6 represents examples of bacterial species commonly found in clinical isolates from RTIs.

Table 1.6. Main bacterial species present in otitis media, tonsillitis, sinusitis and pneumoniae.

RTI type	Clinical manifestation	Bacteria responsible	Ref.
URTIs	Otitis media	<i>S. pneumoniae</i>	210
		<i>H. influenzae</i>	
		<i>M. catarrhalis</i>	
	Tonsillitis	Group A <i>Streptococcus</i>	214
LRTIs	Community-acquired pneumoniae	<i>S. pneumoniae</i>	216
		<i>H. influenzae</i>	
		<i>M. catarrhalis</i>	
		<i>Streptococcus pyogenes</i>	

Although many of those RTIs can be successfully treated with β -lactam antibiotics, for instance, ampicillin or amoxicillin-clavulanic acid, the rise of antibiotic-resistant determinates such as β -lactamases makes bacterial clearance more difficult ²¹⁷. Pneumococcal vaccines have been highly effective but are now under pressure from the emergence of novel serotypes that escape vaccination ^{217,218}. Therefore, new antibiotics are urgently required to meet the twin threats of antimicrobial resistance and ineffective vaccines.

1.8.1. *H. influenzae* strains in RTIs – colonisation and disease

The *H. influenzae* strains are typically divided according to the presence or absence of a polysaccharide capsule. Encapsulated *H. influenzae* strains are called 'typeable', and six capsular types, designated with letters a-f, have been identified. Non-typeable (NTHi) strains are those that do not synthesise the polysaccharide ²¹⁷. Encapsulated type b (Hib) has been

responsible for the most virulent and invasive infections ^{219,220}. However, the Hib vaccine's effectiveness reduced the rate of Hib infections and moved the occurrence towards other serotypes; Around 44% of all *H. influenzae* clinical isolates were NTHi, being considered the most common pathogen ^{221–224}. This strain is now responsible for most URTIs, such as bronchitis, sinusitis and acute otitis media in children, and bacterial pneumonia in elderly patients ^{217,220,225,226}. Moreover, NTHi strains are frequently associated with COPD exacerbations ^{217,226–228}, which are predicted to increase by 2030 (Table 1.5) ²⁰⁶.

1.9. The outer cell wall of Gram-negative bacteria

Bacteria can be broadly classified according to their reaction under gram staining. Some examples of Gram-negative bacteria include species from *Pseudomonas*, *Neisseria*, *Haemophilus* or *Escherichia*. Conversely, species from *Staphylococcus*, *Lactobacillus*, *Bacillus* or *Streptococcus* are classified as Gram-positive bacteria ^{8,229–231}. This distinction is due to significant differences in the outer cell wall, or outer membrane, of the two bacteria groups.

In Gram-negative bacteria, the principal component of the outer membrane is the lipopolysaccharide (LPS) molecules. These amphipathic glycolipids play a crucial role in pathogenesis and are targeted by several antimicrobial agents – perhaps most famously penicillin ²³¹. Within this LPS layer sit pore proteins known as porins that allow nutrient diffusion and the efflux of waste. Underneath the outer membrane sits a thin layer of peptidoglycan (PGN) of around 7 - 8 nm thickness ^{229–231}. This occupies the periplasmic space between the inner and outer membranes. Contrarily, the outer membrane structure in Gram-positive bacteria is markedly different. This group of bacteria have a thicker PGN layer, between 40 to 80 nm, and a periplasm of much smaller volume. The PGN layer consists of alternating units of N-acetyl muramic acid and the disaccharide N-acetyl glucosamine, cross-linked by a pentapeptide side chain (Figure 1.21) ²²⁹.

Moreover, Gram-positive bacteria are also unique in having teichoic acids (TA) or lipoteichoic acid (LTA) in the cell wall, which is sometimes linked via a glycolipid anchor to the plasma membrane ²³². Another critical structural feature of some bacteria is the bacterial capsule. This structure consists of branched or unbranched homo- or heteropolymers, normally polysaccharides, and is located outside the cell membrane or covering the LPS.

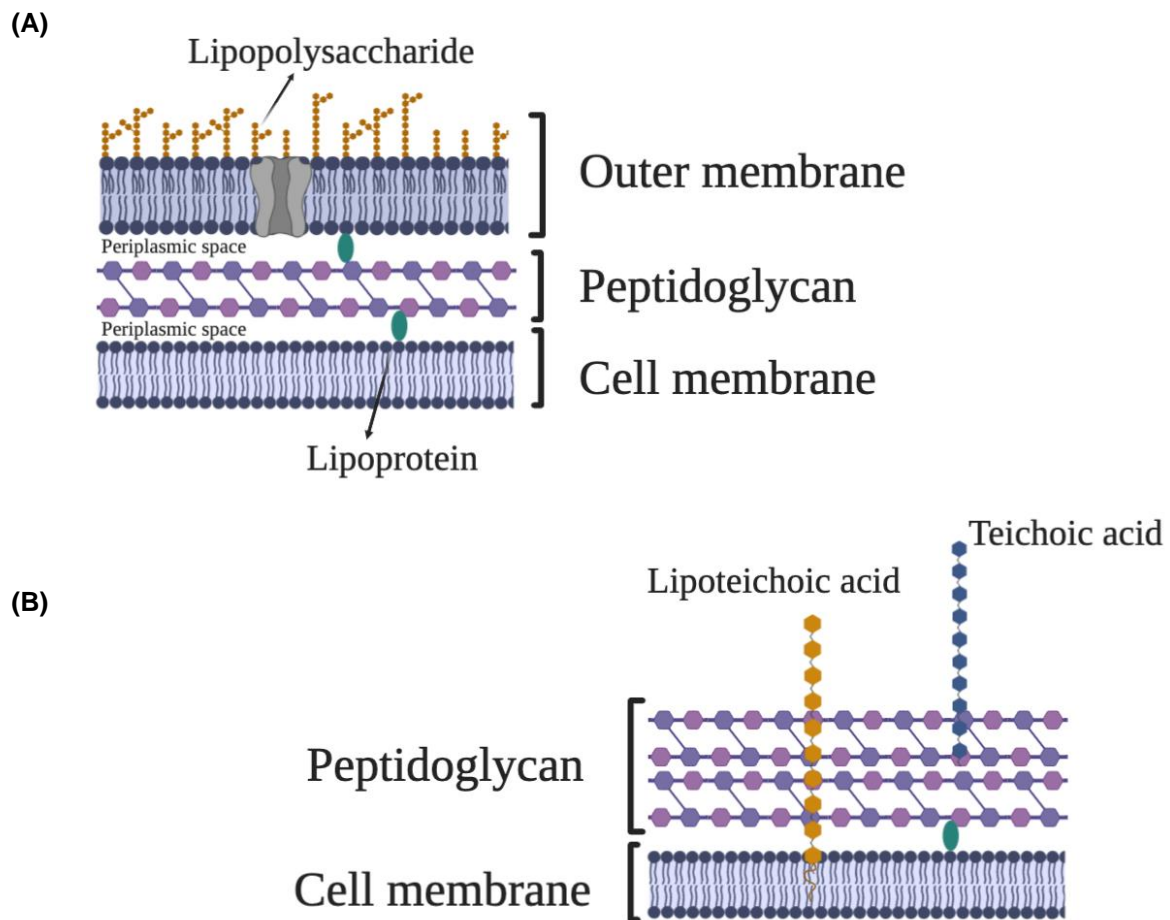


Figure 1.21. Bacteria cell wall structures.

(A) Gram-negative and (B) Gram-positive bacteria. Adapted from Brown *et al.* (2015) ²³².

1.10. Virulence factors in colonising bacteria

Virulence factors are molecules that allow colonising bacteria to successfully adapt to host environments, avoid the host immune system and invade host cells ²⁰⁴. Although virulence factors may vary depending on the colonising bacteria, the most common ones are the polysaccharide capsule, adhesins, LPS structure and outer membrane proteins ^{204,233}. The LPS comprises the Lipid A, the inner core and the O-antigen and is a potent activator of the host pro-inflammatory immune responses through the TLR4-MD2 signalling pathway ^{234–237}. An additional molecule, phosphorylcholine (ChoP), also occurs as a component of LPS structure, is of particular relevance to this thesis, and it will be discussed in greater detail below.

1.10.1. Structural differences between LPS and LOS

The LPS O-antigen is constituted of a variable number of repeating saccharide units and, since it is exposed at the outer surface of the bacteria, is a target for recognition by the host immune system²³⁶. The inner core comprises phosphorylated 2-keto-3-deoxyoctulosonic acid (KDO) linked to heptose units. The outer core of LPS comprises a glucose and galactose heteropolymer (Figure 1.22A)²³⁶. A subset of Gram-negative bacteria (*Neisseria*, *Haemophilus*, *Bordetella*, and *Moraxella* spp.) which inhabit mucosal surfaces express predominantly an analogue of the LPS, the lipooligosaccharide (LOS)^{238,239}. LOS does not contain the O-antigen oligosaccharide units and varies substantially between strains, precisely the number of heptose units (Figure 1.22B)^{240,241}. LOS comprises a lipid A attached to two KDO units and two conserved heptose units in *Neisseria* species. In contrast, the *H. influenzae* LOS consists of three heptose moieties and a single KDO^{238–240}.

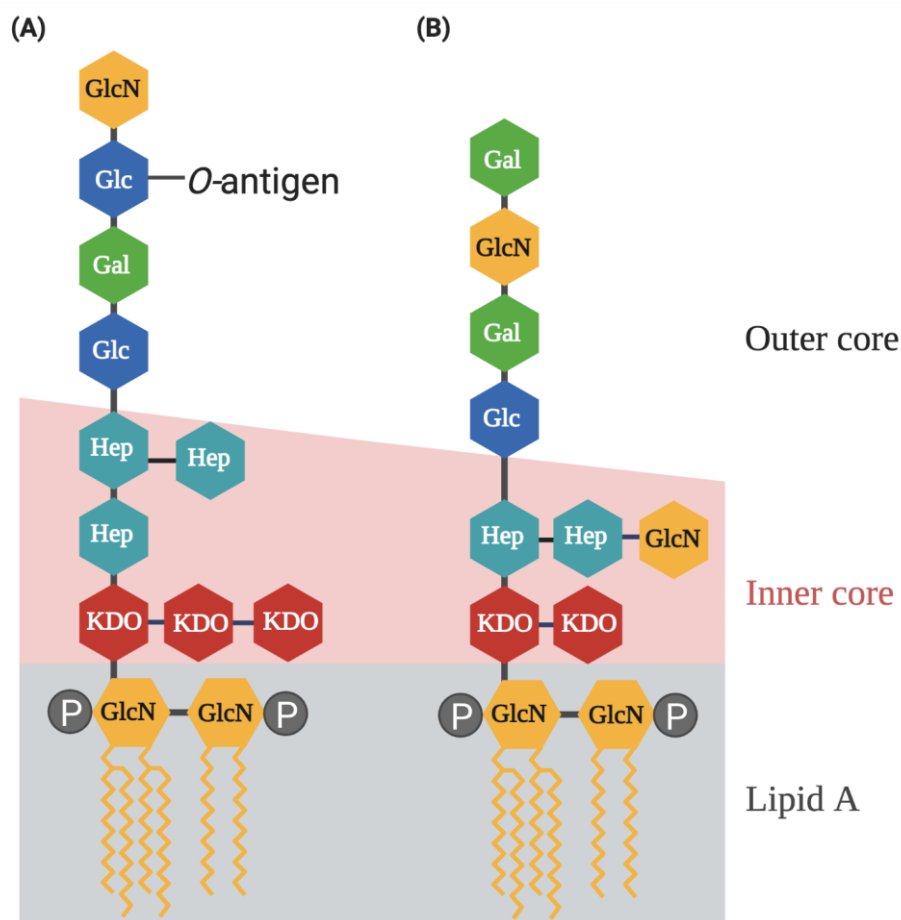


Figure 1.22. General structures of LPS.

(A) *E. coli* LPS structure and (B) *Neisseria gonorrhoeae* LOS structure. GlcN, glucosamine; KDO, 2-keto-3-deoxyoctulosonic acid; Hep, L-glycero-D-mannoheptose; Glc, glucose; Gal, galactose. Adapted from Preston *et. al* (1996)²³⁸.

1.10.2. Main LOS modifications

The glyco-modifications of *H. influenzae* provide a high degree of heterogeneity in LOS, including glucose, galactose, lactose, phosphorylcholine and sialic acid, leading to a heterogeneous population of LOS structures within a single strain^{240,242–245}. As a result, Weiser successfully demonstrated that this is a significant cause of surface antigenic variation²⁴⁶.

To date, the surface variation identified in *H. influenzae* strains is due to *IgtC*, *lic1*, *lic2*, *lic3*, and *lex2* loci, among others^{247–250}. The *lic1* locus (number M37912) is responsible for adding ChoP on LOS^{242,243}, whereas *lic2* and *IgtC* loci are required to insert Gal α 1-4Gal to LOS^{248,251}. The *lic3* locus is involved in adding sialic acid to terminal sugars^{252,253}, and *lex2* is essential in further oligosaccharide extension from Hep_I²⁵⁴. ChoP modification has been identified on all three heptoses, but the position of this modification may be strain-specific^{255,256}. For example, ChoP is linked to chain extension of Hep_I in strain Rd and in the majority of NTHi strains, but to Hep_{III} in Hib and Eagan strains^{240,257}. ChoP is a zwitterionic molecule that binds to LOS covalently via its phosphate group²⁰⁵. Figure 1.23 pinpoints some of the loci responsible for LOS phase variation in NTHi *H. influenzae* strains.

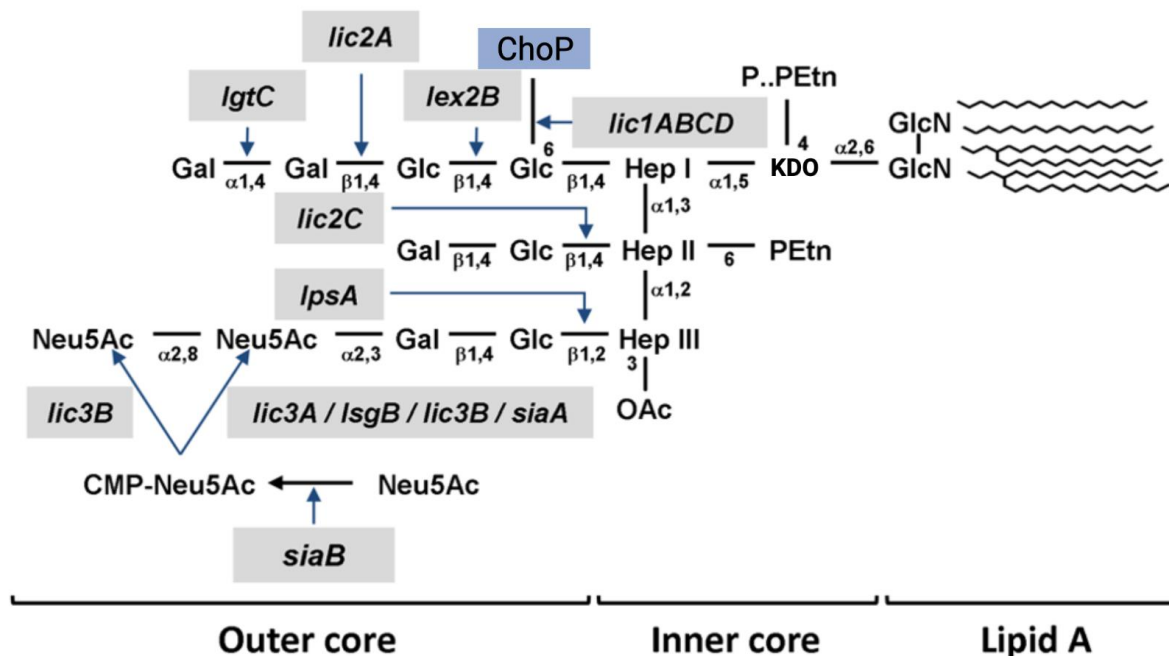


Figure 1.23. Loci responsible for LOS phase variation in NTHi *H. influenzae* strain.

Abbreviations: KDO, 2-keto-3-deoxyoctulosonic acid; Hep, L-glycero-D-manno-heptose; Glc-D-glucose; Gal, D-galactose; PEtn, phosphoethanolamine; P, phosphate; ChoP, phosphorylcholine. Adapted from Martí-Literas *et al.* (2011)²⁵⁰.

1.11. Phosphorylcholine is an important and widespread virulence factor

As well as *H. influenzae*, several other bacteria that colonise and persist in the human RT are able to decorate their cell surface with ChoP^{258–260}. Among these are *Streptococcus pneumoniae*, *Neisseria meningitidis* and *Pseudomonas aeruginosa* (Table 1.7). One of the strategies developed by ChoP expressing bacteria is to vary ChoP location. While in *H. influenzae*, ChoP is found as a modification to the LPS, in *S. pneumoniae*, ChoP is incorporated into TA or LTA, and in *P. aeruginosa*, ChoP is attached to the elongation factor Tu. Additionally, this variation is also observed between different *Neisseria* species. This diversity in the attachment of the ChoP molecule is due to the enzyme LicD, which will be discussed in Section 1.12.2.

Table 1.7. Bacteria that express ChoP modified structures.

Species	Cell structure with ChoP	Ref.
<i>Streptococcus pneumoniae</i>	TA, LTA	258
<i>Haemophilus influenzae</i>	LPS (or LOS)	239
<i>Haemophilus haemolyticus</i>	LPS (or LOS)	261
<i>Pseudomonas aeruginosa</i>	Elongation factor Tu	260
Commensal <i>Neisseria</i> (<i>lactamica</i> , <i>flavescens</i> or <i>subflava</i>)	LPS (or LOS)	262,263
<i>Neisseria meningitidis</i>	Pilus	260
<i>Neisseria gonorrhoeae</i>	Pilus	260
<i>Mycoplasma fermentans</i>	Membrane proteins	264
<i>Morganella morganii</i>	O-antigen	265

Moreover, the expression of ChoP undergoes ‘phase variation’, which is another strategy used by bacteria to adapt to different host environments and persist in the human RT^{239,266,267}. Phase variation is defined by the acquisition or loss of a defined structure spontaneously, and is usually a reversible and random event at high frequency ($>10^{-5}$ per generation)^{247,267,268}. Focusing on *H. influenzae*, ChoP phase variation allows the organism to switch ChoP expression ‘on and off’ during infection within the bacteria population – resulting in stochastic variation in ChoP attachment²⁶⁸. This phenomenon is due to the presence of tandem repeats of the tetramer (5'-CAAT-3')_n in the *licA* gene, and is also observed in commensal *Neisseria*

species^{239,242,243,262,263}. Briefly, when the number of CAAT repeats is in frame, ChoP is phase 'on' and expresses ChoP, whereas ChoP is not expressed when the number of repeats is 'out' of frame²⁶⁸. Furthermore, Weiser *et al.* demonstrated that bacteria expressing the *licA* gene are predominantly isolated in the nasopharynx²⁶⁹. Moreover and despite phase variation, *H. influenzae* expressing ChoP were the dominant population isolated from colonised mice and humans^{250,251,269}.

1.11.1. Phosphorylcholine supports cell adhesion and affects host recognition

Several independent studies have suggested that bacterial ChoP has a distinctive contribution to RT colonisation. Firstly, the ChoP molecule 'anchors' to the epithelial cell surface protein that is the receptor for platelet-activating factor (rPAF). This specific binding event enhances bacterial adherence to the RT epithelial surface and subsequent invasion of deeper epithelial tissue (Figure 1.24A)^{270–272}. This 'molecular mimicry' occurs because ChoP is part of the natural ligand for the PAF receptor (Figure 1.24B).

The rPAF is a G-protein-coupled receptor present in a range of cell types, including endothelial and epithelial cells, monocytes, macrophages and neutrophils. rPAF is linked intracellularly to signal transduction pathways²⁷³. PAF binding elicits a range of pro-inflammatory responses against bacterial infections, including the release of cytokines and toxic oxygen metabolites, as well as the induction of migration and degranulation of monocytes, macrophages and granulocytes^{273–275}.

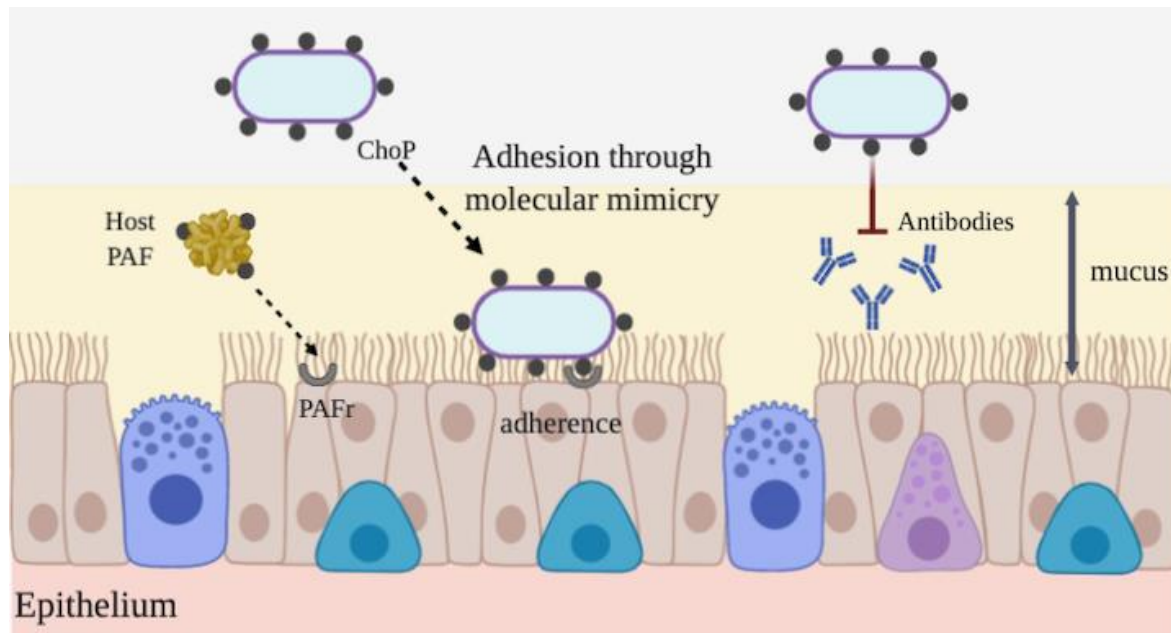
H. influenzae, *S. pneumoniae*, *Aggregatibacter actinomycetemcomitans*, *Histophilus somni*, *P. aeruginosa* and commensal *Neisseria* strains bind to rPAF *in vitro*^{262,270,272,276}. However, *in vivo* experiments in mice deficient in rPAF showed unaltered colonisation levels of NTHi strains from their RT²⁷⁷. Therefore, additional host factors might also be involved in the adherence and invasion of *H. influenzae* strains²⁷⁸.

Secondly, ChoP also appears to have an essential role during disease development. The expression of ChoP by *H. influenzae* was associated with enhanced persistence of otitis media in the chinchilla model system²⁷⁹ and in children²⁸⁰. Moreover, ChoP promotes bacterial resistance in the lungs of infected mice²⁸¹. Similar advantages were observed in other bacteria, such as in *Histophilus somni*, in which ChoP expression improved colonisation of its

bovine host²⁸². Consistent with this, *S. pneumoniae* strains that do not synthesise ChoP-modified TA are less infectious in a murine sepsis model and are unable to colonise the URT in mice²⁸³.

Thirdly, it has been suggested that ChoP expressing bacteria resists to components of the innate immune system such as antimicrobial peptide LL-37/hCAP18²⁸⁴. This is achieved by mimicking the ChoP on phosphatidylcholine in eukaryotic membrane cells, and therefore makes the synthesis of ChoP a potential therapeutic target^{285–287}. Moreover, ChoP-LPS structures modify the physical properties of the outer membrane and consequently decrease cell membrane accessibility and permeability; Thus, reducing antibody binding to LPS (Figure 1.24A)²⁸⁸. Finally, ChoP expressing NTHi strains more readily form antibiotic-resistant biofilms *in vitro* and *in vivo*^{289–291}.

(A)



(B)

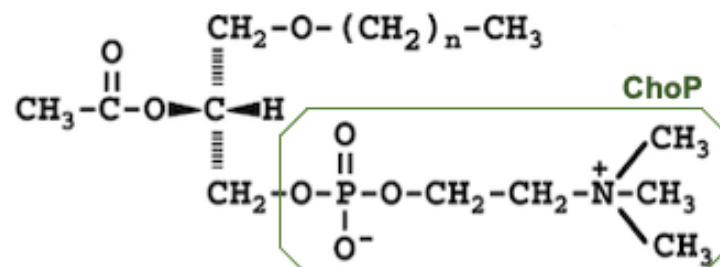


Figure 1.24. Schematic representation of the role of ChoP in *H. influenzae*.

(A) ChoP supports bacterial adherence via a direct interaction with rPAF, and blocks antibody binding. Adapted from Clark and Weiser (2013)²⁰⁵. **(B)** Chemical structures of PAF (1-O-alkyl-2-acetyl-*sn*-glycero-3-phosphocholine); $n = 15 - 17$, in which the green part represents ChoP.

1.11.2. Host defenses against phosphorylcholine expressing bacteria

Despite the several advantages that ChoP confers to colonising bacteria during infection, its expression may only benefit specific host environments. Two major host factors are able to recognise ChoP epitopes in colonising bacteria: the C-reactive protein (CRP) and the anti-ChoP antibodies.

CRP is a constituent of normal serum and is able to recognise both phosphoethanolamine and ChoP in a calcium-dependent manner. CRP is found in the human upper and lower airway surface fluid in quantities capable of bactericidal activity against *H. influenzae* and *S. pneumoniae* expressing ChoP^{269,292,293}. In *H. influenzae* and after binding to ChoP, CRP activates the complement cascade by interaction with the complement component C1q (Figure 1.25)²⁹⁴. Ultimately, this complement activation process leads to opsonisation and clearance of the colonising bacteria from the human RT. Lysenko *et al.* demonstrated that the location of ChoP on the *H. influenzae* LPS influences the degree of CRP-mediated bacterial killing²⁵⁷. Besides, during an infection, when CRP levels are enhanced, ChoP expressing bacteria are more vulnerable to complement-mediated killing than bacteria that lack this epitope^{239,266}. Having ChoP expressing bacteria, on the other hand, is advantageous in mice and rabbit models where CRP remains low^{295,296}. The binding of CRP to ChoP was inhibited in the presence of human surfactant (which contains a substantial quantity of dipalmitoyl phosphatidylcholine), suggesting that abundant surfactant in the terminal airway may diminish the antiadhesive effect of CRP²⁹⁷.

Concerning the second host component, the natural antibodies involved in ChoP recognition are from the B1 lineage (IgM, IgA and IgG). B1 antibodies are part of the innate immune response and have a weaker affinity than antibodies produced after a specific infection²⁹⁸. Originally identified in mice, the most well-known idotype of the anti-ChoP is the T15^{299,300}. Briles *et al.* confirmed that anti-ChoP could offer some defence since *xid* mice are more vulnerable to *S. pneumoniae* infections^{301–303}. Moreover, the same susceptibility was observed in T15-deficient mice³⁰⁴. Although it is unknown if humans produce the T15 idotype³⁰⁴, detectable anti-ChoP IgM and IgG levels have been identified in immunocompetent adults²⁸⁶, and in children after *S. pneumoniae* colonisation³⁰⁵. Moreover, it has been demonstrated that human IgG2 successfully recognised bacteria with ChoP-modified structures, specifically *H. influenzae* and *S. pneumoniae*²⁸⁶. Additionally, antibodies against ChoP might have a protective effect on cardiovascular and atherosclerosis diseases^{306–308}.

In addition to the mechanisms described above, ChoP can modulate host responses during infection. However, all examples to date describe this modulation process in parasites. The most studied example of a ChoP-modified immunomodulatory molecule involves the secretory product of nematodes, ES-62. Although this topic falls outside the scope of this literature review, there are some compelling articles highlighting the role of ChoP towards ES-62^{309,310}.

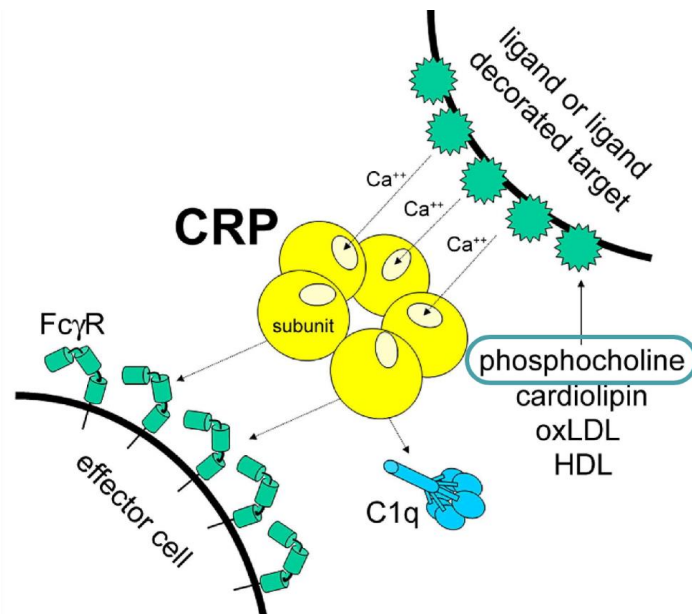


Figure 1.25. Schematic representation of the recognition of ChoP by CRP and the complement activation molecule C1q.

Adapted from Hage and Szalai (2007)³¹¹.

1.11.3. Current and future perspectives on ChoP based vaccines

Currently, there is no available vaccine that targets ChoP directly. There are commercial vaccines, such as the 10-valent pneumococcal non-typeable *H. influenzae* protein D conjugate vaccine (PHiD-CV10), which inhibits the enzyme involved in choline scavenging from the host (Protein D), or the 13-valent pneumococcal conjugate vaccine (PCV13)^{312–314}. The latest research has shown the importance of this protein since inhibition of Protein D function decreased ChoP expression as well as epithelial cell adhesion²⁸⁵. Although remarkable results were achieved^{315–317}, some *S. pneumoniae* or *H. influenzae* strains escaped vaccination. Alternatively, vaccine formulations that effectively induce the production of anti-ChoP could be used to immunise against a variety of bacteria expressing the ChoP epitope³¹⁸. Since immunisation with ChoP combined with a protein carrier, both normal and

xid mice became more resistant to *S. pneumoniae* infection^{319,320}. More recently, Ohori *et al.* demonstrated that an additional immunisation with ChoP after PCV13 could enhance host response against *S. pneumoniae* in mice³²¹. This reflects the potential for the development of antimicrobial vaccines and therapies that directly target ChoP. Vaccines are highly effective but could extinguish the respiratory tract's normal flora and escape phase variation. Small molecule therapies that target pathway enzymes to interfere with ChoP production are thus an increasingly attractive proposition.

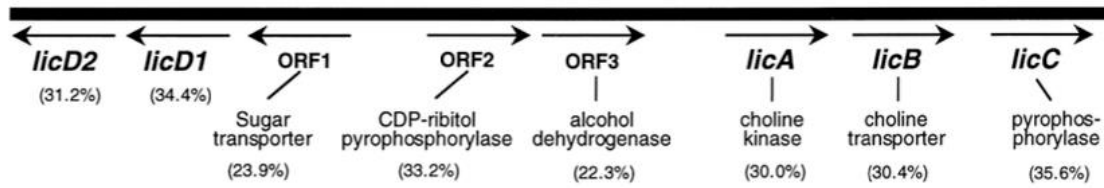
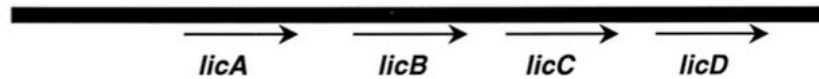
1.12. The biosynthesis of phosphorylcholine modified structures

ChoP synthesis by the human respiratory pathogens requires choline as the primary substrate. However, these pathogens cannot synthesise the choline precursor *de novo* but instead must obtain it from the host environment. When choline is available in the extracellular milieu, it is transported into the cell, enzymatically transformed into ChoP and added to specific bacterial structures through the activity of the gene products of the *lic1* operon. This comprises four genes, *licA*, *licB*, *licC*, and *licD*²⁰⁵.

The *lic1* operon was first described by Weiser *et al.* in *H. influenzae*, and homologues of the four genes were also found in *S. pneumoniae* and commensal *Neisseria* strains^{239,242,263,322}. Weiser *et al.* confirmed the role of *lic1* operon by introducing site-specific mutations in the different loci described in Figure 1.23 and subsequent transformation of *H. influenzae*, resulting in the expression of different oligosaccharides within LPS²⁴². Thus, the expression of ChoP was evaluated by immunoblotting using monoclonal antibodies (MAbs) such as 4C4, 12D9 and 6A2. Kimura and Hansen have been previously described the use of MAbs to study phase variation³²³, and more recently used by Humphries and High to confirm the pathogenesis of Hib strains²⁶⁶.

In *S. pneumoniae*, the *lic* locus includes eight genes (*licA*, *licB*, *licC*, two *licD* homologues and three ORFs), whose functions are still under research (Figure 1.26A)^{322,324}. However, this operon is absent in pathogenic *Neisseria* and *P. aeruginosa*, and the genetic and biochemical basis for the biosynthesis of phosphorylcholine in these organisms undergoes different pathways^{241,262}. Thus, the minimal *lic1* operon from *H. influenzae* has become the model system for understanding ChoP biosynthesis. The overall biosynthetic pathway is represented in Figure 1.26B and briefly discussed below.

(A)

S. pneumoniae*H. influenzae*

(B)

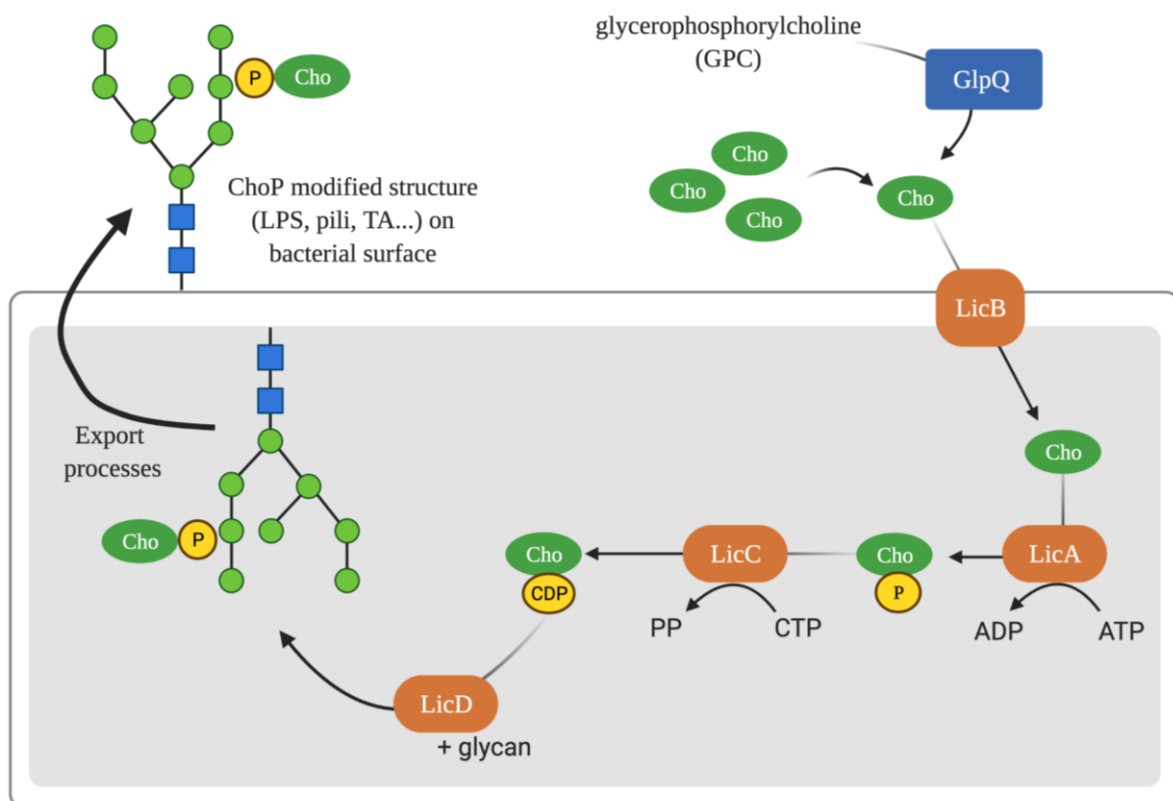


Figure 1.26. The biosynthetic pathway for ChoP.

(A) The organisation of the *lic* locus in *S. pneumoniae* and *H. influenzae*. Adapted from Zhang *et al.* (1999)³²². (B) Schematic of the ChoP pathway in *H. influenzae*. Gene products of the *lic1* operon are shown in orange. Abbreviations: Cho – choline, P – phosphate. Adapted from Young *et al.* (2013)³²⁵.

1.12.1. Phosphodiesterases involved in choline uptake

The first step in the ChoP pathway is dependent on the uptake of choline from the environment. The two primary sources of choline molecules are the glycerophosphorylcholine (GPC) from host membrane cells or free choline from the milieu (Figure 1.26). In order to release choline from GPC, *H. influenzae* uses the *glpQ* gene product, a glycerophosphodiester phosphodiesterase³²⁶. Several studies have demonstrated that this enzyme is required to acquire choline directly from epithelial cells³²⁷. This gene is highly conserved among *H. influenzae* species³²⁸ and is homologous to the *E. coli* enzyme, whose crystal structure has been determined (PDB entries 1YDY and 1T8Q). While the *E. coli* enzyme occurs in the periplasm, the *H. influenzae* homologue (previously named Protein D) is localised to the outer membrane and exposed to the cell surface³²⁹. More recently, Alrousan and Fan have demonstrated that the *glpQ* gene is down-regulated by *H. influenzae* under nutrient-limited conditions but up-regulated when the substrate GPC is present³²⁶. The exact mechanism by which this gene expression is controlled still requires further investigation.

Choline liberated by this enzyme is then used as a substrate by enzymes encoded by *lic1* operon, which are introduced below.

1.12.2. Enzymes of the *lic1* operon

The four proteins encoded by the prototypical *lic1* operon of *H. influenzae* were first appointed based on homology studies of known proteins²³⁹. Three of these proteins - LicA, LicC and LicD - have now been well-characterised *in vitro*, whereas the function of LicB has been hypothesised based on homology analysis and in *in vivo* studies.

1.12.2.1. *licA* gene

The first gene in the *lic* operon encodes the choline kinase LicA. This enzyme phosphorylates choline to form ChoP, with accompanying hydrolysis of ATP^{330,331}. LicA homologues have been identified in the genomes of *S. pneumonia* and *Histophilus somni*³³², among others. The structure of this soluble LicA protein from *S. pneumoniae* has been determined by crystallography³³¹. Intriguingly, in *F. nucleatum*, *C. perfringens* and *T. pallidum*, LicA and LicC result in a bifunctional enzyme³³³.

1.12.2.2. *licB* gene

The second gene, *licB*, is thought to encode a high-affinity choline transporter and is the central focus of this thesis. Although the *licB* gene has some sequence similarity to other choline transporters, it is not closely related to transporters of the Betaine-Choline-Carnitine-Transporter family³³⁴ (see Section 1.13.1). Thus, it appears to represent a novel protein family known as the Choline Uptake Transporter (LicB-T) family³³⁵. Further discussion on this transporter is given in Section 1.13.2.

1.12.2.3. *licC* gene

LicC protein is the most well-characterised in this operon. The *licC* gene encodes a predicted CTP:phosphocholine cytidyltransferase (CCT), producing CDP-choline from ChoP³³⁶. The CCT enzyme was found in cell extracts of *S. pneumoniae* with K_M values for CTP and ChoP of $890 \pm 240 \mu\text{M}$ and $390 \pm 170 \mu\text{M}$, respectively³³⁷. Moreover, its three-dimensional structure has already been determined³³⁸. LicC of *H. influenzae* is 37% identical and 60% similar in their amino acid sequence to that of *S. pneumoniae*, and consequently, they are predicted to have similar functions³³⁸.

1.12.2.4. *licD* gene

The *licD* gene is an NDP-choline transferase that attaches ChoP to different cell structures, like LPS in the case of *H. influenzae*. In this microorganism, the location of ChoP attachment itself is variable, and it has been correlated with the existence of multiple alleles of *licD* genes²⁵⁷. This suggests that the *licD* allele carried by a given *H. influenzae* strain determines the position of ChoP in LPS (e.g., heptose unit as previously mentioned). Moreover, some NTHi strains have a duplicate *licD* gene leading to the attachment of two ChoP molecules per LPS structure³³⁹. Additionally, the two forms of *S. pneumoniae* *licD* gene (*licD1*, *licD2*) have a high sequence similarity to *H. influenzae* *licD*^{322,324,325}. Multiple sequence alignment between these three genes and the *cpsG* gene, which is responsible for the biosynthesis of the polysaccharide capsule, revealed some homology^{322,340,341}.

Additionally, LicD protein contributes to the maturation of stable communities in biofilms since biofilms from chinchillas infected with *H. influenzae* *licD* Δ strain were not discernible compared to WT infection²⁹¹. In parallel, overexpression of LicD has been linked to a higher incidence of respiratory infection³⁴².

1.13. LicB and its relationship with other choline transporters in bacteria

1.13.1. The BCCT family

The BCCT (Betaine-Choline-Carnitine-Transporter) family (2.A.15) can be found in Gram-positive, Gram-negative bacteria as well as archaea, and are the archetypal transporters of choline, L-carnitine or glycine betaine across cellular membranes. This family currently has about 2200 representatives^{40,44,334}. In general, individual BCCT proteins have narrow substrate specificity, in contrast to the broader substrate profile of the ABC or MFS transporter families³³⁴. The BCCT family is generally involved in osmoregulation. Examples of BCCT proteins, the corresponding substrate and energetic driving force are listed in Table 1.8.

Table 1.8. Examples of BCCT family proteins.

Name	Organism	Responds to	Substrate	Powered by	Ref.
BetT	<i>Escherichia coli</i>	Osmotic stress	Choline	Proton motive force, symport	343
BetL	<i>Listeria monocytogenes</i>	Osmotic stress	Glycine betaine	Sodium motive force, symport	344
CaiT	<i>Escherichia coli</i>	Substrate availability	L-carnitine and γ -butyrobetaine	Precursor: product antiport	345
BetP	<i>Corynebacterium glutamicum</i>	Osmotic stress	Glycine betaine	Sodium motive force, symport	346
EctP	<i>Corynebacterium glutamicum</i>	Osmotic stress	Glycine betaine, proline, ectoine	Sodium motive force, symport	347
OpuD	<i>Bacillus subtilis</i>	Osmotic stress	Glycine betaine	Sodium motive force, symport	348
BetT	<i>Pseudomonas syringae</i>	Osmotic stress	Choline, acetylcholine	Proton motive force, symport	349

The BCCT proteins are permeases comprised of between 481 to 706 amino acid residues. They are characterised by having twelve putative TM helices with hydrophilic N- and C-termini of various lengths orientated towards the cytoplasm^{334,335}. Independent studies have suggested that these terminals play a crucial role in osmosensing and/or osmoregulation of

transporter activity, for example, in BetP of *Corynebacterium glutamicum*³⁵⁰, BetT of *E. coli*³⁵¹ and *Pseudomonas syringae*³⁴⁹.

Two well-characterised examples of BCCT proteins are CaiT and BetP, mentioned above. The *E. coli* transporter CaiT catalyses the exchange of L-carnitine with γ -butyrobetaine independently of electrochemical gradient³⁵². On the other hand, BetP of *Corynebacterium glutamicum* is responsible for importing glycine betaine molecules, which is accompanied by the uptake of two sodium ions^{346,353}. Crystallographic studies of BetP (PDB entry 2WIT^{354,355}) and CaiT (PDB entries 3HFX³⁵⁶, 2WSX and 2WSW³⁵⁷) have provided the molecular basis of how this transport is achieved. In most of these transporters, three tryptophan residues form an aromatic cage that engages with the trimethylamine head groups of choline, betaine and L-carnitine, via cation- π and Van der Waals interactions (Figure 1.27)^{334,358}. Additionally, it is thought that salt bridges, hydrogen bonds, and water networks fix and orient the protruding ‘tails’ of these compounds within the binding site^{334,359}.

E. coli BetT is a choline transporter. This substrate is oxidised to produce glycine betaine acting as an osmoprotectant under osmotic stress conditions^{343,351}. Moreover, during aerobic growth, when choline is available and in the presence of high salt concentrations, the *betT* gene is up-regulated³⁶⁰. Fan *et al.* suggested that *betT* in *H. influenzae* may share a similar function as in *E. coli* – choline is used to generate osmoprotectant³⁶¹.

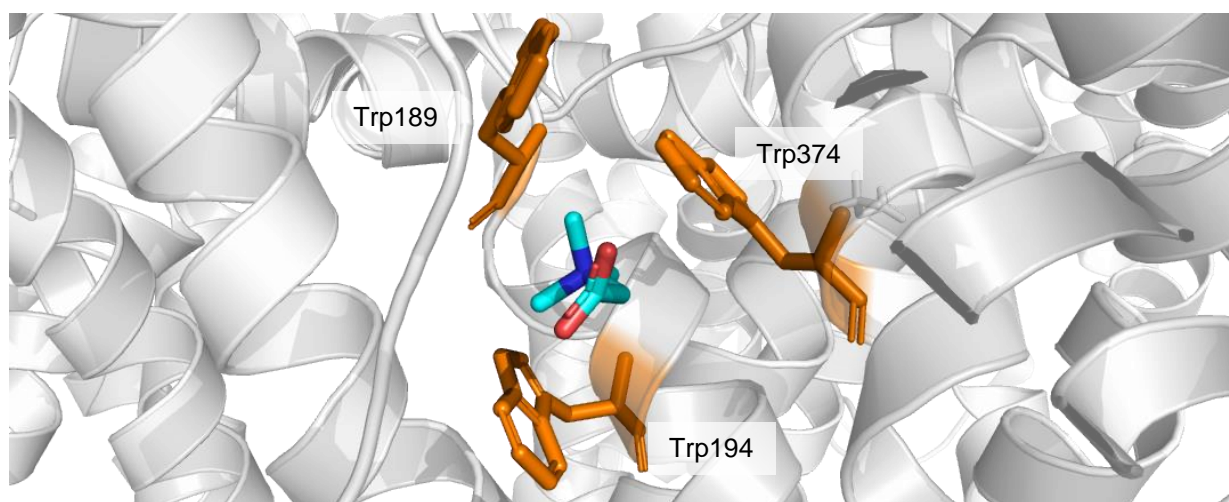


Figure 1.27. Betaine binding site in BetP transporter.

The three tryptophan residues highlighted in orange form the ‘aromatic box’ - Trp 189 and Trp 194 (in transmembrane helix 4), and Trp 374 (in transmembrane helix 8). Figure produced by Pymol software (PDB entry 2WIT).

1.13.2. The uptake of choline by LicB in *H. influenzae*

Although early studies assigned *H. influenzae* LicB to the BCCT family, LicB has a poor identity to BetT in *H. influenzae* (16.1% identity over 47/292 amino acids) and other members^{334,343,361}. Moreover, LicB in *H. influenzae* lacks the Trp motif, which is characteristic of the BCCT family (see Figure 4.14).

Fan *et al.* studied choline uptake by LicB and BetT in *H. influenzae* strains (Figure 1.28)³⁶¹. They confirmed that when choline was present at high levels, these two transporters were not required to synthesise LPS-ChoP. However, LicB was shown to be a high-affinity transporter essential for choline uptake when environmental concentrations were limited. Since choline is expected to be only present at low concentrations in the human nasopharynx, it was suggested that under physiological conditions, LicB is the primary route of choline acquisition by *H. influenzae*. This hypothesis was confirmed when under these restricting conditions, *H. influenzae licB*Δ strains showed a reduced ability to synthesise and attach the ChoP molecule to LPS structure³⁶¹. Additionally, *in vivo* studies have confirmed the prominent virulence factor of LicB during *H. influenzae* infection in a mice lung model²²³. Other work from Liu *et al.* confirmed that *licB* should be considered an essential gene for the proper growth and development of *H. influenzae* in particular and other pathogenic bacteria in general³⁶².

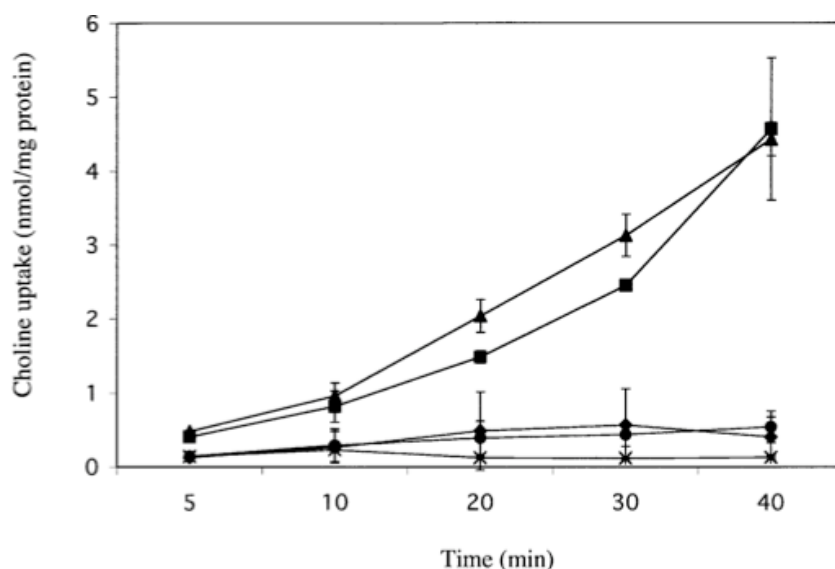


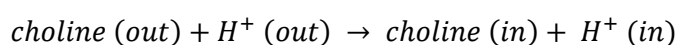
Figure 1.28. The uptake of [¹⁴C]-choline into *H. influenzae* strain was dominated by LicB.

Symbols: ▲ Rd parent strain; ■ *betT*Δ; ◆ *licB*Δ; ● *licB*Δ *betT*Δ; ※ heat-killed Rd. Adapted from Fan *et al.* (2003)³⁶¹.

1.13.3. The LicB transporter

LicB of *H. influenzae* is now recognised as the prototype of the Choline Uptake Transporter (LicB-T) family (2.A.7.18)³³⁵, which is included in the DMT superfamily. LicB has 292 amino acids and ten putative transmembrane alpha-helices identified by bioinformatics. LicB protein has been identified not only in *H. influenzae*, but also in *S. pneumoniae*, commensal *Neisseria* species, *Histophilus sommi*, among others³³³. Moreover, a GFP fusion to LicB supported the localisation of this protein to the periplasmic membrane³⁶³.

Overall, the proposed reaction catalysed by the LicB transporter family is³³⁵:



However, there is no experimental evidence to prove this mechanism. Moreover, no LicB family protein has yet been studied *in vitro*. The only transport assay mentioned in the literature was developed by Fan *et al.*, which was conducted *in vivo*, and the kinetic parameters K_M and V_{max} were calculated by measuring the uptake of [methyl-¹⁴C] choline by *H. influenzae* from the media (Table 1.9)³⁶¹.

Table 1.9. Kinetic parameters determined from transport assays in cell strains.

Data from Fan *et al.* (2003)³⁶¹.

Strain	K_M (μM)	V_{max} ($\text{pmol}\cdot\text{min}^{-1}\cdot\text{mg}_{\text{protein}}^{-1}$)
ChoP ⁺ phase variant of strain Rd (parent strain)	3.6	140
<i>betT</i> ⁺ <i>licB</i> ⁻	27.7	Data not shown

1.14. Project aims

LicB clearly plays an important role in the bacterial colonisation of the human respiratory tract and subsequent pathogen infection. However, little is known about the structure and function of LicB. This project aims to bridge this knowledge gap by carrying out the first biophysical and biochemical characterisation of LicB *in vitro*. This will establish LicB from *H. influenzae* as the model for understanding this protein family. The research questions to be addressed are:

- 1. Does LicB bind to choline with high affinity?** The current model for the physiological role of LicB rests upon this protein binding to choline at low substrate concentrations. Biophysical methods will be used to determine the equilibrium dissociation constant (K_d) for choline. This is a critical test of the current dogma that LicB is a high-affinity choline permease.
- 2. Is LicB a selective choline transporter?** It has been suggested that LicB will be tightly selective for choline, given its particular role in the metabolism of that compound. This assumption will be tested by measuring the binding activity of related substrates to LicB.
- 3. How does LicB recognise choline?** As discussed above, the BCCT transporters feature an 'aromatic box' that is responsible for ligand binding. It is not clear that such a mechanism is exploited by LicB. Site-directed mutagenesis of highly conserved residues will determine which of these residues play critical roles in substrate binding.
- 4. Is LicB a proton/choline symporter?** LicB is assumed to be a proton-driven transporter based on evolutionary relationships, but there is currently no experimental evidence to support this. The influence of bioenergetic proton gradients on choline transport will be dissected in transport assays using reconstituted proteoliposome systems *in vitro*.

Chapter 2. Materials and methods

2.1. Specific laboratory equipment

Table 2.1. Laboratory equipment used.

Centrifuge / Rotor	Supplier
Minispin™ Benchtop	Eppendorf
Optima L-80 XP Ultracentrifuge / 45Ti	Beckman
Optima TLX Ultracentrifuge/ TLS55 swing out	Beckman
Sigma 3-16KL	Sigma
Sorvall RC-6 Plus Centrifuge / SLA-1500 or SLA-3000	Thermo Scientific
Molecular biology equipment	
Owl™ EasyCast™ B2 Mini Gel Electrophoresis Systems	Thermo Scientific
Thermal Cycler 3PRIME	Techne
Protein purification and visualisation equipment	
Amersham™ Imager 600	GE Healthcare Life Sciences
Continuous-Flow Cell Disruptor	Constant Systems
Gel Doc™ EZ System	BioRad
HisTrap™ HP, 1 mL and 5 mL columns	GE Healthcare Life Sciences
Mini-PROTEAN™ tank	BioRad
Minisart™ Plus Syringe Filters (0.45 µm pore size)	Sartorius
Molecular Weight cut-off concentrators	Sartorius
Nitrocellulose membrane	GE Healthcare Amersham
PD-25 and PD-5 Desalting Column	emp Biotech GmbH
XCell SureLock™ tank	Thermo Fisher Scientific
Biophysical techniques equipment	
AKTA Pure Preparative Chromatography System	GE Healthcare Life Sciences
Cary 60 UV-Vis Spectrophotometer	Agilent Technologies
Cary Eclipse Fluorescence Spectrophotometer	Agilent Technologies
Dawn Heleos II Multi-Angle Light Scattering Detector	Wyatt Technology
HP 1100 HPLC System	Agilent Technologies

Biophysical techniques equipment	Supplier
J-1500 Circular Dichroism Spectrophotometer	Jasco
MicroCal iTC-200	GE Healthcare Life Sciences
Mini-Extruder kit	Avanti Polar Lipids
NanoDrop ND-1000 Spectrophotometer	Thermo Fisher Scientific
Optilab rEX Refractometer	Wyatt Technology
Superdex 200 10/300 GL column	GE Healthcare Life Sciences

2.2. Chemicals, reagents and solutions

Table 2.2. List of chemicals, reagents and commercial kits.

Reagent	Supplier
1 kb DNA ladder (#N3232)	New England Biolabs
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE)	Avanti Polar Lipids
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG)	Sigma
Acetic acid (glacial)	Fisher Scientific
Acetylcholine chloride	Sigma
Benzoylcholine iodide	Alfa Aesar
Betaine monohydrate	Sigma
Bis-Tris	Fisher Scientific
Bovine Serum Albumen (BSA)	Sigma
Choline chloride	Acros Organics
Coomassie Brilliant Blue R-250	BioRad
CutSmart® buffer	New England Biolabs
Dansyl and dansyl-choline	Produced by R. Stenner, University of Bristol
DC™ protein assay (#5000111)	BioRad
Detergents (DDM, DM, Cymal-5)	Anatrace
DMSO	New England Biolabs
dNTP mixture	New England Biolabs
Dried skimmed milk	Marvel

Reagent	Supplier
Gel Loading Dye, Purple (6x)	New England Biolabs
Glycerol	Fischer Scientific
Glycine	Severn Biotech
His Tag Monoclonal Antibody (6x), HRP (#R93125)	Invitrogen
HyAgarose™ 3:1 Agarose	ACTGene
Imidazole	Sigma
Isopropyl β-D-thiogalactoside (IPTG)	Generon
L-carnitine	Alfa Aesar
L-cysteine	Sigma
LumniGlo® Ultra Chemiluminescent Substrate	Cell Signalling Technology
Methanol	Riedel-de Haen
NativeMark™ Unstained Protein Standard, 20 to 1,200 kDa (#LC0725)	Invitrogen
NativePAGE™, 4 -16% Bis-Tris precast gel	Thermo Fisher Scientific
NativePAGE™ Sample Buffer (4x)	Invitrogen
NuPAGE LDS Sample Buffer (4x)	Invitrogen
PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa (#26619)	Thermo Scientific
Phusion HF Buffer (5x)	New England Biolabs
Precast 10 cm x 10 cm and 10 cm x 8.5 cm Tris-Glycine 4 – 20% gel	NuSep
Pyranine	MedChemExpress
QIAprep® Spin Miniprep Kit	Qiagen
QIAquick® Gel Extraction Kit	Qiagen
QIAquick® Purification Kit	Qiagen
Quick Ligation Reaction Buffer (2x)	New England Biolabs
SMA2000	Donated by Paul Whitley, University of Bath
S.O.C. media	Invitrogen
Sodium dodecyl sulphate (SDS)	Thermo Fischer Scientific
Sodium phosphate dibasic	Sigma
Sodium phosphate monobasic	Sigma
Sorbitol	Scientific Laboratory Supplies
Sucrose	Sigma

Reagent	Supplier
SYBR® Safe DNA Gel Stain (10,000x)	Invitrogen
Tricine	Fisher
Tween-20	Sigma
Valinomycin	Sigma

2.3. Enzymes

Table 2.3. List of enzymes used.

Enzyme	Supplier
Phusion® High-Fidelity DNA Polymerase (#M0530)	New England Biolabs
Quick Ligase (#M2200)	
Restriction enzymes (<i>DpnI</i> , <i>XhoI</i> , <i>NcoI</i>)	

2.4. Preparation and constituents of buffers

Solutions were prepared in ddH₂O and, if necessary, filter sterilised via a 0.2 µm cellulose nitrate filter and degassed prior to their use. Where required the buffer pH was adjusted using 1 M NaOH or 1 M HCl.

Table 2.4. Buffers used in the thesis, ordered alphabetically by colloquial name.

Buffers designated '*prepare fresh*' were made up in advance of each experiment and used immediately. Other buffers were stored at room temperature.

Buffer	Procedure	Composition and directions for use
Blocking Buffer	Immunoblotting	5% (w/v) milk powder in PBS/T (see below). Prepare fresh
Column Buffer (CB)	Protein purification and biophysical characterisation	50 mM sodium phosphate at pH 7.4, 150 mM NaCl, 5% (v/v) glycerol, supplemented with detergent at working concentrations – prepare fresh
Coomassie staining solution	SDS-PAGE	0.02% (w/v) Coomassie Brilliant Blue R-250, 30% (v/v) methanol and 10% (v/v) acetic acid – filtered

Buffer	Procedure	Composition and directions for use
Destain solution	SDS/BN-PAGE	8% (v/v) acetic acid – prepare fresh
Elution Buffer	IMAC	Column Buffer with 0.5 M imidazole – prepare fresh
Equilibration Buffer	IMAC	Column Buffer with 20 mM imidazole - prepare fresh
Fixing solution	SDS/BN-PAGE	40% (v/v) methanol and 10% (v/v) acetic acid - prepare fresh
IPTG stock solution	Protein expression	10 mM IPTG in ddH ₂ O – prepare fresh and filter sterilise
Membrane Buffer	Protein purification	50 mM sodium phosphate pH 7.4, 150 mM NaCl and 5% (v/v) glycerol - prepare fresh
NativePAGE Cathode Buffer (0.4%)	BN-PAGE	0.4% (w/v) Coomassie Brilliant Blue R-250
NativePAGE Running Buffer (20x)	BN-PAGE	20.92% (w/v) Bis-Tris and 17.92% (w/v) Tricine
Phosphate-Buffered Saline (PBS) (10x)	Protein purification	1.37 M NaCl, 27 mM KCl, 100 mM Na ₂ HPO ₄ and 18 mM KH ₂ PO ₄
PBS/T buffer	'Dot Blot'/WB	1 x PBS with 0.05% (v/v) Tween-20
Tris/Acetate/EDTA (TAE) Buffer	Agarose gel electrophoresis	40 mM Tris base, 20 mM acetic acid and 1 mM EDTA
Tris-Glycine Running Buffer	SDS-PAGE	25 M Tris, 192 mM glycine and 0.1% (w/v) SDS
Wash Buffer	IMAC	Column buffer with 75 mM imidazole - prepare fresh

2.5. Bioinformatic analysis

2.5.1. Computer software

All chemical structures were drawn with ChemBioDraw Ultra 12.0. Enzyme kinetics were analysed with Microsoft Excel and GraphPad Prism 8. Protein structures were visualised, and figures produced with PyMol (version 2.3.5). SnapGene[®] software (version 5.2.4) was used to view, analyse and align DNA sequences. Jalview software (version 2.11.1.3) was used to view multiple sequence alignments and phylogenetic trees. The following online tools were also used to conduct bioinformatic analysis of target protein and nucleic acid, as well as primer design evaluation:

- DNA-DNA Hybridisation: <http://ggdc.dsmz.de/home.php>

- NEB T_m calculator: <https://tmcalculator.neb.com/#!/main>
- Clustal Omega Multi Sequence Alignment: <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- Emboss Needle Pairwise Sequence Alignment: https://www.ebi.ac.uk/Tools/psa/emboss_needle/
- Expasy Translate: <https://web.expasy.org/translate/>
- Expasy ProtParam tool: <https://web.expasy.org/protparam/>
- BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- NCBI protein database: <http://www.ncbi.nlm.nih.gov>
- XtalPred-RF: <https://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl>

2.5.2. Secondary structure and transmembrane topology prediction

Secondary structure was evaluated by PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) and hydrophobicity scale by Expasy ProtScale (<https://web.expasy.org/protscale/>). The transmembrane topology of the LicB recombinant protein was predicted by the online resources listed in Table 2.5. According to the consensus prediction, TOPO2 was used to produce a topology diagram (<http://www.sacs.ucsf.edu/TOPO2/>).

Table 2.5. List of servers and the correspondent links used for TM topology prediction.

Server	Link
TMHMM	http://www.cbs.dtu.dk/services/TMHMM/
TMPred	http://embnet.vital-it.ch/software/TMPRED_form.html
HMMTOP	http://www.enzim.hu/hmmtop/
Philius and OCTOPUS	via TOPCONS server
TOPCONS	http://topcons.cbr.su.se
SCAMPI2	http://scampi.bioinfo.se

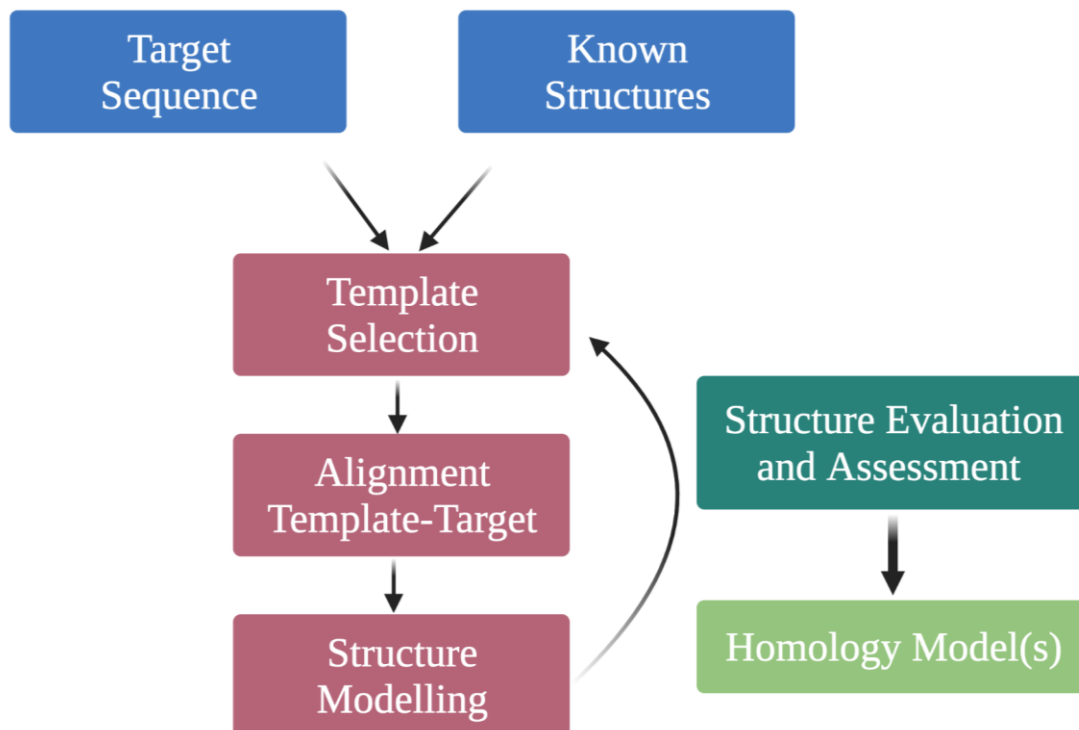
2.5.3. Homology modelling

Template-based homology models were developed using the online servers listed in Table 2.6. In the SWISS-Model server, the best-scored protein template was used to design the LicB homology model.

Table 2.6. List of online resources used for template-based homology.

Server	Link	Ref.
SWISS-Model	https://swissmodel.expasy.org/interactive	364
Memoir/MEDELLER	http://opig.stats.ox.ac.uk/webapps/memoir/php/index.php	365
HHpred	https://toolkit.tuebingen.mpg.de/tools/hhpred	366

Briefly, the mechanism behind homology modelling is represented in Figure 2.1.

**Figure 2.1. General workflow of homology modelling.**

'Known Structures' are derived from the PDB. 'Template Selection' uses BLAST or HMM-HMM profile methods. 'Alignment Template-Target' uses pairwise or multiple alignment sequence. 'Structure Modelling' is based on coordinate building either by template-based fragment assembly (Composer or SWISS-Model) or by the satisfaction of special restraints (MEDELLER). From Kelm *et al.* (2010)³⁶⁷ and Waterhouse *et al.* (2018)³⁶⁴.

2.5.4. Identification of LicB sequence homologues and conserved motifs

Conserved motifs within the LicB sequence were identified using the online tool BLASTP. *Haemophilus influenzae* LicB (RefSeq WP_005672105.1) was compared against proteins in the NCBI protein database. To avoid the return of high numbers of *H. influenzae* strains, sequences from this species were excluded. The BLAST result was analysed as follows. The top 100 sequences were inspected, and partial and redundant sequences were removed. The remaining data thus contain a single full-length representative per species. This resulted in a final number of 30 sequences that were aligned by Clustal Omega.

2.6. General microbiology techniques

2.6.1. *E. coli* strains and vectors

The following *E. coli* strains and vectors were used.

Table 2.7. List of *E. coli* strains and vectors used in LicB project.

Strain	Feature	Use	Supplier
One Shot™ TOP10 Chemically Competent <i>E. coli</i>	Stable replication of high-copy number of plasmids	Plasmid preparation	Invitrogen
T7 Express Competent <i>E. coli</i>	Most popular T7 expression strain; Protease deficient		New England Biolabs
BL21 Star (DE3)	Promotes high mRNA stability and protein yield	Protein expression	Collinson Lab (University of Bristol)
C43(DE3)	Derivative of BL21(DE3) with resistance to expression of toxic proteins		Curnow Lab (University of Bristol)
Vector	Feature	Supplier	
pET28	T7/LacO for IPTG induction in <i>E. coli</i> ; Kanamycin resistance	Novagen	

2.6.2. Media and plates

All powdered media was prepared with ddH₂O and autoclaved at 121°C for 15 minute at 15 psi.

Table 2.8. Media recipes and general preparation.

Media	Preparation	Use	Supplier
Powdered Lysogeny Broth (LB)	2.5% (w/v)	Liquid <i>E. coli</i> culture for protein expression	Fisher BioReagents
Powdered Super Broth (SB) auto-induction media (AIM) including trace elements	7.5% (w/v)		Fomedium™
LB-agar plates	2.5% (w/v) LB powder with 2% (w/v) of agar	<i>E. coli</i> colony culture	Agar from BD Difco™

2.6.3. Antibiotics

Antibiotic solutions were prepared in sterile conditions and filtered via a 0.45 µm syringe filter before storing at 4°C. Antibiotic was added at working concentration to each *E. coli* liquid culture just before inoculating or, after autoclaving, to LB-agar warm solution (approximately 60°C) just before pouring plates.

Table 2.9. List of antibiotics used.

Antibiotic	Stock concentration (1000x stock)	Work concentration	Supplier
Kanamycin	50 mg/mL in ddH ₂ O	50 µg/mL	ITW reagents

2.7. General DNA techniques

All molecular biology protocols used purified and filtered water with a Milli-Q Integral Water Purification System. Unless otherwise stated, all molecular biology reagents are cited in Table 2.2 and enzymes in Table 2.3.

2.7.1. Cloning of *licB* gene into pET28 vector

Purified genomic DNA from *Haemophilus influenzae* was a gift from Simon Newstead, University of Oxford. The *licB* gene was amplified with PCR primers synthesised by Eurofins Genomics, which also introduced unique *NcoI* and *XhoI* sites to facilitate subsequent cloning by restriction digest and sticky-end ligation. The *NcoI* site incorporated the start codon (Table 2.10).

Gene amplification by PCR was carried out in a total volume of 50 μL , containing 2.5 μL of 20 pmol/mL each primer, 1 μL genomic DNA, 1 μL DMSO, 10 μL 5x Phusion HF Buffer, 1 μL of 10 mM dNTPs, 31 μL ddH₂O and 1 μL of Phusion[®] High-Fidelity DNA Polymerase. The reaction proceeded for 25 thermal cycles of 30 second at 95°C, 30 second at 55°C, and 3 minute at 72°C. DNA amplification was confirmed by agarose gel electrophoresis. The amplified gene was recovered from an agarose gel with the QIAquick[®] Gel Extraction Kit according to manufacturer's instructions.

The amplified *licB* gene was then digested simultaneously with restriction endonucleases *NcoI* and *XhoI* for 1 hour at 37°C. The DNA was repurified with a Qiagen PCR clean-up kit prior. Ligation used the Quick Ligase with 1 μL *NcoI/XhoI*-digested pET28 vector and 3 μL digested *licB* insert in a final volume of 20 μL . The reaction was initially mixed on ice, then transferred to room temperature for 1 hour. This temperature ramp is empirically shown to be very effective at promoting ligation, probably offering a compromise between optimal annealing at low temperatures and ligase activity at higher temperatures (P. Curnow, personal communication). Ligations, including vector-only controls, were transformed into commercial ultracompetent TOP10 Competent Cells and the identity of successful transformants confirmed by sequencing.

Table 2.10. Primer sequences for the cloning of *licB* gene in pET28 vector.

*Nco*I and *Xho*I restriction sites are underlined for forward and reverse primers, respectively.

Primer	Sequence (5' → 3')
licB_F	GTTGTT <u>CCATGGG</u> TCGTGGCTATCTCTTTGGCATACTATCC
licB_R	ATGATG <u>CTCGAG</u> TTTTTTGTTCCCTTTGTAATAAAGTGTAACG

2.7.2. Site-directed mutagenesis

Mutations were made using the QuikChange Site-Directed Mutagenesis protocol (Agilent). A pair of primers with the required codon change (GCG for alanine substitution using preferred codon) flanked by fully complementary sequences was used to amplify template plasmid DNA by PCR using Phusion® DNA polymerase. The PCR reaction mixture was as follows: 2.5 µL of 20 pmol/mL each primer, 1 µL of 40 ng/µL genomic DNA, 1 µL of 10 mM dNTPs, 1.5 µL of DMSO, 10 µL of 5x Phusion HF Buffer and 0.5 µL of Phusion® High-Fidelity DNA Polymerase and Milli-Q water to a total reaction volume of 50 µL. PCR reactions were established as follows: 30 second initial denaturation at 95°C, followed by 18 cycles of a 30 second denaturation at 95°C, 30 second of primer annealing (at a specific temperature for each PCR reaction - Table 2.11) and 3 minute extension at 72°C, with a final extension of 10 minute (72°C).

After digestion with 1 µL of *Dpn*I for 1 hour at 37°C, PCR products were analysed by agarose gel electrophoresis. After confirming the amplification, samples containing the gene were further purified by a PCR Purification Kit according to the manufacturer's protocol and transformed into TOP10 Competent Cells.

Table 2.11. List of the forward (F) and reverse (R) primers used for site-directed mutagenesis of LicB, and the correspondent annealing temperature set in PCR reactions.

Alanine mutations (5'-GCG-3') in the DNA sequences are underlined.

Alanine substitution	Primer Sequence (5' → 3')	Annealing temperature
M245A	F: CCAATCAAAGCAG <u>C</u> GGCATTAAATATCACATATTC R: GTGATATTTAATG <u>C</u> CGCTGCTTTGATTGG	48, 50, 52, 54, 60°C
N248A	F: GCAATGGCATTAG <u>C</u> GATCACATATTCTGTATGG R: CCATACAGAATATGTGAT <u>C</u> GCTAATGCCATTGC	50°C
I249A	F: GCAATGGCATTAAAT <u>G</u> CGACATATTCTGTATGGG R: CCCATACAGAATATGT <u>C</u> GCATTTAATGCCATTGC	52°C
T250A	F: GCATTAATATC <u>G</u> CGTATTCTGTATGGGC R: TTGCCCATACAGAATAC <u>G</u> CGATATTTAATGCC	50°C
Y251A	F: CATTAAATATCACAG <u>C</u> GTCTGTATGGGCAATTGG R: CCAATTGCCCATACAGAC <u>G</u> CTGTGATATTTAATG	48°C
W254A	F: CATATTCTGTAG <u>C</u> GGCAATTGGACTTGG R: CCAAGTCCAATTG <u>C</u> CGCTACAGAATATG	48°C

2.7.3. PCR product transformation

Transformation of commercial competent *E. coli* TOP10 cells was performed according to the protocols supplied by the manufacturer. Briefly, 1 μ L PCR product was added to 50 μ L of competent cells and incubated for 30 minute on ice. After the heat shock (42°C for 45 second) and 5-minute incubation on ice, cells were diluted in S.O.C. media and incubated for 1 hour at 37°C and 250 rpm. Different volumes of this mixture (typically between 50 and 200 μ L) were then plated on LB-agar plates with kanamycin and incubated overnight at 37°C. Competent cells with 1 μ L H₂O were used as a negative control.

2.7.4. Plasmid miniprep

At least six successful transformant colonies were selected from LB-agar plates and grown overnight in a shaking incubator (37°C and 250 rpm) in 10 mL LB cultures supplemented with kanamycin. Plasmids were harvested from the overnight cultures using the QIAprep® Spin

Miniprep Kit according to manufacturers' instructions. Purified plasmids were eluted with 50 μ L hot Milli-Q water and quantified with a NanoDrop Spectrophotometer. Concentration was determined from the absorbance spectrum at 260 nm; A260/280 and A260/A230 ratios were used as an indication of purity. Plasmids were stored at -20°C until use.

2.7.5. DNA sequencing

All DNA sequencing was performed commercially by Eurofins Genomics with the standard commercial primers T7 (5'-TAATACGACTCACTATAGGG-3') and T7term (5'-CTAGTTATTGCTCAGCGGT-3'). These targeted the T7 promoter and terminator sequences, respectively, in the pET vectors used for expression. Sequencing results were aligned with the corresponding plasmid sequence using the SnapGene[®] software.

2.7.6. DNA transformation into chemically competent cells

For expression in *E. coli*, the purified plasmid containing the *licB* gene (typically between 1 and 2 μ L) was added into 50 μ L chemically competent expression cell lines on ice. Cells were subsequently incubated on ice for 30 minute, followed by a heat shock based on the manufacturer's instruction (Table 2.12). The mixture was placed again on ice for 5 minute and allowed to recover in 950 μ L of S.O.C. media for 1 hour in an orbital shaking incubator at 37°C and 250 rpm. Successful transformants were selected on LB-agar plates supplemented with kanamycin and incubated at 37°C overnight. Competent cells with 1 μ L H₂O were used as a negative control.

Table 2.12. Heat shock conditions used for each cell line transformation.

Competent cell line	Heat shock (temperature, time)
T7 Express	42°C for 10 second
BL21 Star (DE3)	42°C for 30 second
C43(DE3)	42°C for 45 second

2.7.7. Agarose gel electrophoresis

DNA samples were prepared by mixing 5 μ L PCR reaction product with 2 μ L 6x Gel loading dye. For a 1% agarose gel, 0.5 g agarose was melted in 50 mL of 1x TAE Buffer and poured into a casting tray with a suitable comb. 5 μ L SYBR[®] Safe DNA Gel stain (0.01%) was added when the solution had cooled slightly but not set. DNA samples were run adjacent to a 1 kb reference marker at 100 V for 1 hour. Gels were imaged using a Gel Doc[™] EZ System.

2.8. Protein expression in *E. coli* and protein purification

Primary cultures were inoculated into 10 mL media in a 30 mL universal glass bottle, whereas 1 L secondary cultures were grown in a 2.5 L conical flask. Typically, expression was conducted in an orbital incubator at 250 rpm. OD₆₀₀ was used to indicate the cell density. Transformed strains were stored long-term as concentrated suspensions in a filter sterilised 30% glycerol-70% LB broth at -80°C, unless indicated in the text.

2.8.1. Large scale *E. coli* expression

A single colony from transformation plates was grown overnight at 37°C and 250 rpm in 10 mL LB broth supplemented with kanamycin. This primary culture was subsequently used to inoculate 1 litre of either auto-induction SB or LB broth, supplemented with kanamycin. The secondary culture was then grown under different conditions, as listed in Table 2.13.

Cells were harvested at 5,000 \times g for 20 minute at 4°C and resuspended in 200 mL (for auto-induction cultures) and 100 mL (for IPTG induction cultures) of 1x PBS. The purification of recombinant proteins in detergents was then carried out according to the method of Nancolas *et al.*³⁶⁸ and Curnow *et al.*²⁹ with minor modifications. The protocol used for SMA2000 membrane solubilisation was adapted from that described by Lee *et al.*¹⁶³.

Table 2.13. Conditions used for LicB expression.

Media, *E. coli* strain, induction procedure, expression time and temperature are described. T_{inoc} - inoculation time.

Media	<i>E. coli</i> strain	Induction at OD ₆₀₀	Culture harvesting	Expression temperature
SB (1 L)	T7	Auto-induction	24 hour after T_{inoc}	37°C
			~21 hour after OD ₆₀₀ ~ 0.8 followed by 15-minute incubation on ice	25°C
			~8 hour after T_{inoc}	37°C
	BL21 Star (DE3)	Auto-induction	24 hour after T_{inoc}	37°C
LB (3x 1 L)	T7	0.7 - 0.8 by adding 0.1 mM IPTG	2 hour after IPTG induction	37°C
	C43(DE3)	0.9 - 1.0 by adding 0.1 mM IPTG		

2.8.2. Isolation of membranes

The bacterial cells were lysed in a continuous flow cell disrupter at 25,000 psi. Unbroken cells and large debris were pelleted at 10,000 × g for 12 minute at 4°C. Cell membranes were fractionated at 170,000 × g for 1 hour at 4°C. The resulting membrane pellet was resuspended in 100 mL (for auto-induction cultures) or 40 mL (for IPTG induction cultures) of Membrane Buffer and subjected to at least 30 passes in a glass homogenizer. During cell fractionation, samples of lysate, cytosol and membranes were collected for future analysis.

2.8.3. Membrane fraction solubilisation

The whole volume of the membrane fraction was incubated with different solubilising detergents for 1 hour at room temperature. Insoluble membranes were removed by pelleting at 170,000 × g for 1 hour at 4°C and discarded. For SMA2000, membranes were incubated for 2 hour, and insoluble particles were removed by centrifugation at 100,000 × g for 45 minute at 4°C. Solubilised membrane fraction was collected for future analysis. Table 2.14 lists all solubilising agents used as well as their concentration in solubilisation and working buffers.

Table 2.14. List of chemicals used for membrane solubilisation and subsequent protein purification.

Solubilised agent	% (w/v) used for solubilisation	% (w/v) used in working buffers
DDM	1.0% or 2.0%	0.05%
DM	1.0%	0.20%
Cymal-5	2.4%	0.24%
SMA2000	2.5%	N/A

2.8.4. Purification of recombinant protein

Solubilised membrane proteins were loaded onto a 1 mL 'HisTrap' Ni²⁺-NTA Sepharose column, with a stated binding capacity of at least 40 mg His-tagged protein, equilibrated in 10 column volumes of Equilibration Buffer at a flow rate of 1 mL/minute. The column was washed with 40 volumes of Wash Buffer, and the purified protein was eluted in 2.5 column volumes of Elution Buffer at a flow rate of 0.2 mL/minute. Imidazole was immediately removed using a PD-25 gravity-flow desalting column, and the desalted eluent (3.5 mL) was concentrated in a spin concentrator with an appropriate molecular weight cut-off of either 30 or 50 kDa. Concentration was performed at 3,500 x g at 4°C until the desired volume was achieved (at least 500 µL). Aliquots of purified protein were injected directly to size exclusion chromatography or stored at -80°C. Regarding SMALP purification, the protocol followed was as described above; however, no detergent was added to Equilibration, Wash and Elution Buffers. Samples of column flow-through and wash were collected for further analysis.

2.9. Total protein concentration determination

Total protein concentration in cell fractions was determined using a commercial detergent-compatible Lowry assay, according to the manufacturer's instructions. 25 µL of protein sample were mixed with 125 µL of reagent A+S (alkaline copper tartrate plus sodium dodecylsulphate) and subsequently with 1 mL of reagent B (Folin reagent) in a 1.5 mL cuvette. After 15 minute incubation, the absorbance was measured at 750 nm using a spectrophotometer (Agilent Technologies Cary 60 UV-Vis). Total protein concentration was then calculated based on a calibration curve determined by known concentrations of BSA between 0.2 and 2 mg/mL.

The concentration of purified proteins was determined from the absorbance at 280 nm using NanoDrop Spectrophotometer or Agilent Technologies Cary 60 UV-Vis. The concentration of a given protein of interest within a sample was calculated by the Beer-Lambert Law (Equation 2.1):

$$A = \epsilon \cdot c \cdot l$$

Equation 2.1

Where A is the absorbance, ϵ is the extinction coefficient of the protein ($M^{-1} \cdot cm^{-1}$), c denotes the concentration of the protein (M), and l is the pathlength of the cuvette (cm). The calculated extinction coefficient and the theoretical molecular weight of all LicB variants are listed in Table 2.15.

Table 2.15. Theoretical molecular weight (Da) and extinction coefficient (ϵ) for each protein studied based on ExPASy ProtParam online tool.

LicB variants	Molecular weight (Da)	Extinction coefficient ($M^{-1} \cdot cm^{-1}$)
WT	34,120	64,290
N248A	34,077	64,290
I249A	34,078	64,290
T250A	34,090	64,290
Y251A	34,028	62,800
W254A	34,005	58,790

2.10. Biophysical techniques

2.10.1. Polyacrylamide gel electrophoresis (PAGE)

2.10.1.1. Sodium dodecyl sulphate (SDS)-PAGE

SDS-PAGE was performed with NuSep Tris-Glycine 4-20% acrylamide precast gels. Samples were prepared in 4x NuPage LDS Sample Buffer and boiled at 90°C for 10 minute to guarantee denaturation, except for pure LicB samples. Samples were loaded onto the gel as well as the PageRuler™ Prestained Protein Ladder. Gels were run for 70 minute (NN gels, 10 x10 cm) or

45 minute (NB gels, 10 x 8.5 cm) at 200 V in Tris-Glycine Running Buffer. Fixing and staining procedures were performed for between 15 to 30 minute, and de-staining was performed until the desired contrast was achieved, and gels were visualised using a Gel Doc™ EZ System. Table 2.4 describes the composition of all solutions used. In the following SDS-gel figures, M denotes molecular weight markers in kDa; And lane headings are relevant to cell fractionation (lysate, cytosol, membranes, soluble membranes) and affinity purification (unbound, wash and purified protein).

2.10.1.2. Blue Native(BN)-PAGE

The protocol for Blue Native (BN)-PAGE was adapted from Ma *et al.* ³⁶⁹. Protein samples to be analysed were diluted in 4x NativePAGE™ Sample Buffer. Precast NativePAGE™ 4-16% Bis-Tris gels were run in a Bis-Tris buffer system. Before introducing into the electrophoresis apparatus, the gel was rinsed in ddH₂O, and then each well was filled with 1x Running Buffer. Cathode Buffer containing 0.02% Coomassie Brilliant Blue was loaded into the inner chamber, whereas 1x Running Buffer (without Coomassie) was loaded to the outer reservoir of the gel tank apparatus. The whole gel cassette was placed on an ice bath and running at 150 V for 1 hour, and then raised to 250 V for another 1 hour. Gels were fixed and de-stained as outlined for SDS-PAGE gels.

2.10.2. Direct immunoblot ('Dot Blot')

Direct loading of proteins onto immunoblotting membranes – so-called “dot blotting” - was performed inside a petri dish, in which a filter paper was used to support a dry nitrocellulose membrane. Using a P10 pipette, 2 µL was loaded onto the membrane, forming a wet ‘dot’. After 30 minute of incubation at room temperature, Blocking Buffer was added and left for a 30 minute incubation period. The membrane was then incubated with an anti-His tag antibody in the same buffer for 1 hour. The membrane was then washed five times with short incubation periods in PBS/T Buffer. Chemiluminescence was detected using the LumiGLO reagents on an Amersham™ Imager 600.

2.10.3. MALDI mass spectrometry of LicB protein

Identification of recombinant LicB protein was confirmed using tryptic digest MALDI mass spectrometry. A single gel band apparently corresponding to pure LicB was excised and sent to the Proteomics Facility at the University of Bristol for protein identification. The complete protocol is described in Calvopiña *et al.*³⁷⁰.

2.10.4. Size exclusion chromatography

Size exclusion chromatography (SEC) was performed with a Superdex 200 10/300 GL column. Protein standards used for column calibration were Ovalbumin, Conalbumin, Aldolase, Ferritin and Thyroglobulin (molecular weights and Stokes radius of each standard are listed in Table 2.16). The column void volume (V_0 , 7.9 mL) was previously determined with Blue Dextran, and the total column volume (V_t) was 24 mL. The calibration curve (Figure 2.2) was determined from the particle coefficient value (K_{av} , determined in Equation 2.2) against the logarithmic scale of the molecular weight of each protein standard¹²⁷.

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

Equation 2.2

Where V_e is the elution volume of each protein standard and V_0 is the void volume.

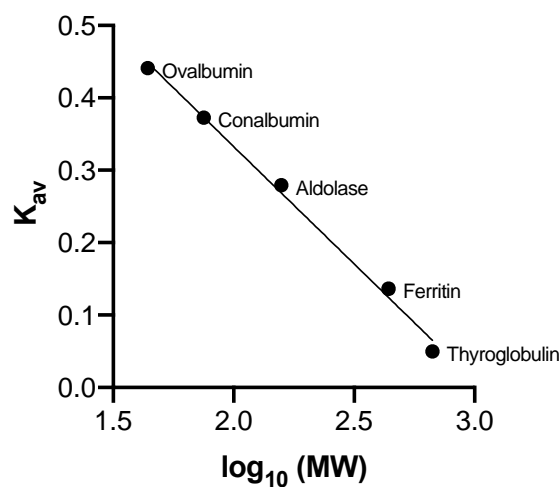


Figure 2.2. Size exclusion chromatography calibration curve.

Calibration curve of the calibrants used to determine the MW_{app} for analysed proteins.

Table 2.16. Values of the molecular weight (kDa) and Stokes radius (nm) for each calibrant and the correspondent elution volume.

	MW (kDa)	log (MW)	V_e	$V_e - V_o$	K_{av}	Stokes radius (nm)
Thyroglobulin	669	2.83	8.70	0.80	0.05	-
Ferritin	440	2.64	10.10	2.20	0.14	6.1
Aldolase	158	2.20	12.40	4.50	0.28	4.8
Conalbumin	75	1.88	13.90	6.00	0.37	3.6
Ovalbumin	44	1.64	15.00	7.10	0.44	3.1

The column was equilibrated in Column Buffer supplemented with detergent at working concentrations (see Table 2.14) before 500 μ L of concentrated protein was loaded and run at 0.5 mL/minute. Column progress was determined by following protein absorption at 280 nm. Column fractions (1 mL) were analysed by DC protein assay and SDS-PAGE. The fractions corresponding to the eluted protein of interest were subsequently concentrated and stored (- 80°C). The absorption data were normalised to compare multiple conditions.

2.10.5. Size exclusion chromatography coupled to multi angle light scattering

Size exclusion chromatography coupled to multi-angle light scattering (SEC-MALS) was set up as follows; a Dawn Heleos II multi-angle light scattering detector with an Optilab rEX refractometer were used in line with a Superdex 200 10/300 GL column. All experiments were performed at room temperature and at a column flow rate of 0.5 mL/minute. The column was equilibrated overnight at approximately 0.05 mL/minute in Column Buffer with 0.05% DDM. Once stable baselines were obtained, 120 μ L samples were loaded, including the BSA sample at 2 mg/mL, the protein sample (at variable concentrations) and a detergent-only sample (DDM) at 5% (w/v). ASTRA 6 software (Wyatt Technology) was used to collect and analyse light scattering (LS), differential refractive index (RI) and absorbance ($UV_{280\text{ nm}}$) according to the manufacturer's instructions. Molecular masses were determined through the three-detector method of Slotboom *et al.*³⁷¹. This calculates the membrane protein molecular weight independent of the micelle size with Equation 2.3:

$$MW, protein = \frac{\Delta LS \times \Delta UV_{280 nm}}{K \times A_{280 nm} \times (\Delta RI)^2}$$

Equation 2.3

Where Δ means the difference between the LicB protein sample and the DDM micelle sample, for the three properties mentioned above – RI , LS and $UV_{280 nm}$. $A_{280 nm}$ is the extinction coefficient in $\text{mL} \cdot \text{mg}^{-1} \cdot \text{cm}^{-1}$ (i.e., molar extinction coefficient divided by molecular mass) and K is a constant, which depends on the properties of the particular column and buffer used, and is calculated using a BSA standard.

The amount of DDM molecules (δ) within the protein-detergent complex (PDC) can be estimated as follows:

$$\left(\frac{dn}{dc}\right)_{\text{apparent}} = \left(\frac{dn}{dc}\right)_{\text{protein}} + \delta \left(\frac{dn}{dc}\right)_{\text{detergent}}$$

Equation 2.4

Rearranging:

$$\delta = \frac{\left(\frac{dn}{dc}\right)_{\text{apparent}} - \left(\frac{dn}{dc}\right)_{\text{protein}}}{\left(\frac{dn}{dc}\right)_{\text{detergent}}}$$

Equation 2.5

Where $\left(\frac{dn}{dc}\right)_{\text{detergent}}$ is 0.1435 mL/g (value from Anatrace), $\left(\frac{dn}{dc}\right)_{\text{apparent}}$ can be calculated if detergent did not absorb light at 280 nm (Equation 2.6) and $\left(\frac{dn}{dc}\right)_{\text{protein}}$ is calculated as described in Equation 2.7.

$$\left(\frac{dn}{dc}\right)_{\text{apparent}} = \frac{A_{280nm} \times \Delta RI}{\Delta UV} \times 1000$$

Equation 2.6

$$\left(\frac{dn}{dc}\right)_{\text{protein}} = \frac{\sum \left(\frac{dn}{dc}\right)_a \times M_a}{\sum M_a}$$

Equation 2.7

Where M_a is the summed mass of each individual amino acid and $\left(\frac{dn}{dc}\right)_a$ is the refractive index of each amino acid (values collected from Zhao *et al.* ³⁷²).

2.10.6. Circular dichroism spectroscopy

Circular dichroism (CD) was performed on a Jasco J-1500 instrument in a 1 mm path length quartz cuvette. Protein concentrations were between 0.1 to 0.7 mg/mL. Buffer backgrounds (Column Buffer with detergent at working concentration) were collected and subtracted from each protein sample experiment. Data collected at wavelengths where the high-tension (HT) voltage exceeded 700 V were excluded. CD scans were carried out for eight repetitions between 190 to 300 nm, and the data interval was 1 nm at 25°C. Thermal melts were performed between 25°C and 95°C at a rate of 1°C/minute. CD scans were taken every 10°C, with each scan repeated four times. Protein stability at 25°C was studied for 16 hour, and CD scans were taken every 0.3 hour. In order to normalise results and correct the signal for protein size, protein concentration, and cuvette pathlength, data were transformed into mean residue ellipticity (*MRE*), calculated according to Kelly *et al.* ³⁷³, by Equation 2.8:

$$MRE = \frac{MRW \times \theta}{10 \times c \times l}$$

Equation 2.8

Where l is the pathlength in cm (thus 0.1 for the 1 mm cell used), θ is the observed ellipticity in mdeg, c is concentration in mg/mL and MRW is the Mean Residue Weight. The MRW can be calculated by dividing the protein MW by the number of peptide bonds; for most of the proteins this is around 110 ± 5 Da.

The percentage of α -helical content was estimated using a linear extrapolation of the mean residue ellipticity at 222 nm according to Equation 2.9 ^{374,375}:

$$\% \alpha - helix = 100 \times \left(\frac{-MRE_{222 \text{ nm}} + 3000}{39000} \right)$$

Equation 2.9

Results were plotted using GraphPad Prism, and melting temperatures were calculated by fitting data in Boltzmann sigmoid function to guide the eye.

2.11. Ligand binding assays

2.11.1. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) experiments were carried out using a MicroCal iTC-200 instrument. Protein and ligand were both diluted in Column Buffer with 0.05% DDM. The reference cell was loaded with this same buffer, whereas the measurement cell was filled with the buffer containing the purified recombinant LicB. Different initial concentrations of ligand were used for serial titrations into the measurement cell (50 μM , 100 μM , 500 μM and 1000 μM). Control experiments were also performed in the absence of ligand and in the absence of protein.

The ITC injection protocol was as follows: initial injection of 0.5 μL for 1 second followed by 20 identical injections of 2 μL over 4 second, with 360 second gap between injections. The stirring rate was 1000 rpm, and the temperature was set to 25°C. Data were visualised and analysed by MicroCal™ Origin Software.

2.11.2. Fluorescence resonance energy transfer

Agilent Technologies Cary Eclipse Fluorescence Spectrophotometer was used for all fluorescence spectroscopy experiments. All assays were carried out at 25°C, excitation was set at 295 nm, and emissions were registered between 305 and 650 nm. Fluorescent dansyl (D) was obtained commercially, and dansyl choline (DC) was synthesised in-house by R. Stenner according to the protocol of Weber and Borris³⁷⁶. All experiments were performed in a reaction buffer that was standard Column Buffer with 0.05% DDM.

2.11.2.1. Ligand binding assay

For ligand binding assays, 1.5 μM purified LicB was pre-incubated in 400 μL of reaction buffer and allowed to equilibrate for 10 minute. Dansyl choline was titrated into this protein solution, and the equilibrium dissociation constant (K_d) of dansyl choline was determined from the change in the emission maxima at around 570 nm. Control assays were performed by adding (i) dansyl choline into reaction buffer (no LicB), (ii) dansyl into reaction buffer with LicB, (iii) dansyl into reaction buffer (no LicB), (iv) reaction buffer only (no ligand) into reaction buffer with LicB. Experiments were performed at least in triplicate.

Emission values were fitted to a hyperbolic binding equation using GraphPad Prism software (Equation 2.10):

$$Y = \frac{B_{max} \times [X]}{K_d + [X]}$$

Equation 2.10

Where Y is the specific binding (here corresponds to the measured fluorescence), B_{max} is the maximum number of binding sites (here corresponds to the maximum of fluorescence units measured), K_d is the equilibrium dissociation constant and $[X]$ is the ligand concentration.

2.11.2.2. Competition assay

For competition assays, pure LicB protein was used at 1.5 μM in 400 μL of reaction buffer. Dansyl choline was introduced at 1.5 times the K_d and allowed to equilibrate for 10 minute. Bound dansyl choline has then competed with a range of other compounds, including choline, acetylcholine, benzoylcholine, L-carnitine, L-cysteine, glycine and betaine. Experiments were performed at least in triplicate. Control assays were performed by (i) adding increasing choline concentrations into LicB with dansyl and (ii) buffer titrations without ligand. Results were fitted to the following four-component inhibition equation using GraphPad Prism (Equation 2.11):

$$Y = Bottom + \frac{Top - Bottom}{1 + \left(\frac{IC_{50}}{[X]}\right)^{Hill\ slope}}$$

Equation 2.11

Or in logarithmic scale:

$$Y = Bottom + \frac{Top - Bottom}{1 + (10)^{\log(IC_{50} - [X]) \times Hill\ slope}}$$

Equation 2.12

Where Y is the measured fluorescence, Top and $Bottom$ are plateaus in the units of the Y axis, IC_{50} is the concentration of competitor that gives a response halfway between $Bottom$ and Top , $Hill\ slope$ is a value that describes the steepness of a curve (or cooperativity) and X is the ligand concentration.

Moreover, inhibition constant (K_i) was calculated from the Cheng-Prussoff equation ³⁷⁷ (Equation 2.13):

$$K_i = \frac{IC_{50}}{1 + \frac{[Ligand]}{K_{d,ligand}}}$$

Equation 2.13

2.12. Transport assay

Proteoliposome-based transport assays were carried out at 25°C in an Agilent Technologies Cary Eclipse Fluorescence Spectrophotometer. Excitation was set at 406 nm and 460 nm, and emission was measured at 510 nm.

2.12.1. Liposome formation

A lipid mixture of 3:1 (w/w) ratio of POPE:POPG was prepared and dissolved in 1:1 (v/v) chloroform:methanol to produce liposomes. After left drying overnight, the lipid mixture was resuspended at 40 mg/mL in so-called 'inside buffer' (5 mM sodium phosphate buffer, 135 mM KCl, 15 mM NaCl and 150 mM sorbitol at pH 8.0) until a smooth suspension was achieved; Then it was subjected to five rounds of freeze-thawing using liquid nitrogen and a 42°C water bath. Liposomes were produced by sequential extrusion (at least eleven times) of the lipid suspension via a 0.4 µm polycarbonate filter, using the Avanti Mini-Extruder according to the manufacturers' instructions at room temperature. Liposomes were either used immediately for reconstitution methods or stored at -20°C overnight.

2.12.2. DDM-mediated reconstitution

When reconstituting by the detergent saturation method, liposomes were first formed by extrusion then incubated for 30 minute with 0.08% DDM. This was confirmed by Abs540 measurements to saturate, rather than solubilise, the liposomes. The permeabilised liposomes were then sonicated for two rounds of 15 second with 5 second rest. LicB was added at 60:1 or 360:1 (w/w) DDM-lipid to protein ratios to a final volume of 300 µL. This mixture was then incubated with 1 mM pyranine for 30 minute at room temperature. DDM-lipid with LicB and

pyranine solution was subsequently diluted 10-fold in 'inside buffer' and subjected to one round of freeze-thawing (as described above) and sonication (two rounds of 5 second with 15 second rest). To remove the excess of detergent and pyranine, proteoliposome solution was passed through the PD-25 column, and eluted proteoliposomes were collected and used immediately. A 'flotation' assay assessed LicB reconstitution.

As part of method development and validation, the presence of pyranine inside proteoliposomes was confirmed by observing excitation spectra between 400 nm and 480 nm.

2.12.3. 'Flotation' assay

The sucrose gradient 'flotation assay' was performed onto a discontinuous gradient. Samples were diluted in 1 volumetric ratio of 50% sucrose and were layered with 8 volumetric ratios of 40% sucrose and 1 volumetric ratio of 0% sucrose (only buffer). Gradients were centrifuged at 200,000 x g for 1 hour, and top, middle and bottom fractions were removed by syringe needle for 'Dot Blot' analysis.

2.12.4. Pyranine transport assay

Choline transport by reconstituted LicB was determined with the fluorescent reporter pyranine, according to the protocols described in Uzdevinys *et al.*³⁷⁸ and Parker *et al.*³⁷⁹. Pyranine reports on the co-transport of protons by the active transport. For this assay, 300 μ L of LicB proteoliposomes were diluted 10-fold into 'outside buffer' (5 mM sodium phosphate buffer, 150 mM NaCl and 150 mM sorbitol at pH 6.5) supplemented with different choline concentrations and mixed by thorough pipetting. After 1 minute incubation, the reaction was initiated by adding 10 μ M of valinomycin. Fluorescence was recorded over 5 minute and the dual excitation method was used, whereby the ratio of the two excitation peaks at 406 nm and 460 nm is determined with emission at 510 nm.

Chapter 3. Recombinant expression and initial biophysical characterisation of the choline permease LicB

3.1. Introduction and aim

LicB is an integral membrane protein first recognised in *H. influenzae*³⁶¹ and since identified in several other colonising opportunistic pathogens. It is expected to function as a high-affinity choline permease that imports choline for the biosynthesis of phosphorylcholine (ChoP). The critical role of LicB is clear since *licB* knockout strains cannot incorporate ChoP into their LPS, which in turn compromises bacteria colonisation, persistence and virulence in the human respiratory tract^{223,361}. This makes LicB a possible target for new antibiotics against pathogenic bacteria.

However, the structure and function of LicB have not yet been characterised. Therefore, the work described in this chapter aimed to recombinantly express and purify LicB to allow the subsequent characterisation of protein secondary structure, oligomeric state and stability *in vitro*. The workflow for this experimental work is shown in Figure 3.1. Additionally, to complement the experimental results, bioinformatic methods were used to produce the first structural model of LicB.

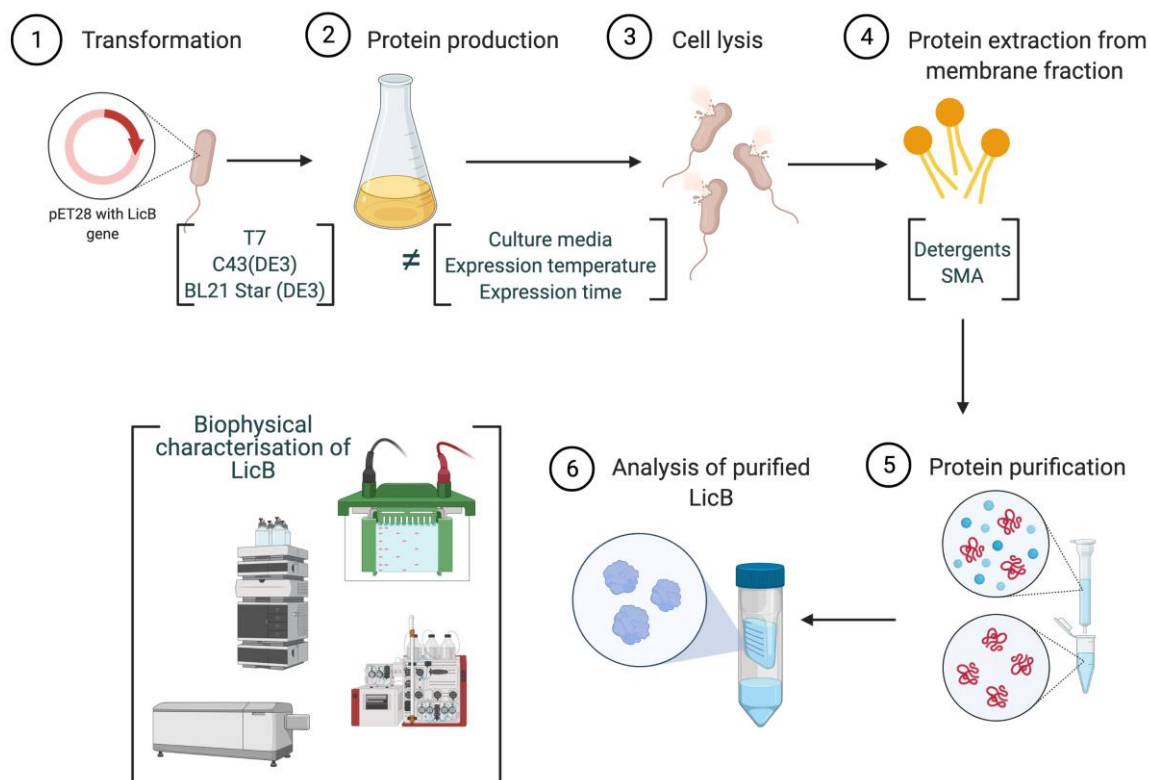


Figure 3.1. Schematic representation of the critical steps from expression to characterisation of LicB.

3.1.1. Special considerations for the recombinant expression of membrane proteins

The primary barrier to mechanistic studies of integral membrane proteins lies in the successful isolation of a stable protein of interest with relatively high yields for further analysis. *E. coli* is still the most widely used system for overexpression of prokaryotic membrane proteins such as LicB. In this chapter, the *E. coli* strains T7 Express, C43(DE3) and BL21 Star (DE3) were screened for optimal expression. These strains have been previously shown to improve biosynthesis of membrane proteins, specifically toxic MP, under the control of the T7lac promoter system^{135,380,381}. In addition, the work here also explores protein expression in auto-induction media (AIM). This is commonly used as a ‘gentler’ induction method that can improve membrane protein production and relies on inducer exclusion between glucose and lactose, avoiding lactose induction prior to glucose intake (Figure 3.2)^{382,383}.

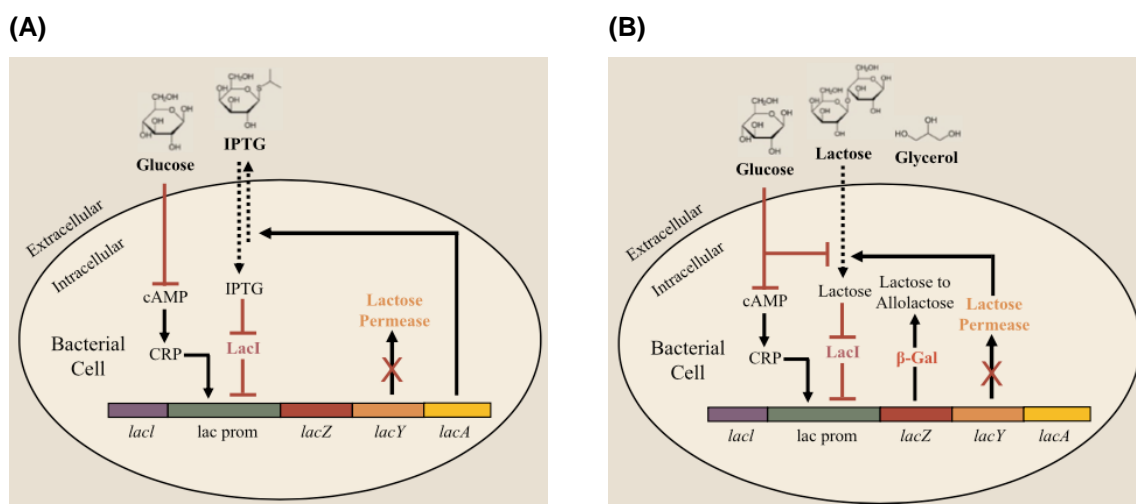


Figure 3.2. Representation of (A) IPTG induction and (B) auto-induction (by using a combination of glucose, lactose and glycerol in the media) for expression of LicB protein.

IPTG – Isopropyl β -D-1-thiogalactopyranoside, CRP – C-reactive protein, cAMP – cyclic adenosine monophosphate, β -Gal – β -galactosidase. Adapted from Crowley and Rafferty (2019) ³⁸⁴.

In order to extract LicB from a membrane environment, membrane fractions were treated with different solubilising chemicals, namely non-ionic detergents (DDM, DM and Cymal-5) and SMA polymer. Their main advantages and disadvantages of these reagents have been described in Chapter 1. The choice of these detergents reflects the dominance of the maltosides in the purification and crystallisation of membrane proteins ^{134,135} and the increasing popularity of SMA in membrane protein studies ^{163,167,171}.

3.1.2. Special considerations in applying biophysical techniques to membrane proteins

An arsenal of techniques including SDS-PAGE, Blue Native-PAGE, SEC-MALS and circular dichroism spectroscopy was used to assess the purity, oligomeric state, aggregation propensity, thermal stability and secondary structure of purified LicB. A brief discussion on how membrane protein properties can affect these analytical methods will be presented.

SDS-PAGE electrophoresis is used to evaluate the purity of purified proteins. The separation and analysis of MP by SDS-PAGE gel electrophoresis can be affected by the amount of detergent used for solubilisation, the presence of modifications (glycosylation or phosphorylation) as well as the number of transmembrane domains (number of hydrophobic

amino acids)³⁸⁵. Typically, most soluble proteins bind 1.4 g SDS/g protein, thus forming negatively charged complexes³⁸⁶. However, the amount of SDS that binds to MP can be significantly diverse; up to 4.5 g SDS/g of membrane protein has been reported, which causes aberrant migration in SDS-PAGE^{385,387}. Moreover, high concentrations of non-ionic detergents can also modify SDS micelles, resulting in unusual gel migration or smeared bands. In some cases, this effect can prevent the fixation of the MP in the gel and prevent effective staining with Coomassie dye. To be more accurate, the determination of molecular weights for MP on SDS-PAGE gel requires calibration with appropriate membrane-soluble standards, which are not commercially available. Similarly, the resolution of membrane proteins in Native-PAGE gel is commonly associated with low clarity and resolution, rendering them difficult to analyse³⁶⁹. This is now starting to be addressed. For example, Pollock *et al.* have recently demonstrated a new separation method to examine MP using SMALP and native gel electrophoresis in their native state³⁸⁸.

Purification of MP through affinity chromatography can be difficult since detergent molecules bound to the protein can entirely or partially occlude protein fractions that mediate binding to the resin. Moreover, when SEC is paired with detergent buffers, the apparent molecular weight of a membrane protein increases due to the detergent micelle weight. However, the amount of detergent (and lipids) associated with the proteins is usually unknown and differs depending on the protein³⁷¹. Therefore, SEC is valid only for determining MP monodispersity and aggregation propensity of membrane proteins. Techniques that determine the amount of detergent molecules within the protein-detergent complex have been developed; some examples include analytical centrifugation³⁸⁹ and MALDI-TOF MS³⁹⁰. Additionally, SEC with in-line light scattering (SEC-MALS) has been considered as a reliable technique to determine the oligomeric state of MP within the micelle^{29,371,391}.

Circular dichroism (CD) spectroscopy can provide insights into the secondary structure of proteins in solution as well as their thermal stability. During CD, beams of left-handed and right-handed circularly polarised light are passed through a protein sample held in a quartz cuvette. The differential absorbance of structural motifs of the peptide backbone, such as α -helices and β -sheets, result in distinctive CD spectra (Figure 3.3). Proteins that are expected to be exclusively α -helical, as LicB, exhibit a characteristic double-dip shaped spectrum with a strong positive band between 190 nm and 200 nm and two strong negative bands at 208 nm and 222 nm^{392,393}. Miles and Wallace highlighted the effects of environmental and biophysical properties of membrane proteins on CD spectra³⁹². Among them are the solvent shift effect, differential light scattering effect, lipid to protein ratio effect and differential absorption flattening effect³⁹². Overall, the combined influence of these effects leads to differences in the

spectra of membrane proteins, either variance of magnitude or wavelength. Mitigation strategies included developing a reference database comprising membrane proteins' spectra and modifying the experimental setup or sample preparation ³⁹².

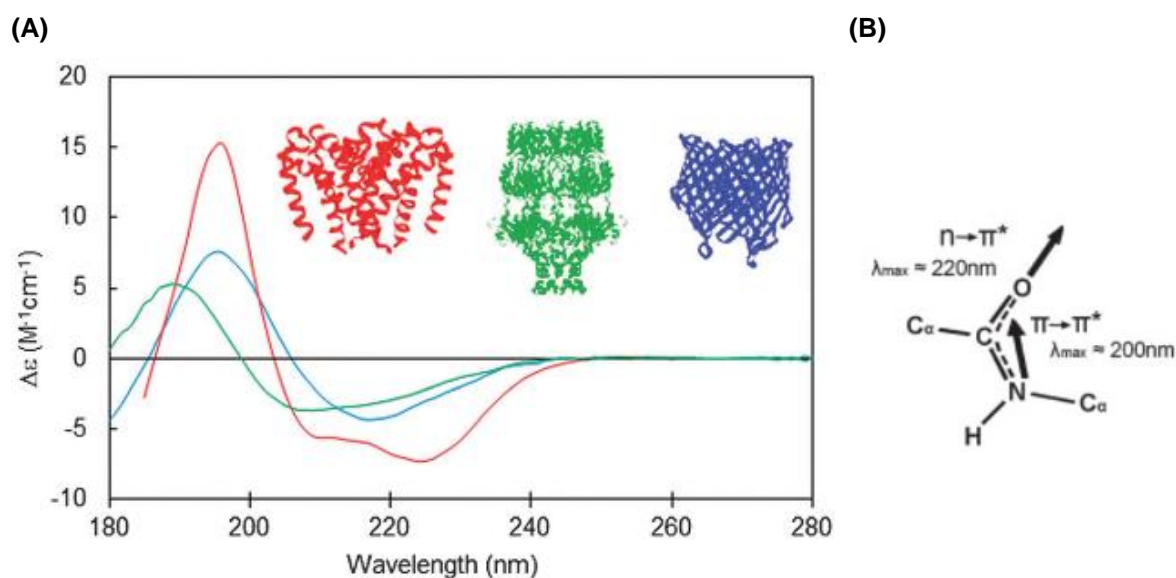


Figure 3.3. Differences in the spectra of proteins with different secondary structures.

(A) Red spectrum corresponds to a predominantly α -helical bundle; Blue spectrum corresponds to a β -barrel; Green spectrum corresponds to a protein exhibiting mixtures of the motifs mentioned above as well as unordered structure. **(B)** Transition dipoles (as thick arrows) of the $\pi \rightarrow \pi^*$ and $n \rightarrow \pi^*$ transitions are underlying the signal at shorter wavelengths. Adapted from Miles and Wallace (2016) ³⁹².

3.2. Results

3.2.1. Bioinformatic analysis of LicB

3.2.1.1. Secondary structure and transmembrane topology

The sequence of recombinant LicB from *H. influenzae* is shown here as a fasta sequence. This sequence of 305 amino acids had a calculated molecular weight of 34,119.69 Da, a theoretical isoelectric point of 8.82, and an extinction coefficient (ϵ_{280}) of 64,290.

>LicB protein sequence

```

MGRGYLFGIL SAVFWALSGL LYNELPLSEY TALGKVISLL FLIDFCSELLV 50
IGITLWRKSA VDFQGVFWQP ALSGILGGPI GMSAYLLSIH YLTIYYAAPL 100
SSLFPVFAAL MSYWILKEKI TKTAQFGFAL AISSSSLAI EVGQEITFNT 150
IGFIFLIICI LGWSSEIVIS SYTMRSLSGL QVYFLRLCGS TIGYLLILFI 200
LSLKNFSLDI LSFNYVQIAG VIIFGALSYC CYYQAIYLLK PIKAMALNIT 250
YSVWAIGLGY LLYKQPIKPI TLLLTLLLSA GVIVTLYYKG EQKLEHHHHH 300
HHHHH* 306

```

>LicB DNA sequence

```

ATGGGTCGTG GCTATCTCTT TGGCATACTA TCCGCTGTAT TTTGGGCATT 50
ATCTGGACTG TTGTATAATG AATTGCCATT AAGTGAATAT ACTGCGTTAG 100
GGAAAGTGAT TTCATTGCTT TTTTAAATG ATTTTGTGCTC GTTATTAGTT 150
ATCGGCATCA CATTATGGCG AAAAAGTGCG GTAGATTTTC AAGGTGTTTT 200
TTGGCAACCT GCGTTATCTG GCATATTAGG CGGACCAATA GGAATGTCTG 250
CCTATTTATT AAGCATTCAT TATTTAACAA TCTATTATGC TGCACCGCTT 300
TCATCTCTCT TTCCTGTTTT CGCTGCGTTA ATGTCTTATT GGATTCTAAA 350
AGAAAAAATT ACCAAAACAG CCAATTCGG ATTCGCCTTG GCTATTATTT 400
CTTCTTCGTT GCTTGCTATT GAAGTAGGAC AAGAAATAAC CTTTAATACT 450
ATCGGTTTTA TTTTTTTAAT AATTTGTATA TTAGGGTGGT CATCTGAAAT 500
AGTGATATCC TCCTATACAA TCGCTTCTCT CTCTGGACTT CAGGTGTATT 550
TTTTAAGACT ATGTGGTTCA ACTATTGGTT ATTTATTGAT TCTATTTATT 600
CTTTCCTTGA AAAATTTCTC ACTAGATATA CTTAGTTTTA ATTATGTACA 650
AATTGCTGGT GTAATAATCT TTGGTGCATT ATCTTACTGC TGCTATTACC 700
AAGCGATTTA TCTGTTAAAA CCAATCAAAG CAATGGCATT AAATATCACA 750
TATTCTGTAT GGGCAATTGG ACTTGTTTAT CTGCTCTATA AACAGCCAAT 800
TAAGCCCATT ACCTTACTGC TCACCTTATT ATTAAGTGCA GGAGTGATCG 850
TTACTTTTA TTACAAAGGG GAACAAAAAC TCGAGCATCA TCATCATCAC 900
CACCACCACC ACCACTGA 918

```

Figure 3.4. Full protein and DNA sequences of LicB expressed in *E. coli*.

Start and stop codons are shown in bold. His₁₀ tag is highlighted in red.

The PSIPRED server was used to predict the secondary structure of LicB from the primary sequence³⁹⁴. The results, shown in Table 3.1, predict that LicB will contain ten α -helices. Transmembrane topology analysis via multiple methods confirms this result, predicting with reasonable confidence that LicB would form ten transmembrane α -helices with N_{in}/C_{in} topology – that is, with both the N- and C-terminus residing in the cytoplasm (Table 3.1). The exception to this came from TMpred predictor, the only method used based on statistical analysis⁸⁵.

TMpred makes predictions using a database of naturally occurring transmembrane proteins and a mixture of multiple matrices for scoring³⁹⁵. Of the methods used, TOPCONS is thought to be generally among the most accurate^{85,88}. The topology generated by TOPCONS was depicted graphically with TOPO2 (Figure 3.5).

An expected Kyte-Doolittle hydrophobicity analysis demonstrates the high overall hydrophobicity of LicB (Figure 3.6A), with hydrophobic amino acids distributed throughout the protein sequence (Figure 3.6B). Methods designed to predict protein crystallizability, such as XtalPred, returned the worst possible score for LicB. Prediction by XtalPred takes into consideration transmembrane helices as well as signal peptides³⁹⁶.

Table 3.1. Output of bioinformatic analyses of the primary structure of LicB.

The methods generally converge in predicting that LicB is a multipass membrane protein comprising 10 transmembrane α -helices, with N_{in}/C_{in} topology.

Online predictor	N _{term}	# Transmembrane										C _{term}
		1	2	3	4	5	6	7	8	9	10	
PSIPRED	ND	3	32	64	94	122	150	179	211	242	269	ND
		-	-	-	-	-	-	-	-	-	-	
TMpred	out	26	55	91	116	139	174	201	238	262	287	in
		4	36	92	124	146	179	215	245	269		
TMHMM	out	-	-	-	-	-	-	-	-	-	-	out
		23	55	87	116	143	172	203	229	263	288	
HMMTOP	in	6	32	70	98	126	152	182	210	244	270	in
		-	-	-	-	-	-	-	-	-	-	
SCAMPI	in	23	55	89	115	143	169	201	233	261	287	in
		5	36	72	96	124	148	188	213	244	268	
OCTOPUS	in	-	-	-	-	-	-	-	-	-	-	in
		25	56	92	116	144	168	208	233	264	288	
Philius	in	4	33	68	95	122	150	182	213	243	268	in
		-	-	-	-	-	-	-	-	-	-	
TOPCONS (consensus)	in	25	53	88	115	142	170	202	233	263	288	in
		5	33	68	95	122	150	182	213	243	268	

3.2.1.2. Template based homology models

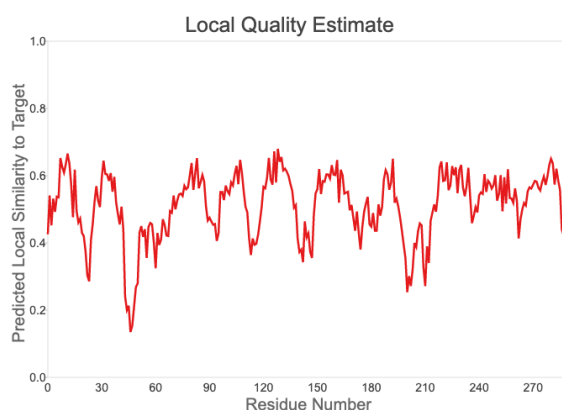
LicB homology modelling was performed using either SWISS-Model, HHpred or MEDELLER/Memoir. The servers HHpred and SWISS-model shared similar template selection from 24 and 42 entries, respectively. In SWISS-Model, the selected templates were aligned based on the higher GMQE (Global Model Quality Estimation) score and percentage of sequence identity. Templates from HHpred were listed according to Score and E-value. The template chosen for homology modelling was the PDB entry 5I20 (toxin efflux protein YddG) since it had a higher percentage of sequence identity and the highest GMQE. However, the model generated from 5I20 had an associated SWISS-Model QMEAN value of -6.47, reflecting poor model quality³⁹⁷ (Table 3.2), as observed in Figure 3.7. Considering MEDELLER/Memoir, the PDB entry 5I20 was used as a template; The model generated by MEDELLER was similar to the SWISS-Model, with minor variations in the helices position within the protein sequence.

Table 3.2. Best scored templates from SWISS-Model and HHpred.

GMQE and QMEAN are SWISS-Model parameters to evaluate the quality of homology models. GMQE score is expressed as a number between 0 to 1, where higher numbers indicate higher reliability. The HHpred E-values and probability scores confirm that the chosen models are sequence homologues of LicB that are suitably close for model-building.

Tool	PDB entry - description (ref.)	GMQE	% of identity	QMEAN
SWISS-Model	5I20 – DMT superfamily transporter ³⁹⁸	0.56	15.11	-6.47
	6OH2 - CMP-sialic acid transporter ³⁹⁹	0.55	11.81	-6.28
		E-value	Score	Probability
HHpred	5I20 (described above)	6.3e-25	180.91	99.96
	5Y78 – Putative hexose phosphate translocator	4.9e-25	184.09	99.96
	6OH2 (described above)	2.0e-24	180.39	99.95

(A)



(B)



(C)

```

Model_01 MGRGYLFGILSAVFWALSGLLYNE--LPLSEYTAGKVISLLFLIDFCSSLVIGITLWRKSAVDFQGVFWQPAL 72
5i20.1.A RSSATLIGFTAILLWSTLALATSSTGAVPQFLLTALTFTIGGAVGI-----AAGLRGVGLSVLRQFVFPVWV 69
Model_01 SGILGGPIGMSAYLLSIHYLTIYYAAPLSSLPVFAALMSYWLKKEIKTKTAQFGFALAISSSLLAIEVGQE-- 145
5i20.1.A HGIGGLPGYHFEYFSAALKLAFPAEAGLVAYLWPLLIVLFSAPLPGERLFAHVAGALMGLAGTVVLLGARAGGFG 144
Model_01 ITFNTIGFIFLIICILGWSSEIVISSYTMRSLSGLQVYFLRLCGSTIGYLLILFILSLKNEFLDILSFNYVQIA 219
5i20.1.A FAEYVPGYLA AAAACAVIWSVYSVA SRRFAR-VFTEVVAGFCLATAALSALCHIL---EPPSVWP-VGSENLAVV 214
Model_01 GVIIF-GALSYCCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYKQPIKPIITLLTLLLSAGVIVTLYYKGEQK 293
5i20.1.A ALGIGPVGIAEYTNDIGMKBGIVRLLGVLISYAAPVLTLLLVVAGFAAPSALAIACALIVCGAAVAQLLAR 287
Model_01 LEHHHHHHHHHHH 305
5i20.1.A -----

```

 Low Quality -> High Quality

Figure 3.7. Quality parameters of the LicB homology model.

(A) Local quality estimate at each residue of the model, expressed as similarity to the target. Typically, residues showing a score below 0.6 are expected to be of low quality. (B) Normalised QMEAN4 score to show the global quality of the model (red star). Higher values mean that the model is of comparable quality to experimental protein structures of similar size. (C) Alignment between target protein sequence (LicB) and the PDB entry 5I20 used for homology modelling. The sequence is coloured based on the QMEAN score scale, orange being associated with a low-quality model.

Overall, the computational model corroborates the TM topology prediction with the presence of ten helices; However, the amino acid range in each TM is distinctive from the interval described in Table 3.1, as well as the presence of random coil within the helix, such as in blue, green and orange TMs (Figure 3.8A). The PDB entry 5I20 corresponds to the YddG transporter responsible for the export of amino acid residues from *Starkeya novella*³⁹⁸. In terms of protein function, these two proteins are responsible for transporting different compounds in different directions: YddG is an exporter while LicB is an importer. Sequence alignment showed low identity and similarity percentages of about 14.5% and 24.3%, respectively. Moreover, essential residues involved in the transport mechanism of YddG, such as Gly71, Gly73, Gly74, Gly77, are not present in *H. influenzae* LicB (Figure 3.8B). In summary, a more appropriate template is required to improve this homology modelling.

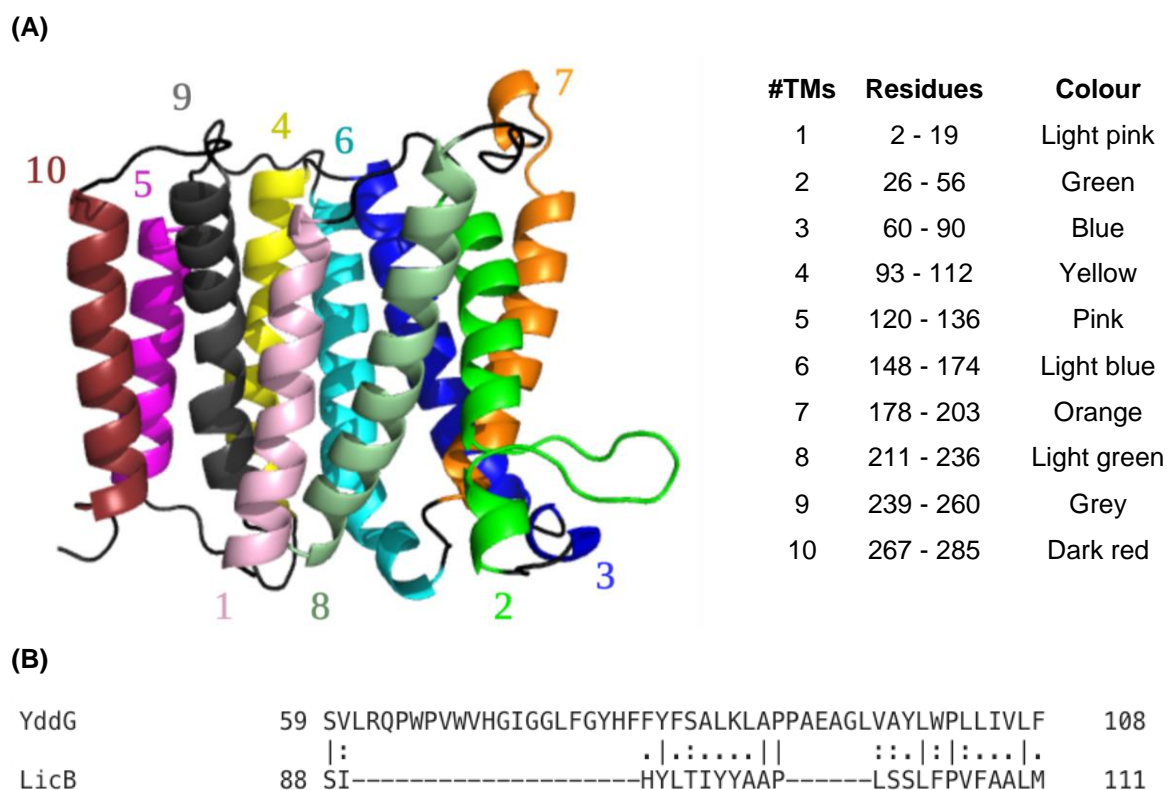


Figure 3.8. Comparison between YddG and LicB transporters.

(A) LicB homology model generated by SWISS-MODEL. The sequence position of the transmembrane helices, and their corresponding colours, are shown. (B) Part of the sequence alignment between *Starkeya novella* YddG and *H. influenzae* LicB, highlighting the relatively weak homology between the two proteins.

3.2.2. Expression and purification of LicB in *E. coli*

The *licB* gene was amplified from *H. influenzae* genomic DNA by PCR and cloned into a commercial expression vector (pET28) to facilitate expression from the T7 promoter. The construct was designed to include a 10-residue polyhistidine tag (His₁₀) at the C-terminal to allow protein purification by immobilised metal affinity chromatography (IMAC). The entire sequence of the LicB expression construct was confirmed by sequencing and is shown in Figure 3.4.

3.2.2.1. Screening different membrane solubilising chemicals for LicB purification

Initial large scale expression trial of LicB was in T7 Express *E. coli* strain and super broth auto-induction media (SB_{AIM}) for expression. Three different mild non-ionic detergents from the maltoside family (DDM, DM, Cymal-5) were tested to see which best-supported protein purification. Alternatively, polymer SMA2000 was used to test a detergent free solubilisation procedure.

After expression, all cultures were treated identically. Cellular fractions were collected by ultracentrifugation and the membrane fraction treated with DDM, DM and Cymal-5 to form soluble protein-detergent complexes or SMA2000 to form SMALPs. Solubilised membranes were applied to a Ni²⁺ affinity column, and the recombinant LicB was eluted from the column by competition with imidazole. The purity of the recombinant LicB was evaluated by staining SDS-PAGE gels with Coomassie Brilliant Blue. To further improve purity as well as a measure of stability, oligomeric state, aggregation propensity and dispersity, protein-detergent complexes were applied to a size-exclusion column (SEC) directly following affinity chromatography. This protocol is shown in Figure 3.9.

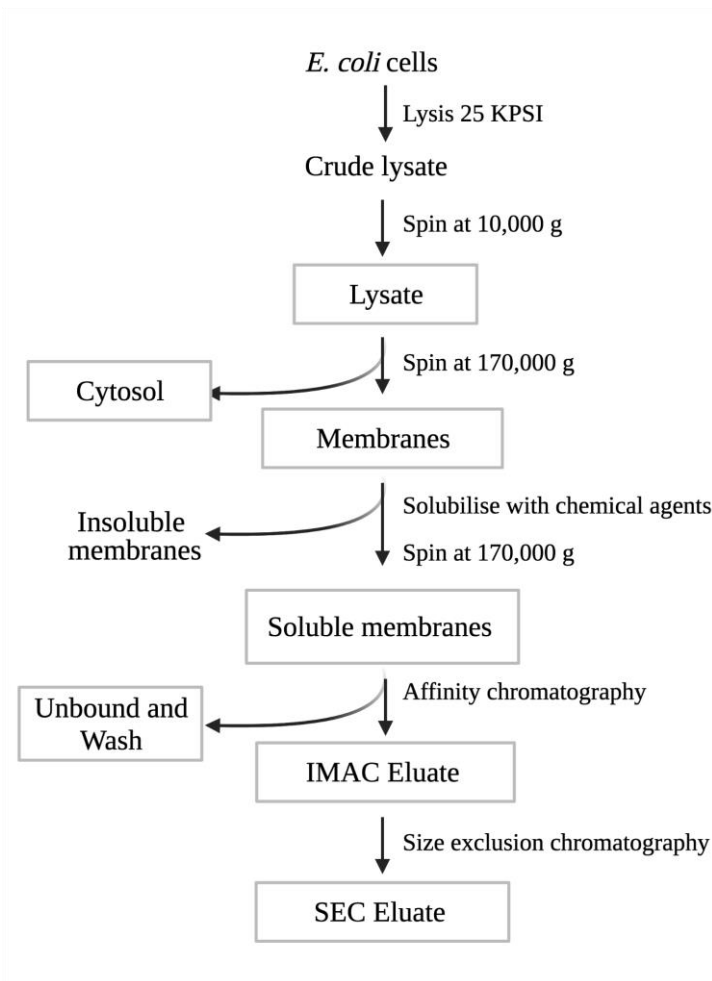


Figure 3.9. Schematic representation of purification protocol.

LicB protein was purified from *E. coli* cellular membranes using a 2-step purification process, affinity chromatography and size exclusion chromatography. Cell fractions boxed in the scheme were collected and analysed by Coomassie-stained SDS-PAGE in later figures.

Table 3.3 and Figure 3.10 show that all of these detergents enabled the purification of LicB with yields after IMAC of 1.17 - 2.37 mg per litre culture. After SEC the corresponding yields reduced significantly.

IMAC eluates were analysed with SDS-PAGE and showed a single major band corresponding to recombinant LicB that migrated close to the theoretical molecular weight of 34 kDa. Some minor contaminants were also evident (Figure 3.10). Alternatively, when SMA2000 was used in purification, no protein was detected on the SDS-PAGE gel (Figure 3.10D); However, a possible band corresponding to LicB was observed between 25 to 35 kDa in the membrane, soluble membrane and unbound fractions, suggesting that resin binding was the limiting factor.

Table 3.3. Yields of recombinant LicB obtained using DDM, DM, Cymal-5 and SMA2000 as solubilising chemicals.

Error represents standard deviation and was calculated from at least triplicate independent experiments.

Chemical	% used for solubilisation	% used in working buffers	Yield after IMAC (mg/L of culture)	Yield after SEC (mg/L of culture)
DDM	1.0%	0.05%	1.17 ± 0.57	0.14 ± 0.04
DM	1.0%	0.20%	2.37	0.21
Cymal-5	2.4%	0.24%	1.65	ND
SMA 2000	2.5%	-	ND	ND

The aggregation propensity and dispersity of purified LicB in each of the three detergents was assessed by SEC (Figure 3.10). The elution profile in Cymal-5 was more heterogenous with extensive protein aggregation to oligomers of >650 kDa that elute in the void volume (V_0). Although the elution profile showed some non-specific aggregation, LicB was eluted as a major peak at around 13.4 mL for DDM and 13.9 mL for DM, corresponding to an apparent molecular weight of 94.1 kDa and 75.3 kDa for DDM and DM, respectively. This difference is most likely due to the micelle size, which is 40 kDa for DM and between 65 – 70 kDa for DDM¹³⁴. Thus, the estimation of the apparent molecular weight of LicB inside the micelle is close to its theoretical molecular weight (Table 3.4). When reapplied to the same column, LicB eluted as a single and defined peak (Section 3.2.3.2), showing that this protein stays stable in DDM without further aggregation. Fractions were collected and ran at SDS-PAGE gel, confirming the elution of the protein. Fractions corresponding to the elution peaks were concentrated and used for further biophysical characterisation.

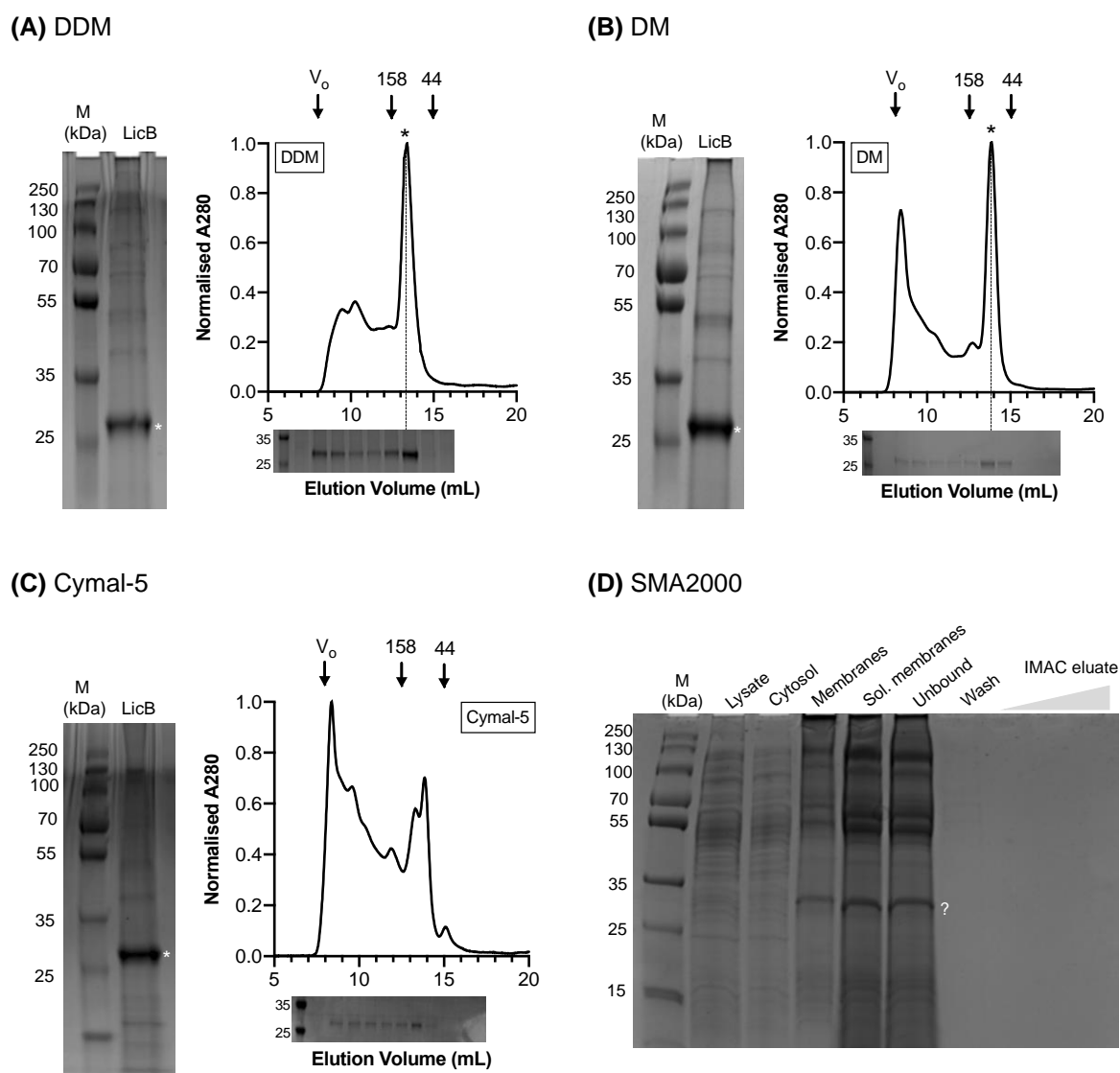


Figure 3.10. Coomassie-stained SDS-PAGE gels demonstrating the successful expression and purification of recombinant LicB after IMAC and, subsequent SEC chromatograms.

(A) DDM; (B) DM; (C) Cymal-5 and (D) SMA2000. The theoretical molecular weight is 34 kDa, (*) corresponds to the position of LicB band. In SEC profile V_0 indicates the void volume. For DDM and DM purifications the major peak marked with a (*) is used for the apparent molecular mass of the protein-detergent complex (MW_{app} PDC).

Table 3.4. Calculated MW_{app} PDC and LicB based on K_{av} values.

K_{av} values were calculated based on the Equation 2.2 and calibration curve is shown in Figure 2.2.

	V_e	$V_e - V_0$	K_{av}	$\text{Log}_{10}(\text{MW})$	MW_{app} PDC (kDa)	MW_{app} LicB (kDa)
DDM	13.4	5.5	0.34	1.97	94.07	24.07 – 29.07
DM	13.9	6.0	0.37	1.88	75.28	35.28

3.2.2.2. Screening different media and host cell

Next, large scale expression trials of LicB screened three different *E. coli* host strains (T7 Express cells, C43(DE3) cells and BL21 Star (DE3) cells) as well as different culture media (LB and SB) using IPTG or auto-induction, depending on the strain (Figure 3.11 and Figure 3.12). DDM was chosen to solubilise membrane fractions.

Overall, LicB was expressed at relatively low levels (Table 3.5). LicB was successfully expressed in T7 Express cells, with higher yields when SB_{AIM} was used versus direct IPTG induction (Figure 3.11 and Figure 3.12A). This might be at least partly due to increased biomass since AIM allowed ten times greater cell density when compared to IPTG induction (Table 3.5). No LicB expression was detected when other host strains were tested, namely BL21 Star (DE3) or C43(DE3), as observed in SDS-PAGE gels recorded in Figure 3.12B and C. Surprisingly, scaling-up of the best condition (T7 Express cells growth in 3 x 1 L SB_{AIM}) did not improve protein yield beyond that of a single 1 L culture, for unknown reasons. As expected, recombinant LicB was localised in *E. coli* membrane fractions in all purifications.

As previously observed, SEC chromatograms of LicB suggested the presence of non-specific protein aggregation due to the presence of a peak corresponding to larger complexes eluted at void volume (in Figure 3.11B as well as in Figure 3.12). Nevertheless, LicB in DDM eluted as a major and homogenous single peak at an elution volume of ~13 mL, confirmed by SDS-PAGE analysis.

Furthermore, the expression of the correct LicB protein was confirmed by mass spectrometry of the major protein band observed on SDS-PAGE. Tryptic digest MS showed a single peptide - SLSGLQVYFLR-. A BLAST search against the NCBI sequence confirmed that the only result corresponded to *H. influenzae* LicB (Table 3.6). It is not unusual for membrane proteins to produce low numbers of tryptic peptides, which prevents full coverage of the protein sequence by this method (P. Curnow, personal communication).

Table 3.5. LicB expression conditions and final purification yields after affinity and size exclusion chromatography.

Expression conditions tested for LicB – culture media and induction system (see Table 2.13 for detailed information on expression conditions). Purification yields (mg protein obtained /L of culture) after IMAC (including protein concentration step) and SEC. The latter was quantified from the elution peak ($V \sim 13$ mL) sample. Most of these conditions were only tested once; therefore, no yield range is described.

Host cell	Culture media	Induction system	Time of expression	Final OD _{600nm}	Yield after IMAC (mg/L)	Yield after SEC (mg/L)
T7 Express	SB	AIM (1 x 1 L)	24 hour	8 - 10	1.17 ± 0.57	0.14 ± 0.04
		AIM scale-up (3 x 1 L)	24 hour	ND	0.340	0.022
T7 Express	LB	0.1 mM IPTG (3 x 1 L)	2 hour	1.05	0.122	0.018
C43(DE3)	LB	0.1 mM IPTG (3 x 1 L)	2 hour	3.50	0.041	0.005
BL21 Star (DE3)	SB	AIM (1 x 1 L)	24 hour	5.30	0.003	ND

*AIM – auto-induction

Table 3.6. Top 5 results from the BLAST search on the peptide sequence -SLSGLQVYFLR-

Description	Scientific Name	Max Score	Total Score	Query Cover	E-value	Accession
lic-1 protein B	<i>Haemophilus influenzae</i> HK1212	37.5	37.5	100%	0.12	EFA28436.1
protein LicB	<i>Haemophilus influenzae</i>	37.5	37.5	100%	0.12	SPX40716.1
LicB protein	<i>Haemophilus influenzae</i>	37.5	37.5	100%	0.12	AAA24972.1
lic-1 protein B	<i>Haemophilus influenzae</i> HK1212	37.5	37.5	100%	0.12	EFA28238.1
Protein licB	<i>Haemophilus haemolyticus</i> M21639	37.5	37.5	100%	0.12	EGT81144.1

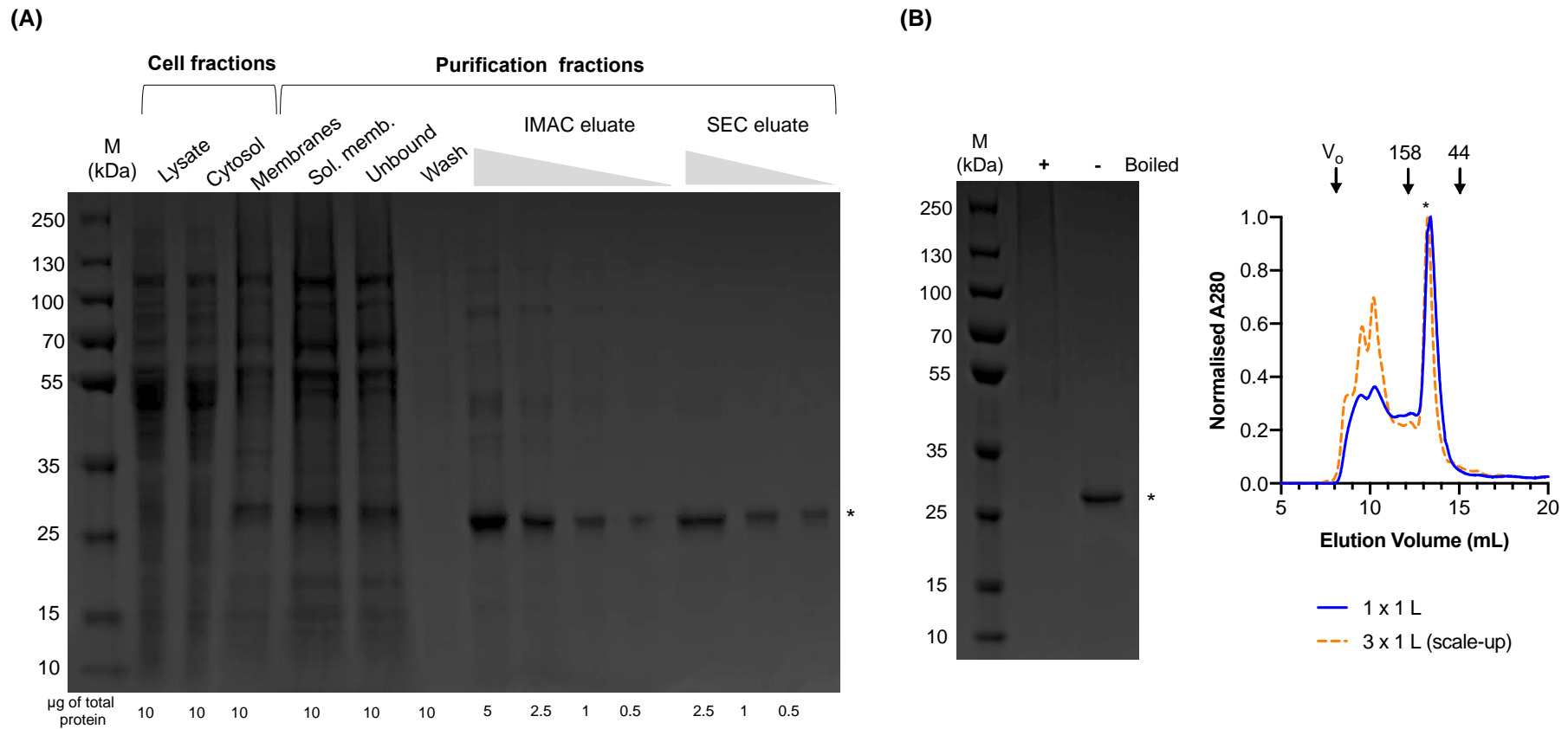


Figure 3.11. Expression of LicB protein in T7 host strain (SB_{AIM}) and purification after solubilisation with DDM.

(A) SDS-PAGE gel of cell fractions and column purification fractions as shown, stained with Coomassie Brilliant Blue. The expected position of LicB on the SDS-PAGE gel based on amino acid composition is indicated by an asterisk. The protein loadings used for Coomassie staining are shown. **(B)** Size exclusion chromatography of LicB, showing a peak at an elution volume of ~13 mL, and SDS-PAGE gel of the eluted protein with and without boiling before loading. The blue line corresponds to purification from 1 L of culture media while the orange line corresponds to LicB purified from 3 L. V_0 indicates the void volume, and numbered arrows indicate the elution volumes of two standard proteins of known molecular weight.

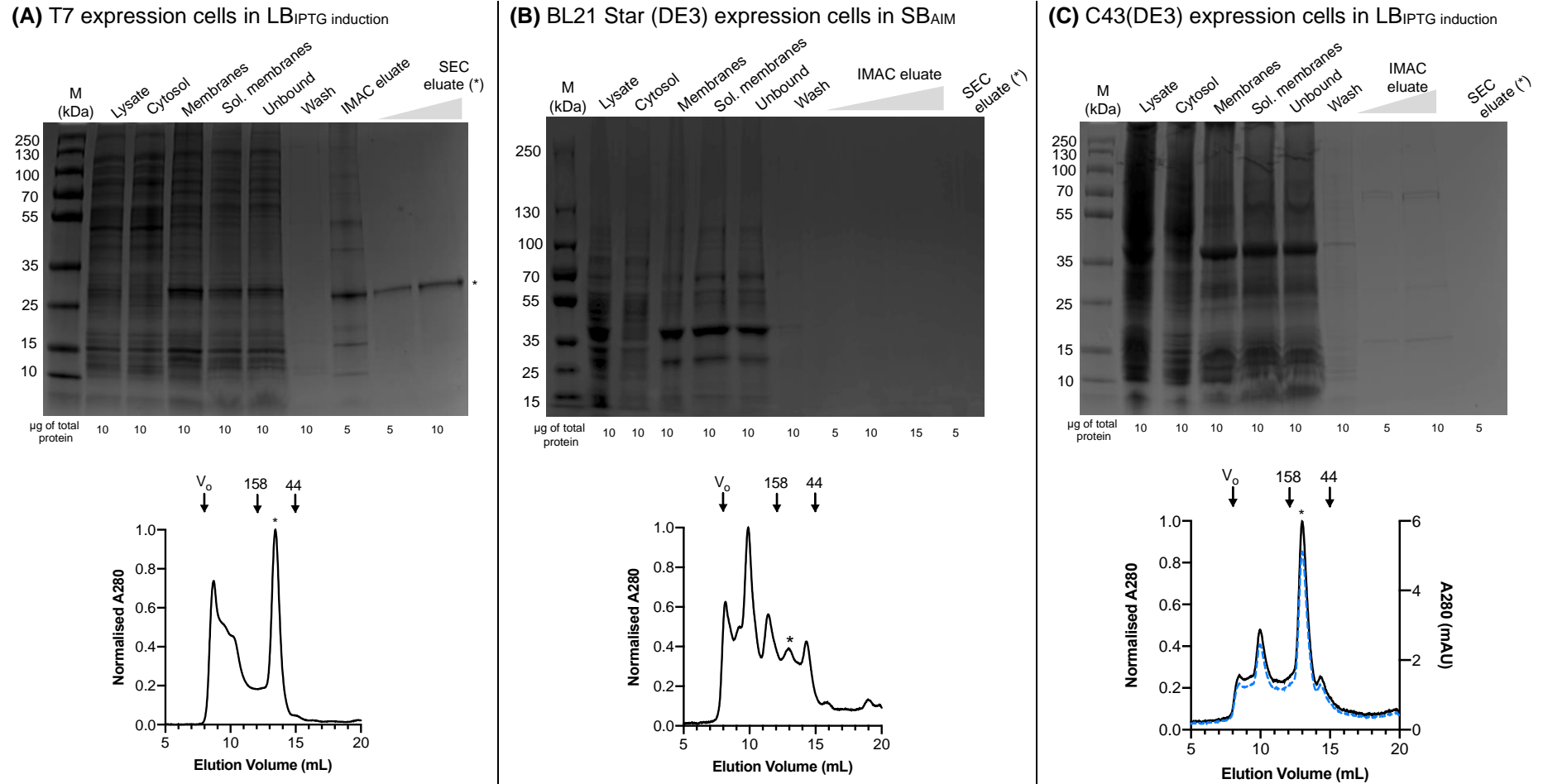


Figure 3.12. Expression of LicB protein in different host *E. coli* strains and culture media.

Each expression system used is labelled on the top of each panel, and all purifications are in DDM. SDS-PAGE gels of cell fractions and column purification fractions stained with Coomassie Brilliant Blue. The protein loadings used for Coomassie staining are shown. The expected position of LicB on the SDS-PAGE gel based on amino acid composition is indicated by an asterisk. Size exclusion chromatography of LicB protein eluted from IMAC. The elution peak (*) was visualised in the SDS-PAGE gels. The blue line in (C) lower panel represents Abs 280 nm (mAU). V_0 indicates the void volume. Collectively, these expression screens determine that the best conditions for recombinant expression of LicB are to use the T7 Express strain in auto-induction media.

3.2.2.3. Optimisation of culture growth conditions and solubilisation for improved LicB production

In order to try and increase LicB solubility and reduce aggregation tendency, three independent experiments were carried out. Changes to the original protocol include different expression temperature and time, as well as using a higher percentage of the solubilising detergent.

In the first experiment, protein expression was induced at 25°C (Figure 3.13B). A shorter expression time was tested in the second experiment, 8 hours at 37°C (Figure 3.13C). The last experiment was performed, where a different percentage of DDM (2%) during membrane solubilisation was used (Figure 3.13D). For a detailed description of these experiments see Table 2.13.

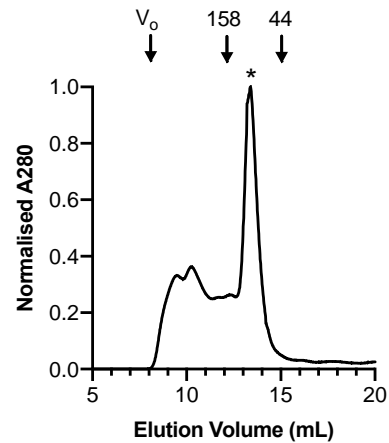
Figure 3.13 shows the SEC profiles of all three independent experiments. The percentage of LicB recovered from SEC was calculated as follows:

$$\% \text{ Recovery of LicB} = \frac{\text{Yield after SEC (elution peak)}}{\text{Yield after IMAC}} \times 100$$

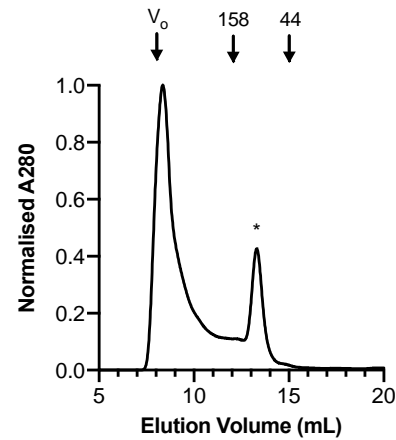
Equation 3.1.

Where the elution peak corresponds to the peak at ~13 mL (marked with an asterisk in Figure 3.13). Overall, these data determine that there was no substantial improvement with any of these variations.

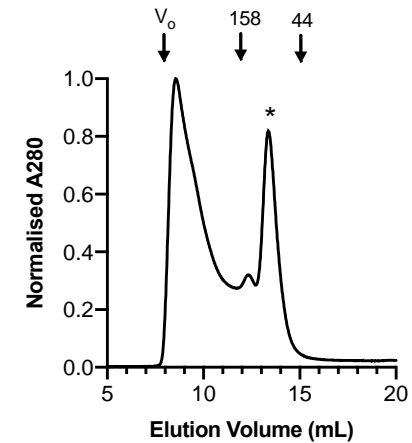
(A) Original protocol - LicB expressed at 37°C for 24 hour and solubilised with 1% DDM. Sample was loaded from IMAC directly to SEC



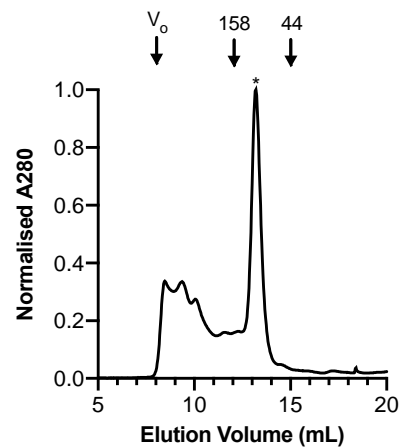
(B) LicB expressed at 25°C for 21 hour and solubilised with 1% DDM



(C) LicB expressed at 37°C for 8 hour and solubilised with 1% DDM



(D) LicB expressed at 37°C for 24 hour and solubilised with 2% DDM.



(E)

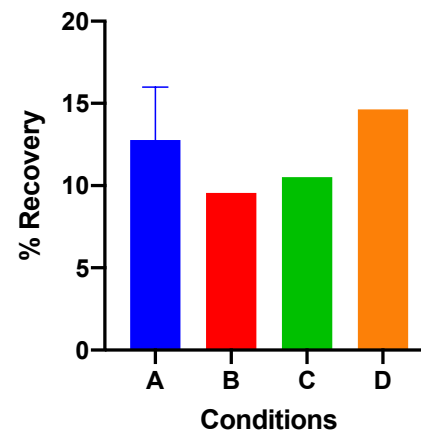


Figure 3.13. SEC profile of LicB protein purified from diverse expression conditions (A-D). (E) % of Recovery of LicB in the different conditions mentioned.

(A) Original protocol. **(B)** Change of temperature and time of expression. **(C)** Change of expression time. **(D)** Change of DDM percentage used. V_0 indicates the void volume, and (*) represents the elution peak of LicB protein.

3.2.3. Biophysical characterisation of purified recombinant LicB

3.2.3.1. Assessing secondary structure by circular dichroism

Circular dichroism (CD) was used to confirm the predicted α -helical structure of LicB. CD was also able to determine the thermal stability (melting temperature) and the long-term stability of LicB at 25°C. CD experiments were run using % (w/v) of detergents at working concentrations, either DDM at 0.05% or DM at 0.20%. Nevertheless, and to distinguish between similar samples, the percentage of solubilisation was used as a reference in the results and discussion section, and in the subsequent figures (DDM at 1% or 2%).

The CD spectrum of SEC-purified protein revealed that LicB was clearly α -helical in both DM and DDM, with characteristic negative deflections at 208 nm and 222 nm (Figure 3.14A). Data recorded in 1% and 2% DDM were indistinguishable. The intensity of the spectra suggested ~62% helix overall (based on Equation 2.9), very close to the expected value of 69% (210 of 305 amino acids within transmembrane α -helices). Helicity was slightly reduced in DM, being only 52% helix overall.

Long-term incubations further confirmed that purified LicB in either DM or DDM was stable at 25°C for at least 16 hour, with no change in the CD spectra over this time (Figure 3.14B).

Thermal denaturation of LicB revealed a major cooperative loss of secondary structure with T_m of around 66°C (DM or 1% DDM), whereas, in 2% DM, the protein was still stable until T_m of 71°C. Above T_m , LicB irreversibly aggregates (Figure 3.15A). Full melting scans of LicB in different micelles are shown in Figure 3.15B to D.

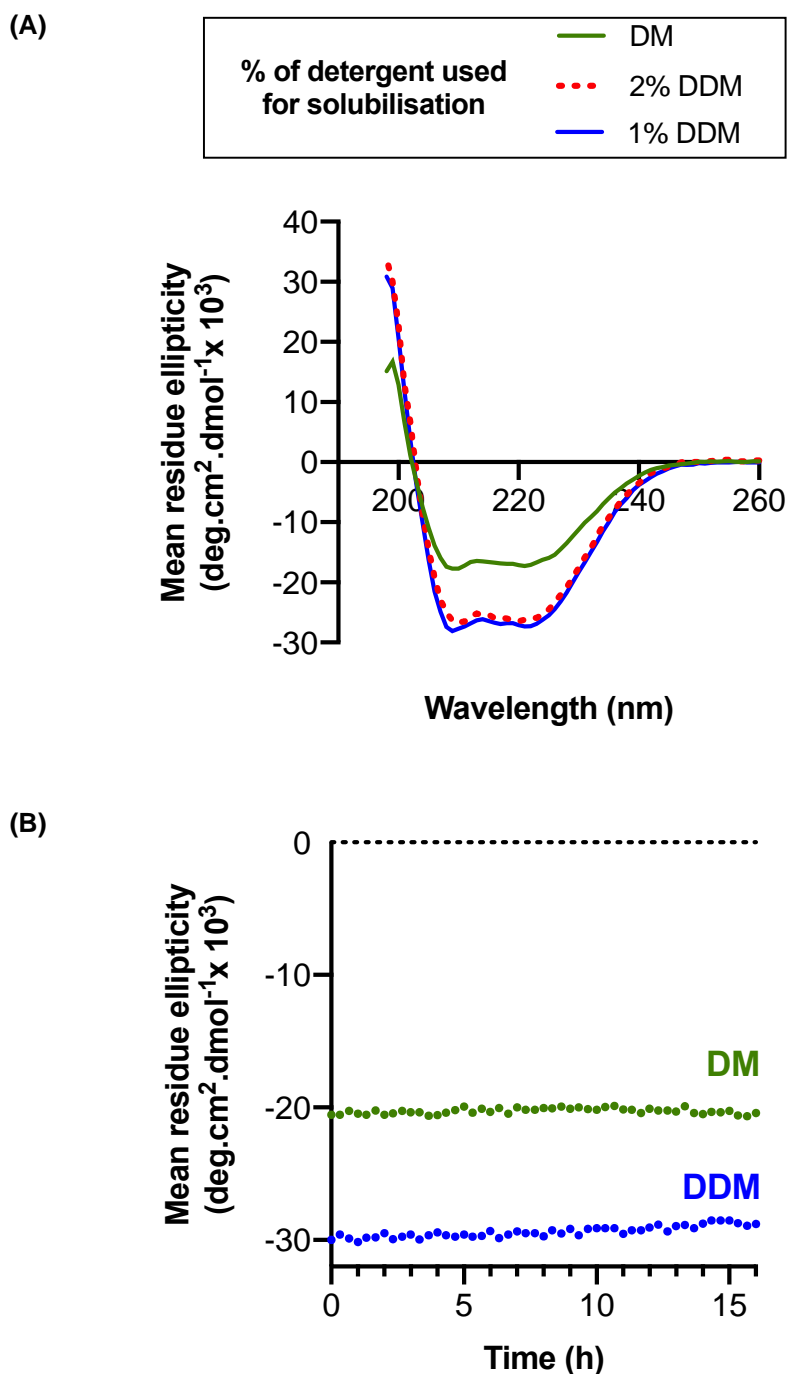


Figure 3.14. Circular dichroism to assess the secondary structure and long-term stability of LicB.

(A) CD spectrum for purified LicB in DM or DDM. Data for values of HT above 700 V were excluded.

(B) Stability of LicB in DDM or DM at 25°C for 16 hour. Data were collected every 0.3 hour. Circular dichroism experiments were run using detergent at working concentrations, DDM at 0.05% or DM at 0.20%. However, and to distinguish between samples, the percentages described in the above figure refer to the percentage of detergent used during the solubilisation step.

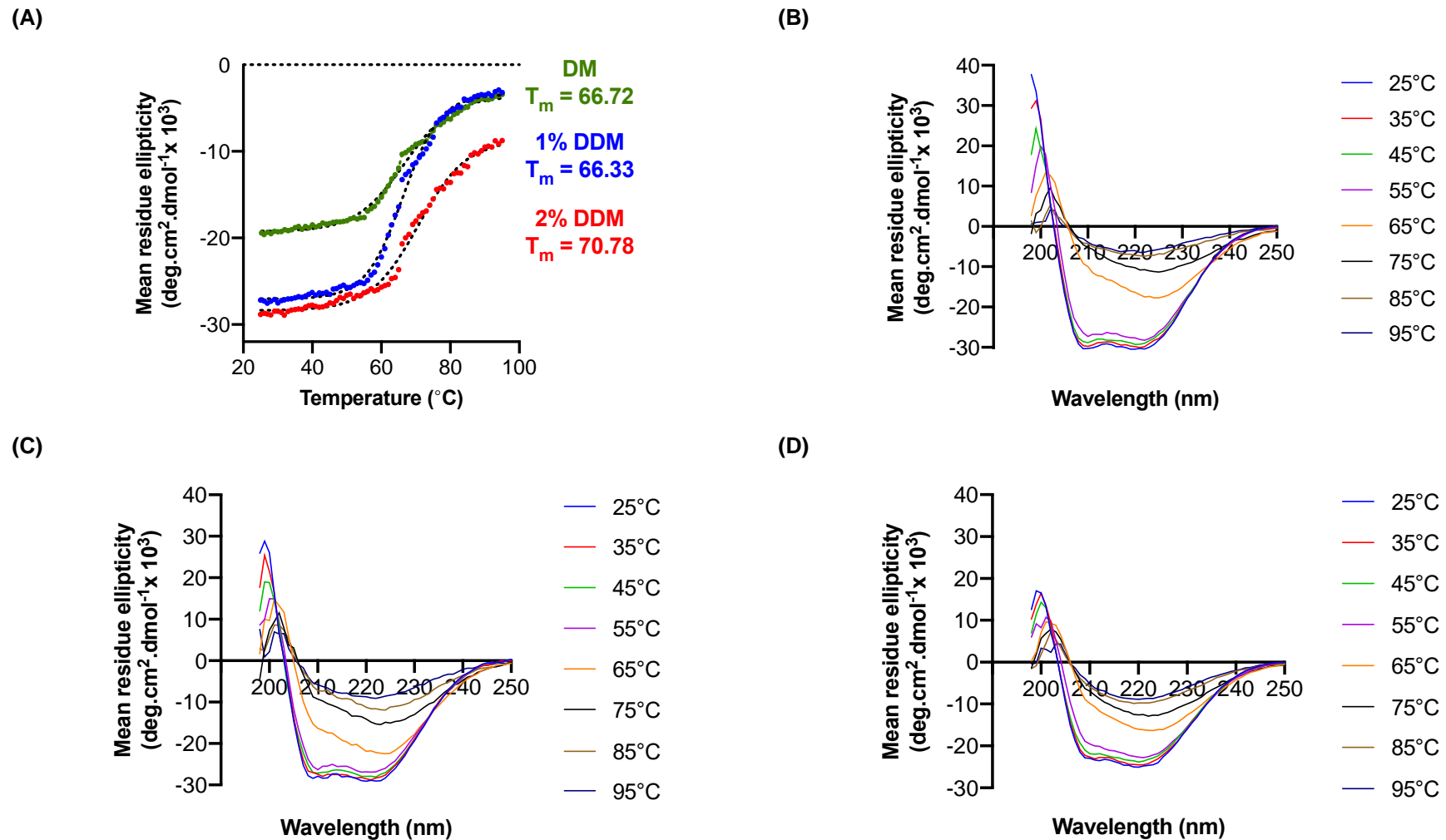


Figure 3.15. Full melting scans of LicB in different detergent micelle between temperatures of 25°C and 95°C.

(A) Thermal stability of LicB in DDM or DM. Data were collected from 25°C to 95°C with 1°C increments. Data were fitted to a Boltzmann sigmoidal function to guide the eye. Melting scans of LicB in **(B)** 1% DDM micelle, **(C)** 2% DDM micelle and **(D)** DM micelle (referring to the % of detergent used for solubilisation). Data for values of HT above 700 V were excluded.

3.2.3.2. Determination of LicB oligomeric state by SEC-MALS

Size exclusion chromatography coupled to in-line multi-angle light scattering (SEC-MALS) was used to calculate the molecular weight of LicB independent of the DDM micelle. The simultaneous measurements of light scattering (LS), refractive index (RI) and absorbance at 280 nm (UV_{280}) are shown in Figure 3.16. This figure combines normalised data from two individual experiments analysing either the LicB protein-detergent complex (blue lines) or DDM micelle alone (grey lines).

Using the three-detector method described by Slotboom *et al.*³⁷¹, the molecular mass for LicB in DDM micelle was found to be 26.1 kDa (Equation 2.3), close to the theoretical monomer weight of 34 kDa. This strongly suggests that LicB is a monomer in the DDM micelle. Additionally, the mass ratio detergent to protein (g/g), known as δ , was found to be 3.42 (Equation 2.5), corresponding to an attached DDM micelle of 89.3 kDa. This was slightly above typical literature values of 65-70 kDa.

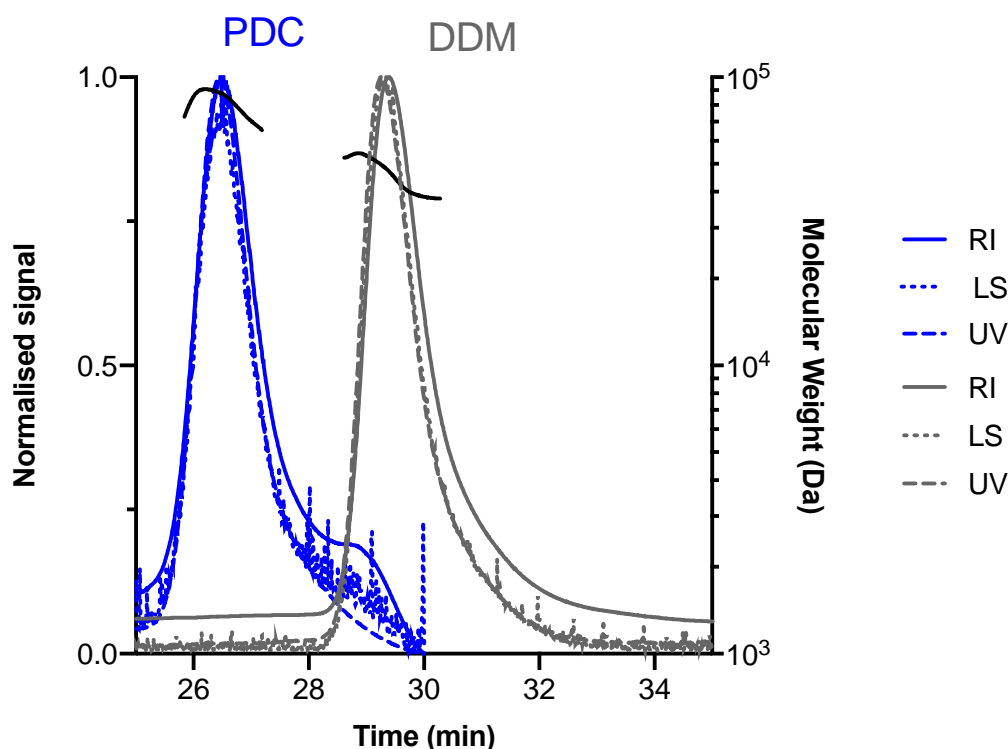


Figure 3.16. SEC-MALS analysis of protein-detergent complex (PDC) of LicB in DDM.

Normalised readings (from RI, LS and UV_{280nm}) are overlapped between the elution time of 25 and 35 minute of the PDC and DDM micelle samples. Black lines represent the molecular weight (Da) calculated across elution peaks.

We further confirmed the oligomeric state of LicB in DDM micelles by Blue Native-PAGE (Figure 3.17). Two samples were inspected, one of purified LicB after IMAC purification and the second sample of purified LicB after SEC. This revealed bands at higher molecular weights, confirming the presence of aggregates in the sample after IMAC. However, after SEC sample showed a single band slightly above 66 kDa (black arrow in Figure 3.17), which broadly agrees with SEC and thus confirmed the monomeric state of LicB in DDM.

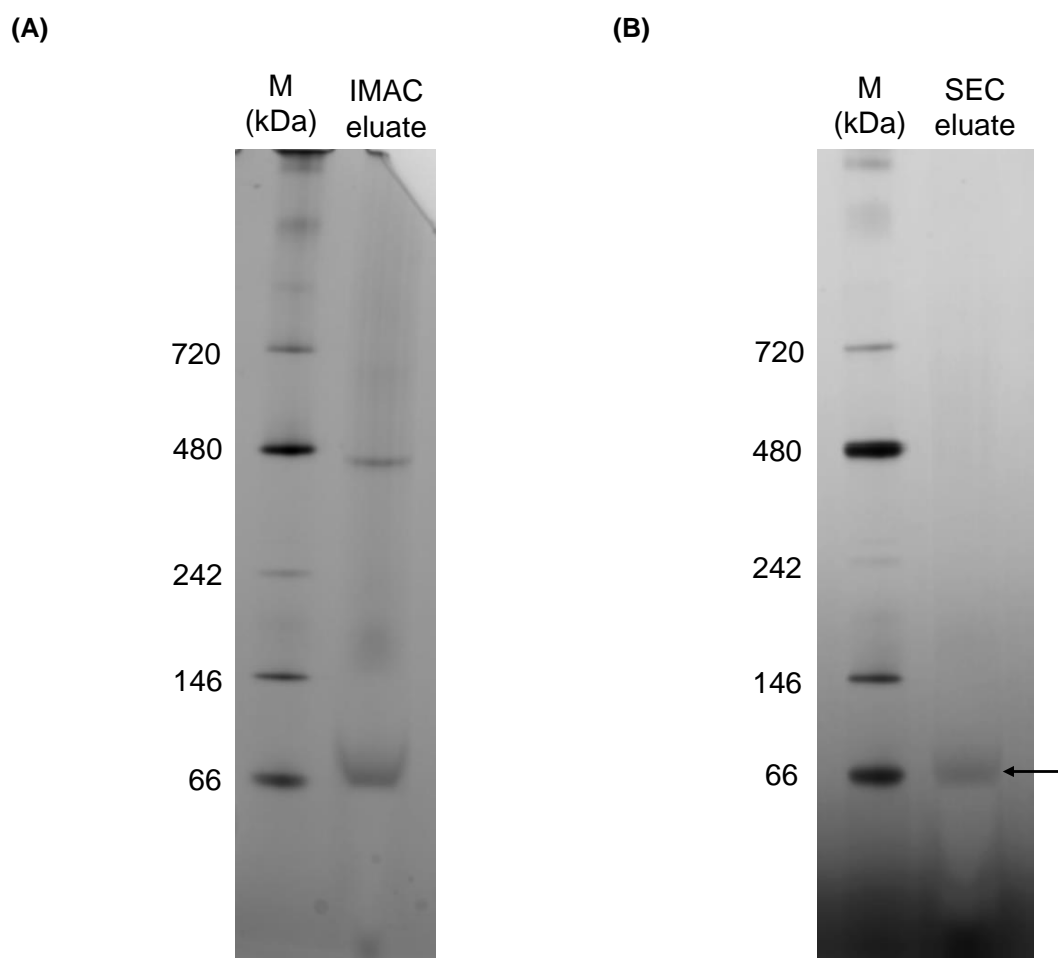


Figure 3.17. Blue Native-PAGE gel of purified LicB samples confirming the presence of aggregates.

(A) Sample of LicB eluted from IMAC resolved in Native-PAGE gel, showing more than a single band at higher molecular weights. **(B)** Sample of purified LicB eluted from SEC in Native-PAGE gel, confirming the monomeric state (represented by the arrow).

3.3. Discussion

3.3.1. Expression and purification results

The work outlined in this chapter sought to develop an expression and purification methodology to enable the optimal production of recombinant LicB. Overexpression is still the major bottleneck in functional and structural studies of membrane proteins, and even when the expression is optimised, membrane protein yields are typically between only 1-5 mg purified protein per litre of *E. coli* culture^{400,401}. Generally, LicB was within this range, being obtained at 1.17 ± 0.57 mg per litre of *E. coli* culture when DDM was used for purification (Table 3.5).

Of the three *E. coli* host strains tested, the T7 Express strain cell performed best, rendering the highest yield (Table 3.5). Contrary to expectations, LicB was not well-expressed either in C43(DE3) or BL21 Star (DE3) expression cells. This result was surprising given that C43 is known to be highly suitable for membrane protein expression^{402,403}, and BL21 Star has enhanced RNA stability that ought to improve expression yields¹³⁵. However, our results resemble other studies that found that membrane proteins' expression is highly strain-dependent³⁸⁰.

The highest yields of LicB were obtained from auto-induction media rather than in IPTG induction media. This result confirms previous findings from Studier *et al.*³⁸² and from Deacon *et al.*⁴⁰⁴, in which Super Broth auto-induction was highly compatible with membrane proteins. The critical advantage of AIM over IPTG induction is that glucose allows higher densities during the first stage of the cell culture, thus avoiding lactose from inducing the target protein. Consequently, more protein per volume is produced.

IMAC purification was largely successful, although LicB exhibited a tendency to aggregate, as seen in the size exclusion chromatograms. Scaling-up of the optimum growth conditions did not result in more protein, leading to the conclusion that aggregation is concentration-dependent. This is consistent with Deacon *et al.*'s findings, in which the yield of target protein was dramatically reduced when higher cell concentrations were obtained⁴⁰⁴. This might be because the increased biomass provides a very high non-specific background that interferes with IMAC column binding.

A screen of different expression conditions varying temperature and time was developed for LicB expressed in T7 Express cells and in SB_{AIM}. This showed that expression of LicB for 24 hour (from the inoculation time) at 37°C achieved the most significant protein yield (Figure 3.13). Likewise, Deacon *et al.* found the highest expression in SB_{AIM} after 24-hour expression at 37°C⁴⁰⁴.

Three solubilising detergents and a SMA polymer were tested for their compatibility with LicB. SMA was unsuccessful in purifying LicB, possibly due to inefficient binding to the Ni²⁺ column. This may be because the tag is occluded by the SMA belt around the protein or could simply require longer incubation times, as described by Lee *et al.*¹⁶³. LicB was successfully solubilised from membrane fractions by DM, DDM and Cymal-5. However, only DM and DDM performed well in size exclusion chromatography, revealing a monodisperse peak at the elution volume of around 13 mL. Some protein aggregates at higher molecular weight were also observed. Nonetheless, purification in DDM was deemed to be successful and suitable for further work.

3.3.2. Bioinformatic and biophysical characterisation

LicB was predicted to have ten transmembrane α -helices. This hypothesis was accurate since LicB was found to have a typical α -helical pattern in CD measurements (Figure 3.14) with the expected degree of helicity in DDM (but not DM). LicB in DDM had a melting temperature of 66°C and was stable at room temperature over long incubations. This suggests that purified LicB is relatively thermostable. SEC-MALS and Blue Native-PAGE both report that a population of purified LicB is monomeric in DDM and that these monomers persist when the protein is reapplied to SEC. Intriguingly, purified LicB gave a slightly lower than expected mass when analysed by SDS-PAGE and SEC-MALS, with both methods reporting ~26 kDa against a theoretical weight of 34 kDa. However, MS appeared to confirm that the purified protein was indeed LicB, and this is confirmed in later chapters by Western blotting against the C-terminal His-tag. Molecular modelling did generate models with the expected number of transmembrane helices but was otherwise less informative because of the weak homology between LicB and the selected template protein.

3.4. Conclusion

This chapter reports the first extensive bioinformatic analyses of the LicB sequence and documents the attempts to recombinantly express and purify LicB. This expression was successfully accomplished, albeit with the issues of low yield and aggregation propensity that are known to be common problems with membrane proteins. The monodispersity and secondary structure content of purified LicB in the detergent DDM provide a suitable platform for further biophysical characterisation, which is pursued in the following chapters.

Chapter 4. Functional characterisation of LicB

4.1. Introduction and aim

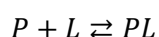
LicB is a putative choline transporter that belongs to the uncharacterised LicB-T family (TCDB 2.A.7.18). Cell culture studies have suggested that LicB is a high-affinity transporter, with K_M for choline of approximately $3.6 \mu\text{M}$ ³⁶¹, but this has not yet been confirmed *in vitro*. Moreover, it is thought that LicB is a proton-driven transporter based on evolutionary relationships, but there is currently no experimental evidence to support this. Therefore, this chapter uses purified recombinant LicB to determine the affinity and selectivity of ligand binding and transport.

This was pursued through two distinctive approaches. Firstly, ligand binding was assessed through Fluorescence Resonance Energy Transfer (FRET) and Isothermal Titration Calorimetry (ITC). The FRET assays made use of a fluorescent choline derivative that was previously used to probe substrate binding by the acetylcholine receptor⁴⁰⁵. Once the binding of this substrate to LicB was established, selectivity was determined through competition assays with non-labelled potential substrates. In the second approach, choline transport was dissected in transport assays using reconstituted liposome systems *in vitro*.

A brief discussion of each technique is presented below.

4.1.1. Protein-ligand interactions

Protein-ligand interactions are the basis of a wide range of biological processes. In the simplest case, the protein-ligand interaction can be described by:



Where, each protein (P) binds to a ligand (L), to form a reversible complex (PL). Thus, the overall reaction depends on the concentration of both the free protein [P] and the free ligand [L].

By definition, at equilibrium, the rate of association is equal to the rate of dissociation. The rate that describes the collision of two molecules is κ_{on} (in $M^{-1} s^{-1}$) and depends upon the size of the molecules involved as well as their microscopic interactions. In ideal cases κ_{on} is diffusion-limited, imposing an upper limit on this value of approximately $1 \times 10^9 M^{-1} s^{-1}$. On the other hand, κ_{off} describes the dissociation constant and only represents the probability of a PL complex dissociating. This lack of concentration-dependence means that κ_{off} has units of s^{-1} , and it is inversely proportional to the strength of the binding interaction. The equivalence of the on-rate and off-rate at equilibrium can be expressed as:

$$\kappa_{on}[L][P] = \kappa_{off}[PL]$$

Equation 4.1

Rearranging Equation 4.1 gives the following expression for the equilibrium binding constant, K_a :

$$K_a = \frac{\kappa_{on}}{\kappa_{off}} = \frac{[PL]}{[P][L]}$$

Equation 4.2

Where K_a is represented as M^{-1} . However, the most commonly used metric of binding affinity is the dissociation constant, K_d (M), which can be defined as:

$$K_d = \frac{\kappa_{off}}{\kappa_{on}} = \frac{[P][L]}{[PL]}$$

Equation 4.3

The equilibrium dissociation constant is typically in the range of μM to nM , with the latter corresponding to higher affinity, in other words, less probability of dissociation.

Furthermore, the thermodynamics of the complex formed can be extrapolated using experimentally defined equilibrium and rate constants from protein-ligand interactions. Under normal pressure and temperature conditions, the binding free energy change is given by:

$$\Delta G_{bind} = G_{solution}^{PL} - (G_{solution}^P + G_{solution}^L)$$

Equation 4.4

And the relation between ΔG_{bind} and K_d is usually given as:

$$\Delta G_{bind} = RT \ln (K_d)$$

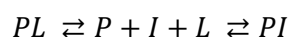
Equation 4.5

Where R is the universal gas constant and T is the experimental temperature. Finally, the difference in free energy can be disintegrated into the entropy change, ΔS_{bind} , and the enthalpy change, ΔH_{bind} , of the reaction:

$$\Delta G_{bind} = \Delta H_{bind} - T \Delta S_{bind}$$

Equation 4.6

When considering a competition assay the inhibitor (or competitor), I , directly competes with the natural ligand for the binding site. This can be expressed as:



At equilibrium, the dissociation constant for the inhibitor, K_i , can be described as:

$$\frac{[PL]}{[PI]} = \frac{K_i [L]}{K_d [I]}$$

Equation 4.7

Ligand-binding interactions can be analysed through a wide range of techniques, and each technique has each own advantages⁴⁰⁶. Two common approaches, which are used here, are Fluorescence Resonance Energy Transfer (FRET) and Isothermal Titration Calorimetry (ITC). These are briefly introduced below.

4.1.2. Fluorescence Resonance Energy Transfer

Fluorescence Resonance Energy Transfer (FRET) is a non-radiative phenomenon which occurs if the excitation peak of the acceptor overlaps with the emission peak of the donor. In the overlapping region, the energy from the molecule with the lower wavelength emission (the donor molecule) is donated to a ground-state acceptor fluorophore – the molecule with the higher excitation wavelength (the acceptor molecule)⁴⁰⁷. This phenomenon is dependent on the Förster distance which in turn is affected by a variety of factors including label orientation

and the donor quantum yield. In order to be viable, FRET pairing molecules are normally in the region of twenty to sixty Angstroms apart. Förster distance is related to FRET efficiency as follows:

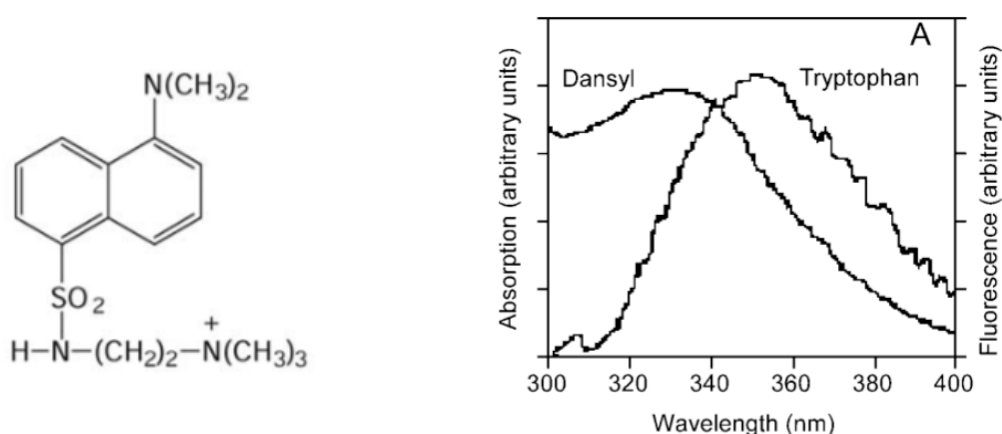
$$E = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6}$$

Equation 4.8

Where R_0 is the distance where the FRET efficiency is 50% and r is the distance between the donor and the acceptor ⁴⁰⁸.

A novel fluorescence binding assay was developed in this project to probe the interaction between LicB and the supposed substrate, choline. This exploits a fluorescent choline derivative, dansyl choline (DC) (Figure 4.1A). The binding of dansyl choline to LicB is anticipated to result in FRET between the intrinsic tryptophan residues of LicB (FRET donor) and this fluorescent substrate (FRET acceptor). This is posited since LicB contains six Trp residues, and in the osmoregulatory choline transporters, these residues are intimately involved in substrate binding ³³⁴. If this FRET phenomenon occurs upon DC binding, it will produce a measurable fluorescence emission signal from the DC acceptor at around 570 nm ⁴⁰⁵. This signal is abolished if the binding of dansyl choline is outcompeted or displaced by other substrate analogues (Figure 4.1B).

(A)



(B)

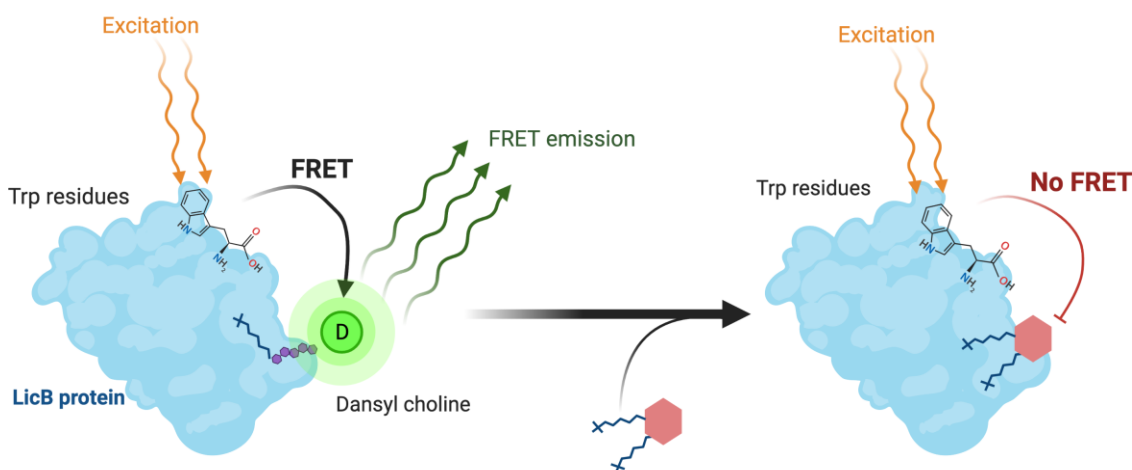


Figure 4.1. Schematic representation of the binding of the fluorescent derivative dansyl choline to LicB.

(A) Chemical structure of dansyl choline. Adapted from Jin (2003)⁴⁰⁵. Graph shows the spectral overlap between the Trp emission spectrum λ_{em} 355 nm (λ_{ex} 280 nm) and the excitation spectrum of dansyl at λ_{ex} 335 nm (λ_{em} 535 nm). Adapted from Gustiananda *et al.* (2004)⁴⁰⁹. **(B)** FRET between intrinsic Trp residues (FRET donor) and the dansyl group (FRET acceptor). FRET is lost when the fluorescent ligand is displaced by another compound (in red).

4.1.3. Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) is a label-free technique that measures the heat produced during a molecular interaction from a single titration⁴¹⁰. It is sometimes called the 'gold standard' of studying ligand binding interactions since ITC produces a complete thermodynamic description, K_d , and binding stoichiometry in a single experiment. The ITC instrument constitutes two identical cells surrounded by an adiabatic jacket (Figure 4.2).

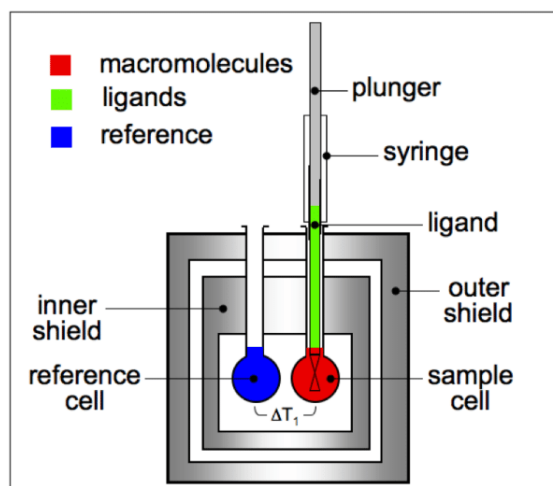


Figure 4.2. Main parts of the ITC instrument.

Adapted from Courtois and Berret (2010) ⁴¹¹.

During a typical experiment, the injection needle is loaded with titrant (ligand solution), which is injected sequentially into the sample cell containing the target protein. Heat is released (exothermic) or absorbed (endothermic) as a result of ligand interaction. Thus, the instrument measures the time-dependent input of power needed to keep the sample and the reference cells at the same temperature ⁴¹². After subtracting any background signal (from dilution of the samples or buffer mixing, for example), the change in enthalpy (or ΔH_{obs}) during the titration is proportional to the fraction of bound ligand. As the protein becomes saturated, the heat signal weakens until only the heat of dilution is detected ⁴¹³. ITC can typically measure an affinity with K_d between 100 μM to one nanomolar ^{414,415}. ITC can determine the enthalpy of the reaction (ΔH), the heat capacity (ΔC), the entropy (ΔS) and the stoichiometry (n) of the protein-ligand interaction. Collectively, these parameters determine the Gibbs free energy (ΔG) and the dissociation constant (K_d), as described from Equation 4.4 to Equation 4.6 above. Moreover, the shape of ITC curves is unique in that it is primarily determined by the c value, defined as ⁴¹⁴:

$$c \text{ value} = \frac{P \times N}{K_d}$$

Equation 4.9

Where, P is the protein concentration used during the experiment and N is the number of binding sites.

When using ITC to study membrane proteins, some limitations are considered, namely the requirement of a relatively large amount of protein volume and the need to keep the protein in a stable and folding state ^{414,416}. Furthermore, the choice of the lipid mimics (such as detergents) can significantly impact the ITC experiment, for instance, the production of unusual peaks shapes or large amplitudes ⁴¹⁴. Despite these challenges, examples of successful use of ITC to study membrane proteins have been reported, e.g., the lactose permease of *E. coli*, LacY ⁴¹⁷.

4.1.4. Transport energetics

In vivo studies suggest that LicB naturally accumulates choline against a concentration gradient ³⁶¹ – that is, intracellular choline concentration is higher than the extracellular concentration. Since it is energetically unfavourable to pump choline against its gradient, another driving force is required. The driving force for secondary active transport by LicB is thus expected to come from electrochemical gradients across the bacterial membrane. Therefore, in order to study LicB transport activity in proteoliposomes, similar gradients should be generated *in vitro*.

LicB is expected to be energised by the electrochemical proton gradient across the bacterial cell membrane – also known as the proton motive force (PMF) and designated $\Delta\mu^{H^+}$. This proton gradient has two components. The first of these is a charge gradient or electrical potential ($\Delta\Psi$), reflecting the uneven distribution of electrical charges across the membrane. To replicate this *in vitro* the gradient across proteoliposome membranes should be inward directed (interior negative) to mirror the physiological situation. The magnitude of $\Delta\Psi$ is calculated using the Nernst equation (Equation 4.10). The second component of the proton gradient is the chemical potential (ΔpH or ΔH^+), which is the different concentration of protons on different sides of the membrane. This can be established *in vitro* by manipulating buffer pH, and the proteoliposome interior should be slightly alkaline (lower proton concentration) relative to the more acidic exterior (higher proton concentration). Generally, a proton-driven transporter in a living cell always experiences the combination of these two components in the proton motive force; however, the transport may not require both gradients ($\Delta\Psi$ and ΔpH), and one alone can provide the driving force for transport ⁴¹⁸.

For any given ion, its movement depends on the transmembrane electrical potential and is given by the Nernst equation ⁵:

$$\Delta\Psi = \left(\frac{RT}{zF}\right) \times \ln \frac{[ion]_{out}}{[ion]_{in}}$$

Equation 4.10

Where T is the temperature, R is the universal gas constant, F is the Faraday constant, z is the formal charge of the molecule, $[ion]_{out}$ is the concentration of the ion outside the vesicle and $[ion]_{in}$ is the concentration of ion inside the vesicle. Calculated values for each ion will be shown in the results section.

Additionally, the electrochemical potential, which is the sum of the chemical component (ΔG_{chem} , Equation 4.11) and the electrical component (ΔG_{elec} , Equation 4.12), determines the energetic cost of transporting an ion⁵, as follows:

$$\Delta G_{chem} = RT \times \ln \frac{[ion]_{out}}{[ion]_{in}}$$

Equation 4.11

$$\Delta G_{elec} = zF \times \Delta\Psi^{ion}$$

Equation 4.12

Where R, T, z, F variables have been previously described and $\Delta\Psi^{ion}$ is the electrical potential of a given ion.

4.2. Results

4.2.1. LicB ligand binding assays

4.2.1.1. Exploring ligand binding through FRET

The binding of a fluorescent choline derivative was first evaluated through FRET assays. Titrating dansyl choline (DC) into micelle solubilised LicB gave rise to a measurable increase in the acceptor (dansyl) emission at 570 nm and a loss of the donor (Trp) signal at 330 nm, consistent with FRET (Figure 4.3). The increase in acceptor emission was only seen in experiments with DC and not in controls with dansyl alone (Figure 4.3B and D). Unexpectedly, control experiments with dansyl also showed a loss in the donor Trp signal at 330 nm (Figure

4.3C). This suggests that free dansyl (and so, presumably, DC) can make non-specific interactions with LicB and that this can quench the Trp fluorescence signal; however, this does not appear to result in a corresponding increase in the dansyl emission. It thus appears that the apparent binding signal is a composite of specific and non-specific interactions but that only specific interactions lead to a FRET enhancement in the acceptor emission ⁴⁰⁵.

An additional set of controls replicated the titration conditions but in the absence of any protein (Figure 4.3A and B). This gave the background signal arising solely from the acceptor under the conditions tested. The emission signal of DC was shifted by 60 nm versus dansyl. Surprisingly, the emission response in these controls was not linear. Instead, it increased rapidly at low dye concentrations and then plateaued. This background signal cannot easily be deconvoluted from the binding signal and so confounds any direct determination of equilibrium dissociation constants, and this is explored in greater detail below.

Figure 4.4 summarises the binding data and shows combined plots of signal change ligand (acceptor) concentration. Figure 4.4A directly compares the scans obtained at 500 μM fluorescent ligand in all cases, including an additional control of LicB only (no acceptor dye). The differences between the complete experiment and controls are noticeable, with a substantial enhancement of acceptor fluorescence when DC is incubated with LicB. This is also evident from the titration curves shown in Figure 4.4B. However, the confounding problem of the ligand background (DC only) is now clear, following the same trend as the full treatment data (LicB + DC). To illustrate this further, both sets of data were fitted to a hyperbolic ligand binding equation. In both cases, the equilibrium dissociation constant, K_d , was approximately 67 μM (Table 4.1).

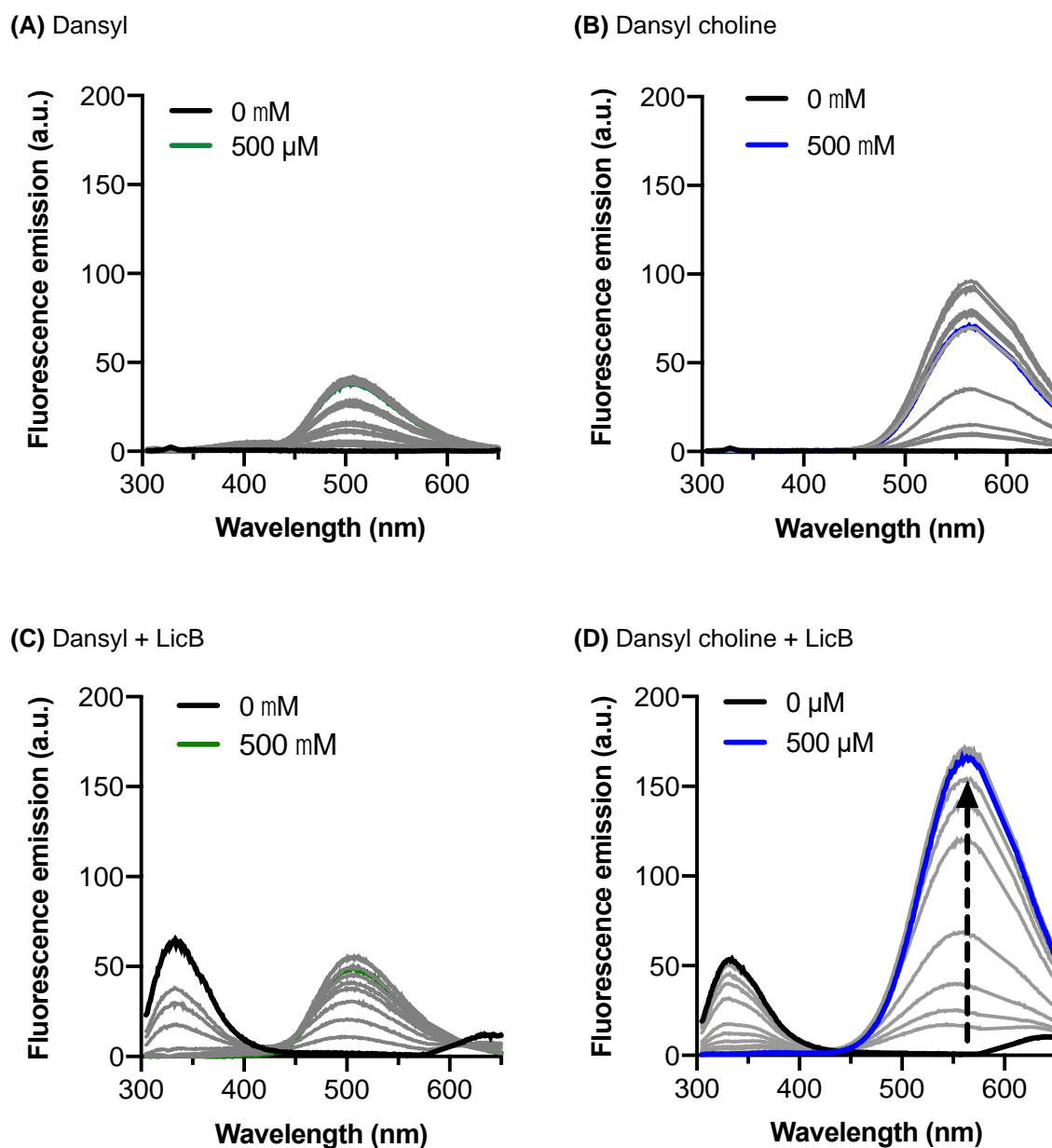


Figure 4.3. Fluorescence emission plotted against wavelength.

Emission spectra of assay components individually - **(A)** dansyl and **(B)** dansyl choline; or in combination - **(C)** dansyl incubated with LicB and **(D)** dansyl choline incubated with LicB, after Trp (donor) excitation at 295 nm.

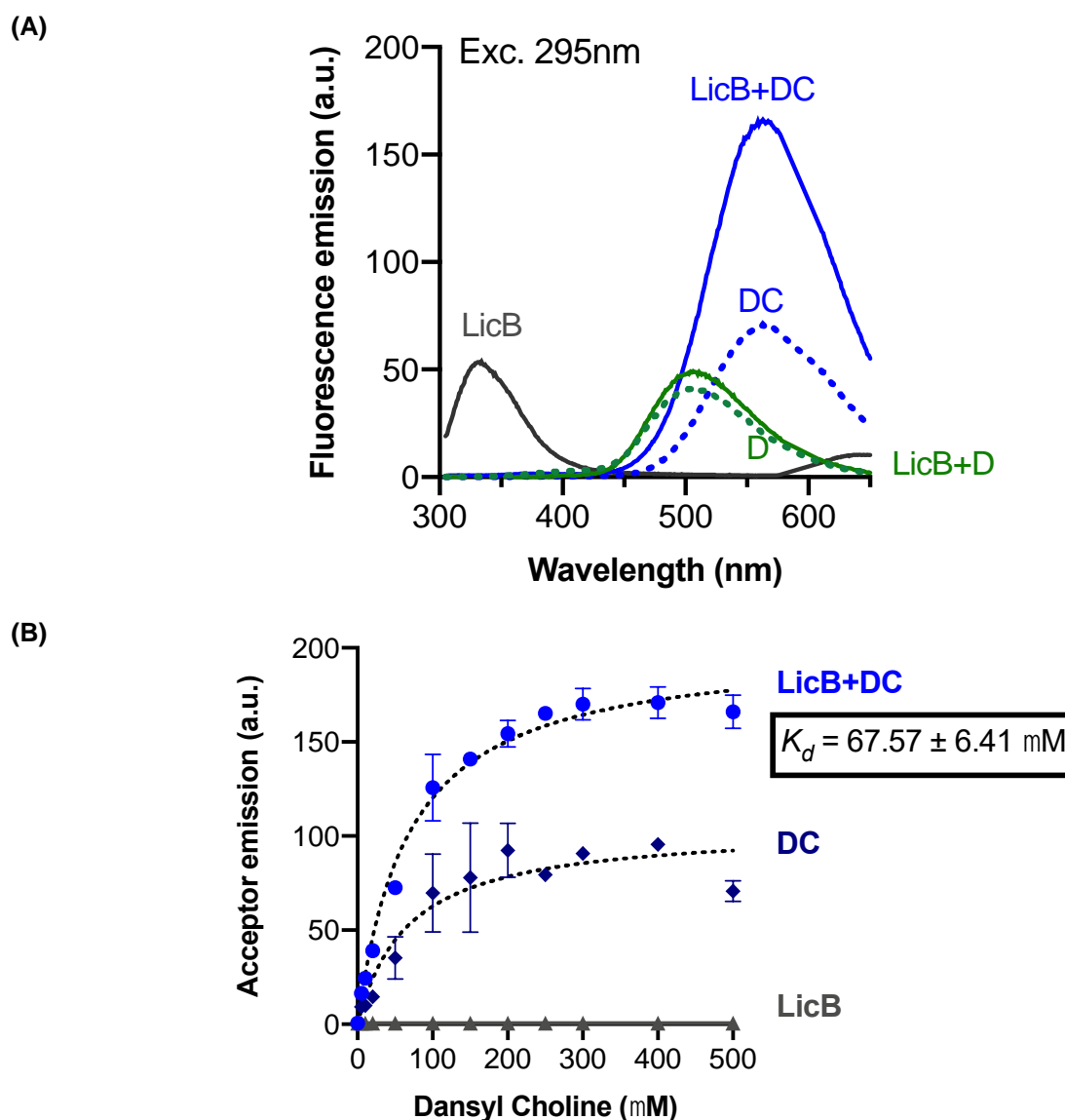


Figure 4.4. Experimental fluorimetry demonstrating the FRET phenomenon.

(A) Summary data from binding experiments. When LicB is incubated with dansyl choline (LicB+DC), an increase in dansyl emission at 570 nm is observed (solid blue lines) that is absent in various controls. Controls with dansyl (D) plus LicB were also performed (green lines). (B) Equilibrium binding of DC to LicB can be distinguished above the background but follows similar trends to control data. Data were collected from triplicate experiments, and the error bars correspond to standard deviation. Data were fitted to a hyperbolic binding equation using GraphPad Prism (black lines).

Table 4.1. Experimental determined binding constants for LicB.

† Errors represent \pm standard deviation from curve fitting.

Experiment	K_d (μM)	Model fits data? / r^2
LicB + DC	$67.57 \pm 6.41^\dagger$	Yes / 0.987
DC	67.07 ± 21.16	Yes / 0.884

4.2.1.2. Competition binding assay and the IC₅₀

The results described above demonstrate the potential for this fluorescence assay to detect ligand binding by LicB, albeit with a complex background signal that prevents simple analysis. In this light, a ligand competition was explored in which non-labelled substrates were used to compete off the fluorescent substrate. This competition ought to only affect the specific component of the assay signal, with the non-specific component unaffected. An advantage of the approach was the ability to screen a panel of related compounds to determine the substrate selectivity of LicB.

For these experiments, LicB protein was saturated with the ligand dansyl choline, with a concentration at 1.5 times the apparent K_d of 67.6 μM . IC₅₀ values were determined by titrating these saturated complexes with a variety of unlabelled ligands. IC₅₀ describes the inhibitor concentration at which the maximal effect produced by reference ligand binding is reduced by 50% (Equation 2.11). Compounds screened were acetylcholine, benzoylcholine, glycine, L-cysteine, L-carnitine and betaine, which chemical structures are represented in Figure 4.6.

We began by validating the assay against the expected native ligand for the transport - choline. The results are shown in Figure 4.5, with excitation at 295 nm in all cases. Introducing choline was found to successfully displace a proportion of the bound label. Titration curves showed the expected sigmoidal response associated with competition experiments, with an IC₅₀ of 444.9 μM (Figure 4.5A). The data were expressed as the ratio between the normalised emission of the acceptor (dansyl choline) and the donor (Trp) to specifically probe changes in FRET. Control experiments showed that adding aliquots of buffer in the same proportion as the competing ligand (choline) did not display DC from the binding site (Figure 4.5B). Controls with dansyl (rather than dansyl choline) showed entirely different behaviour, with a different trend in the FRET ratio and no difference between introducing the ligand and buffer-only controls (Figure 4.5C and D). The addition of choline or equivalent buffer aliquots caused a dramatic increase in dansyl fluorescence, rather than the subtler decrease seen with dansyl-choline. The origins of this behaviour are unknown, but it may result from the slow partitioning of dansyl into detergent micelles over time; dansyl fluorescence is known to increase dramatically in hydrophobic environments. Whatever the cause, there is clearly a marked difference between labelled choline and the dansyl label alone in these experiments. This competition assay thus appears to report on the displacement of specifically bound dansyl choline by the native ligand.

The same assay was then used to assess the ability of a variety of different potential ligands for LicB to produce the same effect (apparent displacement of dansyl choline). The results are shown in Figure 4.6 and Table 4.2. Only choline and acetylcholine were able to displace DC, resulting in loss of FRET signal. As expected, choline bound most tightly, with IC_{50} of 444.9 μ M being nearly three times lower than that of acetylcholine at ~1.3 mM. Benzoylcholine possibly showed a partial loss of signal that could not be well-fit to the competition equation. Other compounds, including betaine and L-carnitine, could not displace DC, and thus no loss of FRET was observed. Hence this experiment concludes that LicB is a relatively specific transporter for choline and, to some degree, acetylcholine. Of particular interest is the discrimination against betaine and carnitine, implying that either a C2 alkyl chain must be adjacent to the trimethylammonium group or that carboxylic acid groups at the molecule 'tail' prevent binding to LicB.

Additionally, K_i values were determined from the Cheng Prussof equation (Equation 2.13). As expected, the K_i is roughly three times higher for acetylcholine, meaning that choline has a higher affinity towards LicB than acetylcholine, and thus, smaller amounts of choline are required to displace DC (Table 4.2).

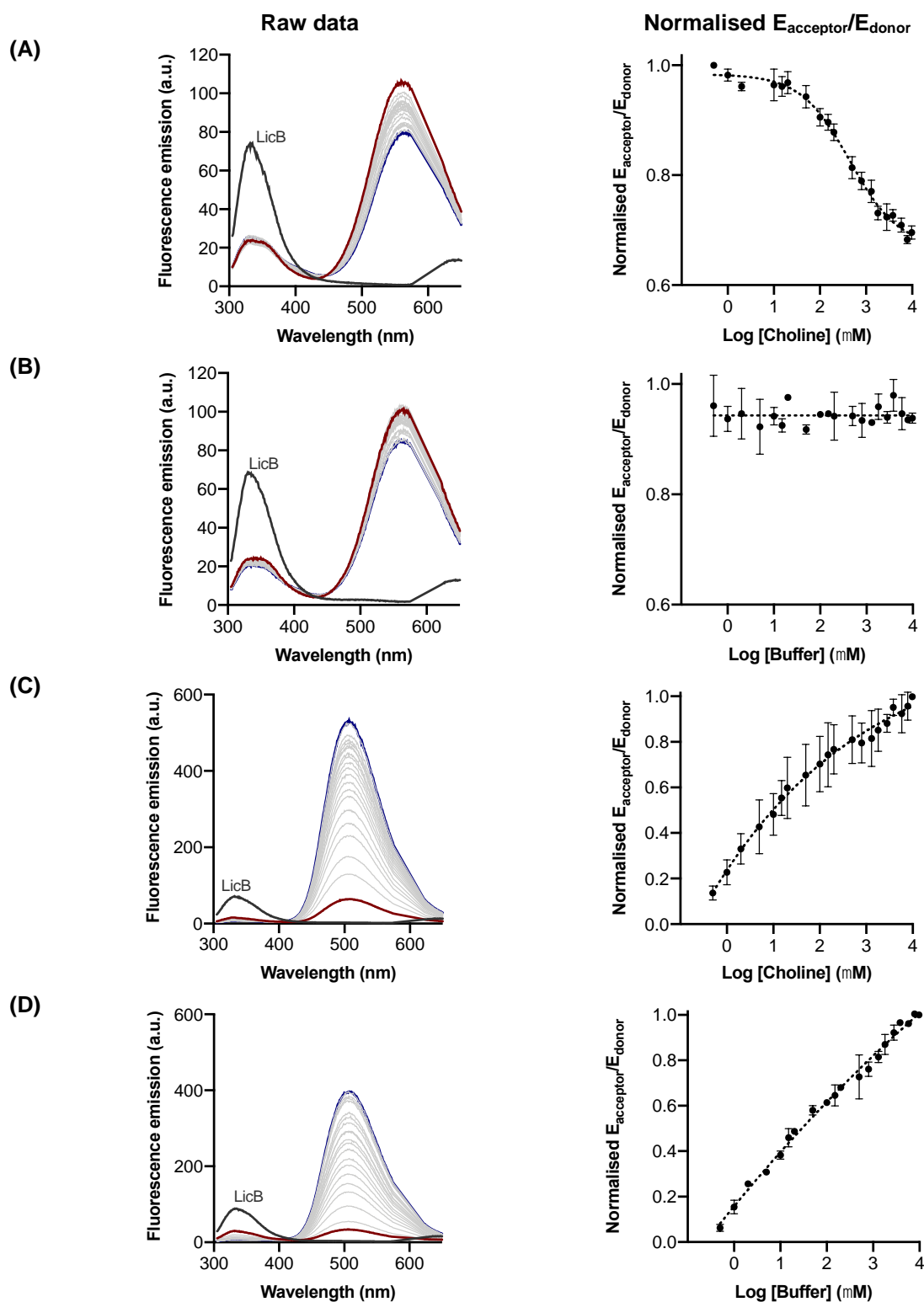


Figure 4.5. Control experiments using increasing concentrations of either choline or the corresponding volume of buffer.

(A) LicB incubated with DC and choline. **(B)** LicB incubated with DC and the corresponding volume of buffer. **(C)** LicB incubated with dansyl and choline. **(D)** LicB incubated with dansyl and the corresponding volume of buffer. The red line represents no choline, while the blue line represents maximum concentration. The concentration of DC and dansyl used was 1.5 times K_d . Panels (B) and (D) represent the addition of equivalent aliquots of the buffer. Data were collected from triplicate experiments, and the error bars show mean \pm standard deviation. Excitation was at 295 nm.

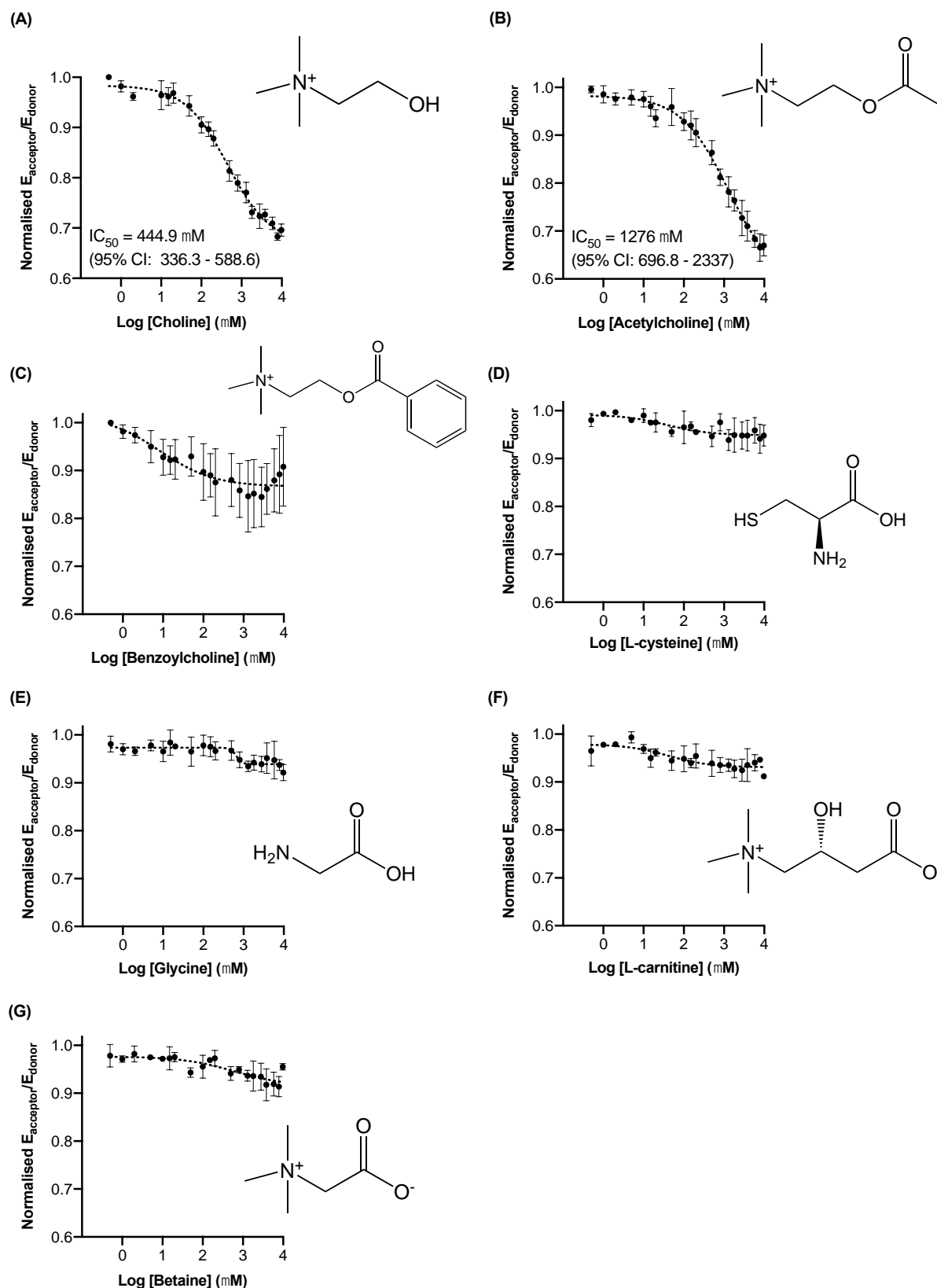


Figure 4.6. Competition experiments with different small molecules.

(A) choline, (B) acetylcholine, (C) benzoylcholine, (D) L-cysteine, (E) glycine, (F) L-carnitine and (G) betaine. The concentration of DC used was 1.5 times K_d . Data were collected from triplicate experiments, and error bars show mean \pm standard deviation. Data were fitted to a four-component inhibition equation using GraphPad Prism (black lines).

Table 4.2. Results of competition assays.

† Errors represent the standard deviation from curve fitting. IC₅₀ (μM) and Hill slope were estimated from the four-component inhibition equation using GraphPad Prism and K_i values were calculated from Equation 2.13.

Compound	Loss of FRET?	Model fits data? / r ²	IC ₅₀ (μM)	Hill slope	K _i (μM)
Choline	Yes	Yes / 0.979	444.9 (95% CI: 336.3 – 588.6) [†]	-0.82	179.4
Acetylcholine	Yes	Yes / 0.966	1276 (95% CI: 696.8 – 2337)	-0.74	514.7
Benzoylcholine	?	Unclear / 0.437	ND	ND	ND
Glycine	No	No / 0.444	ND	ND	ND
L-cysteine	No	No / 0.448	ND	ND	ND
L-carnitine	No	No / 0.473	ND	ND	ND
Betaine	No	No / 0.516	ND	ND	ND

4.2.2. Studying ligand binding through ITC

Isothermal Titration Calorimetry (ITC) was then used to validate the binding of choline to LicB identified by FRET assays.

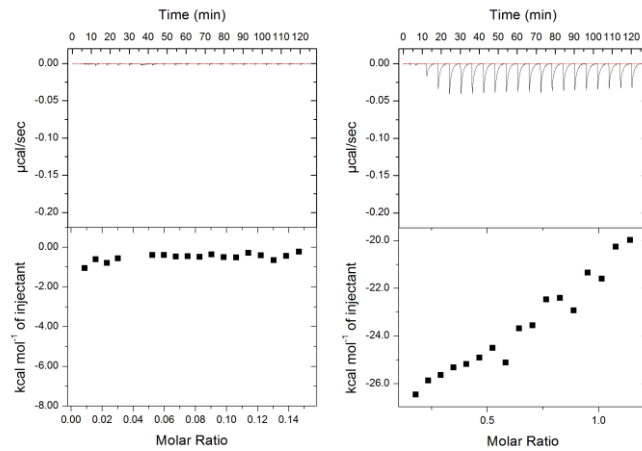
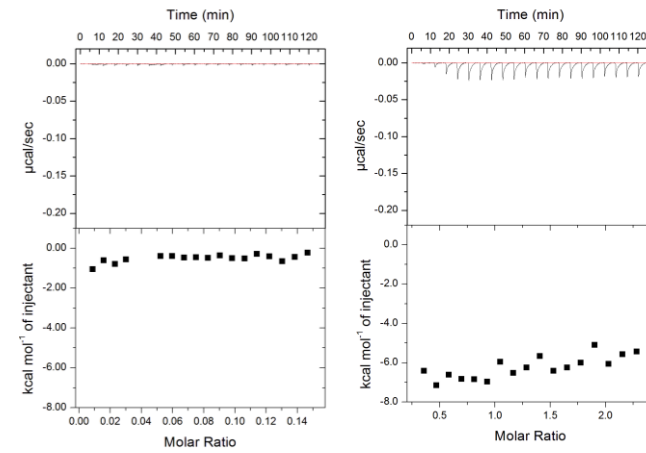
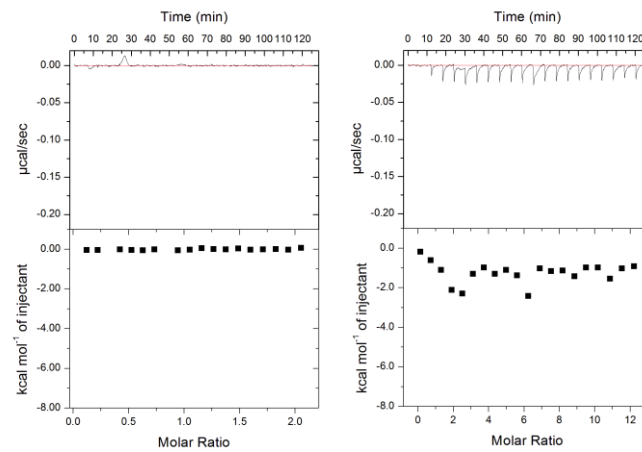
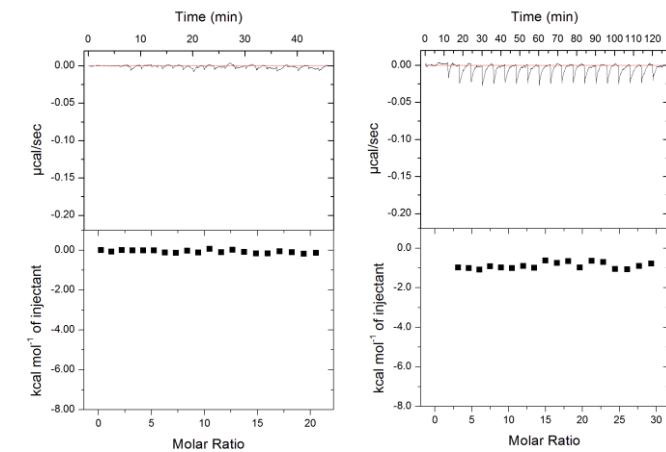
In ITC, the choice of an appropriate ligand and protein concentration is dependent upon the so-called c value, which is the product of the molar protein concentration in the reagent cell and the binding constant. Obtaining c values in the range 1-1000 is generally accepted as providing a suitable regime for ITC data ⁴¹⁹; For $c < 1$, the slope of the ligand binding curve is too shallow for accurate curve-fitting and for $c > 1000$, the binding curve is too steep (effectively, a step-change), and again is not accurately defined by the data. Experimental design for ITC generally involves balancing this ideal range for the c value with having sufficient protein (titrand) to generate a detectable signal and sufficient ligand (titrant) to be in excess at the end of the titration. Since no previous *in vitro* binding assays were reported for LicB, the protein concentration used was between 8 - 9 μM , within the normal range for proteins in ITC of 1-50 μM , and a series of independent experiments were performed with different initial choline concentrations from 50 μM to 1 mM. Control experiments were also performed to determine background noise from buffer mixing or titrant dilution.

ITC experiments with LicB were unsuccessful, showed that choline did not bind at 25°C when titrated into protein samples. Figures of all ITC experiments performed can be seen in Figure 4.7. A summary of all conditions tested is shown in the following table. ITC was not pursued further.

Table 4.3. Summary of choline concentrations tested and the experiment outcome.

c values were calculated based on Equation 4.9 and assuming a 1:1 binding stoichiometry.

[Choline] (μM)	[LicB] (μM)	Does choline bind?	c value for K_d of ~67.6 μM	c value for K_M of 3.6 μM ³⁶¹
50	8.9	No	0.13	2.47
100	9.0	No	0.13	2.50
500	8.4	No	0.12	2.33
1000	8.9	No	0.13	2.47

(A) 50 μM **(B)** 100 μM **(C)** 500 μM **(D)** 1000 μM **Figure 4.7. Isothermal titration calorimetry of LicB**

Titration of different choline concentrations **(A)** 50 μM , **(B)** 100 μM **(C)** 500 μM and **(D)** 1000 μM either into buffer (negative control - left panel) or recombinant LicB (right panel).

4.2.3. Proteoliposome transport assay

Several methods are available for studying the activity of proton-driven transporters, namely the influx of a radiolabeled substrate ⁴²⁰, the influx or binding of a fluorescent derivative (e.g., dansylgalactosides ⁴²¹) or detecting the influx of the co-transported protons with fluorescence methods. LicB is thought to be a proton symporter ³⁶¹, and so an inward-directed proton gradient was applied to drive the uptake of choline from the bulk exterior to the proteoliposome interior. In this specific case, a pH-responsive dye, pyranine, was used to follow the presumed acidification of the vesicle that would accompany choline uptake. Such a system has been used extensively in the past, and the pyranine transport assay used here is identical to that described in Uzdavinys *et al.* ³⁷⁸, Parker *et al.* ³⁷⁹, Lee *et al.* ⁴²² and Minhas and Newstead ⁴²³. Pyranine is a pH-responsive fluorescent dye that is water-soluble and can be trapped within proteoliposomes. The pH response is characterised by a shift in the excitation band from 460 nm above neutral pH to 406 nm at acidic pH. It is generally convenient to measure the ratio of the fluorescence obtained after simultaneous excitation at these wavelengths (F406/F460) to minimise artefacts from, for example, light scattering.

Membrane electrochemical gradients are induced by the addition of a specific potassium channel, valinomycin, to the proteoliposome. Valinomycin is localised in the proteoliposomes membrane and allows the rapid diffusion of potassium across the membrane. This mechanism has been widely reported in the literature ^{379,422}.

4.2.3.1. Reconstitution in synthetic proteoliposomes

Synthetic proteoliposomes provide a well-defined experimental system for studying substrate transport *in vitro*. Liposomes were constituted of POPE and POPG in a ratio of 3:1 (w/w). After extrusion, the lipid membrane was disrupted by the addition of DDM aliquots. The absorbance at 540 nm was measured after the stepwise addition of DDM to monitor the physical state of the liposomes. According to Figure 4.8, DDM seemed to cause partial dissolution of the liposomes. Protein was then added at either 60:1 or 360:1 (w/w) lipid to protein ratio, and excess of detergent removed.

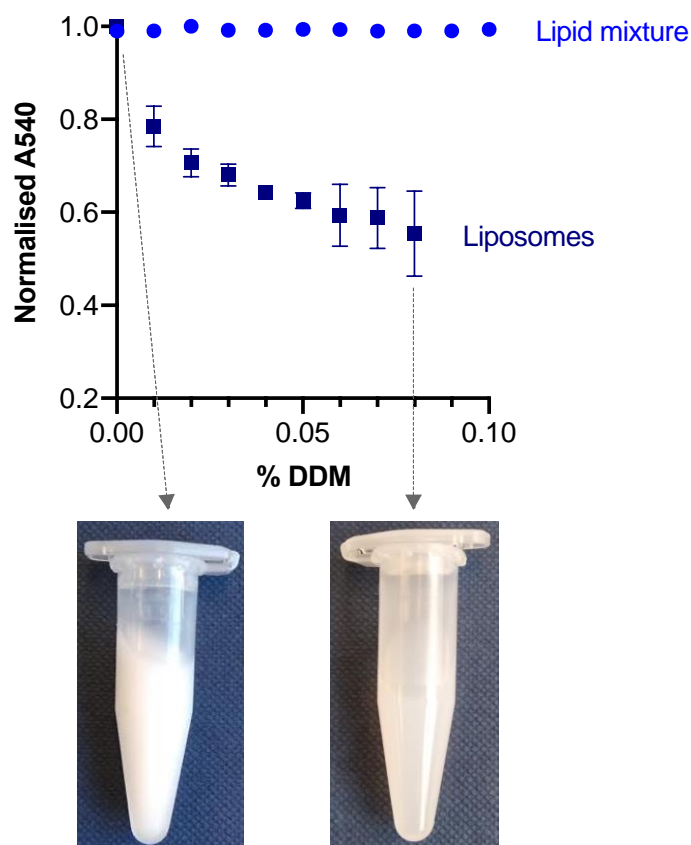


Figure 4.8. Normalised absorbance at 540 nm (A540) of either the lipid mixture before extruding or after extrusion (liposomes) with increasing percentage of DDM.

Images show extruded liposomes before and after detergent treatment.

In order to evaluate the success of the reconstitution procedure, a sucrose ‘flotation’ assay (50%, 40% and 0% sucrose) was performed. This separates intact proteoliposomes from unreconstituted (sedimenting) protein. Gradient samples were directly applied to a nitrocellulose membrane, and the presence of protein detected by immunoblotting against the His-tag. This is often referred to as a ‘Dot Blot’ and can be used to detect the location of the protein on the discontinuous gradient. Figure 4.9 confirms the reconstitution of LicB into proteoliposomes using a 60:1 (w/w) lipid to protein ratio. The protein ‘floats’ with the liposome fraction to the top of the gradient. A control experiment with purified LicB in micelles confirmed that without reconstitution, the protein remains at the bottom of the gradient.

Additionally, the presence of pyranine inside liposomes and proteoliposomes was confirmed by observing the characteristic excitation peaks at 406 and 460 nm (Figure 4.10).

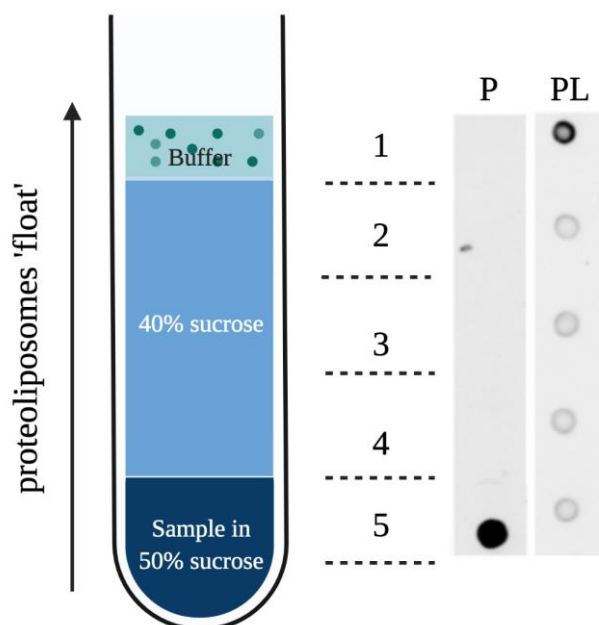


Figure 4.9. Reconstitution of LicB into proteoliposomes.

Immunodetection of the His-tag was used to compare between the position of LicB protein (lane P) and the LicB protein reconstituted in proteoliposomes at 60:1 (w/w) lipid to protein ratio (lane PL) on a 'flotation' assay after ultracentrifugation.

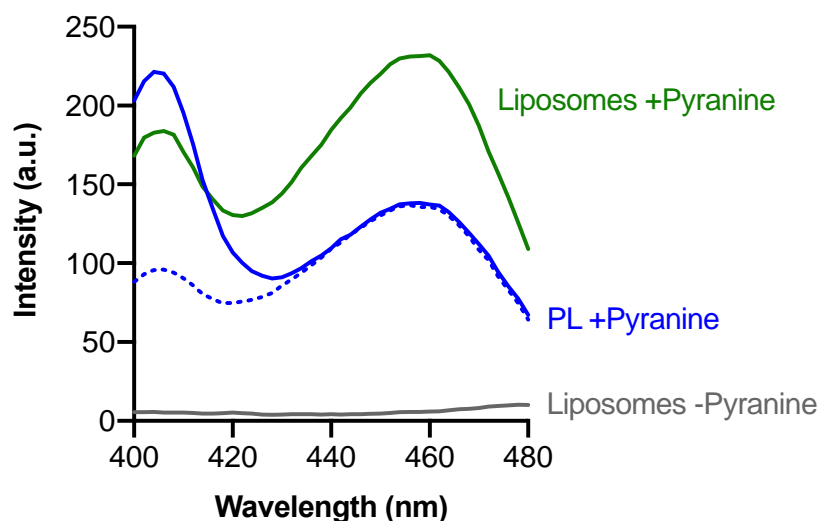


Figure 4.10. Confirmation of pyranine inside liposomes and proteoliposomes.

Excitation spectra of pyranine in proteoliposome (PL) and liposome (L) recovered from the top of the gradient. The green line represents liposomes reconstituted with pyranine, while the grey line represents a no-pyranine control. Blue dashed line represents proteoliposome at 60:1 (w/w) lipid to protein ratio and solid blue line at 360:1 (w/w) lipid to protein ratio.

4.2.3.2. Pyranine transport assay

The chemical proton gradient (ΔpH) was established by reconstituting proteoliposomes at pH 8.0 and diluting them into a buffer at pH 6.5. A charge gradient ($\Delta\Psi$) was established by using a balance of sodium and potassium salts in the following way. Proteoliposomes were formed in high concentrations of potassium salts (135 mM K^+ inside, 0 mM K^+ outside). These potassium-containing proteoliposomes were then diluted into sodium-containing buffers of equivalent osmolarity (15 mM Na^+ inside, 150 mM Na^+ outside). Upon dilution, potassium ions immediately left the vesicle via the valinomycin ionophore and entered the essentially infinite volume of the surrounding buffer. This resulted in an outward-directed Cl^- gradient, where the proteoliposome was full of 'free' Cl^- without a corresponding counterion. This resulted in an uneven charge distribution, where the inside of the proteoliposome was negative. This generated $\Delta\Psi$ (negative inside) with the magnitude of the driving force basically linear with potassium concentrations in the external buffer. Hence, this ought to induce the inward movement of a positively charged proton. All gradients are described in Table 4.4 and Figure 4.11.

Table 4.4. Gradients associated with the outside-to-inside direction for the movement of a proton.

Values were calculated based on Equation 4.10, Equation 4.11 and Equation 4.12. N/A - since we are considering a proton. Sodium is not considered since it is always balanced by an equivalent counterion and makes no contribution to $\Delta\Psi$.

Ion	$\Delta\Psi$ (mV)	ΔG_{elec} (kcal/mol)	ΔG_{chem} (kcal/mol)
Cl^-	+118 (inwards)	-2.7	N/A
H^+	+89 (inwards)	-1.9	-1.9
K^+	negligible	N/A	N/A
Total	+207 (inwards)	-4.6	-1.9
		-6.5 kcal/mol	

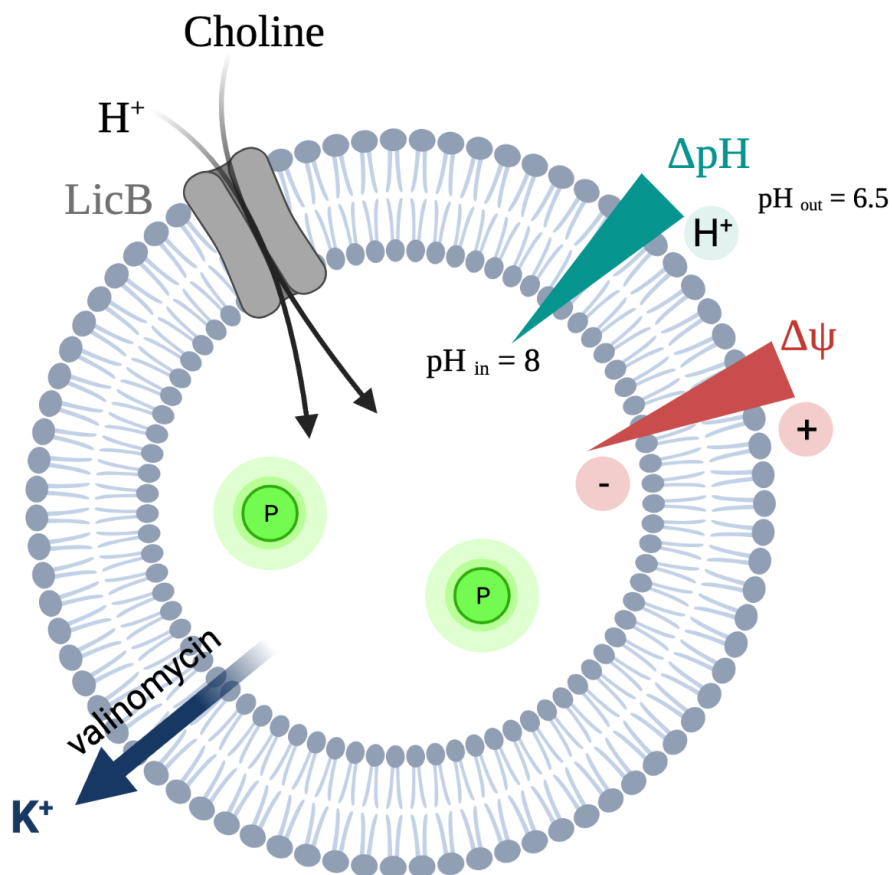


Figure 4.11. Schematic representation of the experimental setup.

The addition of valinomycin produces a Nernst potential (negative inside) that energises proton influx, which is detected by the entrapped pyranine (showing as P).

Preliminary experiments used proteoliposomes reconstituted in a 60:1 (w/w) lipid to protein ratio and conditions discussed above. Figure 4.12A shows control experiments performed in liposomes without LicB present. A background signal is observed in energised liposomes that is essentially independent of the presence of choline, suggesting a substantial background leakage of protons across the membrane into the energised liposome. Such leakage has previously been observed in similar assays. Figure 4.12B shows comparable data from proteoliposomes containing LicB. The trend in these data was slightly different, with leakage apparently occurring more rapidly and a slightly more significant difference between experiments in the presence and absence of choline. It appears that the inclusion of LicB in the proteoliposome membrane renders them more permeable to protons than liposomes. Subtracting the background signal (orange line) showed the impact of choline on the system (Figure 4.12C), revealing a higher signal slope for proteoliposome than for liposome.

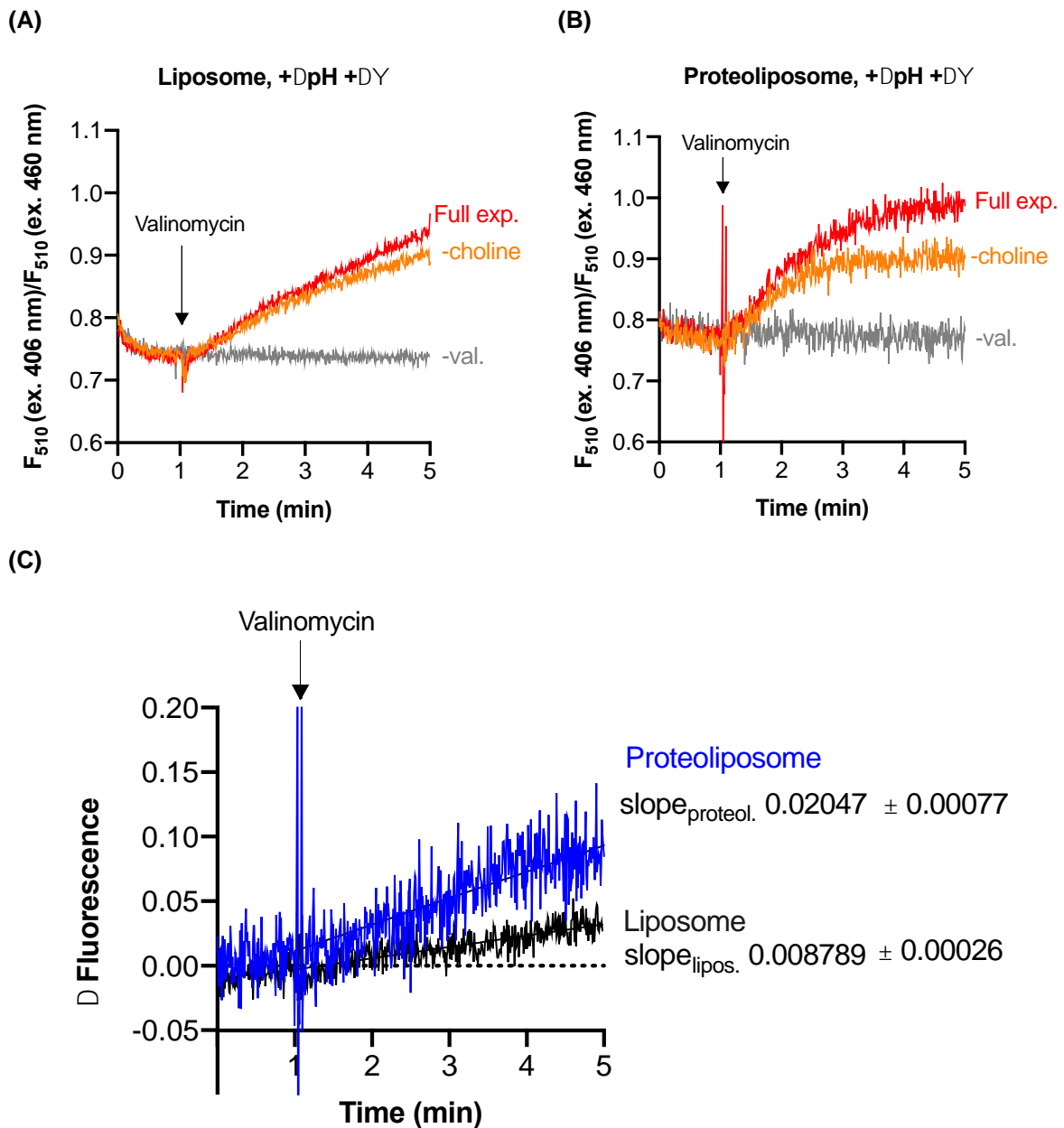


Figure 4.12. Representative pyranine fluorescence traces produced from the proteoliposome transport assays.

Data were collected from **(A)** liposome or **(B)** proteoliposome in 60:1 (w/w) lipid to protein ratio assays. The *full experiment* includes liposome/proteoliposomes, choline at 100 mM and valinomycin at 25°C. Controls were performed without choline or without valinomycin. Data are the average of triplicate experiments. **(C)** Difference of emitted fluorescence between *Full experiment* and *-choline* for liposome (black line) and proteoliposome (blue line). Linear regression was applied to data between 1 to 5 minute, and the slope calculated using GraphPad Prism.

Some adjustments to the buffer composition were performed to dissect the relative contribution of the chemical gradient and the charge gradient to transport. In this case, both ‘outside’ and ‘inside’ buffers were prepared at the same pH (equals 8.0). Figure 4.13 revealed that the background could be minimised if no chemical gradient was applied. There was a visible fluorescence difference in proteoliposome assay between the *full experiment* and in the absence of choline (Figure 4.13B). On the other hand, when choline was added to the liposome, no variance in the fluorescence signal was observed.

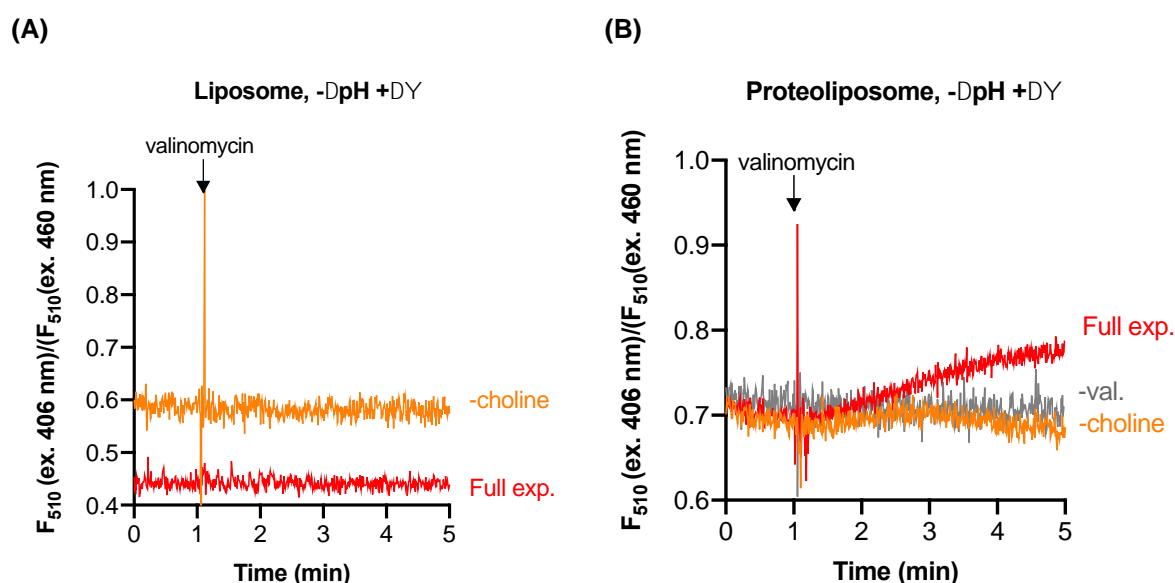


Figure 4.13. Representative pyranine fluorescence traces produced when no chemical gradient (ΔpH) was applied.

Data were collected from (A) liposome or (B) proteoliposome in 60:1 (w/w) lipid to protein ratio assays. *Full experiment* includes liposome/proteoliposomes, choline at 100 mM and valinomycin at 25°C. Controls were performed without choline or without valinomycin. Data are the average of at least duplicate experiments.

4.3. Discussion

LicB is thought to be a high-affinity choline transporter in bacteria from the human respiratory tract. The critical role of a choline derivative, phosphorylcholine, in the viability and persistence of respiratory infections make the study of LicB of considerable interest. This chapter presents the first attempts to study choline binding and transport by LicB *in vitro*.

Both FRET and ITC methods were used to examine choline binding by LicB. Fluorescence methods are attractive because of their sensitivity, which means that they require relatively low amounts of protein. However, as shown here for LicB, such methods depend upon a labelled substrate and can have complex and confounding background artefacts. In contrast, ITC provides a comprehensive characterisation of binding and does not require labelling of substrate or protein; however, it requires a relatively large amount of protein and is generally only valid for relatively tight protein-ligand interactions (K_d at low mM or below).

ITC showed no evidence of choline binding to LicB. Assuming that LicB is indeed a choline transporter, plausible explanations for this negative result are that protein concentration was too low or ligand binding too weak to evolve a detectable binding heat. As shown with the lactose permease (LacY), it might be possible to revisit ITC if a tighter-binding ligand can be identified or a LicB variant that shows higher ligand affinity. Alternatively, higher LicB concentrations might be placed in the sample cell, but this would require the scale-up of LicB purification in a way that has proven challenging (Chapter 3).

A novel fluorescence-based assay was used to determine both the binding affinity and the substrate profile of LicB. The results showed that LicB is highly specific for choline over similar small molecules. Interestingly, these binding studies did not produce values consistent with the relatively low values for K_M of 3.6 μM obtained in cell studies for choline transport by LicB³⁶¹. However, our results are more similar to studies of the osmoprotective choline transporters; for example, BetT from *Pseudomonas syringae* has K_M of 876 μM for the transport of choline, which is surprisingly high compared to other BCCTs³⁴⁹. Clearly, the results of the novel fluorescent assay reported here should be interpreted with caution, and the exact value of K_d must be considered strictly provisional. Nonetheless, the assay does appear to successfully discriminate between different small molecules and their interaction with LicB. Only choline and acetylcholine were found to displace dansyl choline in these competition assays, suggesting that LicB is rather selective in its ligand interaction. Even with the caveats discussed throughout the chapter, these results appear to provide a useful insight into ligand binding by LicB. The fluorometric assay described here could serve as a valuable screen for tight-binding inhibitors of LicB and so may be of interest in drug development. The trimethylammonium headgroup of choline is expected to be important in substrate binding. In osmoregulatory transporters such as BetP, substrate binding occurs via cation- π interactions between the trimethylammonium and multiple tryptophan residues (at different positions depending on the protein)^{354,424}. However, in LicB, the choline headgroup alone was not sufficient to drive tight binding since compounds including benzoylcholine, betaine and L-

carnitine could not displace DC. These compounds preserve these cation- π interactions but have substitutions at the choline hydroxyl 'tail', which is relatively dynamic and can engage in salt bridges and other local interactions in BCCT transporters³³⁴. These data suggest a strict requirement for a 2-carbon unsubstituted aliphatic chain for binding to LicB. A closer look at the alignment between BCCTs (BetT, BetP and CaiT) and LicB showed that the latter does not display the 'aromatic cage' as other BCCTs (showed as blue in Figure 4.14). Nonetheless, it displays a conserved tyrosine (showed as red in Figure 4.14) which is only observed in transporters of quaternary ammonium substrates (for instance, choline or carnitine). Therefore, it can be judged that another binding mechanism is involved in the binding and transport of choline.

Purified LicB was reconstituted into synthetic proteoliposomes in order to try and study substrate transport (Figure 4.9). A well-established pyranine-based assay was used to determine substrate transport by detecting the cotransport of protons indirectly. The results suggest that reconstituted protein is not highly active in transporting H⁺ ions, and by extension, does not transport high levels of choline *in vitro* (Figure 4.12 and Figure 4.13). Choline had caused a slight increase in signal above the background, but this should be treated as provisional and requires further confirmation. Limitations of time and resources prevented establishing a direct method of detecting choline, which would now seem to be an obvious route forward since successful reconstitution methods have been established. Although assay results are entirely consistent with the generation of membrane potential, it would also be helpful to characterise the nature and magnitude of the charge gradient by using potential-responsive dyes such as Oxonol V. Additional work should also attempt to minimise proton leakage. However, the POPE/PG mix used here was previously identified as optimal in this regard⁴²⁵.

Tsai and Miller also identified that a protein concentration higher than 15 μg per mg of lipid significantly accelerate the collapse of pH gradients and cause membrane leakage⁴²⁵. Accordingly, the activity of LacS from *S. thermophilus* depends strongly on the final lipid to protein ratio and decreases rapidly at higher ratios⁴²⁶. The lipid to protein ratio was also found to be critical for the reconstitution of bacteriorhodopsin and H⁺-transporting ATP-synthase from thermophilic *Bacillus* PS3⁴²⁷, the lactose transport system, LacY of *E. coli*⁴²⁸ and for the specific activity of the H⁺-ATPase from chloroplasts⁴²⁹. Thus, the lipid to protein ratio is critical for the final transport activity, and for LicB reconstitution, other ratios should be tested. Concerning the chosen lipid mixture, it has been shown that 3:1 POPE to POPG ratio rather than *E. coli* polar lipid mixture achieved higher performance. Nonetheless, it is known that specific protein-lipid interactions are crucial for their activity⁴³⁰. For example, PE or PC are

essential to obtain a maximal activity of the lactose permease, LacY of *E. coli*⁴³¹, and cardiolipin in the case of cytochrome oxidase⁴³². Accordingly, Warren *et al.* demonstrated that most integral MP require a liquid-crystalline phospholipid environment for function⁴³³. Therefore, reconstituting the membrane protein into a more native environment may influence its transport activity, for instance, in bicelles or nanodiscs.

Another consideration is protein orientation within the proteoliposome membrane. It is unclear if the incorporation of LicB into liposomes, when mediated by DDM, is inserted in the same orientation as that in the native environment. Thus, other mediated detergents or chemicals should be screened. For example, Knol *et al.* demonstrated that the uptake rate of LacS was higher following Triton X-100-mediated reconstitution in comparison with DDM-mediated reconstitution^{434,435}.

LicB	-----	0
CaiT	NEKRKTGIEPKVFFPPLIIVGILCWLTVRDLDAANVVINAVFSYVTNVWG	52
BetP	GEDTQASLNWSVIVPALVIVLATVWVGIGFKDSFTNFASSALSAVVDNLG	100
BetT.Hinfluenzae	-MTKRTSFNPLVIGVTLFFILLMAMIFIAPEQTQALLNQAKSGIFANFS	49
BetT.Ecoli	HSREKDKINPVVFFYTSAGLILLFSLTTLFRDFSALWIGRTLWVSKTFG	55
LicB	-----MRGYLFGILSAVFWALSGLLYN-ELPLSEY TALGKVISLFLID	43
CaiT	WAFEWYVVMVLFQWFWLVFGPYAKKRLG--NEPPEFSTAS-WIFM----M	95
BetP	WAFILFGTVFVFFIVVIAASKFGTIRLGRIDEAPEFRTVS-WISM----M	145
BetT.Hinfluenzae	WFYVLTFSVFLGELLILSVSSLGNIKLGQDEEPEFSFSL-WLAM----L	94
BetT.Ecoli	WYLLAATLYIVFVVCIACSRFGSVKLGPEQSKPEFSLLS-WAAM----L	100
LicB	FCSLLVIGITLWRKSAVDFQGVFWQPAL-SGI-----LGG	77
CaiT	FASCTSAAVLFWGSIETIYY--YISTPPFGLPNSTGAKELGLAYSFLFHG	143
BetP	FAAGMGIGLMFYGTTEPLT--FYRNGVP--GH-DEHNVGVAMSTTMFHWT	190
BetT.Hinfluenzae	FAAGMGVGLMFFGVAEPLT--HYLSDIT-AGS-AEHKQREALHTLTFHWG	140
BetT.Ecoli	FAAGIGIDLMEFVVAEPVT--QYMQPPEGAGQ-TIEAARQAMVWTLFHYG	147
LicB	PIGMSAY---LLSIHYLTIYYAAPL---SSLFVFAALMSYWILKEKIT	120
CaiT	PLPWATYSFLSVAFAFFVVRKMEVIRPSS TLVPLVGEKHV-----	184
BetP	LHPWAIYAIVGLAIAYSTFRVGRKQ-LLSSAFVPLIGEKGAE-----	230
BetT.Hinfluenzae	IHAVAVYGTIALALAYFGFRYKLP-ALRSCFYPLLKDR-I-----	179
BetT.Ecoli	LTGWSMYALMGMALGYFSYRYNLPL-TIRSALYPIFGKR-I-----	186
LicB	KTAQFGFALAIISSSLAIIEVQGEITFNTIGFIFLIICILGWSSEIVISS	170
CaiT	-KGLFGTIVD-----NFYLVALIFAMGTSGLGLATPLVTEC	218
BetP	-EGWLCKLID-----ILAI IATVFGTACSLGLGALQIGAG	264
BetT.Hinfluenzae	-NGKIGDAID-----VMALLATTFGIITTLGFGSSQLGAG	213
BetT.Ecoli	-NGPIGHSVD-----IAAVIGTIFGIATTLGIGVVQLNYG	220
LicB	Y-----TMRSLSGL--QVYFLRLCGSTIGY	193
CaiT	MQWLFGIPH-TLQLDAI IITCWIILNAICVACGLQKGVRIASDVRYSLSF	267
BetP	LSAANI IEDPSDWTIVGIVSVLTLAFIFSAISGVGKGIQYLSNANMVLA	314
BetT.Hinfluenzae	LEQIGWISQNSFALQVGVIVVVMCLAVFSAISGVGKGVKILSEINLTLAF	263
BetT.Ecoli	LSVLFDIP-DSMAAKAALIALSVIIATISVTSGVDKGIRVLSELNVALAL	269
LicB	LLILF ILSL--KNFSLDILS-----FNYVQIAGVIIFGA-----L	226
CaiT	LMLGWVFI VSGASFIMNYFTDSVGM LLMYLPRMLFYT---DPIAKGGFPQ	314
BetP	LLAIFV FVVGPTVSI LNLPGSIGNYLSNFFQ MAGRTAMSADGTAGEWLG	364
BetT.Hinfluenzae	CLLLEVLISGPTLYLLSAFSDNIGNYF SNLVQLSFKT-YAYEQEHTSWFS	312
BetT.Ecoli	GLILEVLFMGDTSFLLNALVLNVGDYVNRFMGMTLNS-FAFDR-PVEWMN	317
LicB	SYCCYYQAIYLLKPIKAMALNITYSVAIGLGYLLYKQ---PIKPITLL	272
CaiT	GWTVFYWA-----WVVIYAIQMSIFLARISRGRTVRELFCFG	350
BetP	SWTIFYWA-----WVISWSPFVGMFLARISRGRSIREFILG	400
BetT.Hinfluenzae	GWTVLYWA-----WVCSWAPFVGLFIARISKGR TIREFIFG	348
BetT.Ecoli	NWTLFEWA-----WVVAWSPFVGLFLARISRGR TIRQFVLG	353
LicB	LTLLLSAGVI-----VTLYYKGEQK-----	292
CaiT	MVLGLTASTWILWTVLGSNTLLMDKNI INIPNLIEQYGVARAI IETWAA	400
BetP	VLLVPAGVSTVWFSIFGGTAIVFEQNGESI WGD---GAAEEQLFGLLHA	446
BetT.Hinfluenzae	VLVIPSLFGILWFTVFGNTAVWLNDGIAAGGLG-EFISSPEILLFKFLNY	397
BetT.Ecoli	TLIIPFTFTLLWLSVFGNSALYEI IHGGAAFAE-EAMVHPERGFYSLLAQ	402

Figure 4.14. Part of the multiple sequence alignment between BCCTs and LicB protein.

Amino acid sequence alignment of *H. influenzae* LicB (UniProt P14182), *Corynebacterium glutamicum* BetP (UniProt P54582), *E. coli* CaiT (UniProt P31553), *E. coli* BetT (UniProt P0ABC9) and *H. influenzae* BetT (UniProt P45335). Conserved residues are highlighted in black, while similar residues (W or Y) are highlighted in grey. Residues highlighted in blue and red are highly conserved among the BCCT family and essential in substrate binding and transport.

4.4. Conclusion

This chapter describes a series of methods that attempt to determine the substrate specificity, selectivity and transport behaviour of LicB. A novel fluorescence assay for ligand binding was developed and, while still needing confirmation by other methods, revealed that LicB could discriminate in binding choline and acetylcholine instead of related small molecules such as glycine, betaine and carnitine. Taken together, these results suggested a strict requirement for a 2-carbon unsubstituted aliphatic chain for binding to LicB. It seems plausible that this assay could be the basis for drug development.

It was possible to confirm LicB reconstitution into proteoliposomes but not to definitively show substrate transport in this system. Additional work on developing a radioactive uptake assay would now be beneficial in understanding transport by LicB.

Chapter 5. Probing the function of LicB by site-directed mutagenesis

5.1. Using mutagenesis to characterise LicB

LicB is a member of the Choline Uptake Transporter Family (TCDB 2.A.7.18), which currently includes only three other proteins (Table 5.1). A multiple alignment sequence of these transporters showed a poor identity percentage, around 5% (Figure 5.1). Close homologues of the archetype LicB from *H. influenzae* have been identified in *S. pneumoniae*, commensal *Neisseria* species, and *Histophilus sommi*, among others³³³ (Figure 5.2). None of these homologues has been characterised, but they are presumed to share a common fold and function as choline permeases. In this respect, it seems reasonable to assume that these homologues will share highly conserved amino acids that play a critical role in protein structure or function. In this chapter, a multiple sequence alignment was used to identify such residues and test their possible role through mutagenesis. As mentioned previously, LicB has poor similarity to the well-characterised BCCTs, such as the BetP or CaiT, missing critical tryptophan residues that promote substrate binding (Chapter 4). Thus, it is not possible to directly translate our understanding of these proteins to LicB.

Table 5.1. Members of the Choline Uptake Transporter Family (TCDB 2.A.7.18).

From Transporter Classification Database³³⁵.

Accession number	Species	Protein name	Length	Number of TMs	Substrate
P14182	<i>H. influenzae</i>	LicB	292	10	Choline
Q8TTA7	<i>Methanosarcina acetivorans</i>	MttP	353	10	Methylamine
Q8TS76	<i>Methanosarcina acetivorans</i>	MttP2	281	9	Methylamine
O50525	<i>Streptomyces coelicolor</i>	Putative integral membrane protein	312	10	Choline

MttP	MEGFSVDLKALKKQESKRKISKGY-MWALFCAVFWGIWYLPGTVVWVLNP	49
MttP2	-----	0
LicB	-----MRGY-LFGILSAVFWALSGLLYNE-----LP	25
PutativeMP	-----MKNGTRGWILYASLLVLFWGVWGAFFSS-----EP	29
MttP	FDEMYGAI AETGGDGTALIIITAVLITAFNALTVM LALMLWNGVLGKFGEL	99
MttP2	-----MITAFNALTVM LALILWNGVLQKFGEL	27
LicB	L-SEYTALGK-----VISLLFLIDFCS-LLVIGITLWRKSAV-----	60
PutativeMP	N-SRYGYPNE-----MIYIIWAF---T-MLIPAYFAMRGNTF-----	61
MttP	VRTMKEFHPCSKWFFLASIFGCPMAILGSFI---AMGFIGGAFAAVAALL	146
MttP2	SRTLKEFNPCSKWFFLASIFGCPAILGSFM---AMGFIGGAFAAVAALL	74
LicB	-----DFQGVFWQPALSGILGCPIGMSAYLL---SIHYLTIYYAAPLSSL	102
PutativeMP	-----DRRP---VAAMYGLIACLTGAGGQLLLQALADGPAYLIFPLVSL	103
MttP	YPVVGSILAYYWYGEKISKRAAIGIAVIVGGVTI--FGGGLLTELSSGN	194
MttP2	YPVIGSVLAYYWYGEKITKRAATGIIVIIILGGITI--FGGGLVTELGAGE	122
LicB	FPVFAALMSYWILKEKITTKTAQGFALAISSSLLAIEVCGQ---EITFNT	149
PutativeMP	SPVITVLMATVLLRERITRLAVVGVVAALVAIVLLSIPSCG---GDSGAH	150
MttP	VQWIGYLGGLMAAAGWGIEGAIAGKGLDISEPDVGLTLRFVGENLIWWVI	244
MttP2	VPWIGYLGGLMAAVGWGIEGAIAGKGLDISEPDVGLTLRFLGENI IWWII	172
LicB	IGF---IFLIICILGWSSEIIVISSYTMRSLSGLQVYFLRLCGSTIGY-LL	195
PutativeMP	GPWLP-MAILIC-VAWGVQA----FCMRKAALVGV-----	179
MttP	IVPILAI VGYPMYSFAFQAIEPLTLLVLIF----AGITFGFCYVCWYKSF	290
MttP2	IVPVLAII GFPMYSYAIQVEPLALLVLIF----AGITFGFCYVSWYKSF	218
LicB	ILFILSLKNF---SLDILSENIVQIAGVII-----FGALS YCCYYQAI	235
PutativeMP	-----NDATTEGWMTISGLLLVPVAFAIMGGFPSGAPWQAP	215
MttP	PIIGVGRG-QGIGNLYGLEAI-----IFIFLFFG	318
MttP2	PIIGVGRG-QGIGNLYGFFAV-----IFIFLFFG	246
LicB	YLLKPIKA-MALNITYSVWAI GLGYLLYKQPIKP-ITLLL TLLL-----	277
PutativeMP	ALTAVTQVLNAV GALFLVMAF SRGKATVVAPITNALAPVLTIVLSLAVYQ	265
MttP	DVPQWTILVGGTLCVIIGSFVMF--TEDTSELETTLR-GE-----	353
MttP2	DVPQWTILLGGTLCIIIGSFVMF--TEEATEIETTLR-GG-----	281
LicB	-----SAGVIVTLYYKGEQK-----	292
PutativeMP	TLPSVWGALGIVLALAGSTLMV-YSDEKSGESSVAPGAADDDVSSVAR	312

Figure 5.1. Multiple sequence alignment between Choline Uptake Transporter Family members.

The alignment was produced based on the protein sequences from Table 5.1. Conserved residues are highlighted in black.

S. pneumoniae	-----MKSKNGVFFGLLSGI F WGLGLTVSAYIFSIFTDLSPFVVAATH-DFLSIFIL	51
H. somni	MRLFSEKQRIIHVGYFCGLLSGLFWAISGIIYKLYTDF----PMLNAFALNII--LSI	54
N. lactamica	-----MRGYLFGAASAVFWAASGLLFDRLP-----FARGAAFQOMVFLFL	41
N. flavescens	-----MRGYLFGAVSAVFWAASGLLFDRLP-----FARGAAFQOMVFLFL	41
N. meningitidis	-----MRGYLFGAVSAVFWAASGLLFDRLP-----FARGAAFQOMVFLFL	41
H. haemolyticus	-----MRGYLFGIISAI FWALSGLLYNELP-----LSGFTALGKVISLFL	41
H. quentini	-----MHGYLFGIISAI FWALSGLLYNELP-----LSGFTALGKVISLFL	41
H. influenzae	-----MRGYLFGIISAVFWALSGLLYNELP-----LSEYTALGKVISLFL	41
H. sp. C1	-----MRGYLFGIISAVFWALSGLLYNELP-----LSEYTALGKVISLFL	41
H. h. M19501	-----MHGYLFGIISAI FWALSGLLYNELP-----LSEYTALGKVISLFL	41
H. aegyptius	-----MRGYLFGIISAVFWALSGLLYNELP-----LSEYTALGKVISLFL	41
S. pneumoniae	LA--FLLVKEG---KVRLSIFLNIRNVSVIIGALAGPIGMQANLYAVKVI GSSSLASSVS	106
H. somni	AEITLLIMAYGYLIKRRKLLPHLNKVSLLAILGGCLGGPVGMICYLQSLQVGVVSYAATIS	114
N. lactamica	IDL-CSLSAVWAGLGRGPRGFRGVFWRPVLVSLGLGGPVGMSAYLLGIHYLTVVYAAPLS	100
N. flavescens	IDL-CSLSAVWAGLGSGRPGFRGVFWRPVLVSLGLGGPVGMSAYLLGIHYLTVVYAAPLS	100
N. meningitidis	IDL-CSLSAVWAGLGSGRPGFRGVFWRPVLVSLGLGGPVGMSAYLLGIHYLTVVYAAPLS	100
H. haemolyticus	IDF-FSLFVIGITLWRKRAVNFQKIFWSPVLSGVLGGPILGMSAYLLSIHYLTIYYAAPLS	100
H. quentini	IDF-FSLFAIGITLWRKRAVNFQKIFWSPVLSGVLGGPILGMSAYLLSIHYLTIYYAAPLS	100
H. influenzae	IDF-CSLLVIGITLWRKSAVDFQGVFWQPALSGILGGPIGMSAYLLSIHYLTIYYAAPLS	100
H. sp. C1	IDF-CSLLVIGITLWRKSAVDFQGVFWQPALSGILGGPIGMSAYLLSIHYLTIYYTAPLS	100
H. h. M19501	IDF-CSLLVIGITLWRKSAVDFQGVFWQPALSGVLGGPIGMSAYLLSIHYLTIYYAAPLS	100
H. aegyptius	IDF-CSLLVIGITLWRKSAVDFQGVFWQPALSGVLGGPIGMSAYLLSIHYLTIYYAAPLS	100
S. pneumoniae	AIYPAISVLLAFFFIKHKISKNTVFGIVLIIGGIIAQ-TYKVEQVNSFYI GILCALVCAI	165
H. somni	ICYPIIASILSIKIFAVKASRRTLFSFVIVLMLMPV-MYQGENSEIRILGIIFAFSCAL	173
N. lactamica	SLFVLAALSAWFIKERTGTAERAGFALAVLSSAGLAADAGGRGGLNALGLVCIACVVS	160
N. flavescens	SLFVLAALSAWFIKERTGTAERAGFALAVLSSAGLAADAGGRGGLNALGLVCIACVVS	160
N. meningitidis	SLFVLAALSAWFIKERTGTAERAGFALAVLSSAGLAADAGGRGGLNALGLVCIACVVS	160
H. haemolyticus	SLFVLAALMSYWIKEKIKTKTAQFGFALAISSSLLAIEVQEI TFNTIGFIFLIICIL	160
H. quentini	SLFVLAALMSYWIKEKISKTAQFGFALAISSSLLAIEVEQEITFNITIGFIFLIICIL	160
H. influenzae	SLFVFAALMSYWIKEKIKTKTAQFGFALAISSSLLAIEVQEI TFNTIGFIFLIICIL	160
H. sp. C1	SLFVFAALMSYWIKEKISKTAQFGFALAISSSLLAIEVEQEITFNITIGFIFLIICIL	160
H. h. M19501	SLFVFAALMSYWIKEKIKTKTAQFGFALAISSSLLAIEVQEI TFNTIGFIFLIICIL	160
H. aegyptius	SLFVFAALMSYWIKEKIKTKTAQFGFVLAISSSLLAIEVQEI TFNTIGFIFLIICIL	160
S. pneumoniae	AWGSESVLSSFAMESELSEIEALLIRQVTSFLSYLVIVLFSHQ-----FTAVANGQ	217
H. somni	CWALEIVISSYVMQH-YPSDQVYLFRRQMGSSSGYFTIILFMFFFYGYDNLIDFNFKPDVN	232
N. lactamica	GWSLEVVL SARTMRT-LSGLQVYCFRLCGSVCGYLLILVCL SARGFSLNPSDLYDARII-	218
N. flavescens	GWSLEVVL SARTMRT-LSGLQVYCFRLCGSVCGYLLILVCL SARGFSLNPSDFYDARII-	218
N. meningitidis	GWSLEVVL SARTMRT-LSGLQVYCFRLCGSVCGYLLILVCL SARGFSLNPSDFYDARII-	218
H. haemolyticus	GWSSEIVISSYTMRS-LSGLQVYFRLCGSTIGYLLILFILSLKDFSLDILSFNYQIA-	218
H. quentini	GWSSEIVISSYTMRS-LSGLQVYFRLCGSTIGYLLILFILSLKDFSLDILSFNYQIA-	218
H. influenzae	GWSSEIVISSYTMRS-LSGLQVYFRLCGSTIGYLLILFILSLKNFSLDILSFNYQIA-	218
H. sp. C1	GWSSEIVISSYTMRS-LSGLQVYFRLCGSTIGYLLILFILSLKDFSLDILSFNYQIA-	218
H. h. M19501	GWSSEIVISSYTMRS-LSGLQVYFRLCGSTIGYLLILFILSLKDFSLDILSFNYQIS-	218
H. aegyptius	GWSSEIVISSYTMRS-LSGLQVYFRLCGSTIGYLLILFILSLKNFSLDILSFNYQIV-	218
S. pneumoniae	LLGLMIVFAAFDMISYLAYYIAINRLQPAKATGLNVSYVVMIVLFAVAVFLGAPLDMTIM	277
H. somni	LWKLEWIVISSGFSYFLYKSIATLNPTRAMALNITYSLMAIVLSAVLLSKEISIFLLM	292
N. lactamica	-----GIIAFGALS YACYQAIYLLKPLKAMALNITYPVWAMVLGFFVYRHPVPPSAII	272
N. flavescens	-----GIIAFGALS YACYQAIYLLKPLKAMALNITYPIWAMVLGFFVYRHPVPPSAII	272
N. meningitidis	-----GIIAFGALS YACYQAIYLLKPLKAMALNITYPIWAMVLGFFVYRHPVPPSAII	272
H. haemolyticus	-----GVIIFGALS YCCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYQPIKIPITLL	272
H. quentini	-----GVIIFGALS YCCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYQPIKIPITLL	272
H. influenzae	-----GVIIFGALS YCCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYQPIKIPITLL	272
H. sp. C1	-----GVIIFGALS YVCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYQPIKIPITLL	272
H. h. M19501	-----GVIIFGALS YCCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYQPIKIPITLL	272
H. aegyptius	-----GVIIFGALS YCCYQAIYLLKPIKAMALNITYSVWAIGLGYLLYQPIKIPITLL	272
S. pneumoniae	TSLVVIAGVYIIKE-----	292
H. somni	IGIIVLMNIF--IVAKGEEK	310
N. lactamica	WAVLLGTGAALS LCGRKGEI	292
N. flavescens	WAVLLGTGAALS LYGREGEI	292
N. meningitidis	WAVLLGTGAALS LYGREGEI	292
H. haemolyticus	LTLLLSAGVIVTLYYKGEQK	292
H. quentini	LTLLLSAGVIVTLYYKGEQK	292
H. influenzae	LTLLLSAGVIVTLYYKGEQK	292
H. sp. C1	LTLLLSAGVIVTLYYKGEQK	292
H. h. M19501	LTLLLSAGVIVTLYYKGEQK	292
H. aegyptius	LTLLLSAGVIVTLYYKGEQK	292

(see caption next page)

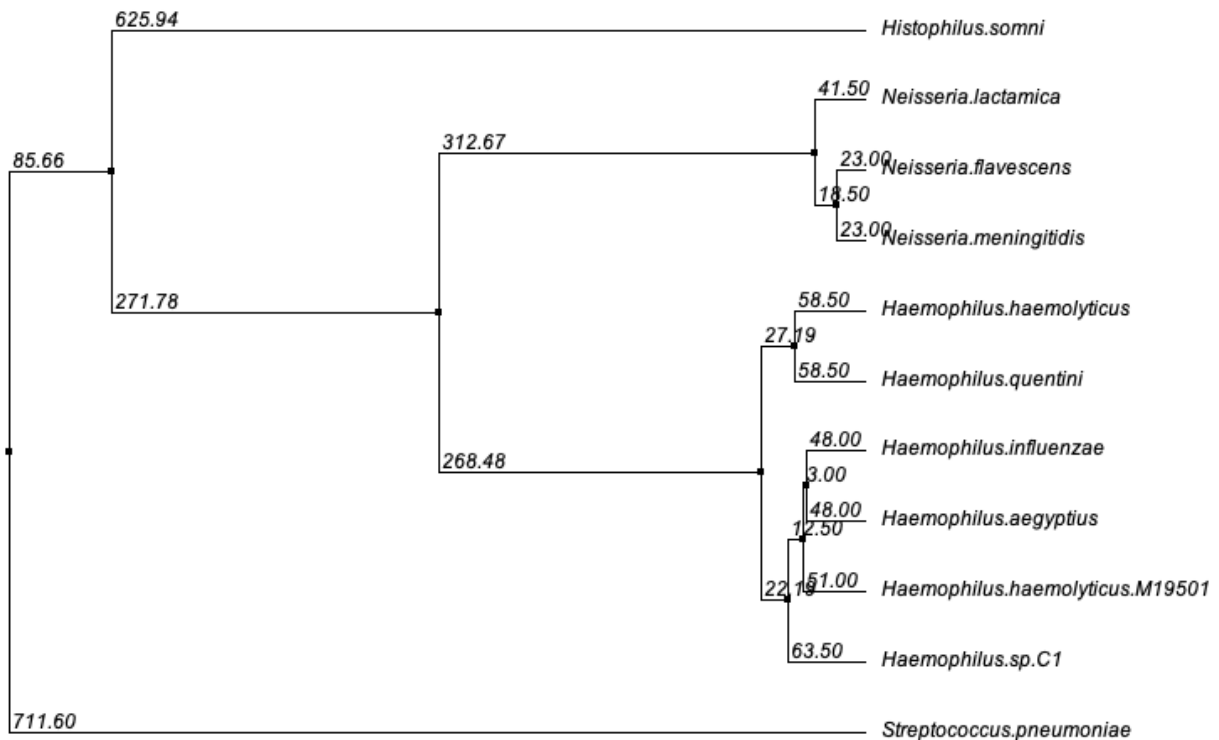


Figure 5.2. Multiple sequence alignment of LicB in selected opportunistic pathogens of the human nasopharynx and the corresponding phylogenetic tree.

Homologues were selected from UniProt database. *Haemophilus influenzae* (UniProt entry P14182), *Haemophilus haemolyticus* M19501 (F9GQV8), *Haemophilus aegyptius* (A0A3S4TRV6), *Haemophilus* sp. C1 (A0A0M1GMP2), *Haemophilus haemolyticus* (A0A0M3G111), *Haemophilus quentini* (A0A657KXA9), *Neisseria lactamica* (A0A378VI20), *Neisseria flavescens* (C0EQM0), *Histophilus somni* (Q0I5L5), *Neisseria meningitidis* (SPY01485.1), *Streptococcus pneumoniae* (A0A0H2UQH5). Numbers in the figure correspond to the evolutionary distance between the sequences and were calculated by maximum likelihood. Phylogenetic tree was generated using Jalview software.

5.1.1. Site-directed mutagenesis

Site-directed mutagenesis has been used extensively in protein science due to its powerful ability to probe protein structure and function. Alanine-scanning mutagenesis has been particularly successful in systematically identifying protein positions important for function or ligand binding⁴³⁶; indeed, alanine is generally taken as the default mutation, on the basis of substitution removing all side-chain atoms at the β -carbon, resulting in a 'chemically inert' residue. This avoids introducing side-chain bulk or functionality (such as hydrogen-bonding groups) that might disrupt the protein, and the small cavities generated by alanine substitution are generally tolerated. Thus, the effects of individual alanine mutations can be used to infer the roles of individual side chains in structure and function.

5.2. Results

5.2.1. Identifying conserved residues through multiple sequence alignment

Conserved sequence motifs were identified using the online tool BLASTP. This used the sequence LicB from *H. influenzae* to query the NCBI protein database. Results were filtered to remove the very high number of hits returned from different strains of *H. influenzae*. The top 100 sequences returned from this BLASTP query were inspected and redundancy further minimised by excluding all but one representative from each species. Finally, partial sequences were excluded so that only full-length proteins were included. This resulted in a final collection of 30 homologous sequences that were aligned by Clustal Omega (BLAST result and full alignment are represented in Table 5.2 and Figure 5.3, respectively). As expected, the majority of the BLAST results were designated as uncharacterised DMT family transporters. Figure 5.5 reveals an illustrative phylogenetic tree of the alignment.

This multiple sequence alignment revealed that only 20 amino acid residues were completely conserved among all sequences (highlighted in blue in Figure 5.4 top panel). A highly conserved motif emerged in putative helix 9, with the sequence:

- (I/L)(K/R)AM*xxNITYxxW -

in which A, N, I, T, Y and W amino acid residues were absolutely conserved in all sequences, and M is conserved in 29 of 30 sequences (Figure 5.4 bottom panel). In LicB, this motif corresponds to the following sequence -IKAMALNITYSVW-.

The position of this motif within the LicB homology model generated in Chapter 3 is shown in Figure 5.6. The motif runs down the internal face of helix 9, as would be expected of residues involved in maintaining structural integrity via helix-helix interactions or playing a role in substrate transport.

Table 5.2. List of the 30 sequences used for the multiple sequence alignment.

Reference	Transporter	Specie
WP_005672105.1	LicB protein	<i>Haemophilus influenzae</i>
CVQ29934.1	drug/metabolite transporter (DMT) superfamily permease	<i>Streptococcus pneumoniae</i>
WP_065250943.1	DMT family transporter	<i>Haemophilus aegyptius</i>
WP_005632346.1	LicB protein	<i>Haemophilus haemolyticus</i>
WP_053465541.1	LicB protein	<i>Haemophilus</i> sp. C1
WP_070647074.1	LicB protein	<i>Haemophilus</i> sp. HMSC066A11
WP_044233009.1	DMT family transporter	<i>Haemophilus quentini</i>
EGT81144.1	LicB protein	<i>Haemophilus haemolyticus</i> M21639
WP_032803484.1	DMT family transporter	<i>Haemophilus</i> sp. oral taxon 851
WP_118847083.1	DMT family transporter	<i>Neisseria lactamica</i>
WP_025457124.1	DMT family transporter	<i>Neisseria polysaccharea</i>
WP_118792478.1	DMT family transporter	<i>Neisseria bergeri</i>
WP_003682075.1	DMT family transporter	<i>Neisseria flavescens</i>
AGB93814.1	LicB protein	<i>Avibacterium paragallinarum</i>
VTX86427.1	EamA-like transporter family protein	<i>Neisseria subflava</i>
WP_046338537.1	DMT family transporter	<i>Pasteurella multocida</i>
WP_101774900.1	DMT family transporter	<i>Pasteurella oralis</i>
WP_077478779.1	DMT family transporter	<i>Rodentibacter trehalosifermentans</i>
WP_090921251.1	DMT family transporter	<i>Pasteurella skyensis</i>
WP_081253531.1	DMT family transporter	<i>Eikenella corrodens</i>
WP_115322638.1	DMT family transporter	<i>Pasteurella canis</i>
WP_011148873.1	DMT family transporter	<i>Photorhabdus laumondii</i>
WP_049584926.1	DMT family transporter	<i>Photorhabdus luminescens</i>
WP_112895251.1	DMT family transporter	<i>Photorhabdus bodei</i>
WP_038239927.1	DMT family transporter	<i>Xenorhabdus szentirmaii</i>
WP_065388896.1	DMT family transporter	<i>Photorhabdus namnaonensis</i>
NBI12435.1	EamA family transporter	<i>Haemophilus felis</i>
WP_054477567.1	DMT family transporter	<i>Photorhabdus heterorhabditis</i>
WP_075294186.1	DMT family transporter	<i>Histophilus somni</i>
WP_015836391.1	DMT family transporter	<i>Photorhabdus asymbiotica</i>

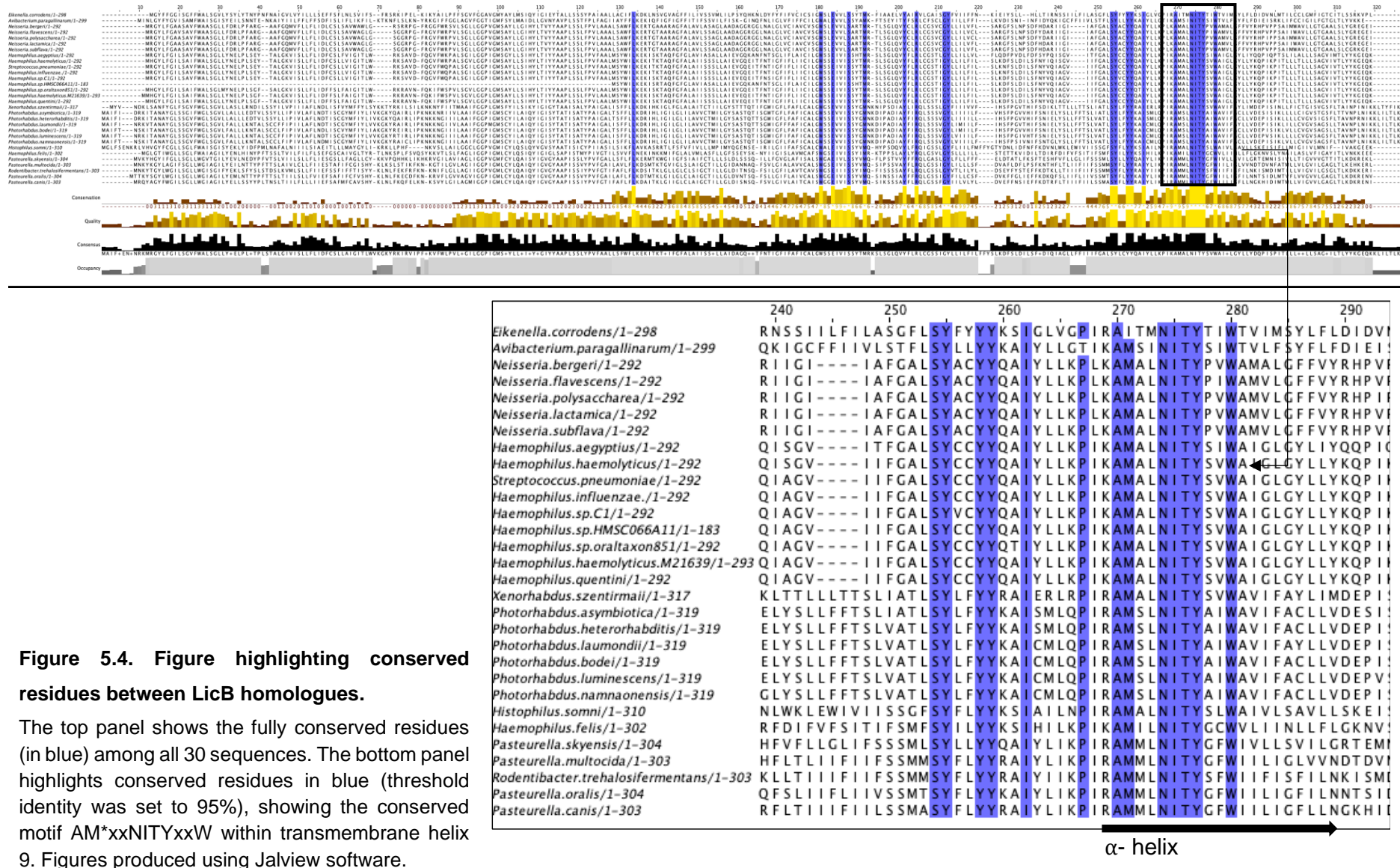
Chapter 5. Probing the function of LicB by site-directed mutagenesis

Eikenella.corrodens	-----MGYFFGGISGFFWALSGLVLSYLYTNPFNFAIGVLYVILLSEFFSF	48
Avibacterium.paragallinarum	-----MINLRYEYGVISAMFWAISGISYELLSNNTTE--NKAIYIILFFLFFSDFISL	50
Neisseria.bergeri	-----MRGYLFGAASAVFWAASGLLFDRLPFARG--AAFQGMVFLFLDLDLCSL	47
Neisseria.flavescens	-----MRGYLFGAASAVFWAASGLLFDRLPFARG--AAFQGMVFLFLDLDLCSL	47
Neisseria.polysaccharea	-----MRGYLFGAASAVFWAASGLLFDRLPFARG--AAFQGMVFLFLDLDLCSL	47
Neisseria.lactamica	-----MRGYLFGAASAVFWAASGLLFDRLPFARG--AAFQGMVFLFLDLDLCSL	47
Neisseria.subflava	-----MRGYLFGAASAVFWAASGLLFDRLPFARG--AAFQGMVFLFLDLDLCSL	47
Haemophilus.aegyptius	-----MRGYLFGILSAVFWALSGLLYNELPLSEY--TALGKVISLFLDIFCPSL	47
Haemophilus.haemolyticus	-----MHGYLFGILSAIFWALSGLLYNELPLSEY--TALGKVISLFLDIFCPSL	47
Streptococcus.pneumoniae	-----MRGYLFGILSAVFWALSGLLYNELPLSEY--TALGKVISLFLDIFCPSL	47
Haemophilus.influenzae.	-----MRGYLFGILSAVFWALSGLLYNELPLSEY--TALGKVISLFLDIFCPSL	47
Haemophilus.sp.C1	-----MRGYLFGILSAVFWALSGLLYNELPLSEY--TALGKVISLFLDIFCPSL	47
Haemophilus.sp.HMSC066A11	-----	0
Haemophilus.sp.oraltaxon851	-----MHGYLFGILSAIFWALSGLMYNELPLSGF--SALGKVISLFLDIFCPSL	47
Haemophilus.haemolyticus.M21639	-----MMHGYLELGLSAIFWALSGLLYNELPLSGF--TALGKVISLFLDIFCPSL	48
Haemophilus.quentini	-----MHGYLFGILSAIFWALSGLLYNELPLSGF--TALGKVISLFLDIFCPSL	47
Xenorhabdus.szentirmaii	--MYV---NDKLSANFYGLFSGVFWLGSGLVLLASLNRNDILSSYLVPVIAFLNDLIS	55
Photorhabdus.asymbiotica	MAIFI---DRKITANAYGLSSGIFWGLSGVLLALLELDTVLSSYLLIPVLAFLNDTISC	57
Photorhabdus.heterorhabditis	MAIFI---DRKITANAYGLSSGIFWGLSGVLLALLELDTVLSSYLLIPVLAFLNDTISC	57
Photorhabdus.laumondii	MAIFI---NRKVITANAYGLSSGIFWGLSGVLLALLELDTVLSSYLLIPVLAFLNDTISC	57
Photorhabdus.bodei	MAIFT---NSKITANAYGLSSGIFWGLSGVLLALLELDTVLSSYLLIPVLAFLNDTISC	57
Photorhabdus.luminescens	MAIFT---NRKITANAYGLSSGIFWGLSGVLLALLELDTVLSSYLLIPVLAFLNDTISC	57
Photorhabdus.namnaonensis	MAIFT---NSKITANAYGLSSGIFWGLSGVLLALLELDTVLSSYLLIPVLAFLNDTISC	57
Histophilus.somni	MGLFSENKRLVHVGYFCGLLGLFWAISGISYELLYDFFPMLNAFALNIIILSIAEITLL	60
Haemophilus.felis	-----MGLGTIWGLLGLFWAISGISYELLYDFFPMLNAFALNIIILSIAEITLL	50
Pasteurella.skyensis	-----MVKYHYGIFGLLGLLWGVGTILYEVLNEDYFPVTSLVIIISLFLIESGSL	52
Pasteurella.multocida	-----MNKYKGYLAGIFGLLWGIAGIILYEILNTYPTTSLAIVLCLLFIESTAF	52
Rodentibacter.trehalosifermentans	-----MNKYKGYLAGIFGLLWGIAGIILYEILNTYPTTSLAIVLCLLFIESTAF	52
Pasteurella.oralis	-----MTTKYSGLWGLLGLLWGIAGIILYEILNTYPTTSLTIIISLFLVIEFSAF	53
Pasteurella.canis	-----MRQYAGYFWGLLGLLWGIAGIILYEILNTYPTTSLTIIISLFLVIEFSAF	52
Eikenella.corrodens	LNLSVIFS--FRSRKIPEL-KIKYAILPFFSFGVFCGAVGMAYLMSIQYIGIEYATLSS	105
Avibacterium.paragallinarum	IFLIKFIL-KTKNFLSLKN-YRKGIFPFGLAGVFGGTIGMFSYLMATDILLVNVAFLSS	108
Neisseria.bergeri	SAVWAGL----RSRRPG-FRGGVFRSGLVGLGGPVGMSAYLLGHLHYLTVYYAAPLSS	101
Neisseria.flavescens	SAVWAGL----SGGRPG-FRGGVFRSGLVGLGGPVGMSAYLLGHLHYLTVYYAAPLSS	101
Neisseria.polysaccharea	SAVWAGL----SGGRPG-FRGGVFRSGLVGLGGPVGMSAYLLGHLHYLTVYYAAPLSS	101
Neisseria.lactamica	SAVWAGL----SGGRPG-FRGGVFRSGLVGLGGPVGMSAYLLGHLHYLTVYYAAPLSS	101
Neisseria.subflava	SAVWAGL----SGGRPG-FRGGVFRSGLVGLGGPVGMSAYLLGHLHYLTVYYAAPLSS	101
Haemophilus.aegyptius	FVIGITLW----RKSVD-FQGVFWKPALSGLGGPIGMSAYLLSIHYLTVYYAAPLSS	101
Haemophilus.haemolyticus	LVIGITLW----RKSVD-FQGVFWKPALSGLGGPIGMSAYLLSIHYLTVYYAAPLSS	101
Streptococcus.pneumoniae	LVIGITLW----RKSVD-FQGVFWKPALSGLGGPIGMSAYLLSIHYLTVYYAAPLSS	101
Haemophilus.influenzae.	LVIGITLW----RKSVD-FQGVFWKPALSGLGGPIGMSAYLLSIHYLTVYYAAPLSS	101
Haemophilus.sp.C1	LVIGITLW----RKSVD-FQGVFWKPALSGLGGPIGMSAYLLSIHYLTVYYAAPLSS	101
Haemophilus.sp.HMSC066A11	-----	0
Haemophilus.sp.oraltaxon851	FAIGITLW----RRAVN-FQKIFWSPVLSGLVGLGMSAYLLSIHYLTVYYAAPLSS	101
Haemophilus.haemolyticus.M21639	FAIGITLW----RRAVN-FQKIFWSPVLSGLVGLGMSAYLLSIHYLTVYYAAPLSS	102
Haemophilus.quentini	FAIGITLW----RRAVN-FQKIFWSPVLSGLVGLGMSAYLLSIHYLTVYYAAPLSS	101
Xenorhabdus.szentirmaii	VYMFYLYKYYKTKRILSLKNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	115
Photorhabdus.asymbiotica	GYMFIYLVKGYQAIRLIPKNNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	117
Photorhabdus.heterorhabditis	GYMFIYLVKGYQAIRLIPKNNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	117
Photorhabdus.laumondii	GYMFIYLVKGYQAIRLIPKNNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	117
Photorhabdus.bodei	GYMFIYLVKGYQAIRLIPKNNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	117
Photorhabdus.luminescens	GYMFIYLVKGYQAIRLIPKNNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	117
Photorhabdus.namnaonensis	GYMFIYLVKGYQAIRLIPKNNKNIIVLAAIFGGPIGMSAYLLSIHYLTVYYAAPLSS	117
Histophilus.somni	LMAYGLI-KRKLPHF---NKVSLALIGGCLGGPVGMICYLQSIQVGVYAATIS	115
Haemophilus.felis	AIVGLTYR-TLNKSPVSVQYKKTLSLFLAGLGGPIGMLCYLQSIQVIGVGLSAPIS	109
Pasteurella.skyensis	LFAGLLCY-KVQPHKHLKRVGLAVIAGLGGPVGMFCYLQAIQYIGVYAAPIS	111
Pasteurella.multocida	IFCGISHY-KLKLSTIKFKN-KGTLGLVLAGIIGGPIGMLCYLQAIQYIGVYAAPIS	110
Rodentibacter.trehalosifermentans	IFFTISYI-KLNLFEKFRFN-KNIFLGLLAGIIGGPIGMLCYLQAIQYIGVYAAPIS	110
Pasteurella.oralis	IFCFVSHY-KLNLFEKCFKN-KRVFLGVAGVIGGPIGMLCYLQAIQYIGVYAAPIS	111
Pasteurella.canis	MFCVSHY-KLNLFEKQELKN-KSVYLLGILAGMIGGPIGMLCYLQAIQYIGVYAAPIS	110
Eikenella.corrodens	SYPAIAALLACIFLKDKNLSVGVAGFFIIVSSVMLILPSYQHNLDYFYFIVFCICISIG	165
Avibacterium.paragallinarum	TFPLFAGIIAYFLPEKIQIFIGFFITIFSSVILFISK-GINQFNILGLVFFFCIGL	167
Neisseria.bergeri	LFPVLAALSAWFLKERTGTAARAGPALAVLSSAGLAADAGRGGLNALGLVCAVCVSG	161
Neisseria.flavescens	LFPVLAALSAWFLKERTGTAARAGPALAVLSSAGLAADAGRGGLNALGLVCAVCVSG	161
Neisseria.polysaccharea	LFPVLAALSAWFLKERTGTAARAGPALAVLSSAGLAADAGRGGLNALGLVCAVCVSG	161
Neisseria.lactamica	LFPVLAALSAWFLKERTGTAARAGPALAVLSSAGLAADAGRGGLNALGLVCAVCVSG	161
Neisseria.subflava	LFPVLAALSAWFLKERTGTAARAGPALAVLSSAGLAADAGRGGLNALGLVCAVCVSG	161
Haemophilus.aegyptius	LFPVFAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Haemophilus.haemolyticus	LFPVFAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Streptococcus.pneumoniae	LFPVFAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Haemophilus.influenzae.	LFPVFAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Haemophilus.sp.C1	LFPVFAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Haemophilus.sp.HMSC066A11	-----MSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	52
Haemophilus.sp.oraltaxon851	LFPVLAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Haemophilus.haemolyticus.M21639	LFPVLAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	162
Haemophilus.quentini	LFPVLAALMSYWLKEKISKTAQFGFGLAVIASALLAIEVQKANFNTSGLIFLAICILG	161
Xenorhabdus.szentirmaii	LYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	175
Photorhabdus.asymbiotica	TYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	177
Photorhabdus.heterorhabditis	TYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	177
Photorhabdus.laumondii	TYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	177
Photorhabdus.bodei	TYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	177
Photorhabdus.luminescens	TYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	177
Photorhabdus.namnaonensis	TYPAIGALISFFLLKDRHILGILGLLILAVVCTMILGYSASTQTISGWIGFLFAFICALG	177
Histophilus.somni	CYPIIASLISKFLAVKASRRTLFSVFLVLLMIMYQGENSE-IRILGIIFAFSCALC	174
Haemophilus.felis	MYPIVGTILSVVFLNEKINRKMIFGLALVMSFLGFSSEYSK-NYIIGSLAVMCAFS	168
Pasteurella.skyensis	LYPVFALLSFLPKERMKTWGIIGFSIAI FCTLLSLDLSSSQ-ILLFVGLFAFISALS	170
Pasteurella.multocida	VYPVFGALLAVLFLKDSMTKTVIGLSLAI GCTILLGIDANNAQ-FSVLGIALAVVCAIS	169
Rodentibacter.trehalosifermentans	IYPVFGTIFAPFLKDSITKGLLGLCLSIGCTILLGLDITNSQ-FSILGFLAVTCAVS	169
Pasteurella.oralis	VYPVFGALLAVLFLKDSMTKTVIGLSLAI GCTILLGIDANNAQ-FSVLGIALAVVCAIS	170
Pasteurella.canis	IYPVFGTIFAPFLKDSITKGLLGLCLSIGCTILLGLDITNSQ-FSVLGIALAVTCAIS	169
Eikenella.corrodens	WSLEVVLSSYTMK-FIAAEVLYAIRVLAISGYFVIFFN---KIEIYSL--HLLTIRN	219
Avibacterium.paragallinarum	WALEVVLSSYAMK-FTSEYIYFYSRLCFSCLYIILFFI---LKVDSINI-IFNIDYQK	222
Neisseria.bergeri	WSLEVVLARTMR-TLSGLQVYCLRLCGSVCGYLLILVFL---SARGFSLNPSDFHDARI	217
Neisseria.flavescens	WSLEVVLARTMR-TLSGLQVYCLRLCGSVCGYLLILVFL---SARGFSLNPSDFHDARI	217
Neisseria.polysaccharea	WSLEVVLARTMR-TLSGLQVYCLRLCGSVCGYLLILVFL---SARGFSLNPSDFHDARI	217
Neisseria.lactamica	WSLEVVLARTMR-TLSGLQVYCLRLCGSVCGYLLILVFL---SARGFSLNPSDFHDARI	217

Neisseria.subflava	WSLEVLSARTMR-TLSGLQVYCLRRCGSSVCGYLLILVFL---SARGFSLNPSDFHDARI	217
Haemophilus.aegyptius	WSSEIVISSHTMR-SLSGLQVYFLRLCGSTLGYLLLLVL---FLQDFPVDLDFDSYPOI	217
Haemophilus.haemolyticus	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKDFSLDILSFNYIQI	217
Streptococcus.pneumoniae	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKNFSLDILSFNYVQI	217
Haemophilus.influenzae	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKNFSLDILSFNYVQI	217
Haemophilus.sp.C1	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKDFSLDILSFNYVQI	217
Haemophilus.sp.HMSC066A11	WSSEIVISSHTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKDFSLDILSFNYVQI	108
Haemophilus.sp.oraltaxon851	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKDFSLDILSFNYVQI	217
Haemophilus.haemolyticus.M21639	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKDFSLDILSFNYVQI	218
Haemophilus.quentini	WSSEIVISSYTMR-SLSGLQVYFLRLCGSTIGYLLILFVL---SLKDFSLDILSFNYVQI	217
Xenorhabdus.szentirmaii	WGSEVVISSYGMKNKIPSDIAYLIRQLSSSLGYFIIIVVF---IHSFFGVTHIFSDIKL	231
Photorhabdus.asymbiotica	WSSEVVISSYGMNKDIPADIAFYFIRQLSSSVGYLIIILF---VHSFFGVVHIFSNIEL	233
Photorhabdus.heterorhabditis	WSSEIVISSYGMNKDIPADIAFYFIRQLSSSVGYLIIILF---IHSFFGVVHIFSNIEL	233
Photorhabdus.laumondii	WSSEVVISSYGMNKDIPADIAFYFIRQLSSSVGYLIMILF---IHSFFGVTHIFSNIEL	233
Photorhabdus.bodei	WSSEVVISSYGMNKDIPADIAFYFIRQLSSSVGYLIMILF---IHSFFGVVHIFSNIEL	233
Photorhabdus.luminescens	WSSEVVISSYGMNKDIPADIAFYFIRQLSSSVGYLVIIILF---IHSFFGVHIFSNIEL	233
Photorhabdus.namnaonensis	WSSEVVISSYGMNKDIPADIAFYFIRQLSSSVGYLVIIILF---IHSFFSIVNIFSNTEL	233
Histophilus.somni	WALEIVISSYVMQ-HYPSDQVYLFQIGSSSLGYFLIILEMFFYGTYNLIDFNFFKQDNL	233
Haemophilus.felis	WGSEIVISSYIMR-KTPPSLAYLYRQLGSLVGLSLLILL---STETTKVIDLIDIRF	223
Pasteurella.skyensis	WGAEIVLSSYVMQ-FLPSTVVYFFRQLGASLGYLFLFFF---ELDTATFLKSFTEHSF	224
Pasteurella.multocida	WGSEIVISSYVMQ-SIPSSVAYFLRQLGSSIGYLLILLSF---DVAFLDPLPSFKNTHF	225
Rodentibacter.trehalosifermentans	WGSEIISSYIMQ-FISSSSAYFLRQLGSSLGYVTLILYL---DSEYFYSTEFFKDTKL	225
Pasteurella.oralis	WGGEIVISSYVMQ-SINSSIAYFFRQLGSSIGYFILLCSF---DVKFFGLIEFFKDKQF	224
Pasteurella.canis	WGGEIVISSYVMQ-SIKSSSAYFLRQLGSSPGYIILVLYL---DVEFFNSIEFFKDRF	224
	*. *::*: * . * * : * * :	:
Eikenella.corrodens	SSIILFILASGFLSYFYKYKISGLVGPRAITMNIYTIWTVIMSFLDIDVNLMTILC	279
Avibacterium.paragallinarum	IGCFPIIVLSTFLSYLLYYKAIYLLGTIKAMSINIYTSIWTVLFYSYFLDIEISRKLIFC	282
Neisseria.bergeri	IGI----IAFGALSYACYQAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIMW	273
Neisseria.flavescens	IGI----IAFGALSYACYQAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	273
Neisseria.polysaccharea	IGI----IAFGALSYACYQAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIMW	273
Neisseria.lactamica	IGI----IAFGALSYACYQAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	273
Neisseria.subflava	IGI----IAFGALSYACYQAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIMW	273
Haemophilus.aegyptius	SGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSIWAIGLGYLTYQQPIQPTSLIL	273
Haemophilus.haemolyticus	SGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	273
Streptococcus.pneumoniae	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	273
Haemophilus.influenzae	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	273
Haemophilus.sp.C1	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	273
Haemophilus.sp.HMSC066A11	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	164
Haemophilus.sp.oraltaxon851	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	273
Haemophilus.haemolyticus.M21639	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	274
Haemophilus.quentini	AGV----ITFGALSYCCYQAIYLLKPKKAMALNIYTSVWAIGLGYLTYKQPIKPIITLLL	273
Xenorhabdus.szentirmaii	TTLLLTSLIATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISINLLFI	291
Photorhabdus.asymbiotica	YSLFFFTSLVATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Photorhabdus.heterorhabditis	YSLFFFTSLVATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Photorhabdus.laumondii	YSLFFFTSLVATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Photorhabdus.bodei	YSLFFFTSLVATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Photorhabdus.luminescens	YSLFFFTSLVATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Photorhabdus.namnaonensis	YSLFFFTSLVATLSYLFYKAIERLRPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Histophilus.somni	WKLEWIVIISSGFSYFLYKSIHILNPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	293
Haemophilus.felis	DIWFVSIITIFSMFSYLYYKSIHILNPIRAMALNIYTSVWAVIFAYLIMDEPISIKLILL	283
Pasteurella.skyensis	VFLGLIFSSMLSYLLYYQAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	284
Pasteurella.multocida	LTLIIIFIIFSSMMSYFLYRATYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	285
Rodentibacter.trehalosifermentans	LTIIFIIFSSMMSYFLYRATYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	284
Pasteurella.oralis	SLIIFLIIVSSMTSYFLYKAIYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	285
Pasteurella.canis	LTIIFIIFSSMMSYFLYRATYLLKPKKAMALNIYTPVWAMVLGFFVYRHPVPSAIIW	284
	** *::*: * : :::: ***** * : : . :	:
Eikenella.corrodens	CLGMFIGTCTITLSSRKVPL-----	298
Avibacterium.paragallinarum	CIGLFGTGLTYVYKKE-----	299
Neisseria.bergeri	AVLLGTGAALSPLYGREGEI-----	292
Neisseria.flavescens	AVLLGTGAALSPLYGREGEI-----	292
Neisseria.polysaccharea	AVLLGTGAALSPLYGREGEI-----	292
Neisseria.lactamica	AVLLGTGAALSPLYGREGEI-----	292
Neisseria.subflava	AVLLGTGAALSPLYGREGEI-----	292
Haemophilus.aegyptius	ALLLSAGVIMTYLYKGEQK-----	292
Haemophilus.haemolyticus	TLLLSAGVIMTYLYKGEQK-----	292
Streptococcus.pneumoniae	TLLLSAGVIMTYLYKGEQK-----	292
Haemophilus.influenzae	TLLLSAGVIMTYLYKGEQK-----	292
Haemophilus.sp.C1	TLLLSAGVIMTYLYKGEQK-----	292
Haemophilus.sp.HMSC066A11	TLLLSAGVIMTYLYKGEQK-----	183
Haemophilus.sp.oraltaxon851	TLLLSAGVIMTYLYKGEQK-----	292
Haemophilus.haemolyticus.M21639	TLLLSAGVIMTYLYKGEQK-----	293
Haemophilus.quentini	TLLLSAGVIMTYLYKGEQK-----	292
Xenorhabdus.szentirmaii	CTGIVSGSFLTAVNPNINIKKLTYPFIK	317
Photorhabdus.asymbiotica	CIGVSTGSFLTAVNPNINIKKLTYPFIK	319
Photorhabdus.heterorhabditis	CVGVSTGSFLTAVNPNINIKKLTYPFIK	319
Photorhabdus.laumondii	CVGVSTGSFLTAVNPNINIKKLTYPFIK	319
Photorhabdus.bodei	CVGVSTGSFLTAVNPNINIKKLTYPFIK	319
Photorhabdus.luminescens	CVGVSTGSFLTAVNPNINIKKLTYPFIK	319
Photorhabdus.namnaonensis	CVGVSTGSFLTAVNPNINIKKLTYPFIK	319
Histophilus.somni	GIIVLMNFI--IVAKGEEK-----	310
Haemophilus.felis	ASLVFIGAYITLVNKKKEE-----	302
Pasteurella.skyensis	TIGVVVGTITLTKDKREKL-----	304
Pasteurella.multocida	VLGVILGAGLTLKHEKREKL-----	303
Rodentibacter.trehalosifermentans	VGVVILGAGLTLKDKREKL-----	303
Pasteurella.oralis	VGVVILGAGLTLKDKREKL-----	304
Pasteurella.canis	VGVVILGAGLTLKDKREKL-----	303
	:	:

Figure 5.3. Multiple sequence alignment of LicB from *H. influenzae* against sequences selected from BLAST search.

Identification of each sequence is listed in Table 5.2



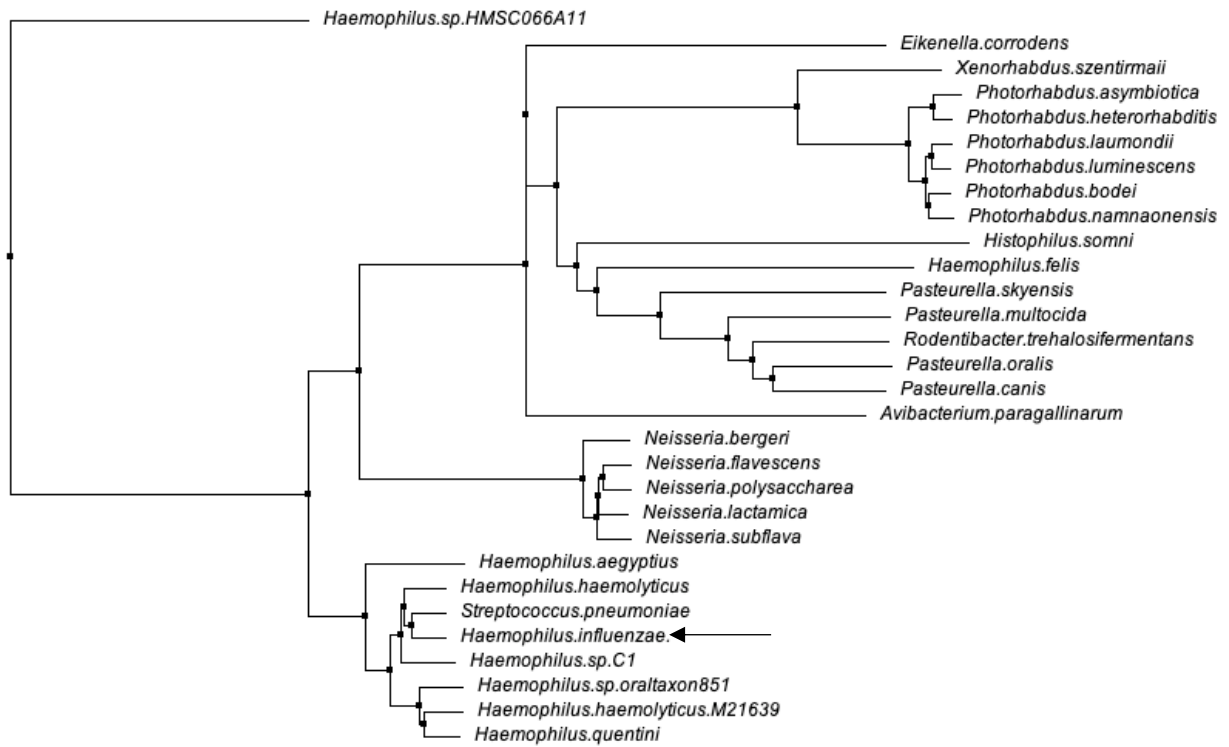


Figure 5.5. Phylogenetic tree based on the multiple sequence alignment.

LicB highlighted with an arrow corresponds to the protein from *H. influenzae* studied in this project. Figure generated using Jalview software.

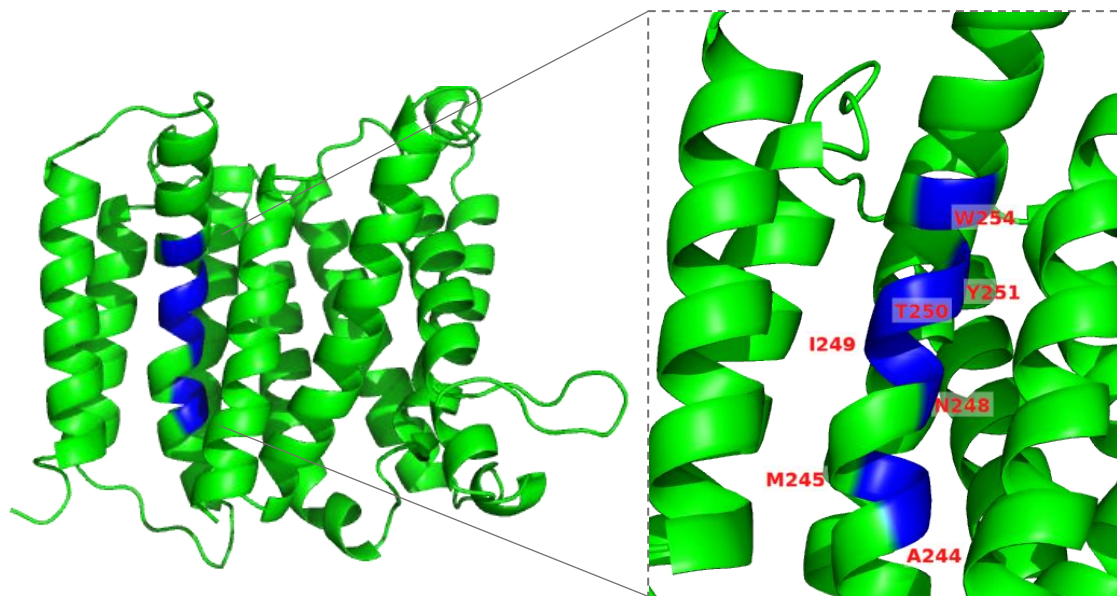


Figure 5.6. LicB homology model generated in Chapter 3.

Conserved residues A244, M245, N248, I249, T250, Y251 and W254 are highlighted in blue. Figure produced using Pymol.

5.2.2. Expression and purification of LicB variants

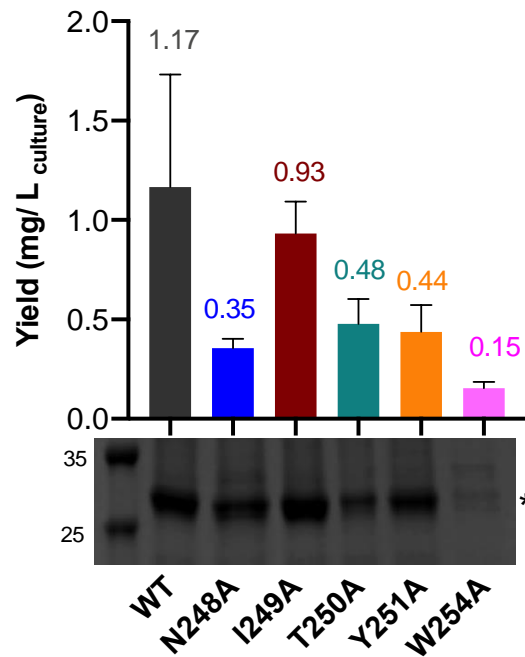
The specific site-directed mutations to the conserved motif in TM9 were introduced into wild-type LicB, which were confirmed by sequencing. Mutations were identified with conventional nomenclature, where N248A means that Asp at sequence position 248 was mutated to Ala.

Despite extensive efforts, it was not possible to produce mutation M245A for unknown reasons. All other variants were expressed from pET28 in *E. coli* T7 Express cells as described in Chapter 3. The purification of each variant followed the same protocol as for the WT protein (SB_{AIM} media growth for 24 hour at 37°C) and purified using 1% DDM. The expression level of each variant was, in general, lower than the WT (1.17 ± 0.57 mg/L culture), with the best-expressed being the I249A variant (0.93 ± 0.16 mg/L culture). The expression variant W254A was barely detectable, with a purification yield of only 0.15 ± 0.032 mg/L culture. This variant was not pursued further. Otherwise, SDS-PAGE confirms the successful expression and purification of each variant, with a band at the expected molecular size of around 34 kDa in each case (Figure 5.7A and B). Bands on the gel slightly shifted during SDS-PAGE, probably due to anomalous SDS binding³⁸⁵.

Each variant was analysed with size exclusion chromatography to determine the stability and aggregation propensity of each of these four homologues in DDM micelles. Variants N248A, I249A and Y251A, showed a similar SEC profile to WT, eluting as a homogeneous single peak at ~13 mL (Figure 5.8). This meant that protein recovery after SEC was consistent with the recovery of the WT. In contrast, T250A showed substantial heterogeneity, peak broadening and aggregation (green chromatogram in Figure 5.8).

The LicB variants purified from SEC were subjected to secondary structure analysis as well as competition assay through FRET.

(A)



(B)

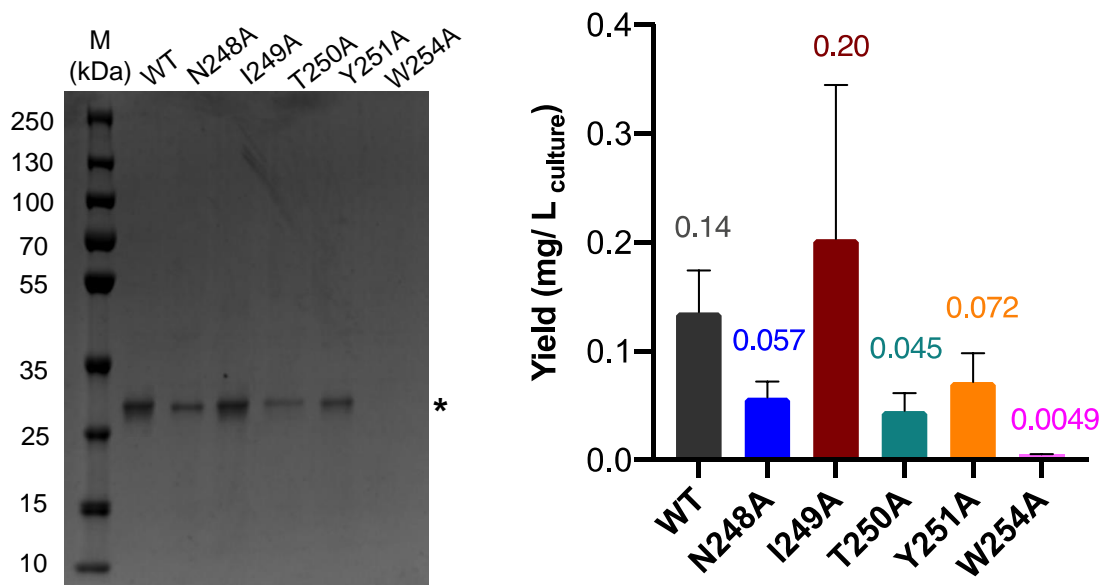


Figure 5.7. Purification of LicB variants.

Protein purification yields and representative SDS-PAGE gel of purified proteins from **(A)** IMAC eluate and **(B)** from SEC eluate. Values are mean \pm standard deviation from at least three individual experiments. Bands in SDS-PAGE gels corresponding to purified LicBs are marked (*). Each lane contains in **(A)** 5 μ g or in **(B)** 1 μ g of purified protein.

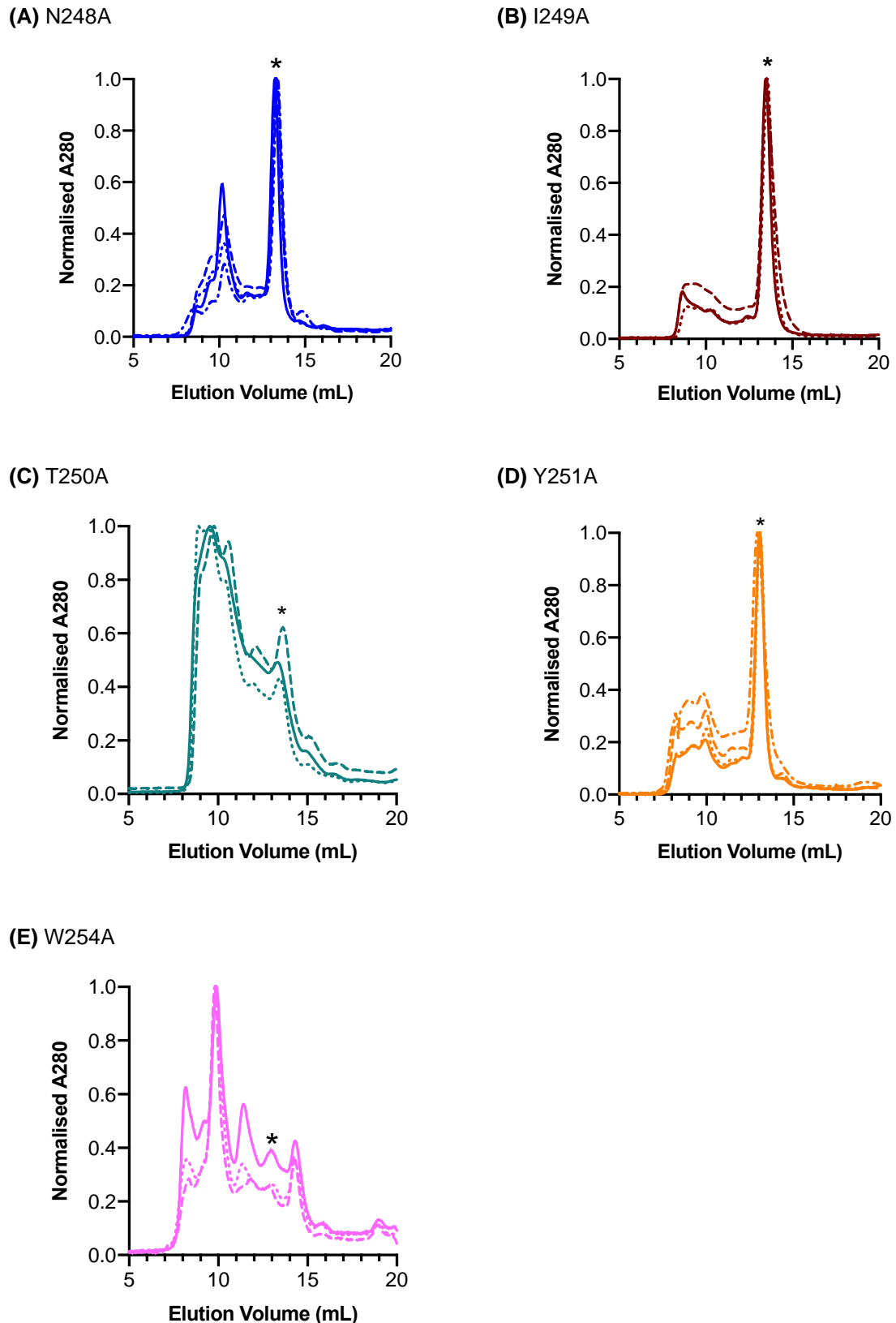


Figure 5.8. Size exclusion profiles of LicB variants solubilised in 1% DDM.

N248A, I249A and Y251A were reasonably similar to WT (the 13 mL fraction remains a single peak (*)), while T250A showed increased heterogeneity and non-specific aggregation. Each line corresponds to an individual SEC run.

5.2.3. Structural characterisation of LicB variants

Circular dichroism was used to evaluate if LicB variants retained the overall helicity of the WT protein, as well as the sensitivity of these variants to temperature.

The circular dichroic spectrum of N248A, T250A and Y251A did not show strong negative deflections at 222 and 208 nm (Figure 5.9B, D and E, respectively), implying a substantially lower percentage of α -helix content in these variants (Figure 5.11A). In the case of T250A, this is consistent with the variant disrupting protein structure sufficiently to cause aggregation in SEC (Figure 5.8). The CD spectrum of I249A was the most similar to the WT protein (Figure 5.9C), albeit with a slight reduction in apparent helicity to 49.5% (WT of 62.3%). It seems likely that such minor differences in helicity values are due, at least in part, to difficulties in accurately determining the concentration of the variant proteins because of their low yield.

The variants N248A, I249A and Y251A, were similar to WT in showing a cooperative unfolding transition in thermal melts (Figure 5.10B, C and E, respectively). I249A and Y251A showed a significant transition with T_m ~66°C (as WT), while N248A had a slightly higher T_m of around 73°C (Figure 5.11B). In contrast, T250A showed a distinctive non-cooperative melting curve (Figure 5.10D). It is generally considered that cooperative melting is a hallmark of well-folded, compact proteins where the thermal unwinding of secondary structure is correlated to the simultaneous loss of tertiary structure. The melting curve of T250A is thus again consistent with substantial disruption to protein structure in this variant.

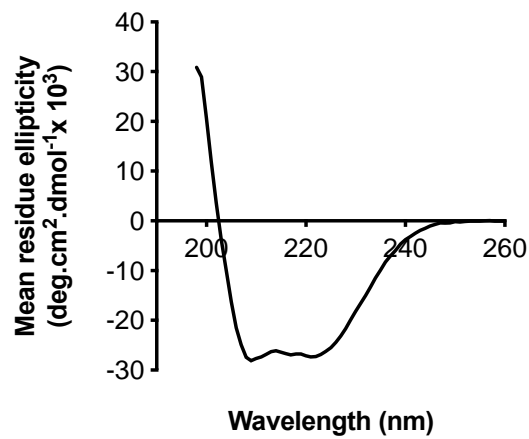
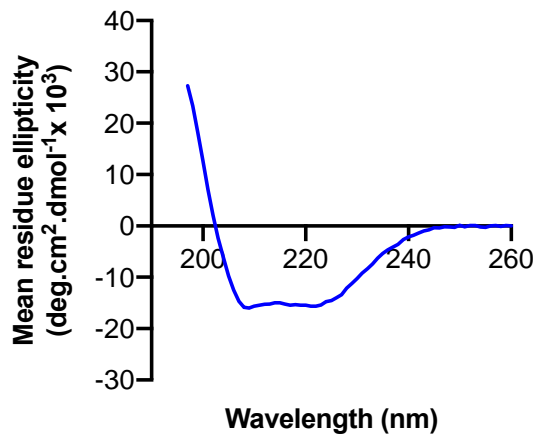
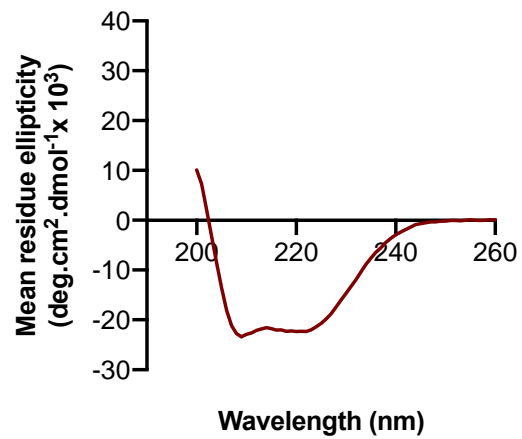
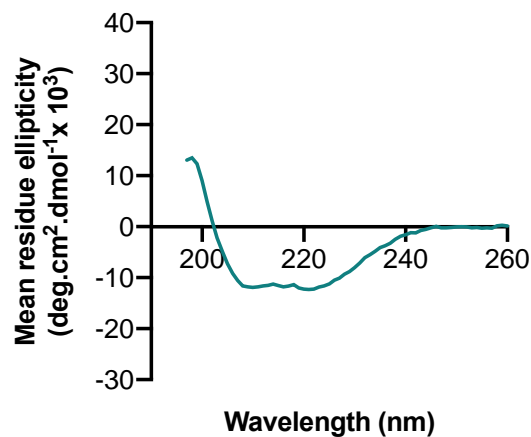
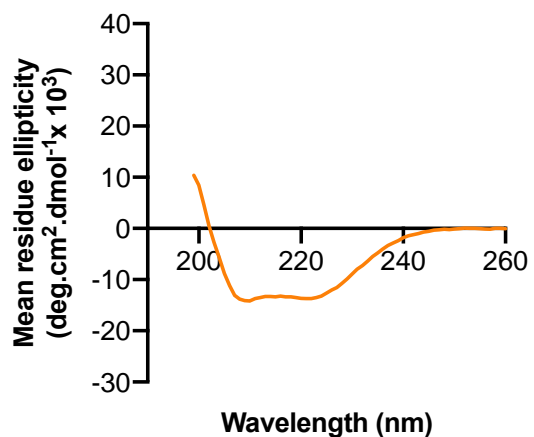
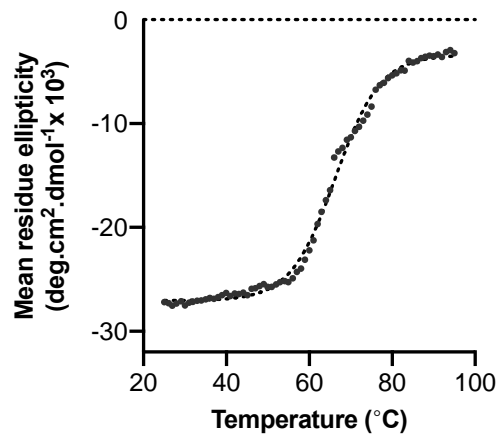
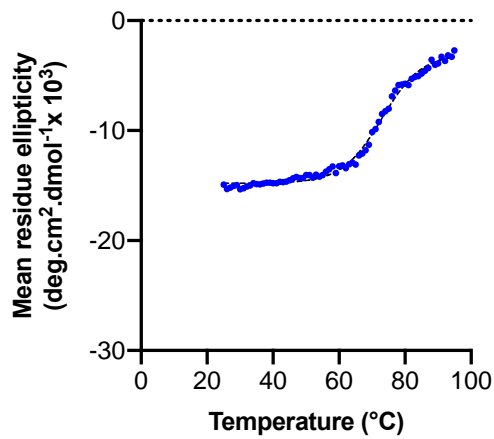
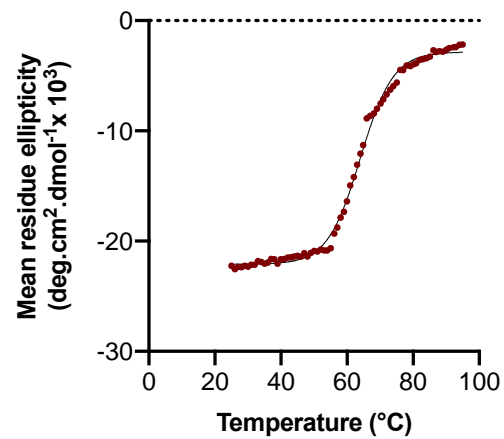
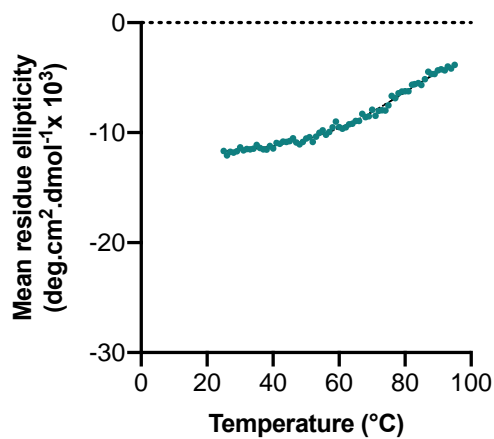
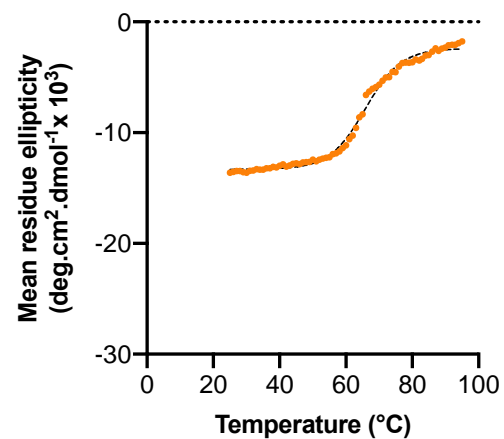
(A) LicB WT**(B)** N248A**(C)** I249A**(D)** T250A**(E)** Y251A

Figure 5.9. Circular dichroism spectroscopy of LicB variants in 1% DDM.

CD spectra of **(A)** WT LicB compared to variants as shown in panels **(B)** to **(E)**. All LicB variants showed a reduced ellipticity in comparison with WT. Data for values of HT above 700 V were excluded.

(A) LicB WT**(B)** N248A**(C)** I249A**(D)** T250A**(E)** Y251A**Figure 5.10. Thermal melts of LicB variants in 1% DDM.**

Thermal stability of **(A)** WT LicB and **(B-E)** LicB variants as shown. Data were collected from 25°C to 95°C with 1°C increment. Data were fitted to a Boltzmann sigmoidal function to guide the eye.

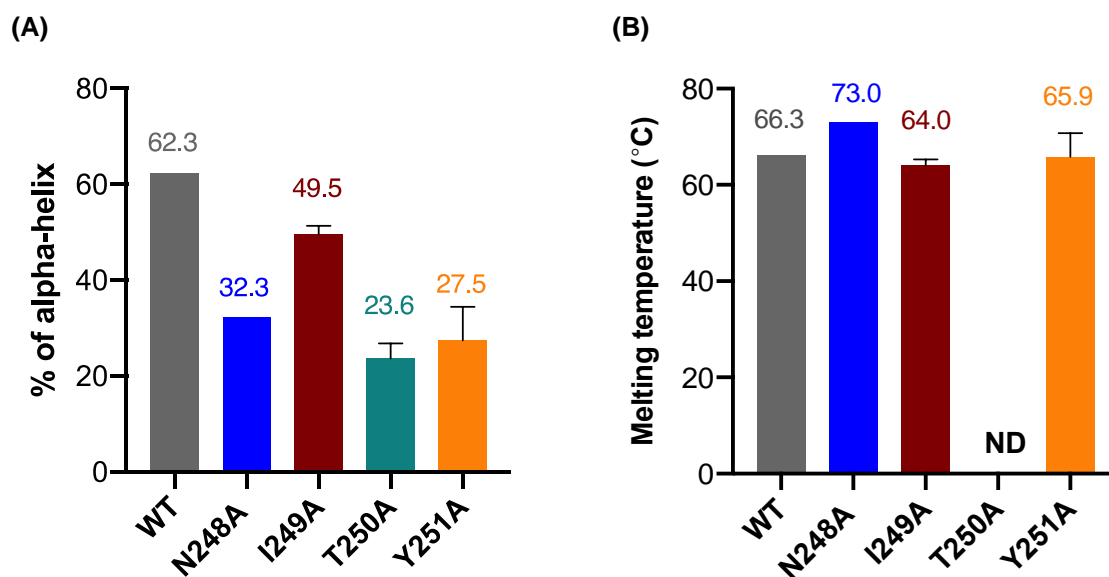


Figure 5.11. Summary of biophysical properties of LicB variants.

(A) Percentage of α -helix of each LicB variant determined from Equation 2.9. **(B)** Melting temperature (T_m) of each LicB variant. Values are mean \pm standard deviation from duplicates where possible.

5.2.4. Competition assay

To further investigate the implication of each substitution on the ligand binding, a competition assay was performed as described in Chapter 4 using the fluorescence ligand dansyl choline (DC). As before, DC was introduced at 1.5 times the apparent K_d value of WT LicB before being competed off with choline. The FRET signal was followed to determine the displacement of DC from the protein. The results are shown in Figure 5.12A to D and summarised in Figure 5.12E.

Experiments with N248A, I249A and Y251A all showed loss of FRET signal consistent with choline binding, albeit with weaker affinity than that observed for WT (Figure 5.12E). Even in the best case, for I249A, the IC_{50} value was around four times higher than WT; in the worst case (N248A), the IC_{50} was increased by orders of magnitude. Since variant T250A appears to be structurally compromised, it was predicted that no binding would be seen in this variant, and indeed, that was the case (Figure 5.12C). T250A thus effectively serves as a further negative control for this experimental system, increasing confidence that the DC assay reports an authentic binding event.

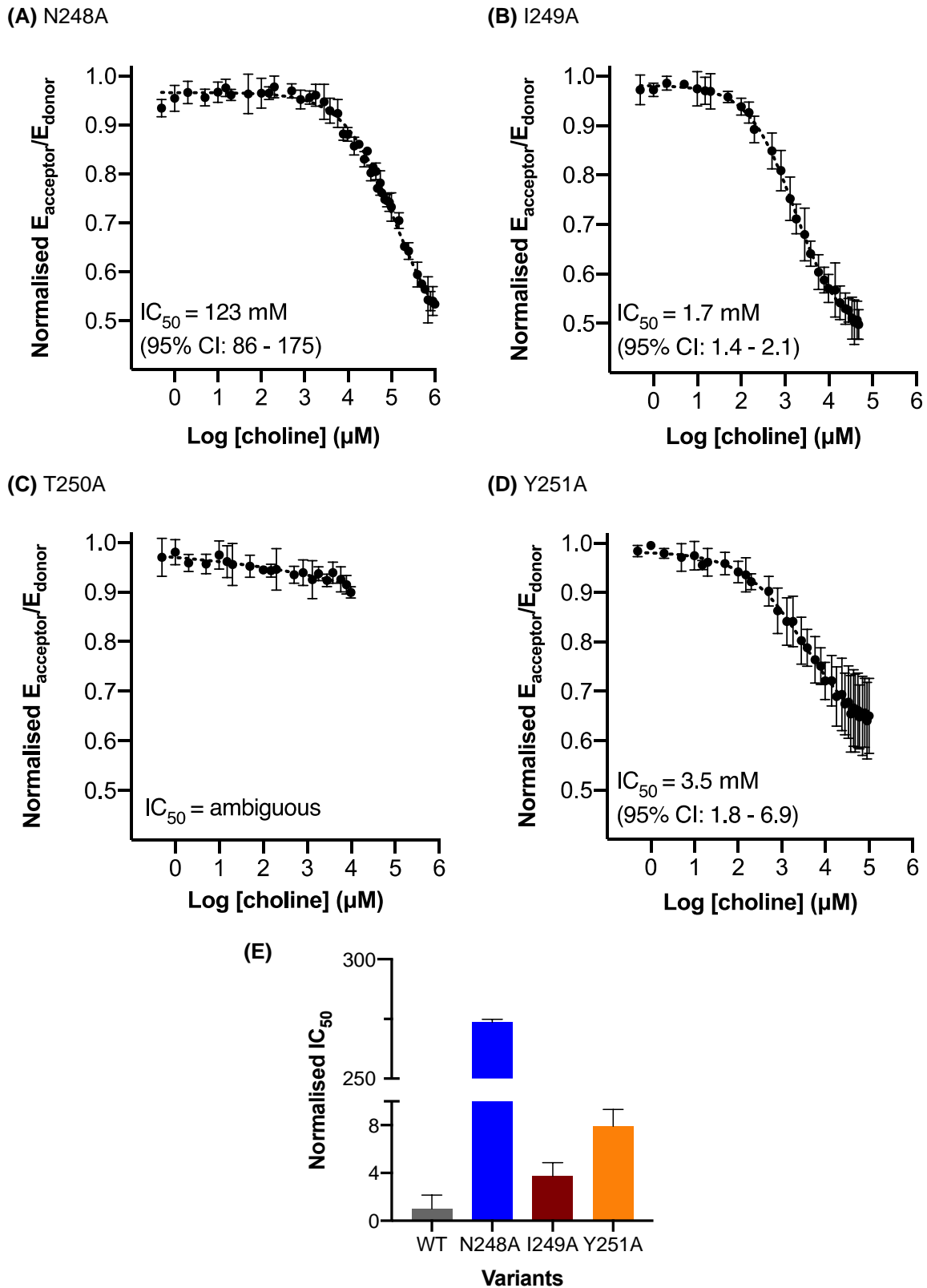


Figure 5.12. Competition assays of each LicB variant.

(A-D) Competition of choline against dansyl choline (DC) for the different variants as shown. The concentration of DC was 1.5 times the K_i . Data were collected in triplicate, and mean \pm standard deviation is shown. Data were fitted to a four-component inhibition equation (lines), and standard errors of IC_{50} values were calculated using GraphPad Prism. **(E)** Normalised IC_{50} values of WT LicB and variants as shown. Data are mean \pm standard deviation from three independent experiments.

5.3. Discussion

Multiple sequence alignment of LicB homologues revealed a highly conserved motif (AM*xxNITYxxW) within transmembrane helix nine. The online tool ConSurf server confirmed the presence of highly conserved amino acid residues (-NITY-) within the motif (Figure 5.13). Five of these conserved residues were successfully mutated to Alanine in turn. This chapter has described the recombinant expression, structural integrity, and ligand binding function of these variants.

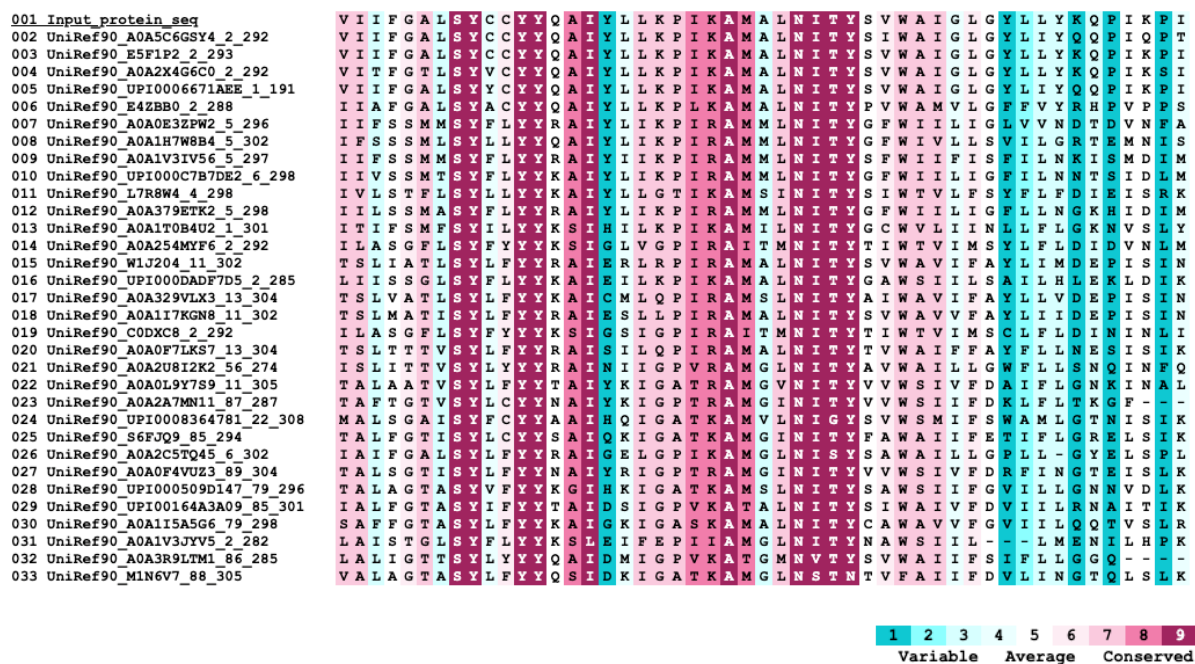


Figure 5.13. Part of the multiple sequence alignment between LicB homologues coloured by a degree of conservation.

-NITY- motif was recognised as a highly conservative motif - alignment and conservation scores (1 to 9) calculated using the online tool ConSurf server⁴³⁷.

All of the variants were purified at lower yields than WT LicB. Some of the variants (W254A, T250A) tended to aggregate. T250A had much lower helicity than WT and did not show a cooperative transition in thermal unfolding experiments. The remaining three variants (I249A, N248A, Y251A) were more similar to WT. These all displayed similar SEC profiles to WT and cooperatively unfolded with the same T_m . However, all three variants showed a substantial reduction in ligand binding affinity. Collectively, these data support the importance of these conserved residues in either or both of LicB structure and function.

Regarding the impact of these substitutions on circular dichroism, all variants showed a reduced helicity compared to WT. This was somewhat surprising since a single mutation typically does not disrupt secondary structure to the degree shown here. This is particularly true of integral membrane proteins because this loss of helicity would correspond to an unwinding of transmembrane helices within the hydrophobic micelle. This is energetically highly unfavourable since helix unwinding exposes main-chain hydrogen bonding groups, which cannot be satisfied in the micelle interior. However, W254A and T250A clearly do cause significant disruption in protein structure. This goes against the initial assumption that W254 could play a role in substrate binding, analogous to the Trp cage found in other choline transporters. However, it supports the hypothesis that this sequence conservation of these residues reflects their importance in LicB. The lower helicity observed in other variants should be interpreted cautiously. Accurate helicity can only be known if protein concentration is known very accurately. This becomes increasingly difficult with the low protein yields obtained when these variants are purified because background scattering at shorter wavelengths interferes with determining A280. It thus seems reasonable to suppose that I249A has a relatively similar structure to WT. This is consistent with the similar melting behaviour of this variant. Variants N248A and Y251A seem to have genuinely reduced helicity yet still display the same melting trends as the WT protein. Although this will clearly require further characterisation, these data again confirm the likely importance of these latter residues in LicB.

Additionally, based on the homology model of LicB, it is possible to visualise possible polar interactions between mutated residues and other molecules. A possible interaction is observed in the asparagine residue (N²⁴⁸) that could form hydrogen bonds with the hydroxyl group in the 'tail' of the choline ligand (Figure 5.14). This interaction was also observed in ChoX from *Sinorhizobium meliloti*, which is part of the choline-inducible transport system ChoVWX³⁵⁹.

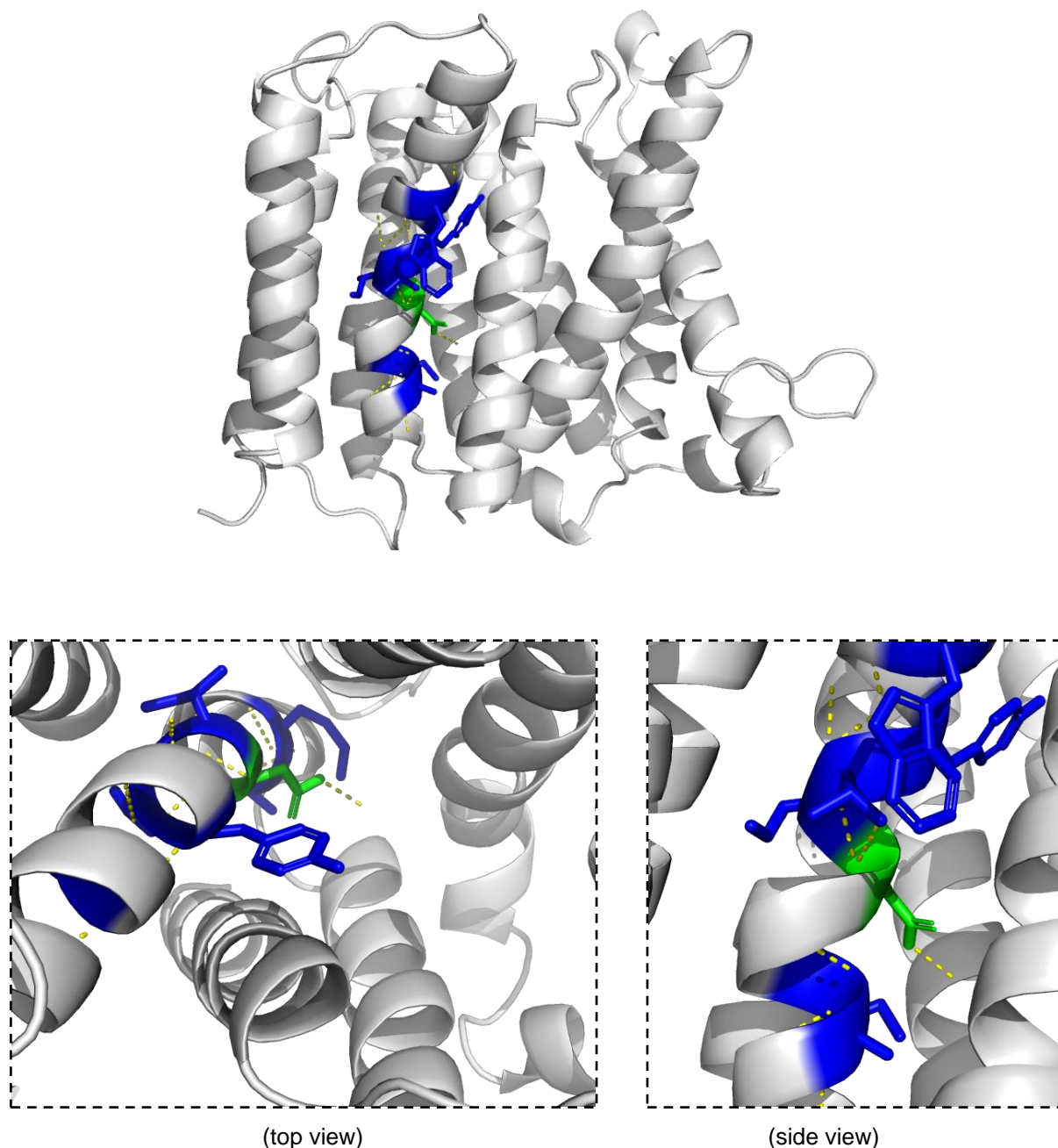


Figure 5.14. Illustration of the possible polar interactions performed by mutated residues.

Mutated residues are highlighted in blue, while asparagine (N²⁴⁸) is highlighted in green. Possible polar interactions of mutated residues are in yellow.

Four of these variants were taken through ligand binding studies using the FRET-based competition assay. As expected, the fluorescence signal of T250A did not change. This is entirely consistent with the structural characterisation of this variant, suggesting that the protein fold is severely compromised.

Competition assays of N248A, I249 and Y251 showed that higher choline concentrations were needed to display dansyl choline compared to WT, leading to higher IC_{50} values. It seems reasonable to conclude that all of these residues contribute somehow to the formation of the ligand binding site. Binding was moderately reduced (about 4-fold) in I249A but was notably severely compromised in N248A. It seems plausible that the asparagine side-chain might be involved in binding choline; this conclusion is tempered by the reduction in the helicity of this variant, with mutation N248A having an impact on the structural integrity of LicB. Although it was possible to observe a reduction in the helicity values of all variants, the possibility of an incorrect measurement of the sample concentration should not be disregarded. Thus, the helicity values of the variants are not 'real', and the protein structure might be intact; since melting curves showed a cooperative melting behavior usually associated with well-folded proteins (except the T250A variant).

Further work should use more conservative mutations (N248T, N248S) to try and dissect the functional role of this residue without disrupting the structure. N248D is isosteric but should be avoided because of the negative charge on aspartic acid. Likewise, it would be worth trialling conservative mutations at other positions such as I249V and Y251F.

5.4. Conclusion

A highly conserved motif was identified in LicB and the importance of these conserved residues explored by mutagenesis. LicB was sensitive to these mutations with noticeable changes in both structure and function, consistent with the potential importance of these conserved residues. In particular, the fluorescence competition assay revealed IC_{50} of N248A of around 274-fold higher than WT, leading to the conclusion that this polar residue might have a critical role in forming the ligand binding pocket.

Chapter 6. Summary and future work

LicB is a membrane protein that plays an essential role in bacterial colonisation of the human respiratory tract. However, we know almost nothing about the form and function of this protein. Biophysical and biochemical techniques were used to characterise LicB for the first time to bridge this knowledge gap. The main topics described in this thesis concern the overexpression, purification, biophysical and functional characterisation of the choline transport protein, LicB, of *H. influenzae*.

Chapter 3 reports the first extensive bioinformatic analyses of the LicB sequence and documents attempts to recombinantly express and purify LicB. This expression was successfully accomplished in *E. coli* T7 Express cells, though with the issues of low yield and aggregation propensity that are known to be common problems with membrane proteins. The monodispersity and secondary structure content of purified LicB in the detergent DDM provided a suitable platform for further biophysical characterisation pursued in the following chapters. Although DDM has been recognised as the most used detergent for either solubilisation or crystallisation of α -helical membrane proteins^{134,135}, future work should prioritise the screening of other solubilising chemicals. Moreover, LicB's yield could be improved by the screen of different expression systems and induction conditions.

Chapter 4 describes a series of methods that attempt to determine the substrate specificity, selectivity and transport behaviour of LicB. A novel fluorescence assay for ligand binding was developed and, while still needing confirmation by other methods, it revealed that LicB could discriminate in binding choline and acetylcholine instead of related small molecules such as glycine, betaine and carnitine. It seems plausible that this assay could ultimately be the basis for drug development. We will exploit this new understanding to develop novel research tools that will underpin a future drug screening programme. Regarding the development of a transport assay, this would be considered in further work. A radioactive uptake assay of choline by LicB could be helpful in understanding the transport mechanism, which has been performed *in vivo* by Fan *et al.*³⁶¹.

Homologues of *H. influenzae* LicB have been found in pathogens of the human respiratory tract, such as in *S. pneumoniae* or *Neisseria* species. In Chapter 5, a highly conserved motif was identified in LicB and the importance of these conserved residues explored by mutagenesis. LicB was sensitive to these mutations with apparent changes in both structure

and function, consistent with the potential importance of these conserved residues. In particular, the fluorescence competition assay revealed IC_{50} of N248A of around 274-fold higher than WT, leading to the conclusion that this polar residue might have a critical role in forming the ligand binding pocket. Additionally, further efforts can be made to study LicB homologues, including their expression and biophysical characterisation. These studies may enable us to a better understanding of the transport mechanism.

To conclude, the work described in this project provides a step towards a more detailed understanding of the molecular mechanism of LicB protein. However, it remains significant outstanding questions regarding the transport mechanism, structure and function.

PART II - The study of acyl-CoA:alcohol O-acyltransferases from yeast and fruit

Chapter 7. Literature review

Enzymes from the broad family of acyl-CoA:alcohol O-acyltransferases (AATases; EC 2.3.1.84) are able to synthesise volatile ester products from acyl-CoA and alcohol. Volatile esters are essential secondary metabolites produced naturally by yeast during fermentation and in flowers and fruits during fruit ripening^{438–440}. The activity of these enzymes is responsible for many of the flavour-active compounds found in fruits and industrial fermented alcoholic drinks, such as beer, wine and sake, which are of both commercial and cultural significance^{438–441}. Therefore, understanding their molecular biochemistry is of broad interest since they can impact fruit maturation and flavour⁴⁴² and generate renewable biofuels, fine chemicals or fragrances by metabolic engineering^{443–445}. This chapter will present the first attempt to comprehensively characterise the AATases from brewer's yeast, *Saccharomyces cerevisiae*, and commercial fruits.

7.1. Volatile esters in the aroma of food and drink

Aroma is an essential component in the taste of food and drinks⁴⁴⁶. It often arises from a subtle mixture of many different compounds and is regarded as perhaps the most complicated quality sensation associated with flavour. Multiple interdependent factors can determine the impact of an individual aroma compound, such as the pH, ethanol level, salt concentration, association to other compounds (fats/oils, proteins or starch) and temperature⁴⁴⁷. The main six classes of fruity and floral aroma compounds are volatile esters, higher alcohols, polyfunctional thiols, lactones, furanones and terpenoids⁴⁴⁸. The first group, volatile esters, are relevant to this thesis and will be discussed in more detail in this review. The chemical structures of the most aroma-active esters in food and drink are highlighted in Figure 7.1.

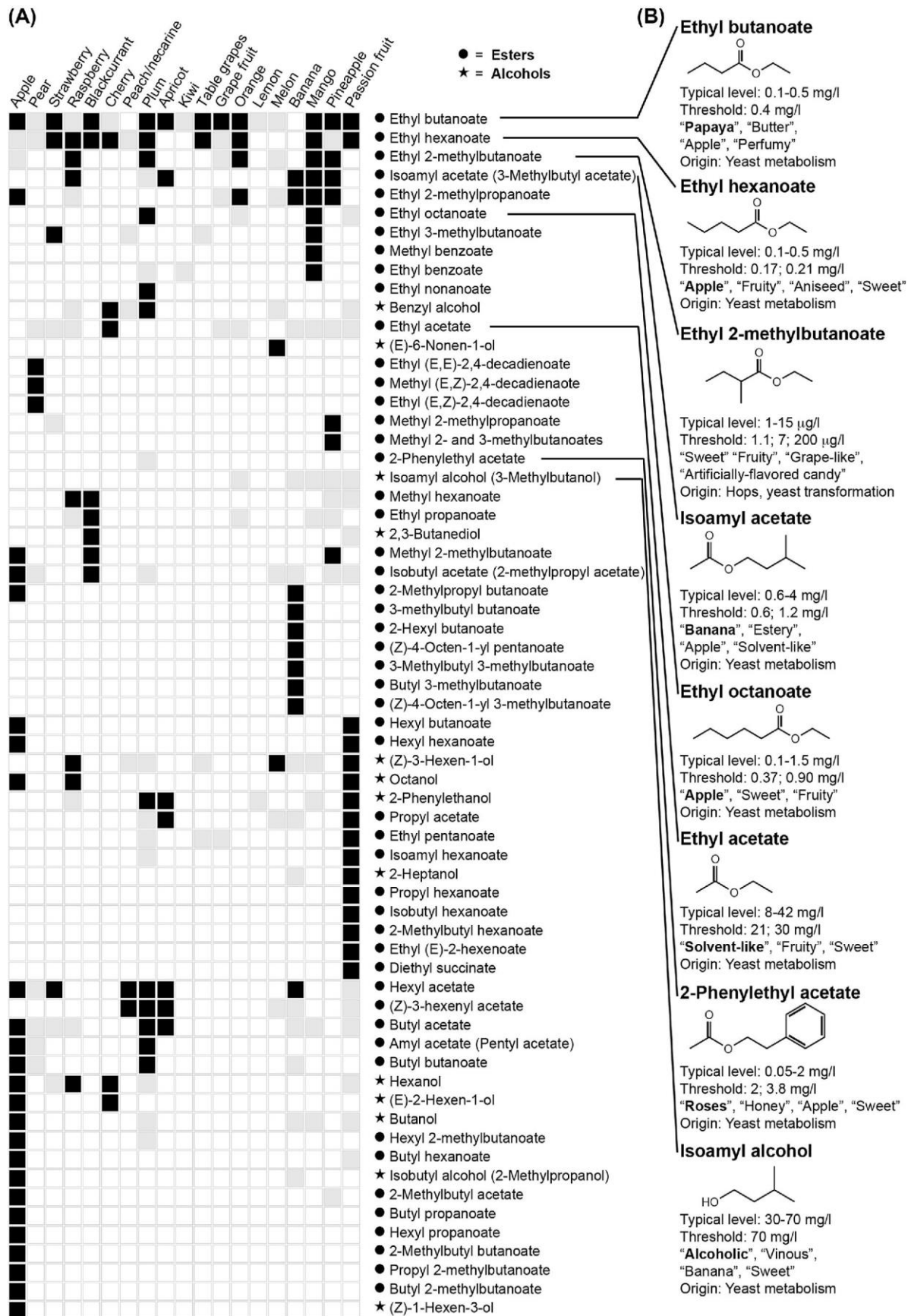


Figure 7.1. List of the main esters and alcohol compounds in (A) fruits and (B) beers.

Adapted from Holt *et al.* (2019) ⁴⁴⁸.

The brewing industry - worth over £150 bn in Europe alone - is making a continuous effort to develop new beers by increasing or diversifying the profile of fruity and floral aromas. In yeast-derived beverages, major contributions to flavour come from the volatile esters isoamyl acetate (banana), ethyl hexanoate (apple, aniseed) and ethyl octanoate (fruity, apple) ^{449,450}. The complex flavour profiles of commercial fruits such as apple, banana, strawberry and cantaloupe melon are also influenced by both straight-chain and branched-chain volatile esters (Figure 7.2) ⁴⁵¹. As an example of this, 86.4 million metric tons of apples were harvested in 2018 ⁴⁵², and more than 300 volatile esters have been identified as part of apple flavour, with dominant notes from the ethyl, butyl and hexyl acetates ⁴⁵¹. The diversity and quantity of esters produced varies between apple cultivars, for instance the ester-profile of Granny Smith is different from Fuji, giving rise to completely different aroma profiles ⁴⁵³.

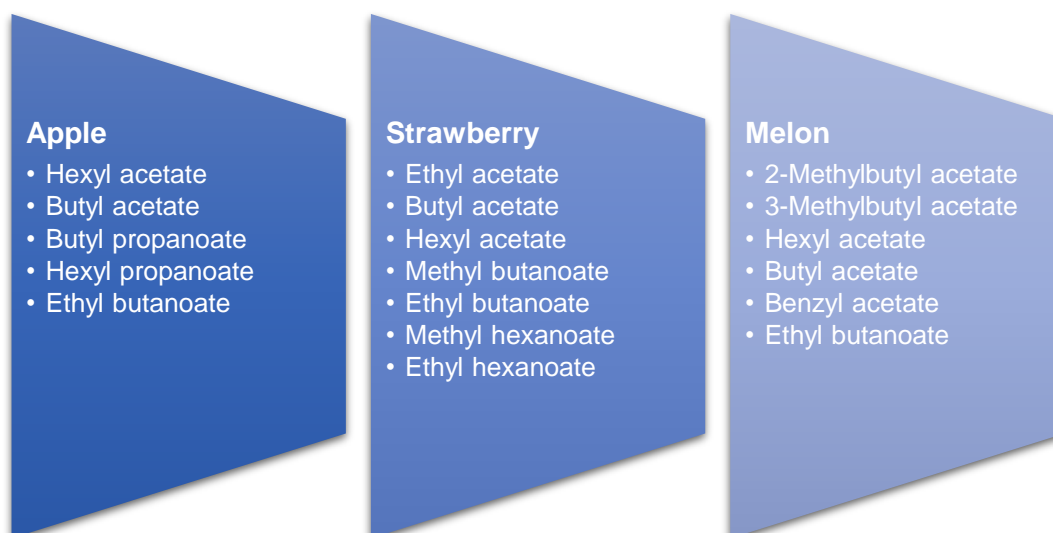


Figure 7.2. List of some esters produced by apple, strawberry and melon fruits.

Adapted from Song and Forney (2008) ⁴⁵¹.

This short review will introduce the enzymes involved in the production of esters in the yeast *S. cerevisiae* and in commercial fruits, their regulatory pathways, and prior *in vitro*, *in vivo* and *in silico* studies. Yeast AATs will be firstly discussed, followed by plant AATases, particularly the best-studied examples from melon, strawberry and papaya.

7.2. General considerations on AATases

The AATases are members of a broad and functionally varied family of enzymes known as CoA-dependent acyltransferases^{439,454}. This family includes well-studied archetypes, for instance, the chloramphenicol acetyltransferase (EC 2.3.1.28)⁴⁵⁵, the carnitine acetyltransferase (EC 2.3.1.7)⁴⁵⁶ and the choline acetyltransferase (EC 2.3.1.6)⁴⁵⁷.

The AATases are bisubstrate enzymes responsible for transferring an acyl group from an acyl-CoA donor to an alcohol acceptor (Figure 7.3). A degree of substrate promiscuity means that a single enzyme may produce a variety of acyl esters, including the volatile compounds with acyl chain lengths categorised as “short” or “medium” length. While this designation is somewhat subjective, “short” chains are typically taken as C2-C6 and “medium” chains as C6-C12, as for the fatty acids^{438,439,458}.

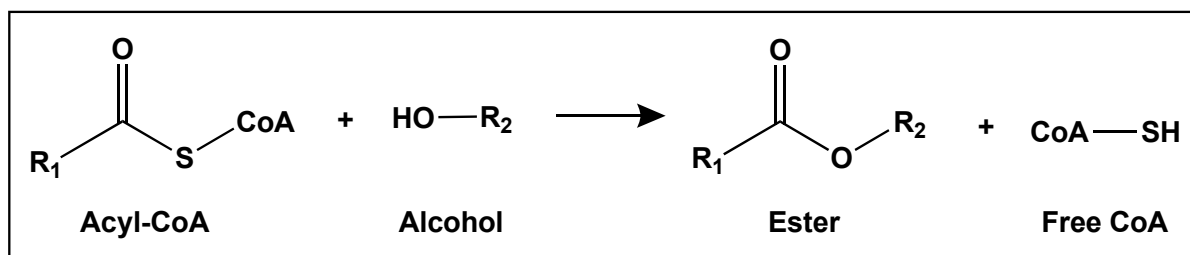


Figure 7.3. General overview of the synthesis of esters by alcohol acyltransferases.

Adapted from Saerens *et al.* (2010)⁴³⁸.

7.3. Ester production in yeast *S. cerevisiae*

Esters are formed intracellularly by fermenting yeast cells under anaerobic conditions. The diversity of substrates available within the cell means that two main groups of volatile esters are produced - the acetate esters and the medium-chain fatty acid (MCFA) ethyl esters. These two groups are produced by two major families of acyl-CoA:alcohol O-acyltransferases in the brewing yeast *Saccharomyces cerevisiae*: the acyl-CoA:ethanol O-acyltransferases (AEATs) and the alcohol acetyltransferases (AATs)⁴³⁸. Figure 7.4 summarises the main characteristics of AEATs and AATs and the main products of their enzymatic reactions. The two major AATs in *S. cerevisiae* are Atf1 and Atf2 (Figure 7.4A); the two major AEATs are Eht1 and Eeb1 (Figure 7.4B).

Family	Examples (Uniprot ID)	Main reaction substrates <i>in vivo</i>	Reaction products and aroma description
<u>Acetate esters:</u>			
AATs	Atf1 (P40353) Atf2 (P53296) Eat1 (P53208)	Acetyl-CoA and different alcohol derivatives	Ethyl acetate (nail polish, alcohol) Isoamyl acetate (banana, pear) Isobutyl acetate (fruity) Phenylethyl acetate (roses) Hexyl acetate (sweet, perfume)
A			
<u>MCFA ethyl esters:</u>			
AEATs	Eht1 (P38295) Eeb1 (Q02891)	Acyl-CoAs and ethanol	Ethyl butanoate (floral, fruit) Ethyl hexanoate (green apple) Ethyl octanoate (sweet soap) Ethyl decanoate (floral, soap)
B			

Figure 7.4. Representation of the two major enzyme families in *S. cerevisiae* responsible for the synthesis of volatile esters.

Example of the enzymatic reaction catalysed by **(A)** AATs and **(B)** AEATs. Adapted from Pires *et al.* (2014) ⁴⁵⁹ and Saerens *et al.* (2010) ⁴³⁸.

7.3.1. Ester biosynthesis pathways

During fermentation, the overall production of these esters and the ratio of different ester chain lengths can vary drastically among yeast strains⁴⁶⁰. For more than 40 years, a considerable effort has been made to elucidate the ester biosynthesis pathways, genes involved and the factors influencing ester synthesis rates. The main factors controlling ester production during fermentation are now thought to be the concentration of either substrate within the cell and the balance between the enzymatic activities involved in forming and breaking the ester linkage⁴³⁸. The biochemical pathways involved in the production of acetate esters and MCFA ethyl esters in *S. cerevisiae* are represented in Figure 7.5. This highlights the contribution of the major AATs (Atf1, Atf2) and AEATs (Eht1, Eeb1). It should be noted that a third AAT, known as Eat1, was recently identified and that triple mutant strains in which all of the known AATs are deleted (*atf1Δ atf2Δ eat1Δ*) still have residual production of ethyl acetate and isoamyl acetate⁴⁶¹. It is thus expected that there are AATs or alternative ester synthesis mechanisms still to be discovered.

Acetyl-CoA is the most common short-chain acyl-CoA substrate and is mainly formed by the oxidative decarboxylation of pyruvate in the mitochondria (Figure 7.5). Although a definitive physiological role for volatile esters in yeast remains unclear, esters might control cell membrane fluidity⁴⁶² and promote the dissipation of yeast and fruit by attracting feeding animals^{438,463,464}. As discussed below, an emerging concept is that AATs are also important in recycling CoA from medium-chain acyl-CoAs, metabolic 'dead ends' which are released from the yeast fatty acid synthase during fermentation.

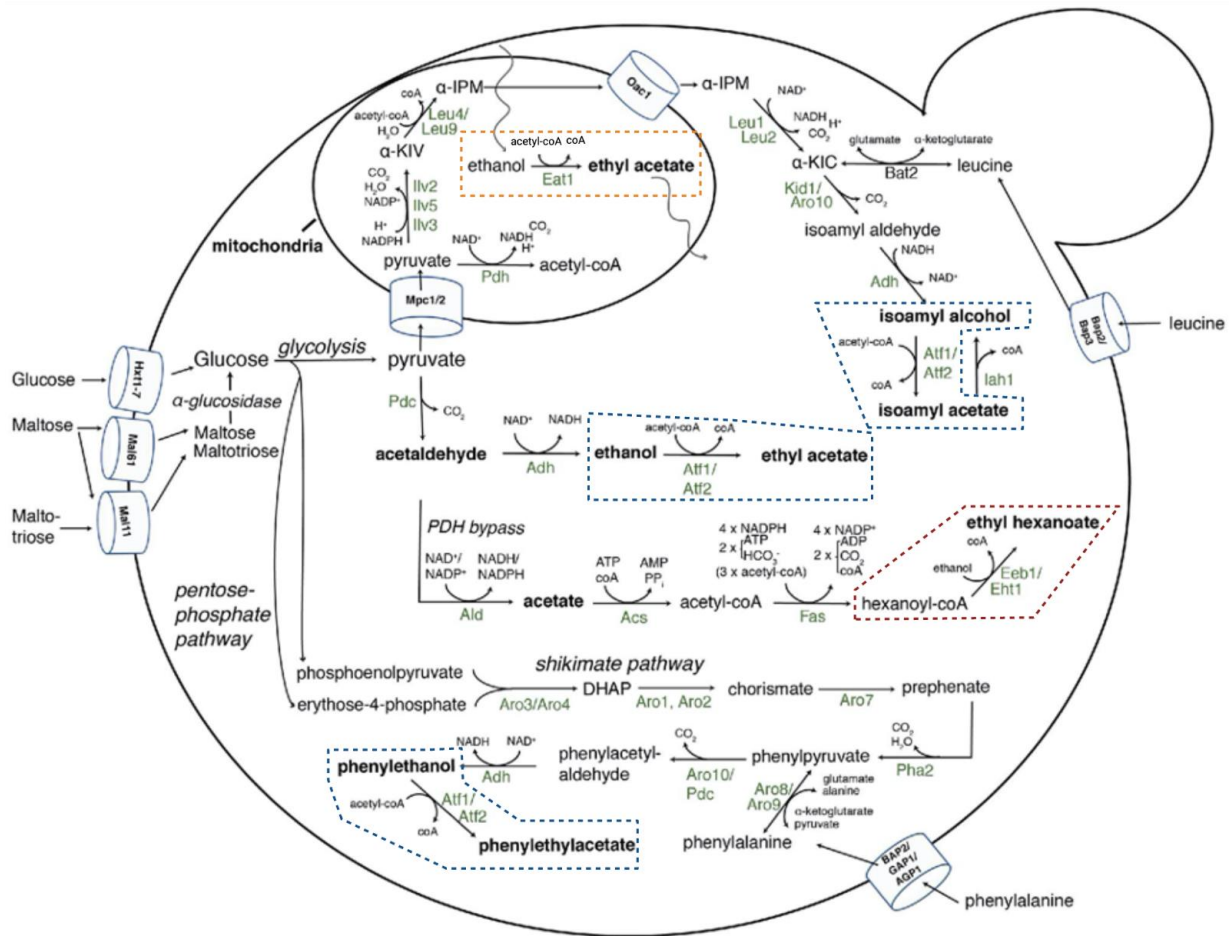


Figure 7.5. Main pathways involved in the formation of esters.

Isoamyl acetate and phenylethyl acetate produced by the AATs Atf1p and Atf2p (blue box), ethyl hexanoate produced by AEATs Eht1p and Eeb1p (red box) and ethyl acetate produced by Atf1, Atf2 (blue box) or Eat1 (orange box), in the yeast *S. cerevisiae*. Adapted from Holt *et al.* (2019) ⁴⁴⁸.

7.3.2. AATs genes - *atf1* and *atf2*

In brewer's yeast *S. cerevisiae*, the most well-characterised AATs are the gene products of *atf1* and *atf2*, known as Atf1 and Atf2, respectively (Table 7.1). Some strains have additional enzymes, such as the homologue Lg Atf1 found only in the lager strain *S. pastorianus* ^{465,466}. It is assumed that Atf1 and Lg Atf1 have diverged from a common ancestor and that their functions are conserved ^{467,468}. Both Atf1 and Atf2 catalyse the production of short- and medium-chain aliphatic (C2-C12) esters from a range of alcohols ^{368,458,466,469,470}.

Table 7.1. Molecular characteristics of known AATs genes.Adapted from Mason and Dufour (2000) ⁴³⁹.

Gene	<i>atf1</i>	<i>lg atf1</i>	<i>atf2</i>
Source	<i>S. cerevisiae</i>	<i>S. pastorianus</i>	<i>S. cerevisiae</i>
ORF designation	YOR377w	Yscatf1b	YGRI77c
Chromosome	XV	850 kb	VII
ORF size (bp)	1575	1635	1605
Product MW (kDa)	61.1	63.2	61.9

7.3.2.1. Regulation of *atf1* and *atf2* gene expression

Several studies have been carried out in order to understand the role of the Atf1 and Atf2 proteins in acetate ester synthesis. The majority of these studies are at the cellular level by using genetic knockout or overexpression of *atf1* and *atf2* genes in brewing yeast or wine yeast strains ⁴⁷¹. Generally, these studies suggest that the production of isoamyl acetate and ethyl acetate are directly related to the expression levels of *atf1* and *atf2* ^{469,472,473}. Single genetic knockouts of each of the two AATs suggested that Atf1 is crucial in acetate ester synthesis ⁴⁷¹. Double knockout of these two genes (*atf1*Δ *atf2*Δ) eliminated the synthesis of isoamyl ester but only reduced the synthesis of ethyl and propyl acetates by 50% ⁴⁷¹. This strongly suggests that these latter products are also produced by different ester synthases, which are currently unknown. Conversely, the overexpression of Atf1 and Atf2 leads to an increase in acetate ester production ^{470,471,474}.

The regulation of the *atf1* gene has received further interest. Fujii *et al.* showed that expression of Atf1 is repressed by adding unsaturated fatty acids (UFAs) to the culture medium and by highly aerobic conditions ^{475,476}. Two different regulatory pathways are involved in the *atf1* transcription repression. The repression by oxygen is umpired through the Rox1–Tup1–Ssn6 hypoxic repressor complex and *ole1* gene ^{477–479}. In contrast, the repression by UFAs is regulated by a ‘Low-Oxygen Response Element’ (LORE) in the *atf1* promotor sequence (Figure 7.6) ⁴⁸⁰. However, there is still some ambiguity in the repressor pathway, which needs to be further investigated. Verstrepen *et al.* suggested that *atf1* transcription is induced by sugars regulated through the Ras/cAMP/PKA signalling pathway ^{458,481}. Moreover, *atf1* transcription via the ‘Fermentable Growth Medium-induced’ pathway seems to be triggered by adding maltose and nitrogen compounds to yeast media ⁴³⁸.

Other studies suggest that Atf1 activity and, consequently, the volatile ester profile can be significantly altered by fermentation conditions such as temperature, ethanol concentration, CO₂/hydrostatic pressure or pitching rate^{481,482}. Concerning temperature, the highest temperature reported by a thermotolerant yeast during ester production processes was 42°C⁴⁸³. Therefore, finding new yeast strains or thermostable AATs with maximum enzyme activity at elevated temperature is crucial for the industrial production of esters.

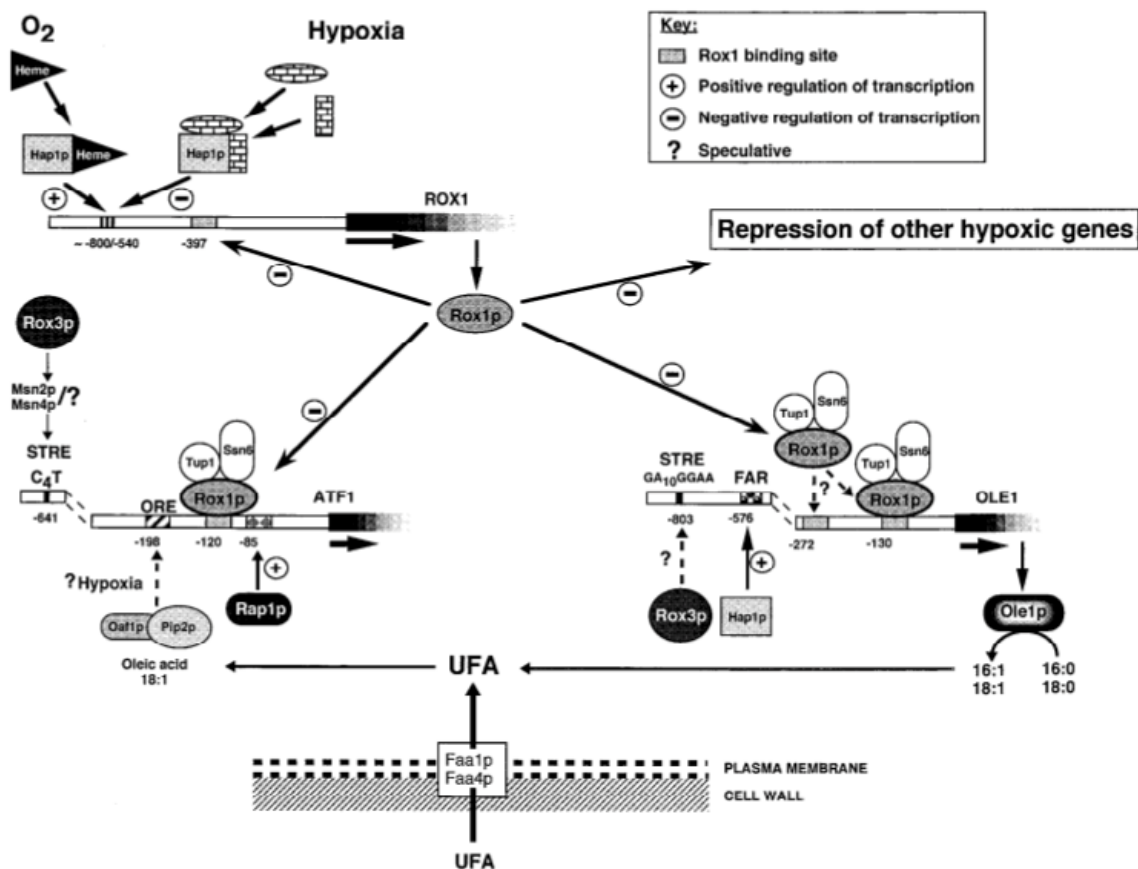


Figure 7.6. Schematic pathway regulation of *att1* and *ole1* genes in aerobic conditions and in presence of UFAs.

Adapted from Mason and Dufour (2000)⁴³⁹.

7.3.3. Atf1, Atf2 and Lg Atf1 proteins – homology and cellular localisation

This section describes the sequence features and localisation profile of AATs relevant to the thesis. The sequences of Atf1 from *S. cerevisiae* and Lg Atf1 from *S. pastorianus* are closely conserved, with 81% sequence identity⁴³⁹. Atf1 and Atf2 have only modest sequence identity, being around 37%⁴⁷³. However, the calculated global hydropathy profiles of Atf1 and Atf2 are similar, around -0.34 and -0.36, respectively (values calculated based on Kyte and Doolittle method)^{472,473}. The hydrophobicity analysis of Atf1 shows only short runs of hydrophobic residues, which leads to the conclusion that the Atf1 protein does not have a membrane-spanning region^{472,473}. However, there is growing evidence that Atf1 is a peripheral membrane protein localised in the membrane of lipid particles that bud from the endoplasmic reticulum⁴⁸⁴. Lin and Wheeldon suggested that this localisation may be due to amphipathic helices at the C- and N-terminal of Atf1⁴⁸⁵. The same authors tested this hypothesis by making Atf1 variants featuring an N-terminal deletion (Δ 2-48) and a C-terminal deletion (Δ 503-525). The results of cell imaging studies suggested that both the N- or C-terminal amphipathic helices of Atf1 are required for membrane association. Without either one of these domains, the protein is released to the cytosol (Figure 7.7)⁴⁸⁵. Another study suggested that Atf2p is localised in the ER membrane, being an integral membrane protein with both termini orientated toward inside the compartment (Figure 7.8)⁴⁸⁶.

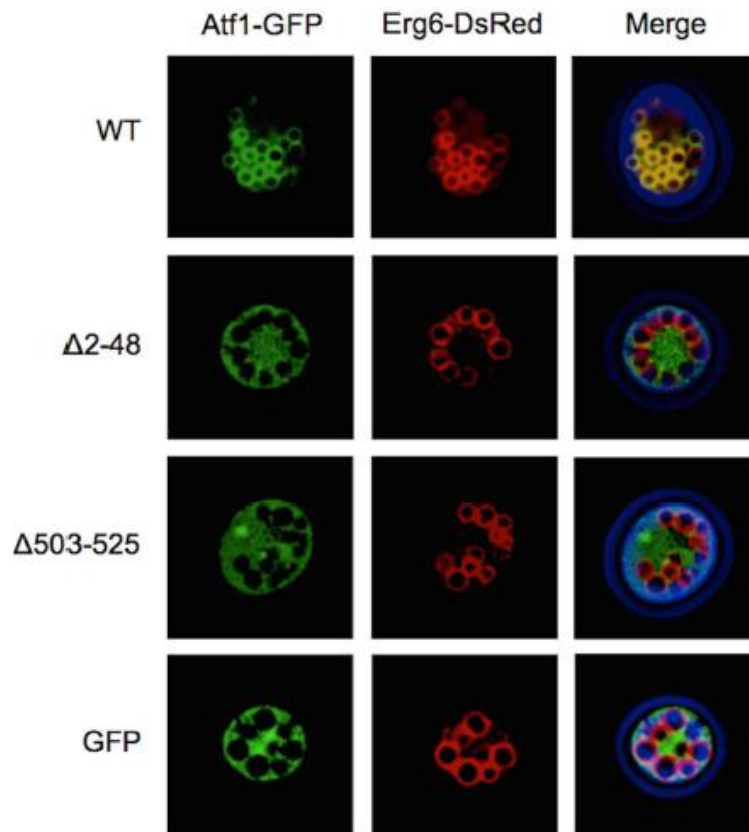


Figure 7.7. Cellular localisation of co-expressing Atf1-GFP (in green) and the lipid particle marker Erg6-DsRed (in red).

Fluorescent microscopy images showed that wild-type (WT) Atf1-GFP is localised to lipid particles (first row), while truncations at N- or C- termini ($\Delta 2-48$ and $\Delta 503-525$, respectively) result in cytosolic GFP signal (second and third rows). Adapted from Lin and Wheeldon (2014) ⁴⁸⁵.

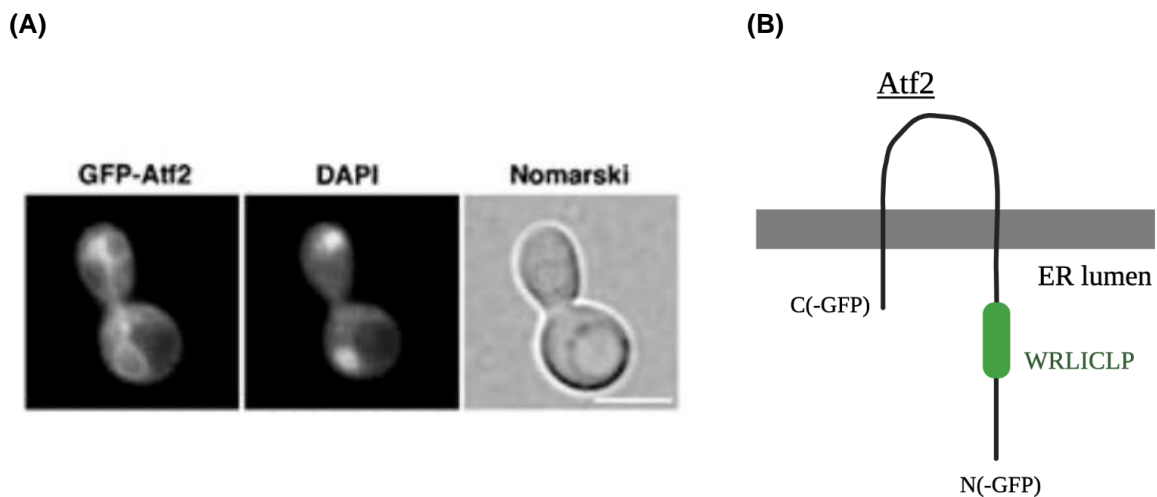


Figure 7.8. Cellular localisation of Atf2 protein.

(A) Fluorescent microscopy images of GFP-Atf2, showing the apparent localisation to the yeast ER. Nuclei were stained with DAPI. **(B)** Schematic representation of Atf2 insertion in the membrane of ER with the conserved sequence in green. Adapted from Tiwari *et al.* (2007) ⁴⁸⁶.

7.3.4. *In vitro* studies on Atf1 and Atf2 enzymes

The biochemistry of Atf1 and Atf2 is relatively well studied compared to their structure, of which little is known. No crystal structure of Atf1 is yet available, and confident homology modelling is complex since there are no suitable templates. The next part of this introduction summarises current knowledge of the structure and function of these proteins.

Atf1 was previously isolated from yeast in a semi-pure state after being liberated from cell membrane fractions via treatment with the detergents MEGA9, Triton X-100 and DDM^{466,487,488}. However, these methodologies were laborious and involved multiple purification steps. More recently, Nancolas *et al.* expressed a C-terminal His-tagged Atf1 in recombinant yeast cells and were able to purify this protein in a single affinity column step³⁶⁸. This allowed them to characterise Atf1 *in vitro* for the first time. Attempts to express Atf1 and Atf2 as recombinant proteins in *Escherichia coli* led to aggregates with reduced activity^{368,489}. The same is true of related plant AATases⁴⁸⁹.

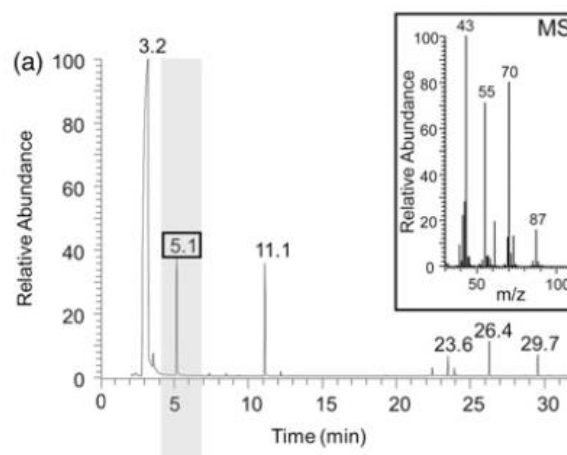
Many previous studies on yeast cell cultures and partially-purified Atf1 have strongly suggested that Atf1 prefers acetyl-CoA to other acyl-CoAs⁴⁷⁶. Interestingly, there is evidence that Atf1 protein is slightly more active against straight-chain alcohol substrates than against branched-chain alcohols. That activity of Atf1 is actually lowest with ethanol^{476,487}; This was surprising since this protein is thought to catalyse ethyl ester synthesis. Although recent studies on purified Atf1 have primarily confirmed these observations³⁶⁸, they have also provided unexpected results. These suggest that Atf1 is active as both an acetyltransferase and a thioesterase, meaning that this protein simply hydrolyses the thioester bond of the acyl-CoA substrate rather than transferring the acyl group to an alcohol. This hydrolysis is dominant *in vitro*, and activity is highest toward medium chain length acyl-CoAs, with the highest activity towards octanoyl-CoA³⁶⁸. This is despite the same study showing that only acetyl-CoA is used in acyl transfer (Figure 7.9).

Further comparison of Atf1 and Atf2 with other acyltransferases suggests that a specific histidine residue will be critical for catalysis, being highly conserved in a -HxxxD- active site motif. However, mutating the conserved catalytic histidine residue only halves the activity of Atf1 (Figure 7.9B), whereas, in most related enzymes, this mutation abolishes catalytic activity altogether³⁶⁸. These recent studies suggested that the detailed function and structure of Atf1 protein, including the specificity towards cosubstrates, has not been fully understood, and plausibly that current purifications do not remove contaminant proteins that contribute to a

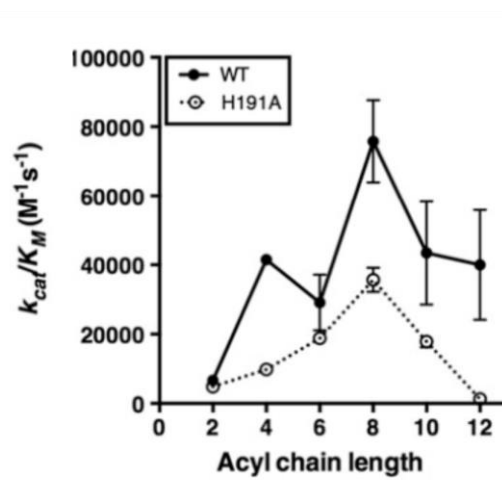
high background signal, masking the actual effect of Atf1 mutations. Therefore, and in order to explain these results, a high-resolution structure of this protein is required. However, a previous attempt to crystallise full-length recombinant Atf1 was performed without success (*Paul Curnow, personal communication*).

Additionally, sequence analysis of Atf1, Atf2 and Lg Atf1 shows that they all contain a novel conserved motif (-WRLICLP-) which might be involved in the active site. Although this hypothesis has never been experimentally tested, both Atf1p and Atf2p are strongly inhibited by sulfhydryl compounds, implying that the conserved cysteine may play a catalytic role ⁴⁶⁶. Again, this is yet to be confirmed by experiment.

(A)



(B)



(C)

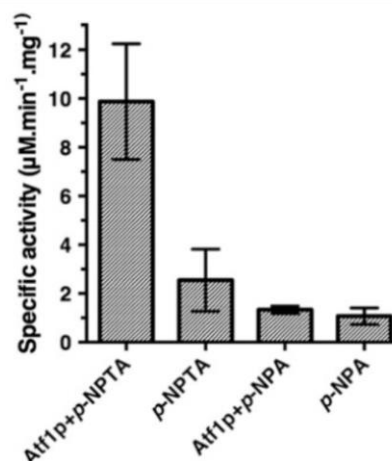


Figure 7.9. Biochemical assays of Atf1 protein performed by Nancolas *et al.* (2017) ³⁶⁸.

(A) Production of isoamyl acetate by purified Atf1 confirmed by GC-MS, peak at 5.1 min (boxed). **(B)** k_{cat}/K_M values showed the highest catalytic efficiency toward C8 (octanoyl) substrates. The active site mutation H191 decreased the overall activity only by half. **(C)** Thioesterase activity of Atf1 was confirmed by using a generic thioesterase substrate (*p*-NPTA) during enzyme assays.

7.4. Ester production in fruits

Plants comprise large families of AATases whose activity has been recognised in either flowers (e.g., *Clarkia breweri* or rose), leaves (e.g., *Arabidopsis thaliana*) or fruits during ripening⁴⁹⁰. Cultivars with high AATase activity produce more ester and, as a result, a more enriched aroma. In the ripe fruit, volatile esters play a dual role since they protect against pathogens as well as attract animals⁴⁹¹. Some of the most relevant plant AATases are listed in Table 7.2., including strawberry, melon and papaya AATases which will be the main focus of this review.

7.4.1. Ester biosynthesis pathways in fruits

Although the biosynthesis pathway of volatile esters in fruits have been previously studied – especially in apple, melon and strawberry – we still have a limited understanding of the mechanisms involved in ester biosynthesis⁴⁹¹.

Like yeast, fruit AATases catalyse the last step in ester formation by transacylation from an acyl-CoA to alcohol. Fellman and Pérez have suggested that the principal source of substrates for AATase reactions in fruits results from the metabolism of amino acids^{492,493}. Free amino acid pools will determine the aroma profile patterns in each fruit by dictating the variety and extent of ester products⁴⁹⁴. In strawberry, the major free amino acid is alanine, whose concentration decreased together with an increase in ethyl esters during ripening⁴⁹⁵. The biochemical conversion of amino acids into esters seems to occur via deamination followed by decarboxylation⁴⁹³. Myers and Tressl have already demonstrated this pathway in bananas and by Yu *et al.* in tomato fruit using radioactive labelled L-leucine^{496–498}.

However, it has been shown that fruit esters are also dependent on other biogenic pathways, namely from ripening related processes, for instance, alterations in cell wall composition⁴⁹⁴. Lipoxygenase (LOX) contributes to the breakdown of fatty acids to C6 aldehydes, later converted to alcohols by alcohol dehydrogenase (ADH)⁴⁴². This biochemical pathway has been demonstrated in transgenic tomato when the aroma profile changed dramatically by altering the levels of fatty acids⁴⁹⁹. Moreover, two ADH genes in Cantaloupe Charentais melon have been isolated and shown to be up-regulated during ripening⁵⁰⁰.

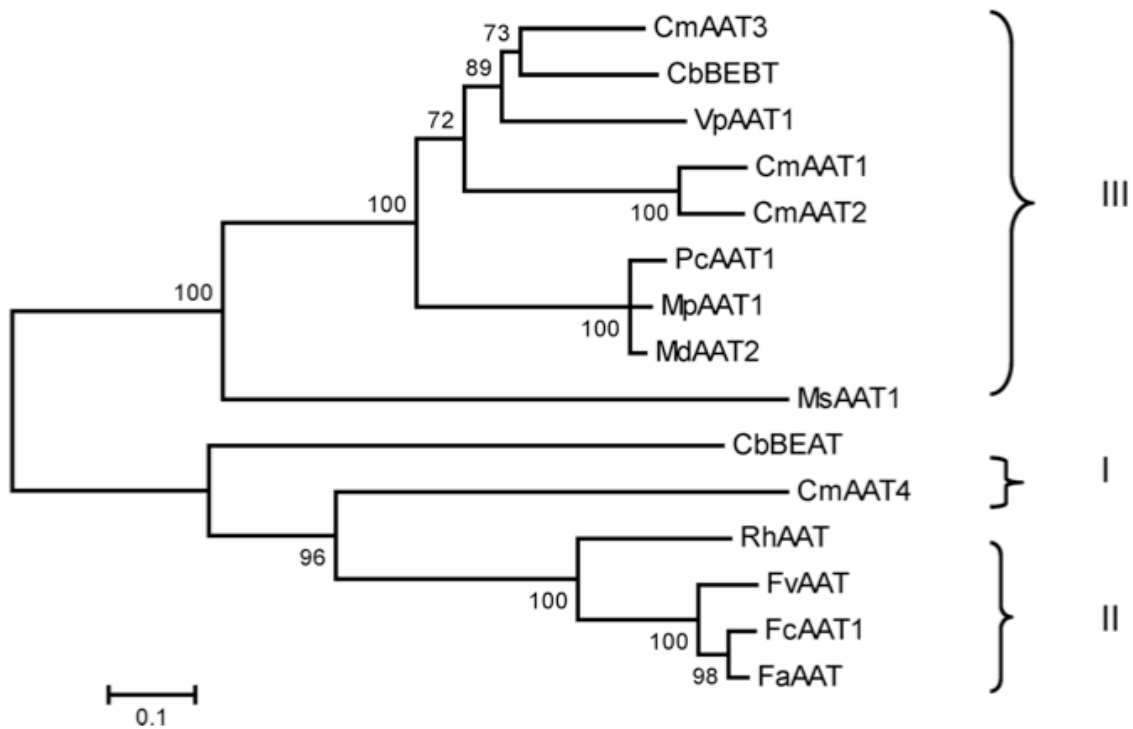
7.4.2. Main characteristics of fruit AATases

Based on sequence homology, plant acyltransferases dependent on CoA have been included in the BAHD¹ superfamily⁵⁰¹. By the end of this review, more than 46 distinct BAHD acyltransferases have been recognised based on biochemical experiments. To present, the BAHD members are thought to be monomeric enzymes with an MW from 48 to 55 kDa. Moreover, on average, a BAHD protein has around 445 amino acids⁴⁹⁰. Table 7.2 compares important characteristics of fruit AATases, including the major esters produced by each enzyme.

BAHD elements can also be divided into three main subgroups or clades. The phylogenetic tree produced is based on the multiple sequence alignment between some relevant plant AATases (Figure 7.10). Ultimately, this association reflects the enzymatic activity of each subgroup and, consequently, structural and functional similarities. For example, VpAAT1, CmAAT3 and CbBEBT showed the highest enzymatic activity towards the production of benzyl acetate⁵⁰².

Although the phylogenetic tree shows a close relationship between FaAAT1 and FvAAT1, their affinity towards alcohols are relatively different since FvAAT1 is more active towards short-chain alcohols than FaAAT1⁴⁹¹. On the other hand, AATases from different fruits, for instance, banana (MsAAT1) and melon (CmAAT1), exhibit comparable alcohol preferences and consequently activity profiles^{491,503}; they are quite distant in the phylogenetic tree. Therefore, it has been recognised as convergent evolution, in which uncharacterised AATases with similar ester production profiles independently appeared in different fruits.

¹ "BAHD family was named according to the first letter of each of the first four biochemically characterised enzymes of this family (BEAT, AHCT, HCBT, and DAT)" - definition from D'Auria (2006)⁴⁹⁰.



Plant	Protein abbr.
Apple	MpAAT1, MdAAT2
Banana	MsAAT1
<i>Clarkia breweri</i>	CbBEAT, CbBEET
Melon	CmAAT1 – T4
Papaya	VpAAT1
Pear	PcAAT1
Strawberry	FvAAT(1), FaAAT(1), FcAAT1
Rose	RhAAT

Figure 7.10. Phylogenetic tree analysis of fruit AATases.

Analysis includes the most relevant fruit AATases – the correspondent fruit for each protein abbreviation is described. Adapted from González *et al.* (2009) ⁵⁰⁴.

7.4.3. Cellular localisation and conserved motifs in fruit AATases

To date, it has been assumed that BAHD members are cytosolic since none appear to include hydrophobic sequences that would lead to organelle localisation or secretion ⁴⁹⁰. Although the BAHD superfamily is highly diverse and comprises enzymes with only 10-30% similarity at the residue level, they must contain two conserved motifs ⁵⁰¹. One of these motifs is the -HxxxD-

which is highly conserved among AATases from plants and fungi, as mentioned earlier. This motif is found at the centre of the protein sequence; therefore, it might be responsible for catalysis. The second motif, which AATases from *S. cerevisiae* do not share, is the -DFGWG-motif. This conserved region is located near the carboxyl terminus, and it seems to be critical for protein stability and folding ⁴⁹⁰. Independent investigations had found a significant decrease in enzyme activity when one or both motifs were deleted or modified ⁵⁰⁵.

Furthermore, a recent investigation on an alcohol acyltransferase from *Cucumis melo* revealed that a specific amino acid might be relevant for catalysis as well. The threonine change to an alanine at position 268 in CmAAT2 led to a non-functional enzyme, whose enzymatic activity was reestablished when the change was reversed ^{503,506}. Nevertheless, the significance of Ala268 residue remains unclear ⁵⁰³. A better insight into the protein's catalytic mechanism and the significance of common motifs came with the first crystal structure of a BAHD member, the vinorine synthase (Figure 7.11) ⁵⁰⁷.

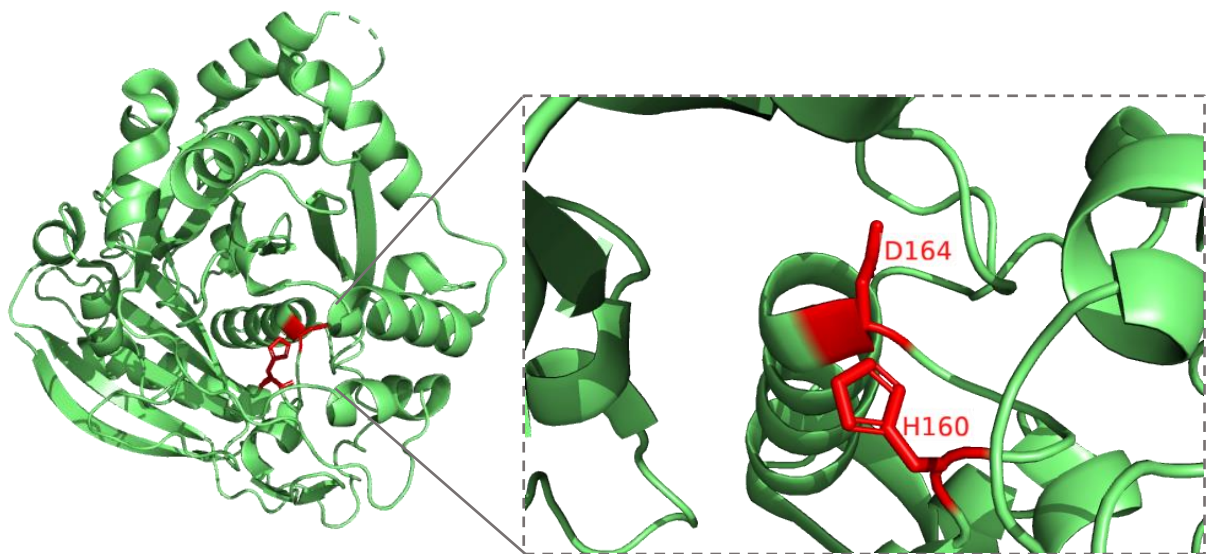


Figure 7.11. Vinorine synthase crystal structure (PDB entry 2BGH).

The residues His160 and Asp164 from the conserved motif -HxxxD- are highlighted. Image reproduced using PyMOL software.

Table 7.2. Characterisation of BAHD acyltransferases from different commercial fruits - papaya, melon, strawberry, apple and kiwi.Adapted from D'Auria (2006) ⁴⁹⁰.

AATase name	Species	Tissue	Acyl-CoA	Examples of the major esters produced	Protein size (kDa)	Accession number
VpAAT1	<i>Vasconcellea pubescens</i> (mountain papaya)	whole fruit	hexanoyl	butyl butanoate, methyl hexanoate, ethyl hexanoate, butyl hexanoate, ethyl butanoate, octyl acetate	51.6	FJ548611
FaAAT1 (or SAAT)	<i>Fragaria x ananassa</i> (cultivated strawberry)	whole fruit	acetyl	methyl hexanoate, hexyl acetate, hexyl butyrate, octyl acetate	50.7	AAG13130
FvAAT1 (or VAAT)	<i>Fragaria vesca</i> (wild strawberry)	whole fruit	acetyl	ethyl acetate, ethyl butanoate, ethyl hexanoate, octyl acetate	51.3	AAN07090
MpAAT1	<i>Malus x pumila</i> (apple)	whole fruit	short- to medium-chain CoAs	hexyl acetate, 2-methyl-butyl acetate, hexyl-2-methyl-butanoate, butyl-butanoate	50.6	AAU14879
CmAA T1 – T4	<i>Cucumis melo</i> (melon)	whole fruit	short- to medium-chain CoAs	ethyl acetate, ethyl butanoate, butyl acetate, butane-2,3-diol diacetate, 2-methylbutyl acetate, hexyl acetate	50.9 (average)	CAA94432 AAL77060 AAW51125 AAW51126
AcAT9 AcAT16	<i>Actinidia</i> spp. (kiwifruit)	whole fruit	short-chain CoAs benzoyl	butyl benzoate, methyl benzoate, ethyl butanoate, methyl butanoate, butyl acetate	47.8 51.9	KJ626344 HO772640
BanAAT	<i>Musa sapientum</i> (banana)	pulp	acetyl	cinnamyl acetate and other medium-chain aliphatic acetate esters	45.8	CAC09063

7.4.4. *In vivo* studies using fruit crude extract

Capillary gas chromatography coupled with mass spectrometry (GC-MS) has been used for *in vivo* analysis of plant esters. Yamashita and colleagues used this methodology to identify seventy esters in strawberries incubated with aldehydes, in which acetate ester was the most dominant⁵⁰⁸. The same approach has been applied to analyse the ester profile of papaya^{509,510} and kiwifruit⁵¹¹.

Different research groups have been able to partially purify strawberry AATases from crude fruit extracts. Olías *et al.* showed an AATase from *Fragaria ananassa* X *Duchessne* var. Chandler has maximal activity against acetyl-CoA and hexyl alcohol as substrates^{512,513}. González *et al.* demonstrated that FcAAT1 from *Fragaria chiloensis* is most active towards acetyl-CoA and short-chain alcohols⁵⁰⁴. In general, these results correlate with the findings from *in vitro* studies mentioned below. However, studying AATases in crude extracts can be misinterpreted because their activity profiles' can be influenced by the mixture of different AATases rather than a specific AATase. Likewise, banana, melon, apple and olive AATases have also been investigated in the crude extract^{514–518}. Activity assays of those AATases confirmed that acetyl-CoA is the most used substrate by plant AATases. Furthermore, in these experiments, the use of Triton X-100 to extract AATases from fruit pulp was essential since, without detergents, fractions showed very low activity levels. Consequently, it has been suggested that fruit AATases could be membrane-bound proteins, as already observed for Atf1.

Moreover, *in vivo* studies have shown that fruit AATases were found to be regulated by ethylene and 1-methylcyclopropene, which act as inducer and inhibitor, respectively, in banana⁵¹⁹, papaya^{502,520} or apple⁵²¹, to name a few.

7.4.5. *In vitro* experiments with plant AATases

Several AATase proteins have been recombinantly expressed in either *S. cerevisiae* or *E. coli* from fruit species with a high commercial interest in recent years.

Balbontín *et al.* successfully cloned and expressed VpAAT1 in *S. cerevisiae*. The VpAAT1 protein was purified from the soluble fraction and appeared to be most active in producing benzyl acetate⁵⁰². Each of the melon AATases (CmAATs) were also successfully

recombinantly expressed in *S. cerevisiae*, and their activities were tested towards different substrates *in vitro*. Although no activity was observed for CmAAT2, the remaining AATs (CmAAT1, CmAAT3 and CmAAT4) could synthesise esters^{503,506,522}. Recombinant CmAAT1 and CmAAT4 are highly active in producing (*E*)-hex-2-enyl acetate, whereas benzyl acetate is preferably synthesised by CmAAT3^{503,523}. Kiwi AATase genes (*AcAT9* and *AcAT16*) have also been recombinantly expressed in *S. cerevisiae* and were purified from lysate fraction and found to be active^{524,525}. Both enzymes preferred butanol as an alcohol acceptor molecule in the presence of a variety of acyl-CoAs. Nonetheless, *AcAT16* is more selective towards benzoyl-CoA, while *AcAT9* prefers acetyl-CoA and propionyl-CoA^{524,526}.

Strawberry, banana and apple AATases have been recombinantly expressed in *Escherichia coli*, using the C43(DE3), the BL21 Gold (DE3) or the BL21 CodonPlus-RIL strains^{491,527,528}. Three types of strawberry AATases have been bacterially expressed, and their ester profile determined, namely FaAAT1⁵²⁹, FvAAT1⁴⁹¹ and FaAAT2⁵²⁷. Strawberry AATases were purified from the soluble fraction using either a His-tag (FvAAT1 and FaAAT1) or a GST-tag (FaAAT2). Moreover, their activity towards different substrates was studied, e.g., FvAAT1 showed a higher affinity towards short-chain alcohols (C6-C8) than FaAAT1 and FaAAT2⁵²⁷. As for experiments conducted with recombinant yeast, the purified recombinant enzymes were not characterised at any level of detail.

7.4.6. *In silico* studies on plant AATases

The crystal structure of vinorine synthase and other similar proteins have been used to generate structural models of fruit AATases, namely VpAAT1^{530–533}, FaAAT1⁵³⁴ and CmAAT⁵³⁵. However, few experimental investigations have been conducted to corroborate those models. The main goal for these *in silico* analyses was to investigate their catalytic mechanism by molecular dynamics and comparative modelling. *In silico* studies revealed that the substrate binding mechanism is supported by forming a non-covalent ternary complex⁵³⁴. Additionally, molecular dynamics simulations were used to test the affinity of fruit AATases towards a wide range of alcohols and acyl-CoA substrate. Furthermore, amino acid substitutions in the highly conserved motifs mentioned earlier revealed a significant reduction in the enzymatic activity and structural modifications in the indispensable solvent channel^{531,533,535}. Once again, experimental support for these conclusions is lacking.

7.5. Perspectives on the industrial use of AATases

The AATases are now being used in synthetic biology to produce fragrances or chemicals in cellular 'factories' ⁴⁴⁵. For example, Atf1 protein has been introduced in the AckA-Pta metabolic pathway to produce isobutyl acetate in *E. coli*, which is an industrial solvent and food additive ⁵³⁶. In the same context, Bohnenkamp *et al.* engineered an *E. coli* strain to produce ethyl acetate, an industrial solvent used for coatings, resins or paints, using the AAT pathway ⁵³⁷. The most suitable enzyme was chosen among Atf1 from *S. cerevisiae* and Eat1 from *Kluyveromyces marxianus* and *Wickerhamomyces anomalus* ⁵³⁷. More recently, Atf1 protein was used to produce synthetic biology education kit – BioBits™, mainly used for olfactory outputs ⁵³⁸.

7.6. Aims of this chapter

Despite modelling and experimental attempts, the mechanism of the AATases remains to be fully understood. No experimental structures have been reported, and fundamental characterisation of the semi-purified AATases was largely absent. For example, the purity or composition of these protein fractions was not determined by SDS-PAGE, and there was no evidence for the dispersity or folded state of the purified proteins. Additionally, the methodology applied to determine K_M or V_{max} is generally through GC-MS, by stopping the reaction at a specific time point (e.g., an hour) and determining how much product has been produced. A more thorough approach to enzyme kinetics will require continuous assays that measure the initial linear part of the enzyme progress curve to determine the reaction velocity accurately. Consequently, being able to understand the molecular mechanism and the nature of the enzyme-substrate interaction in various AATases from commercial fruits will broaden our knowledge about this enzyme family. This would, in turn, enable us to improve fruit flavour, in particular during storage, transport, and processing.

This chapter proposes to explore the recombinant expression of AATases from yeast and fruits to reveal the fundamental basis of their structure and function.

Research Strand 1. Building on previous work, obtain structural and functional information on yeast Atf1 and consequently validate the supposed catalytic mechanism. Research questions to be addressed include:

- What is the basis of the protein-substrate interaction?

- What causes acetyl-CoA hydrolysis to dominate over acyl transfer?
- Why does Atf1 hydrolyse medium-chain length acyl-CoAs?
- Can hydrolysis be disfavoured to generate a more productive enzyme?
- Why do active site mutations only have a modest functional impact on Atf1?

Research Strand 2. Clone, express, purify and characterise plant AATases from papaya (VpAAT1), cultivated strawberry (FaAAT1) and melon (MAAT). This includes defining the substrate profile of each enzyme as well as progressing to structural biology. Research questions to be addressed include:

- How can we optimise the recombinant expression of plant AATases?
- What is the substrate preference of the plant enzymes *in vitro*?
- How are different substrates discriminated?
- Do the plant enzymes proceed by the classical ternary complex mechanism?
- How do the enzyme constants (V_{\max} , K_M) relate to other BAHD enzymes and Atf1?
- Is the structure of plant AATases similar to other BAHD enzymes, as assumed?
- Do the plant AATases also exhibit thioesterase activity?

Chapter 8. Materials and methods

All general methods were described in Chapter 2, with any deviation from these methods highlighted in the text. Experimental methods and specific materials, including enzymes, media and expression strains to the study of alcohol acyltransferases are outlined below.

8.1. Chemicals, reagents and solutions

In addition to the reagents described in Table 2.2, specific reagents for the study of alcohol acyltransferases are listed below.

Table 8.1. List of reagents and commercial kits used.

Reagent	Supplier
α -ketoglutarate	Sigma
Anti-V5-HRP Monoclonal Antibody (#R961-25)	Invitrogen
Arabinose	Sigma
BugBuster [®] Protein Extraction Reagent	Novagen
Detergents described on Table 8.17	Anatrace
EDTA	Fisher
Coenzyme A (CoA)	Sigma
Galactose	Sigma
Gibson Assembly Master Mix (2x)	New England Biolabs
Glucose	Sigma
Guanidine-Hydrochloride (Gnd-HCl)	Sigma
Isoamyl alcohol	Sigma
Lithium Acetate	Sigma
Nicotinamide adenine dinucleotide (NAD ⁺)	Sigma
Octanoyl-CoA	Sigma
PEG 3350	Sigma
<i>p</i> -nitrophenylthioacetate (<i>p</i> -NPTA)	Sigma
Single stranded carrier DNA (ssDNA)	Thermo Fisher Scientific

Reagent	Supplier
Thesit	Sigma
Thiamine pyrophosphate (TPP)	Sigma
Urea	Sigma

8.2. Enzymes

Table 8.2. List of enzymes used in the AATases project.

Reagent	Supplier
α -Ketoglutarate Dehydrogenase from porcine heart	Sigma (commercial) or kindly supplied by Prof. Dick Denton (in-house)
In-Fusion HD Enzyme Premix	Takara Bio
Restriction enzymes (<i>Xho</i> I, <i>Bam</i> HI, <i>Kpn</i> I, <i>Hind</i> III and <i>Xba</i> I)	New England Biolabs

8.3. Prepared buffers

Table 8.3. List of prepared buffers.

Buffer	Use	Composition
Arabinose	Protein expression	10% w/v in ddH ₂ O – prepare fresh and filter sterilise
Galactose	Protein expression	20% (w/v) in ddH ₂ O – prepare fresh and filter sterilise
Glucose	Protein expression	20% (w/v) in ddH ₂ O – prepare fresh and filter sterilise
PEG 3350	Yeast transformation	50% (w/v) in ddH ₂ O – prepare fresh and filter sterilise
Transfer Buffer	WB	0.3% (w/v) Tris, 1.44% (w/v) glycine, 0.037% (w/v) SDS and 10% (v/v) methanol

8.4. General microbiology techniques

8.4.1. Strains and expression vectors

Table 8.4. *E. coli* strains used in AATase project.

Strain	Feature	Use	Supplier
One Shot™ TOP10 Chemically Competent <i>E. coli</i>	Stable replication of high-copy number of plasmids	Plasmid preparation	Invitrogen
Stellar Competent Cells			Takara Bio
BL21-AI Chemically Competent <i>E. coli</i>	Tight regulation and expression of toxic proteins based on arabinose promoter	Protein expression	Invitrogen
BL21(DE3) Chemically Competent <i>E. coli</i>	Protease deficient strain for T7 RNA polymerase expression of proteins under IPTG-inducible lacUV5 promotor		New England Biolabs

Table 8.5. *S. cerevisiae* strain used in AATase project.

Strain	Feature	Use	Supplier (Ref.)
FGY217	Protease-deficient ($\Delta pep4$) (this mutation also knocks down the activity of other vacuolar hydrolases)	Protein expression	Homemade (Curnow lab); originally supplied by Prof. Simon Newstead ⁵³⁹

Table 8.6. Expression vectors used in AATase project.

Vector	Feature	Supplier
pET28	T7/LacO for IPTG induction in <i>E. coli</i> ; Kanamycin resistance	Novagen
pOPINF pOPINM pOPINNUSA	T7/LacO for IPTG induction in <i>E. coli</i> ; Ampicillin resistance	OPPF*-UK
pYES2	<i>URA3</i> gene for auxotrophic selection in yeast and 2 μ origin for high-copy maintenance; Ampicillin resistance	Invitrogen

* Oxford Protein Production Facility

8.4.2. Media and agar plates

All powdered media was prepared with ddH₂O and autoclaved at 121°C for 15 min at 15 psi. Either SC-agar or YPD-agar (agar from BD Difco™) plates were prepared by adding 2% (w/v) to the liquid broth before autoclaving and, glucose was added at 2% (v/v) to warm broth after autoclaving.

Table 8.7. Media and general preparation.

Media	Preparation	Use	Supplier
Powdered AIM including trace elements	3.5% (w/v) of LB 4.6% (w/v) of 2YT	Liquid <i>E. coli</i> culture for protein expression	Fomedium™
Selective culture (SC) broth	0.67% (w/v) of Yeast Nitrogen Base Media and 0.19% (w/v) of Yeast Synthetic Drop-out Medium without URA	Liquid <i>S. cerevisiae</i> culture for protein expression	Fomedium™
YPD rich media	1% (w/v) of Yeast Nitrogen Base Media and 2% (w/v) of Peptone	Liquid <i>S. cerevisiae</i> culture for yeast transformation	Fomedium™

8.4.3. Antibiotics

Antibiotic solutions were prepared in sterile conditions and filtered via a 0.45 µm syringe filter and were stored at 4°C. Antibiotic was added at working concentration to each *E. coli* liquid culture just before inoculating or, after autoclaving, to LB-agar warm solution.

Table 8.8. List of antibiotics used.

Antibiotic	Stock concentration (1000x stock)	Work concentration	Supplier
Kanamycin	50 mg/mL in ddH ₂ O	50 µg/mL	ITW Reagents
Carbenicillin	50 mg/mL in ddH ₂ O	50 µg/mL	Sigma

8.5. Bioinformatic analysis

Software and online tools used for the AATase project were as described in Section 2.5. Additionally, the XtalPred web server (<https://xtalpred.godziklab.org/XtalPred-cgi/xtal.pl>) was used to predict protein crystallizability³⁹⁶. By using this bioinformatic tool, the most promising predictions were selected for further investigation. The following fruit AATases were considered, papaya (VpAAT1), cultivated strawberry (FaAAT1), kiwi (AcAT16 and AcAT9), apple (MpAAT1) and melon (MAAT).

8.6. General DNA techniques

All molecular biology protocols used purified and filtered water with a Milli-Q Integral Water Purification System. Unless otherwise stated, all molecular biology reagents are cited in Table 2.2 and enzymes in Table 2.3.

8.6.1. Polymerase chain reaction

Gene amplification was performed using polymerase chain reaction (PCR). PCR reaction mixes were established as outlined in Table 8.9. Unless otherwise specified, Phusion® High-Fidelity DNA Polymerase was used for PCR reactions, employing the amplification conditions listed in Table 8.10. All primers were synthesised by Eurofins Genomics.

Table 8.9. Typical PCR reaction mix composition.

Component	Volume	Concentration
Each primer (20 pmol/mL)	2.5 µL	1 pmol
DMSO (optional)	1.5 µL	3%
Template DNA	typically, 1 µL	< 250 ng
10 mM dNTPs	1 µL	200 µM
5x Phusion HF Buffer	10 µL	1x
Phusion® High-Fidelity DNA Polymerase	0.5 µL	1.0 units/50 µL PCR
Total	50 µL	-

Table 8.10. Typical PCR amplification conditions.

Step	Temperature	Time
Initial denaturation	98°C	30 second
25 cycles	98°C	10 second
	Annealing temperature	30 second
	72°C	30 second
Final extension	72°C	10 minute
Hold	10°C	-

8.6.2. Cloning of modified *atf1* gene constructs into pYES2 vector

The galactose-inducible pYES2 yeast expression plasmid containing the *atf1* gene was provided by Paul Curnow³⁶⁸ (Figure 8.1). This was used here for protein expression and as DNA template to perform truncations in Atf1 at the N-terminal (*atf1* ΔN), at the C-terminal (*atf1* ΔC) and a dual truncation of both terminal domains (*atf1* ΔN/ΔC). The entire plasmid pYES2 containing the *atf1* gene was amplified by PCR using the primers listed in Table 8.11. The PCR product pYES2|*atf1* ΔC means that the *atf1* gene was amplified with a deletion between positions 503 to 525, whereas in pYES2|*atf1* ΔN the *atf1* gene has a deletion between positions 2 to 48. For the double deletion (pYES2|*atf1* ΔN/ΔC), the DNA template used was pYES2|*atf1* ΔN alongside with the primers Cterm_F and Cterm_R.

PCRs reactions were carried out as cited in Table 8.9. The reaction proceeded as described previously (Table 8.10) at annealing temperature of 50°C. A 1% agarose gel in TAE buffer was used to confirm the correct size of the relevant PCR products (as described in Section 2.7.7). PCR products were purified using the QIAquick PCR Purification Kit, and purified PCR products were then ligated via a Gibson Assembly reaction according to the manufacturer instructions (incubation for 1 hour at 50°C). 5 μL of the ligation mix were transformed into TOP10 Chemically Competent *E. coli* according to the protocol described in Section 2.7.3. Successful transformants were selected from LB-agar plates supplemented with carbenicillin. Plasmid Miniprep procedure and DNA sequencing were performed as described in Chapter 2, before the inserts were confirmed by sequencing.

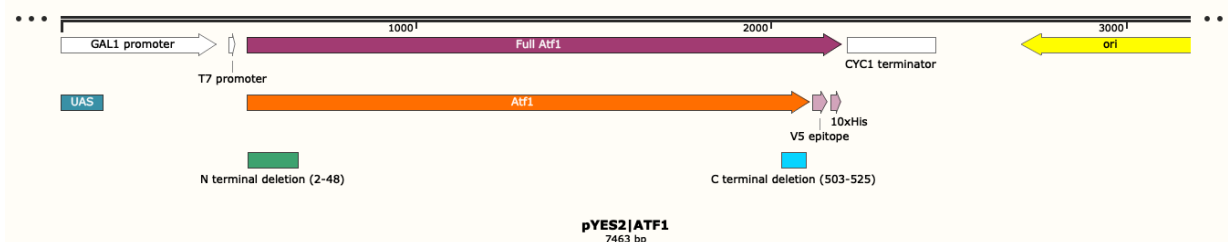


Figure 8.1. Representation of a part of the galactose-inducible pYES2 yeast expression plasmid cloned with the *atf1* gene.

Plasmid includes His₁₀ tag and V5 epitope (GKPIPPLLGLDST). Deletions Δ 2-48 and Δ 503-525 are labeled and represented in green and blue, respectively. Figure constructed using SnapGene software.

Table 8.11. DNA template, name and sequence of reverse and forward primers used for PCR amplification of the entire plasmid containing the *atf1* gene.

Template	Primer	Sequence 5' → 3'	PCR product
pYES2 <i>atf1</i>	Cterm_F	TTCAACAAAGTCTAGAGGGCCCTTCGAAG	pYES2 <i>atf1</i> Δ C
	Cterm_R	GCCCTCTAGACTTTGTTGAAGCAACAACAA	
	Nterm_F	CGGATCCATGTGCACATATGGAGAATTGAGTGATT	pYES2 <i>atf1</i> Δ N
	Nterm_R	CATATGTGCACATGGATCCGAGCTCG	
pYES2 <i>atf1</i>	Cterm_F	(described above)	pYES2 <i>atf1</i>
Δ N	Cterm_R	(described above)	Δ N/ Δ C

8.6.3. Cloning of *atf1* gene into pET28 vector

The *E. coli* plasmid pET28 was digested with *Bam*HI and *Xho*I, and the small fragment (40 bp) excised from the multiple cloning site was removed by gel electrophoresis (as described in Section 2.7.7). The digested plasmid was recovered using a QIAgen Gel Extraction Kit. *atf1* inserts were obtained through the amplification of full-length and partial *atf1* genes in pYES2 by PCR. DNA template and selected primers are described in the following table. To facilitate cloning, unique restriction sites for *Bam*HI and *Xho*I were included at the forward and reverse primers, respectively (Table 8.12). A stop codon was also included at the C-terminus, which was missing in the target gene because of the presence of C-terminal tags. PCR reactions were carried out as cited in Table 8.9. The reaction proceeded according to Table 8.10 at annealing temperatures of 50°C or 52°C.

As before, PCR products were separated from primers and partial products on a 1% agarose gel and recovered. After restriction digest with *Bam*HI/*Xho*I, the PCR products were inserted into pET28 by cohesive end ligation. Ligation used the Quick Ligase with 1 μ L *Bam*HI/*Xho*I-digested pET28 vector and 6 μ L digested insert in a final volume of 20 μ L. The reaction was initially mixed on ice, then transferred to room temperature for 1 hour. Table 8.12 describes all plasmids produced after the cloning process. After being transformed into TOP10 Chemically Competent *E. coli* cells, colonies containing the plasmid were selected on LB agar plates containing kanamycin (see Chapter 2 for full protocol description). Colonies were regrown, and plasmids were extracted, purified and sequenced as described previously.

Table 8.12. Sequence of the forward (F) and reverse (R) primers used to amplify *atf1* gene.

*Bam*HI and *Xho*I restriction sites are underlined. DNA template and primers used to obtain the inserts by PCR.

Primer	Sequence 5' → 3'
WT_F	ATGCCTT <u>GGATCC</u> ATGAATGAAATCGATGAG
WT_R	ATGCCTT <u>CTCGAG</u> CTAAGGGCCTAAAAGGAGAGCTTTG
NTD_F	ATGCTT <u>GGATCC</u> ATGTGCACATATGGAGAATTGAGTG
CTD_R	ATGCCTT <u>CTCGAG</u> CTACTTTGTTGAAGCAACAACAATATTCATCC

Template	Primer	PCR product	Final plasmid
pYES2 <i>atf1</i>	WT_F and WT_R	<i>atf1</i>	pET28 <i>atf1</i>
pYES2 <i>atf1</i> Δ C	WT_F and CTD_R	<i>atf1</i> Δ C	pET28 <i>atf1</i> Δ C
pYES2 <i>atf1</i> Δ N	NTD_F and WT_R	<i>atf1</i> Δ N	pET28 <i>atf1</i> Δ N
pYES2 <i>atf1</i> Δ N/ Δ C	NTD_F and CTD_R	<i>atf1</i> Δ N/ Δ C	pET28 <i>atf1</i> Δ N/ Δ C

8.6.4. Synthetic genes of AATases

Based on the bioinformatic analysis, specific fruit AATases genes were synthesised by Thermo Fisher with optimised codons for *E. coli* expression, namely MAAT (GenBank protein ID EU431334), VpAAT1 (FJ548611) and FaAAT1 (AAG13130).

8.6.5. Cloning of AATase genes in pYES2 vector

For *S. cerevisiae* expression, alcohol acyltransferase synthetic genes were amplified by using primers stated in Table 8.13. Primers had *Hind*III (forward primer) and *Xba*I (reverse primer) recognition sites for specific ligation into pYES2 vector. PCR reactions were carried out as cited in Table 8.9. Reactions proceeded according to Table 8.10 at annealing temperatures of 52°C, or 54°C. PCR products were purified using the QIAquick PCR Purification Kit. All inserts were then cloned into a *Hind*III/*Xba*I-digested pYES2 vector (Table 8.13) through the Quick Ligase enzyme, as previously described. This strategy ensured that the V5 epitope and the His₁₀ tag were incorporated at the C-terminus of all inserts. The ligation product was then transformed into TOP10 Competent Cells, as described in Section 2.7.3.

Table 8.13. Sequence of forward (F) and reverse (R) primers used for alcohol acyltransferase gene amplification by PCR.

Restriction sites (*Hind*III and *Xba*I for forward and reverse, respectively) are underlined. Final plasmid description is also represented.

Gene template	Forward (F) and reverse (R) primers (5' → 3')	PCR product
<i>MAAT</i>	F: AGTCCT <u>AAGCTT</u> AAAAAATGGAAACCATGCAGACC R: AGTCCT <u>TCTAGAC</u> CAGTGCGCTCAGAATTG	<i>MAAT</i>
<i>FaAAT1</i>	F: AGTCCT <u>AAGCTT</u> AAAAAATGGAAAAAATCGAGGTGAG R: AGTCCT <u>TCTAGA</u> AATCAGTGTTTTCGGGC	<i>FaAAT1</i>
<i>VpAAT1</i>	F: AGTCCT <u>AAGCTT</u> AAAAAATGGCAGAAAAAGCAAGC R: AGTCCT <u>TCTAGA</u> AATGGCGCTCACAATAAA	<i>VpAAT1</i>

Vector	PCR product	Final plasmid
	<i>MAAT</i>	pYES2 <i>MAAT</i>
pYES2	<i>FaAAT1</i>	pYES2 <i>FaAAT1</i>
	<i>VpAAT1</i>	pYES2 <i>VpAAT1</i>

8.6.6. Cloning of AATase genes in pOPIN vectors

The pYES2 vector was used as a template to amplify all synthetic genes using the primers listed in Table 8.14. For this approach, primers were designed with a homologous sequence complement to linearised vector ends, which are coloured in red and yellow in Figure 8.2.

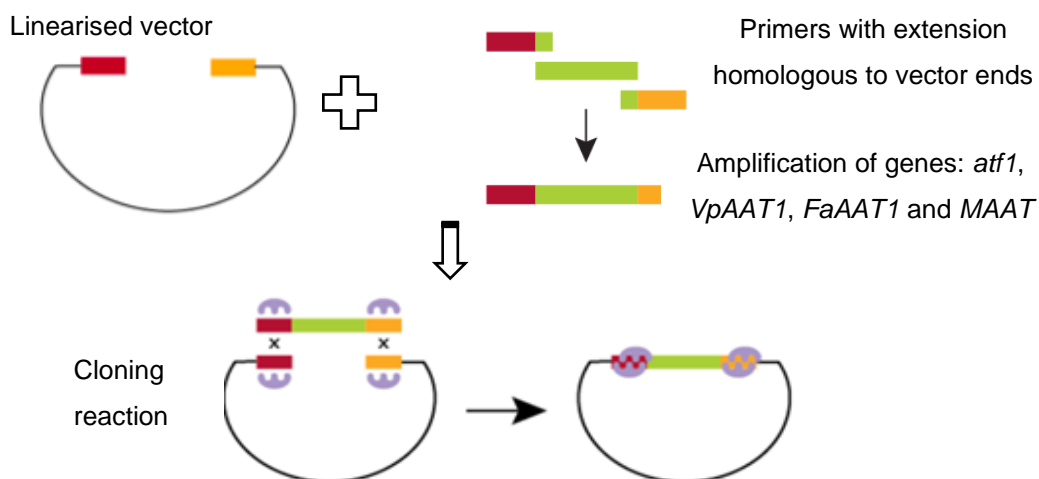


Figure 8.2. Representation of the strategy employed to clone *atf1*, *VpAAT1*, *FaAAT1* and *MAAT* genes into pOPIN vectors.

PCR reactions were carried out as cited in Table 8.9. The reaction proceeded at different annealing temperatures, according to Table 8.10. PCR products were purified using the QIAquick PCR Purification Kit. After vectors were linearised with *KpnI* and *HindIII*, all inserts were cloned into the pOPINF, pOPINM and pOPINNUSA via the In-Phusion® HD Cloning protocol. The ligation reaction was set using 2 μ L of 5x In-Fusion HD Enzyme and a 1:1 molar ratio of insert to vector until a final volume of 10 μ L. Reaction mixtures were incubated for 20 minute at 50°C, and the products were used to transform Stellar Competent Cells or stored at -20°C. Features of each construct were described in Table 8.15.

Table 8.14. Sequence of forward (F) and reverse (R) primers used for gene cloning into pOPIN vectors.

The extensions homologous to vector ends are underlined.

Gene template	Forward (F) and reverse (R) primers (5' → 3')
<i>atf1</i>	F: <u>AAGTTCTGTTTCAGGGCCCG</u> ATGAATGAAATCGATGAGAAAAATC
<i>VpAAT1</i>	F: <u>AAGTTCTGTTTCAGGGCCCG</u> ATGGCAGAAAAAGCAAGC
<i>FaAAT1</i>	F: <u>AAGTTCTGTTTCAGGGCCCG</u> ATGGAAAAATCGAGGTGAG
<i>MAAT</i>	F: <u>AAGTTCTGTTTCAGGGCCCG</u> ATGGAAACCATGCAGACC
Reverse primer is universal for all genes since it will amplify the V5 epitope sequence present in DNA template	
	R: <u>ATGGTCTAGAAAGCTTTA</u> CGTAGAATCGAGACCGAG

Table 8.15. Features of the pOPIN vectors.

All pOPIN vectors have ampicillin resistance, His₆ tag at N-terminal and V5 epitope (GKPIPPLLGLDST) at C-terminal. pOPINM vector has a Maltose Binding Protein (MBP) tag and pOPINNUSA vector has a N utilisation substance protein A (NusA) tag, both at N-terminal.

Vector	Fusion protein	PCR product	Final plasmid
pOPINF	(His) ₆ -3C-POI-V5	<i>atf1</i>	pOPINF <i>atf1</i>
		<i>VpAAT1</i>	pOPINF <i>VpAAT1</i>
		<i>FaAAT1</i>	pOPINF <i>FaAAT1</i>
		<i>MAAT</i>	pOPINF <i>MAAT</i>
pOPINM	(His) ₆ - MBP -3C-POI-V5	<i>atf1</i>	pOPINM <i>atf1</i>
		<i>VpAAT1</i>	pOPINM <i>VpAAT1</i>
		<i>FaAAT1</i>	pOPINM <i>FaAAT1</i>
		<i>MAAT</i>	pOPINM <i>MAAT</i>
pOPINNUSA	(His) ₆ - NusA -3C-POI-V5	<i>atf1</i>	pOPINNUSA <i>atf1</i>
		<i>VpAAT1</i>	pOPINNUSA <i>VpAAT1</i>
		<i>FaAAT1</i>	pOPINNUSA <i>FaAAT1</i>
		<i>MAAT</i>	pOPINNUSA <i>MAAT</i>

8.6.7. PCR product transformation

Two different cell strains were used for PCR product transformation, either *E. coli* TOP10 or Stellar Competent Cells. Transformation protocol for *E. coli* TOP10 has been described previously (Section 2.7.3), whereas transformation of Stellar Competent Cells is briefly described here. 2.5 µL of the cloning reaction of alcohol acyltransferase genes into pOPIN vectors were transformed into Stellar Competent Cells according to manufacturer's instructions. Heat shock was performed at 42°C for 45 second; transformants were selected from LB-agar plates supplemented with carbenicillin. Competent cells with 1 µL H₂O were used as a negative control.

8.6.8. Plasmid miniprep and DNA sequencing

Plasmids were prepared as described in Chapter 2 and quantified before sending to sequencing. Plasmids were sequenced via a commercial service using primers listed in Table 8.16, and results analysed by SnapGene software.

Table 8.16. List of primers used for sequencing service.

Plasmid used	Forward primer (5' → 3')	Reverse primer (5' → 3')
pYES2	Gal1 (AATATACCTCTATACTTTAACGTC)	CYC1 (GCGTGAATGTAAGCGTGAC)
pET28	T7 (TAATACGACTCACTATAGGG)	T7term (CTAGTTATTGCTCAGCGGT)
pOPINF	T7	Bglob-pA-R (TTTTGGCAGAGGGAAAAAGA)
pOPINM	MBP_F (AAACGCCAGAAAGGTGAAATC)	Bglob-pA-R
pOPINNUSA	NUSA_F (AAAGCCGGAGCACTGATTATG)	Bglob-pA-R

8.6.9. DNA transformation into chemically competent cells

8.6.9.1. *S. cerevisiae* transformation

After confirming the constructs by sequencing, plasmids were transformed into *S. cerevisiae* FGY217 strain as described in Gietz and Schiestl⁵⁴⁰. The transformed yeast was grown on YPD-agar plates lacking uracil for the selection of uracil auxotrophs.

8.6.9.2. *E. coli* transformation

For expression in *E. coli*, sequenced plasmids were added into calcium-competent BL21(DE3) or BL21-AI cells that had been thawed on ice. After incubation on ice for 30 minute, cells were then heat shocked at 42°C for 30 second (for both strains). The cells were placed on ice for five minute before being allowed to recover in S.O.C media at 37°C and 250 rpm for 1 hour. Finally, the successful transformants were selected on LB agar plates with specific antibiotics - kanamycin and carbenicillin for pET28 and pOPIN plasmids, respectively.

8.7. General protein expression and purification protocols

The purification of recombinant proteins was carried out according to the method of Nancolas *et al.*³⁶⁸ and similar to the LicB purification as described in Section 2.8.

For *E. coli* and *S. cerevisiae* expression, 10 mL cultures were inoculated in a 30 mL universal bottle, small scale cultures as 100 mL media in a 250 mL glass conical flask, while large scale culture as 1 L media in 2.5 L conical flask. For *S. cerevisiae* expression, all cultures were incubated at 30°C with agitation at 250 rpm. Glucose and galactose solutions were prepared as described in Table 8.3. For *E. coli* expression, all cultures were incubated at 37°C with agitation at 250 rpm and supplemented with the appropriate antibiotic as described in Table 8.8. OD₆₀₀ was used to indicate the cell density. Stocks of viable cells for inoculation were kept at -80°C, as described in Chapter 2.

8.7.1. Expression in *E. coli*

8.7.1.1. Small scale protein expression and solubility analysis

The initial expression experiments were carried out in small scale cultures as outlined below.

A single colony from transformation plates was collected and used to inoculate 10 mL of LB-medium supplemented with the appropriate antibiotic. Overnight cultures were grown at 37°C with shaking at 250 rpm and were used at 1% (v/v) to inoculate 100 mL fresh LB media. Incubation was carried out at the same conditions as above until the required OD had been reached, OD₆₀₀ ~ 0.5. For BL21(DE3) strain, the cultures were subjected to a cold shock on ice for 15 minute before being induced with 1 mM IPTG and incubated overnight (~20 hour) at 20°C. For BL21-AI strain, IPTG and arabinose were added to each culture at final concentrations of 0.1 mM and 0.1%, respectively. Induced cultures were grown for 2 more hour at 37°C and 250 rpm.

Cells were harvested at 16,000 x g for 10 minute and subsequently resuspended in 1 mL of BugBuster reagent. Protein extracts were incubated for 20 minute at room temperature with gentle agitation. Soluble and insoluble fractions were separated by centrifugation at 16,000 x g for 20 minute. Supernatants were collected for further analysis, while pellets were solubilised in urea (6 M or 8 M), 6 M Gnd-HCl or 1% SDS to evaluate the formation of inclusion bodies.

For testing expression in auto-induction media, 1 mL of the overnight culture was diluted in 100 mL of LB_{AIM}, 2YT_{AIM} and SB_{AIM} (supplemented with the appropriate antibiotic) and incubated at 37°C for 24 hour. Cells were harvested and treated as above, except that the cell pellet was resuspended in 5 mL of BugBuster reagent.

8.7.1.2. Large scale protein expression

IPTG induction

For expression using BL21(DE3) cell strain: Single colonies from transformation plates were used to inoculate 10 mL LB broth supplemented with the carbenicillin as above. These start cultures were then used to inoculate 1 L of LB media supplemented with the same antibiotic. Cultures were then grown at 37°C with vigorous shaking at 250 rpm to an OD₆₀₀ of ~0.5. Once this OD had been reached, cultures were induced with 1 mM IPTG after being cold shocked on ice for 15 minute. Cultures were then grown for 20 hour at 20°C and 250 rpm.

Expression using BL21-AI cell strain: Transformed *E. coli* strains were grown overnight at 37°C with constant agitation of 250 rpm in LB media supplemented with kanamycin. 6 flasks containing 1 L LB medium supplemented with kanamycin were each inoculated with 10 mL overnight culture. Cultures were incubated at 37°C with constant agitation of 250 rpm until OD₆₀₀ was 0.7. To induce protein expression, IPTG and arabinose were added to each flask at final concentrations of 0.1 mM and 0.1%, respectively. Induced cultures were grown for 2 hour (at 37°C and 250 rpm) and OD₆₀₀ was measured to see if growth paused post-induction, which was assumed to be associated with the expected metabolic load.

Auto-induction

A single colony was collected from the same transformation plate and grown overnight in 10 mL LB media as above; These cultures were then used to inoculate 1 L of LB_{AIM}, 2YT_{AIM} and SB_{AIM} (supplemented with the appropriate antibiotic) and incubated at 37°C for 24 hour.

8.7.2. Protein expression in *S. cerevisiae*

8.7.2.1. Large scale protein expression

A single colony from *S. cerevisiae* transformation plates was used to inoculate 10 mL SC media without uracil and supplemented with 2% glucose. These small cultures were grown overnight at 30°C with constant agitation of 250 rpm. The recombinant protein expression was induced by inoculating this starter culture into 1 L SC media containing 0.1% glucose and 2% galactose and incubating for 24 hour at 30°C.

8.7.3. Isolation of membranes

Unless otherwise stated, isolation of membranes was performed as described in Chapter 2. Yeast and *E. coli* cells were lysed at 35,000 psi and 25,000 psi, respectively. The membrane pellet was resuspended in 40 mL of Membrane Buffer.

8.7.4. Detergent screening assay

200 μ L of *E. coli* and *S. cerevisiae* membrane fractions were incubated with 50 μ L of 5% detergent solution (all detergents used are listed in Table 8.17) for 1 hour at room temperature. Insoluble membranes were pelleted at 60,000 rpm for 1 hour at 4°C. Both pellet and supernatant were retained for western blotting against the recombinant V5 epitope.

Table 8.17. List of detergents used for detergent screening assay and the correspondent critical micelle concentration (CMC) ^{105,130,134}.

Detergent	CMC (mM)	% CMC
<i>n</i> -dodecyl- β -D-maltopyranoside (DDM)	0.17	0.0087
<i>n</i> -decyl- β -D-maltopyranoside (DM)	1.80	0.087
<i>n</i> -nonyl- β -D-glucoopyranoside (NG)	6.50	0.20
2,2-didecylpropane-1,3-bis- β -D-maltopyranoside (Lauryl Maltose Neopentyl Glycol, LMNG)	0.01	0.001
Thesit (hydroxypolyethoxydodecane)	0.10	0.005
5-cyclohexyl-1-pentyl- β -D-maltoside (Cymal-5)	2.50	0.12
<i>N,N'</i> -bis-(3-D-gluconamidopropyl)cholamide (Big CHAP)	1.40	0.12
3-[(3-cholamidopropyl)dimethylammonio]-2-hydroxy-1-propanesulfonate (CHAPSO)	8 – 10	0.50
Fos-choline-14 (FC-14)	0.12	0.0046
Sodium cholate	9 – 15	0.41 - 0.60
Decanoyl- <i>N</i> -methylglucamide (MEGA-10)	6 – 7	0.21
<i>n</i> -octyl- β -D-thioglucoopyranoside (OTG)	9	0.28
Sodium dodecyl sulphate (SDS)	6 – 8	0.17 - 0.23
Triton 10%	0.20	0.01
Tween-20	0.06	0.0074

8.7.5. Membrane fraction solubilisation

Unless otherwise stated, the membrane fractions of large scale expression cultures were solubilised as described in Section 2.8 by introducing different detergents into Column Buffer (Table 8.18).

Table 8.18. List of detergents used to solubilise *E. coli* and *S. cerevisiae* membrane fractions solubilisation.

Solubilised agent	% used for solubilisation	% used in working buffers
Thesit	2.0%	0.02%
Cymal-5	2.4%	0.24%
FC-14	1.0%	0.10%
DDM	1.0%	0.05%

8.7.6. Purification of recombinant protein

Unless otherwise stated, protein purification was performed by IMAC as described in Section 2.8. All buffers (Equilibration, Wash and Elution) were supplemented with the appropriate detergent at working concentration. Molecular weight cut-off concentrators of 30 kDa, 50 kDa and 100 kDa were used to concentrate purified AATs. Concentration was performed until the desired volume of 1 mL was achieved. Aliquots of purified proteins were used directly or stored in -80°C.

8.8. Total protein concentration determination

Determination of total protein concentration and pure recombinant proteins were performed as described in Section 2.9. Molecular weights and extinction coefficients of all alcohol acyltransferases studied here are listed in Table 8.19.

Table 8.19. Theoretical molecular weight and extinction coefficient (ϵ) for each alcohol acyltransferase protein calculated based on ExPASy ProtParam online tool.

Fusion protein (plasmid)	Protein	Molecular weight (Da)	Extinction coefficient (M⁻¹.cm⁻¹)
POI - V5 - (His)₁₀ (pYES2)	Atf1	64,799	83,770
	Atf1 Δ N	59,253	80,790
	Atf1 Δ C	62,354	82,280
	Atf1 Δ N/ Δ C	56,808	79,300
	FaAAT1	54,464	68,870
	MAAT	55,559	34,380
	VpAAT1	55,394	34,380
(His)₆-3C - POI - V5 (pOPINF)	Atf1	65,254	83770
	FaAAT1	54,919	68,870
	MAAT	56,014	34,380
	VpAAT1	55,849	34,380
(His)₆ – MBP - 3C - POI - V5 (pOPINM)	Atf1	105,807	150,120
	FaAAT1	95,472	135,220
	MAAT	96,567	100,730
	VpAAT1	96,402	100,730
(His)₆ – NusA - 3C - POI -V5 (pOPINNUSA)	Atf1	120,005	114,710
	FaAAT1	109,671	99,810
	MAAT	110,766	65,320
	VpAAT1	110,600	65,320

8.9. Biophysical techniques

8.9.1. Western blotting

Western blotting was performed with nitrocellulose membranes that were prepared by soaking for 1 hour in Transfer Buffer. The transfer was performed for 90 minute at 25 V (approximately 300 mA). The membrane was blocked with Blocking Buffer for 30 minute and then incubated with monoclonal anti-V5 conjugated with horseradish peroxidase (HRP) for 1 hour. Finally, after rinsed with PBS/T Buffer, the membrane was developed by chemiluminescence was detected using the LumiGLO reagents on an Amersham™ Imager 600.

8.9.2. Size exclusion chromatography

SEC was performed with a 24 mL Superdex S200 10/300 column as described in Section 2.10.4. The column was equilibrated in Column Buffer supplemented with the appropriate detergent before 1 mL sample was loaded and run at 0.5 mL/minute.

8.9.3. SEC-MALS and circular dichroism

SEC-MALS and circular dichroism procedures were performed as described in Section 2.10.5. The buffer used for SEC-MALS and circular dichroism of Atf1 recombinant protein was constituted by Column Buffer supplemented with 0.1% FC-14.

8.10. Enzymatic activity assays

8.10.1. Acyltransferase activity

A coupled assay was used to monitor acyltransferase activity at 25°C (see Figure 9.7). The assay was performed as described in Knight *et al.*⁵⁴¹. Briefly, purified Atf1 WT or Atf1 $\Delta N/\Delta C$ were incubated with 400 μ L of reaction buffer, containing 50 mM sodium phosphate at pH 7.4, 0.4 mM NAD⁺, 0.4 mM TPP, 2 mM α -ketoglutarate, 1 mM EDTA and 0.39 U α -ketoglutarate dehydrogenase. The compounds octanoyl-CoA (substrate) and free CoA (positive control) were at 2.5 mM and isoamyl alcohol at 9.2 mM. Measurements were performed using a Cary 60 UV-VIS instrument. The produced NADH concentration (μ M) was plotted against time (minute) using GraphPad Prism software.

8.10.2. Thioesterase assay

Thioesterase activity was assessed by incubating 250 μ M *p*-nitrophenylthioacetate (*p*-NPTA) with purified Atf1 in 1 mL reaction buffer (50 mM sodium phosphate buffer at pH 7.4, 150 mM NaCl and 5% glycerol) supplemented with detergent at working concentration, at 25°C. The concentration of the thioesterase activity product, *p*-nitrophenol, was determined by absorbance at 400 nm relative to a standard curve.

Chapter 9. Expression and characterisation of AATases

Chapter 9 is a combination of results from both yeast Atf1 as well as for AATases from fruits (melon, papaya and strawberry). Specific objectives are outlined in each subchapter.

9.1. Engineering yeast alcohol acetyltransferase Atf1 for structural biology

9.1.1. Introduction and aim

Atf1 is one of a small family of alcohol acetyltransferases in the yeast *Saccharomyces cerevisiae*. This enzyme catalyses the synthesis of acetate esters from alcohol and acetyl-CoA. Such esters play an important role in the flavour and aroma of fermented beverages, including beer, sake and wine. However, the catalytic mechanism of Atf1 is not well understood. Atf1 is a peripheral membrane protein, and previous studies have suggested that amphipathic helices at the N- and C-termini localise Atf1 to the endoplasmic reticulum and lipid particles. Although Atf1 has been recombinantly expressed and purified, Atf1 tends to aggregate, thought to be driven by the amphipathic domains³⁶⁸.

The present work aimed to engineer Atf1 protein that could be suitable for X-ray crystallography. The hypothesis tested was that if the hydrophobic N- and C-terminal domains of Atf1 drive protein aggregation (Figure 9.1), then removing these domains should limit aggregation and so provide more appropriate targets for crystallography. In order to test this, WT recombinant Atf1 and two truncation variants were expressed and purified using the same methodology as described in Nancolas *et al.*³⁶⁸. Atf1 recombinant proteins were expressed in *S. cerevisiae* as previously and in *E. coli* as an alternative approach that might provide higher protein yields. All proteins were then purified and characterised. The methodology followed during the project is represented in Figure 9.2.

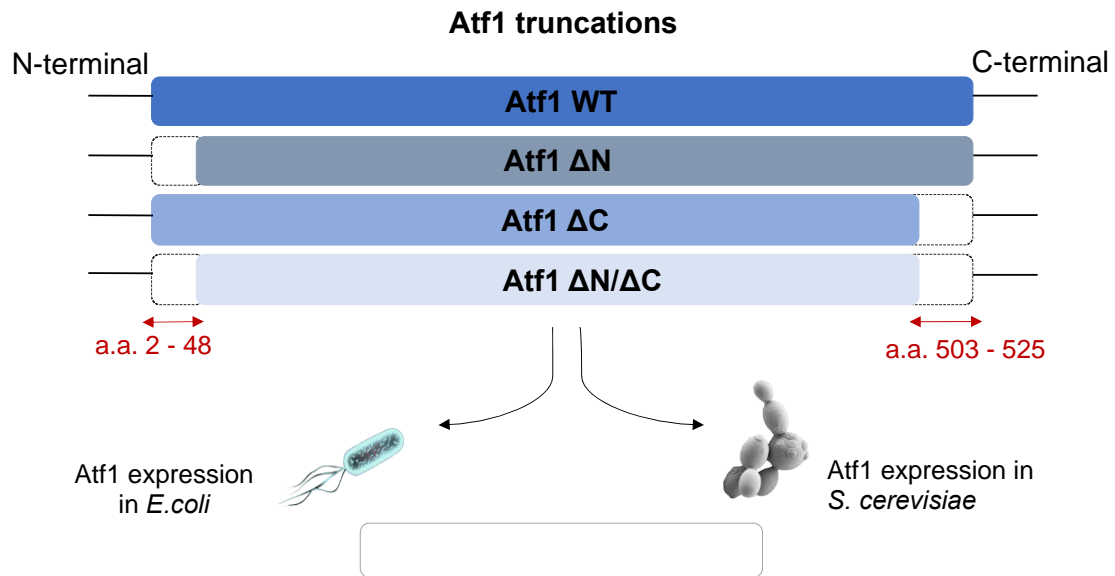


Figure 9.2. Representation of the project workflow and methodology.

Three truncation variants were constructed featuring deletions of the N-terminal, C-terminal, or both. All constructs were expressed in both *E. coli* and *S. cerevisiae*. The amino acids (a.a.) deleted at the N-terminal were between positions 2 to 48, and the C-terminal truncation was between positions 503 to 525. Variants were transformed into both yeast and bacterial plasmids for recombinant expression and purification.

9.1.2. Results

9.1.2.1. *atf1* gene deletions and protein expression in *S. cerevisiae*

The first cloning strategy intended to introduce the terminal truncations already identified by Lin and Wheeldon⁴⁸⁵. This strategy prevents the membrane association of Atf1 – essentially turning this peripheral membrane protein into a soluble cytosolic protein. Truncations were introduced into a previously-cloned *atf1* gene via PCR, with the success of this approach confirmed by sequencing (Figure 9.3). Otherwise, sequences were identical to WT Atf1, including the V5 and His-tag for detection and purification, respectively.

Atf1	MNEIIDEKNQAPVQQECLKEMIQNGHARRMGSVEDLYVALNRQNLRYRNFCTYGELSDYCTR
Nterm_del	-----MCTYGELSDYCTR
Cterm_del	MNEIIDEKNQAPVQQECLKEMIQNGHARRMGSVEDLYVALNRQNLRYRNFCTYGELSDYCTR
Double_del	-----MCTYGELSDYCTR :*****
Atf1	DQLTLALREICLKNPTLLHIVLPTRWPNHENYYSSEYYSRPHPVHDYISVLQELKLSGV
Nterm_del	DQLTLALREICLKNPTLLHIVLPTRWPNHENYYSSEYYSRPHPVHDYISVLQELKLSGV
Cterm_del	DQLTLALREICLKNPTLLHIVLPTRWPNHENYYSSEYYSRPHPVHDYISVLQELKLSGV
Double_del	DQLTLALREICLKNPTLLHIVLPTRWPNHENYYSSEYYSRPHPVHDYISVLQELKLSGV *****
Atf1	VLNEQPEYSAVMKQILEEFKNSKGSYTAKIFKLTTLTIPIYFGPTGPSWRLICLPEEHTE
Nterm_del	VLNEQPEYSAVMKQILEEFKNSKGSYTAKIFKLTTLTIPIYFGPTGPSWRLICLPEEHTE
Cterm_del	VLNEQPEYSAVMKQILEEFKNSKGSYTAKIFKLTTLTIPIYFGPTGPSWRLICLPEEHTE
Double_del	VLNEQPEYSAVMKQILEEFKNSKGSYTAKIFKLTTLTIPIYFGPTGPSWRLICLPEEHTE *****
Atf1	KWKKFIFVSNHCMSDGRSSIHFFHDLRDELNNIKTPPKLDYIFKYEEDYQLLRKLEPEPI
Nterm_del	KWKKFIFVSNHCMSDGRSSIHFFHDLRDELNNIKTPPKLDYIFKYEEDYQLLRKLEPEPI
Cterm_del	KWKKFIFVSNHCMSDGRSSIHFFHDLRDELNNIKTPPKLDYIFKYEEDYQLLRKLEPEPI
Double_del	KWKKFIFVSNHCMSDGRSSIHFFHDLRDELNNIKTPPKLDYIFKYEEDYQLLRKLEPEPI *****
Atf1	EKVIDFRPPYLFIPKSLLSGFIYNHLRFSSKGVCMRMDVEKTDVVTEIINISPTFEFQA
Nterm_del	EKVIDFRPPYLFIPKSLLSGFIYNHLRFSSKGVCMRMDVEKTDVVTEIINISPTFEFQA
Cterm_del	EKVIDFRPPYLFIPKSLLSGFIYNHLRFSSKGVCMRMDVEKTDVVTEIINISPTFEFQA
Double_del	EKVIDFRPPYLFIPKSLLSGFIYNHLRFSSKGVCMRMDVEKTDVVTEIINISPTFEFQA *****
Atf1	IKANIKSNIQGKCTITPFLHVCWFVSLHKWGKFFKPLNFEWLTDFIPADCRSQLPDDDE
Nterm_del	IKANIKSNIQGKCTITPFLHVCWFVSLHKWGKFFKPLNFEWLTDFIPADCRSQLPDDDE
Cterm_del	IKANIKSNIQGKCTITPFLHVCWFVSLHKWGKFFKPLNFEWLTDFIPADCRSQLPDDDE
Double_del	IKANIKSNIQGKCTITPFLHVCWFVSLHKWGKFFKPLNFEWLTDFIPADCRSQLPDDDE *****
Atf1	MRQMYRYGANVGFIDFTPWISEFDMNDNKENFWPLIEHYHEVISEALRNKKHLHGLGFNI
Nterm_del	MRQMYRYGANVGFIDFTPWISEFDMNDNKENFWPLIEHYHEVISEALRNKKHLHGLGFNI
Cterm_del	MRQMYRYGANVGFIDFTPWISEFDMNDNKENFWPLIEHYHEVISEALRNKKHLHGLGFNI
Double_del	MRQMYRYGANVGFIDFTPWISEFDMNDNKENFWPLIEHYHEVISEALRNKKHLHGLGFNI *****
Atf1	QGFVQKYVNI DKVMCDRAIGKRRGGTLLSNVGLFNQLEEPDAKYSICDLAFGQFQGSWHQ
Nterm_del	QGFVQKYVNI DKVMCDRAIGKRRGGTLLSNVGLFNQLEEPDAKYSICDLAFGQFQGSWHQ
Cterm_del	QGFVQKYVNI DKVMCDRAIGKRRGGTLLSNVGLFNQLEEPDAKYSICDLAFGQFQGSWHQ
Double_del	QGFVQKYVNI DKVMCDRAIGKRRGGTLLSNVGLFNQLEEPDAKYSICDLAFGQFQGSWHQ *****
Atf1	AFSLGVCSTNVKGMNIVVASTKNNVVGVSQESLEELCSIYKALLLGP SRGPFEGKPIPNPLL
Nterm_del	AFSLGVCSTNVKGMNIVVASTKNNVVGVSQESLEELCSIYKALLLGP SRGPFEGKPIPNPLL
Cterm_del	AFSLGVCSTNVKGMNIVVASTK-----SRGPFEGKPIPNPLL
Double_del	AFSLGVCSTNVKGMNIVVASTK-----SRGPFEGKPIPNPLL *****
Atf1	GLDSTRTGHHHHHHHHHHH
Nterm_del	GLDSTRTGHHHHHHHHHHH
Cterm_del	GLDSTRTGHHHHHHHHHHH
Double_del	GLDSTRTGHHHHHHHHHHH *****

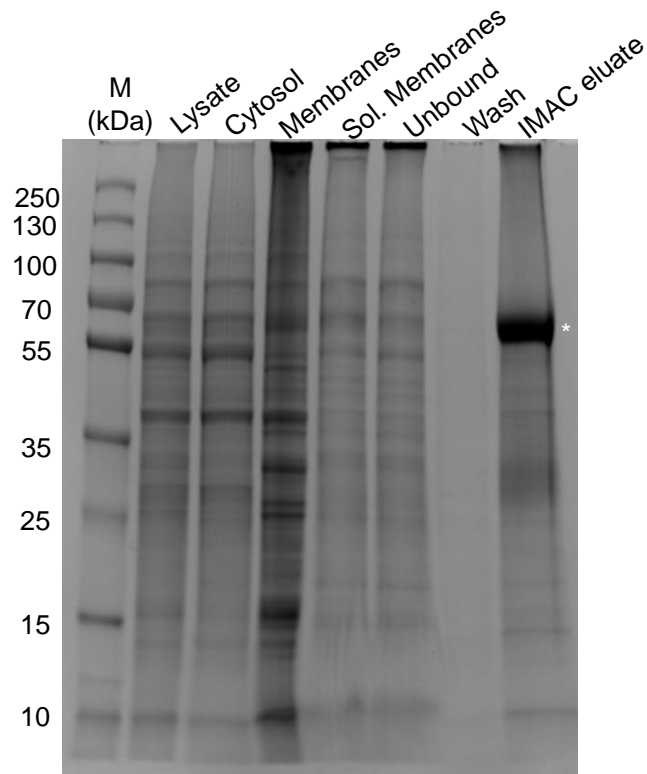
Figure 9.3. Multiple sequence alignment of WT and truncated variants of Atf1.

The N-terminal and C-terminal domains removed in the deletion variants are highlighted in grey.

Plasmids carrying these mutant *atf1* genes were used to transform the *S. cerevisiae* strain FGY217 prior to protein expression tests. The plasmid containing the *atf1* genes has two additional components necessary for its purification and detection. The first is a 10-residue polyhistidine tag (His₁₀) to perform protein purification by immobilised metal affinity chromatography (IMAC). The second is a V5 epitope to allow protein detection through the anti V5 antibody conjugated with horseradish peroxidase (HRP). The recombinant WT Atf1 protein was successfully expressed by *S. cerevisiae* (Figure 9.4), as already demonstrated in Nancolas *et al.* ³⁶⁸. A series of ultracentrifugation steps were performed to separate cell fractions, and western blotting was used to confirm that WT Atf1 protein was localised in sedimenting membranes (Figure 9.4). Membrane fractions were solubilised with the detergent Thesit, and soluble protein-lipid-detergent complexes were formed. Insoluble material was removed by centrifugation, and the soluble supernatant used for western blotting. Figure 9.4 shows that Atf1 was found within these soluble membranes and so had been efficiently solubilised by Thesit. The soluble membranes were applied to the Ni²⁺ affinity resin, and Atf1 protein was bound tightly, being depleted from the unbound fraction (Figure 9.4). After washing, purified Atf1 was ultimately eluted with imidazole.

The purified protein was observed as a single band on SDS-PAGE after Coomassie staining (Figure 9.4A). This band ran at a position close to 65 kDa, the calculated theoretical molecular weight of Atf1. This band was also confirmed as Atf1 by western blotting (Figure 9.4B). The purification yield was 2 mg Atf1 per litre yeast expression culture. All of these results are in agreement with the previous publication on this recombinant protein ³⁶⁸.

(A)



(B)

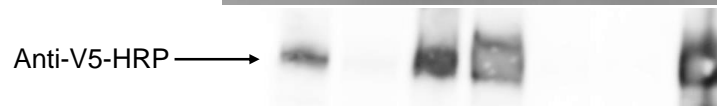


Figure 9.4. Expression and purification of Atf1 in *Thesit*.

M, molecular weight markers in kDa as shown; lane headings are relevant to cell fractionation (lysate, cytosol, membranes, soluble membranes) or affinity purification (unbound, wash and purified Atf1) **(A)** Coomassie-stained SDS-PAGE gel of cell fractions and purified fractions from Ni²⁺ column. Atf1 protein ran at the expected molecular weight of 65 kDa (*). 20 µg of protein were loaded for all fractions except for the purified Atf1 (10 µg). **(B)** Western blotting of cellular fractions confirmed the membrane localisation of Atf1 protein (10 µg of protein were loaded).

The same methodology was used to purify Atf1 ΔN, ΔC and ΔN/ΔC. Western blot analysis of lysate, cytosol and membranes fractions showed, against expectation, that all variant proteins were still being localised in the sedimenting membrane fraction (Figure 9.5). This was not consistent with the cytosolic expression of truncated proteins observed by Lin and Wheeldon⁴⁸⁵. The reason for this discrepancy is unclear, but maybe because that study utilised cell imaging rather than centrifugal fractionation or because they made use of the constitutive PGK1 promoter for expression rather than the inducible GAL promoter of the pYES2 vector.

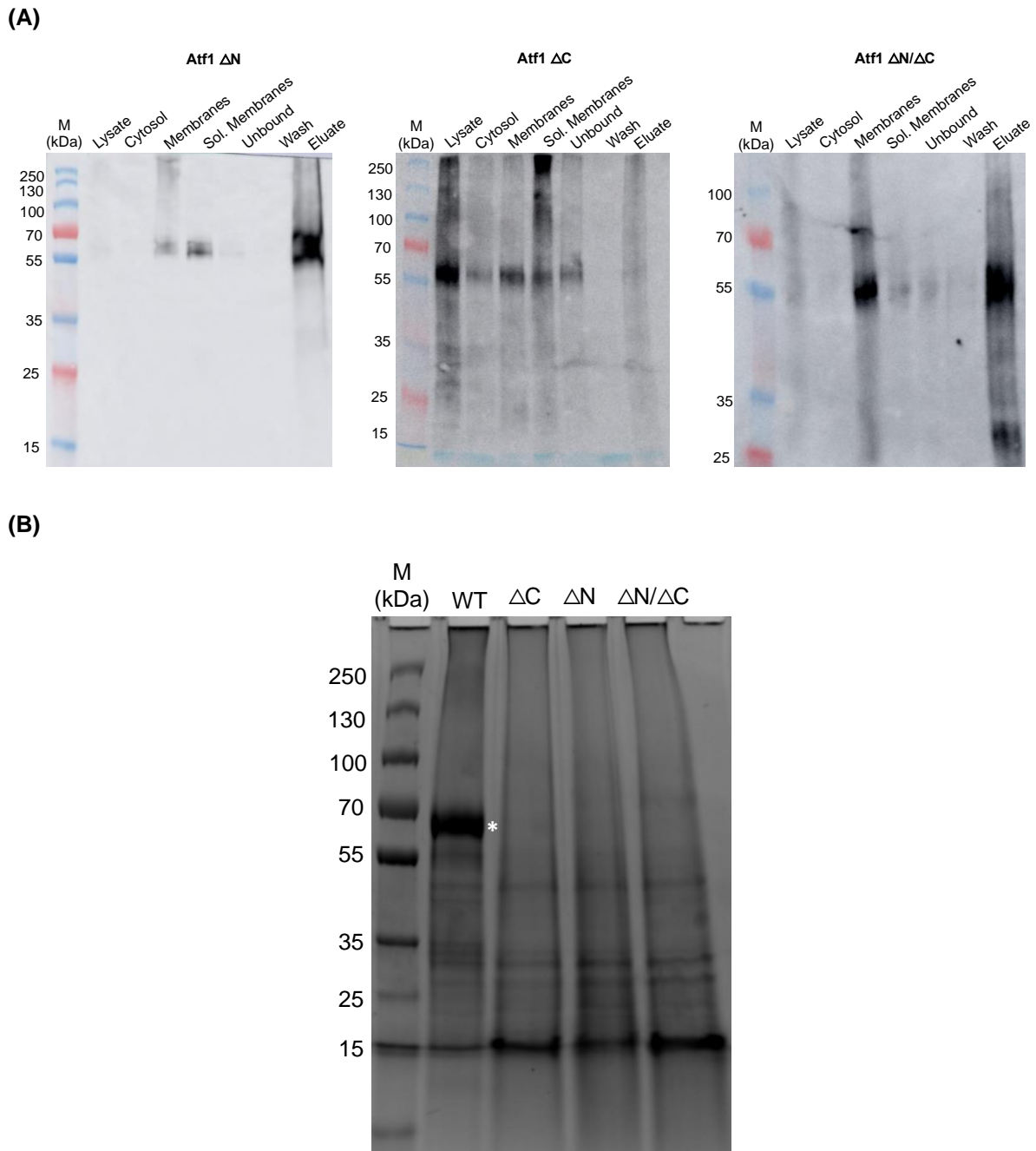


Figure 9.5. Truncated Atf1 proteins were still localised in membrane fraction.

(A) Western blotting of variant proteins in lysate, cytosol and membranes confirmed that Atf1 with truncations in N- and C-terminal was expressed into sedimenting membranes. 20 μg of protein were loaded. **(B)** Coomassie-stained SDS-PAGE gel of purified Atf1 and variants as shown. 20 μg of protein were loaded.

Each of the variant proteins was purified as described for WT Atf1. However, it appeared that variant proteins are only weakly overexpressed compared to the WT protein (Figure 9.5; Table 9.1). This was again unexpected given the study of Lin and Wheeldon⁴⁸⁵ in which expression levels of identical truncations remained at the same level as wild-type. It may be that this arises

because Atf1 is endogenous to *S. cerevisiae*, so that quality control checkpoints mark these aberrant polypeptides for degradation⁹⁸. The yield of the variant proteins was so low that although they could be detected by Western blotting, no bands were visible on Coomassie-stained SDS-PAGE gels (Figure 9.5). This strongly suggests that no significant expression of variant proteins was achieved.

Table 9.1 Purification yield for the Atf1 truncated proteins after IMAC.

Protein	Yield (mg/L of culture)
Atf1	2.04
Atf1 Δ C	0.03
Atf1 Δ N	0.05
Atf1 Δ N/ Δ C	0.02

9.1.2.2. Size exclusion of Atf1 WT and Atf1 Δ N/ Δ C

Even though truncated proteins were not well expressed, further characterisation was attempted on the small amount of recovered protein. Size exclusion chromatography (SEC) was performed for purified WT Atf1 and Atf1 Δ N/ Δ C. SEC of purified WT resolved a single peak that eluted in the void volume of the column (>600 kDa; Figure 9.6A). It is known that Atf1 is aggregation-prone, and the poor SEC profile was unsurprising, although previous work³⁶⁸ did achieve slightly better results, resolving a low molecular weight peak as well as a void peak. In any case, this was clear evidence that the WT protein tends to form non-specific hydrophobic interactions and aggregate in the detergent used, and so is not a suitable starting point for crystallography. However, these interactions appeared to be reversible since SEC fractions were loaded in an SDS-PAGE, and a band on the expected size was observed (65 kDa). This phenomenon has also been reported by Knight *et al.* for the AATase Eht1⁵⁴¹ and also by Nancolas *et al.* for Atf1³⁶⁸.

The SEC profile of Atf1 Δ N/ Δ C showed only a void peak (Figure 9.6B); it seemed that no protein was eluted from the column, or it was at undetectable low concentrations. Very faint bands were observed on a Coomassie-stained SDS-PAGE gel of early fractions. This confirms the low yield and unsuitable characteristics of the Atf1 deletion variant.

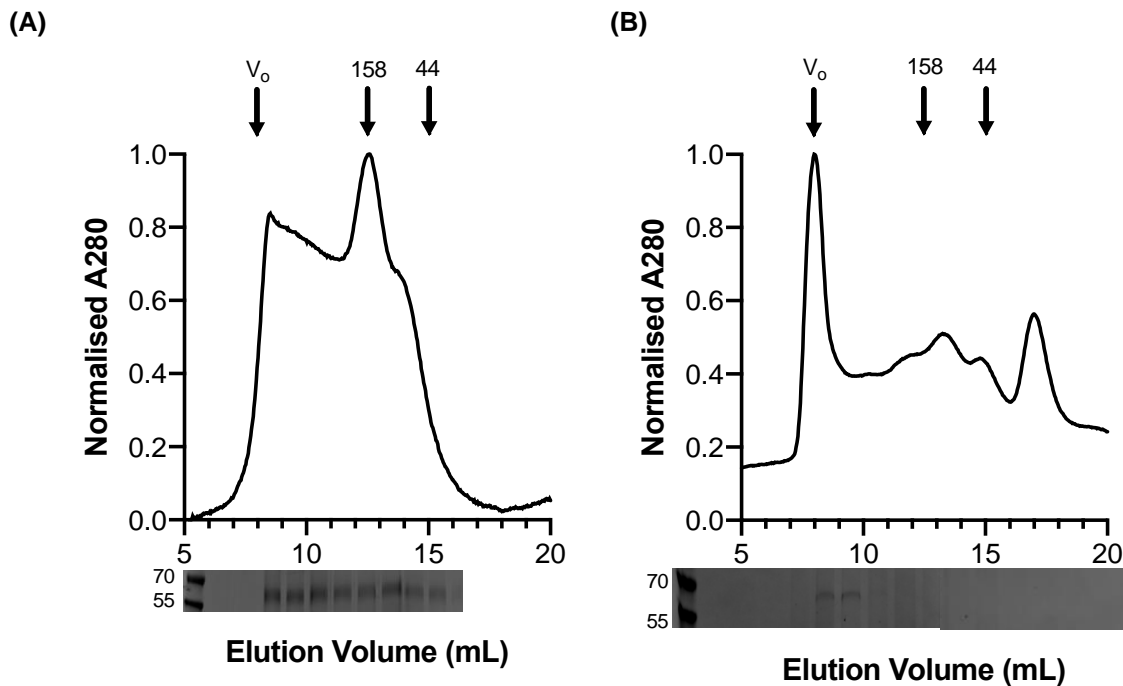


Figure 9.6. Size exclusion chromatogram of Atf1 variants.

(A) Purified WT Atf1 and (B) Atf1 $\Delta N/\Delta C$ in Thesit. Void volume is represented as V_0 . Underneath each SEC profile is a Coomassie-stained SDS-PAGE gel loaded with samples of each fraction. 20 μg of protein were loaded. Arrows show the elution volume of two standard proteins, with molecular weight as shown.

9.1.2.3. Acyltransferase activity of WT and Atf1 $\Delta N/\Delta C$

A coupled assay was used to determine the activity of Atf1 WT and Atf1 $\Delta N/\Delta C$ proteins, as defined by Dunn *et al.*⁵⁴². The acyl transfer reaction catalysed by Atf1 generates free coenzyme A (CoA), which is the substrate for the second reaction catalysed by α -ketoglutarate dehydrogenase. This coupled enzyme reduces NAD^+ to NADH, generating an absorbance signal at 340 nm (Figure 9.7).

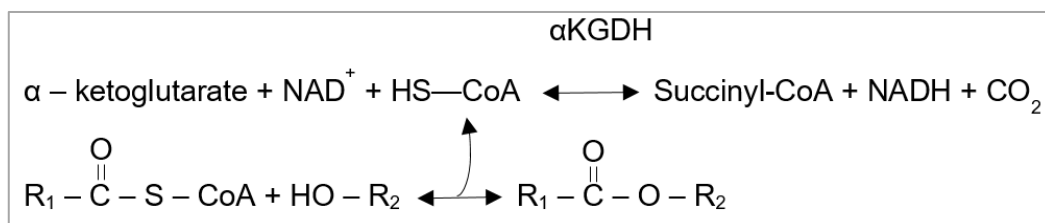


Figure 9.7. Coupled reaction used for acyltransferase activity.

Atf1 activity is linked to α -ketoglutarate dehydrogenase (αKGDH), which converts NAD^+ to NADH generating a spectroscopic signal detectable at 340 nm.

The coupled assay allows the measurement of the initial (linear) enzyme rate and avoids the drawbacks of single time-point assays used elsewhere in the literature ^{442,470,513,524,529}. The results are shown in Figure 9.8. All experiments were carried out at 2.5 mM octanoyl-CoA, 9.2 mM of isoamyl alcohol and 8 μ g of Atf1 WT or Atf1 Δ N/ Δ C. Positive control was provided by using 2.5 mM CoA-SH, and negative controls were performed by omitting an assay component. Overall, the rate of formation of NADH is slower than previously reported ³⁶⁸. This may be because of the increased aggregation of this protein preparation compared with the samples in that study. Concerning WT reactions, it was observed that the formation of NADH is only slightly higher in the presence of the alcohol (21.6 μ M over 10 minute of reaction) than in its absence (18 μ M over 10 minute of reaction). This result was expected from previous work showing that the presence of the alcohol cosubstrate has little effect on acetyl-CoA hydrolysis *in vitro* ³⁶⁸. As expected, protein activity was still detected over negative controls. Atf1 Δ N/ Δ C variant appeared to be inactive.

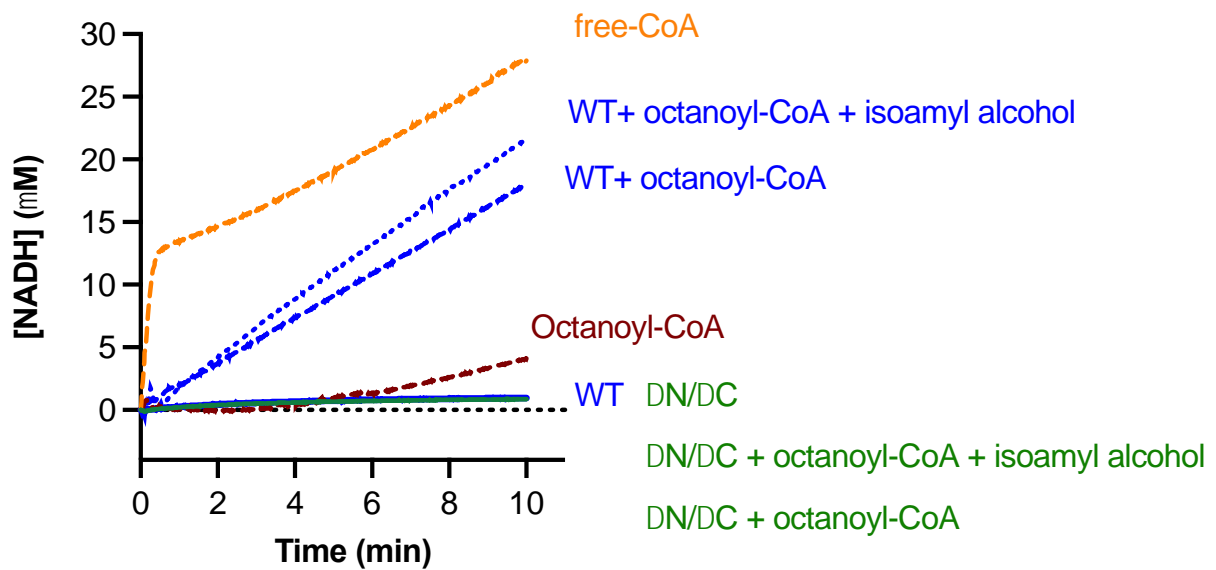


Figure 9.8. Coupled enzyme assay for Atf1 WT and Atf1 Δ N/ Δ C activities.

The turnover of octanoyl-CoA by Atf1 monitored by the coupled assay. A positive control was also performed using only CoA-SH (*free-CoA*) as a direct substrate for α KGDH. The formation of NADH relative to reactions with octanoyl-CoA, isoamyl alcohol, WT and Atf1 Δ N/ Δ C were also considered.

9.1.2.4. *atf1* gene deletions and protein expression in *E. coli*

Given the unexpected failure to express Atf1 deletions in yeast, recombinant expression was trialled in *E. coli*. Recombinant Atf1 has already been expressed in *E. coli*, although it is generally found to form insoluble aggregates with low enzymatic activity^{489,543}. However, the truncated proteins described here have never been expressed and might be more amenable since the deletion of the hydrophobic domains could prevent aggregation. Here we aim to express truncated Atf1 in *E. coli* as an alternative approach to the yeast expression system.

Truncated Atf1 was cloned into pET28 and expressed in *E. coli* BL21-AI. The success of this cloning was confirmed by sequencing. Small scale expression trials were then performed. 100 mL cultures were induced with 0.1 mM IPTG and 0.1% arabinose, and cells were lysed with BugBuster reagent. The lysates were analysed by SDS-PAGE and Western blotting to evaluate protein expression. The results are shown in Figure 9.9. No protein expression could be detected.

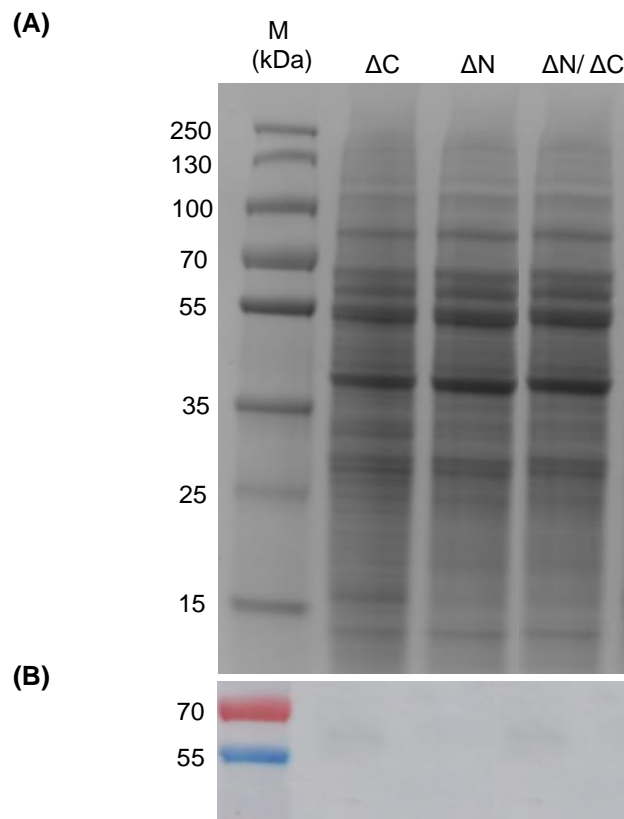


Figure 9.9. *E. coli* does not express Atf1 truncations.

(A) Coomassie-stained SDS-PAGE gel of lysate. 20 μ g of protein were loaded. **(B)** Western blotting of variant proteins in the lysate. 10 μ g of protein were loaded.

9.1.3. Discussion

Atf1 from *S. cerevisiae* is of interest because of its role in the flavour of fermented beverages and a potential metabolic engineering tool. In order to establish the complete catalytic mechanism of this protein, the experimental crystal structure would be a significant achievement. In this context, truncations in N-terminal (Atf1 Δ N), C-terminal (Atf1 Δ C) and in both termini (Atf1 Δ N/ Δ C) of Atf1 were performed to try and generate a construct that could be suitable for structural biology. Lin and Wheeldon previously demonstrated that the deletion of these residues in either terminal released the protein to cytosol⁴⁸⁵, and therefore, we supposed that Atf1 carrying these mutations could potentially become soluble and suitable for crystal trials.

Here, Atf1 WT and Atf1 truncated proteins were recombinantly expressed in recombinant yeast and bacterial hosts. The results unexpectedly showed that truncated proteins were expressed only at very low levels. Where expression was detectable in yeast strains, the truncation variants still sedimented with cell membranes; therefore, other parts of Atf1 protein could be involved in membrane association, or small amounts of truncated protein could be escaping cellular quality control mechanisms that otherwise tag them for degradation.

This data suggested that the approach considered to achieve the Atf1 crystal structure was not suitable and should be discontinued. Instead, further work should focus on finding suitable detergents or alternative protocols for successfully purifying WT Atf1 from recombinant yeast strains and limiting the aggregation of this protein in the following purification. Such an approach is described in the next section of the thesis.

9.2. Detergent screen of Atf1 expressed in *S. cerevisiae*

9.2.1. Introduction and aim

Considering the results described in 9.1 above, the main aim of this section was to find a better detergent that could extract and stabilise Atf1 as a monodisperse sample and, consequently, more suitable for biophysical and structural studies. Atf1 was expressed in *S. cerevisiae* as before, and the purification procedure was subsequently adapted.

9.2.2. Results

9.2.2.1. Detergent screening

A detergent screen was employed to try and identify conditions for purifying monodisperse Atf1. Ultracentrifugation Dispersity Sedimentation (UDS) is a rapid and high-throughput screen widely used in membrane protein studies. The protein is diluted into a panel of detergents and subject to centrifugation at 60,000 rpm; this will pellet aggregated protein while soluble and non-aggregated protein will be retained in the supernatant ¹⁰¹. A panel of 15 different detergents was employed, using the V5 antibody to quantify the amount of solubilised Atf1. The results are shown in Figure 9.10.

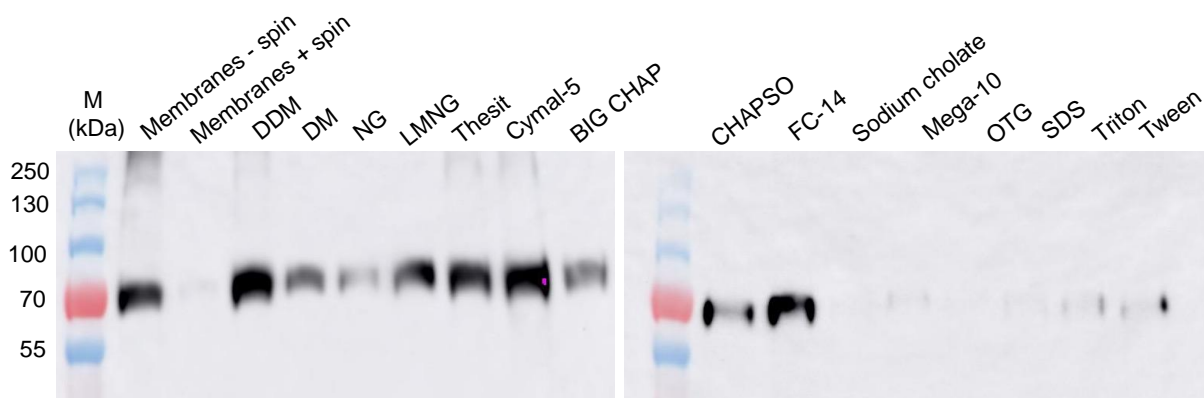


Figure 9.10. Detergent screening assay of Atf1 protein in 15 different detergents.

Membrane – spin, untreated membranes prior to centrifugation. *Membrane + spin*, supernatant from the centrifugation of untreated membranes. Otherwise, membranes were treated with detergents, as shown. Images are qualitative Western blots with anti-V5-HRP, subject to the same exposure time. 10 µg of protein were loaded.

The detergent screening revealed that the most promising protein-detergent combinations were DDM, LMNG, Thesit, Cymal-5 and FC-14. These results were in close agreement with previous data showing that Atf1 can be solubilised in Thesit and DDM ³⁶⁸. The fos-choline detergents are known to be very effective at solubilising membrane proteins but can disrupt protein structure somewhat similar to SDS. LMNG and Cymal-5 are ‘gentler’ detergents that have been successfully used with membrane proteins and so are encouraging new ‘hits’ from the screen.

9.2.2.2. Large scale expression and purification of Atf1 protein

Based on the information from detergent screening, Atf1 was purified by Ni⁺ IMAC in the detergents Cymal-5 and FC-14. The purified protein was analysed by SDS-PAGE, Western blotting and SEC. FC-14 gave the highest yields of 1.4 mg Atf1 per litre culture (Table 9.2), and both detergents gave a single band at the expected molecular weight of 65 kDa on SDS-PAGE.

The SEC profile of Atf1 solubilised with Cymal-5 showed only a large peak at the void volume, suggesting significant aggregation of this purified protein (Figure 9.11A). In contrast, the SEC profile of Atf1 in FC-14 (Figure 9.11B) gave a single peak at an elution volume of 10.9 ml. This single peak corresponds to an apparent molecular weight of 283 kDa (based on Equation 2.2), although such estimates are unreliable for membrane proteins because of the influence of micelle shape on protein elution. Fractions corresponding to this peak were concentrated and used for further characterisation by SEC-MALS and circular dichroism.

Table 9.2. Yields obtained from Atf1 purification using Cymal-5 and FC-14 as solubilising agents.

Detergent	% used for solubilisation	% used in working buffers	Yield after IMAC (mg/L of culture)
Cymal-5	2.4%	0.24%	0.88
FC-14	1.0%	0.10%	1.40

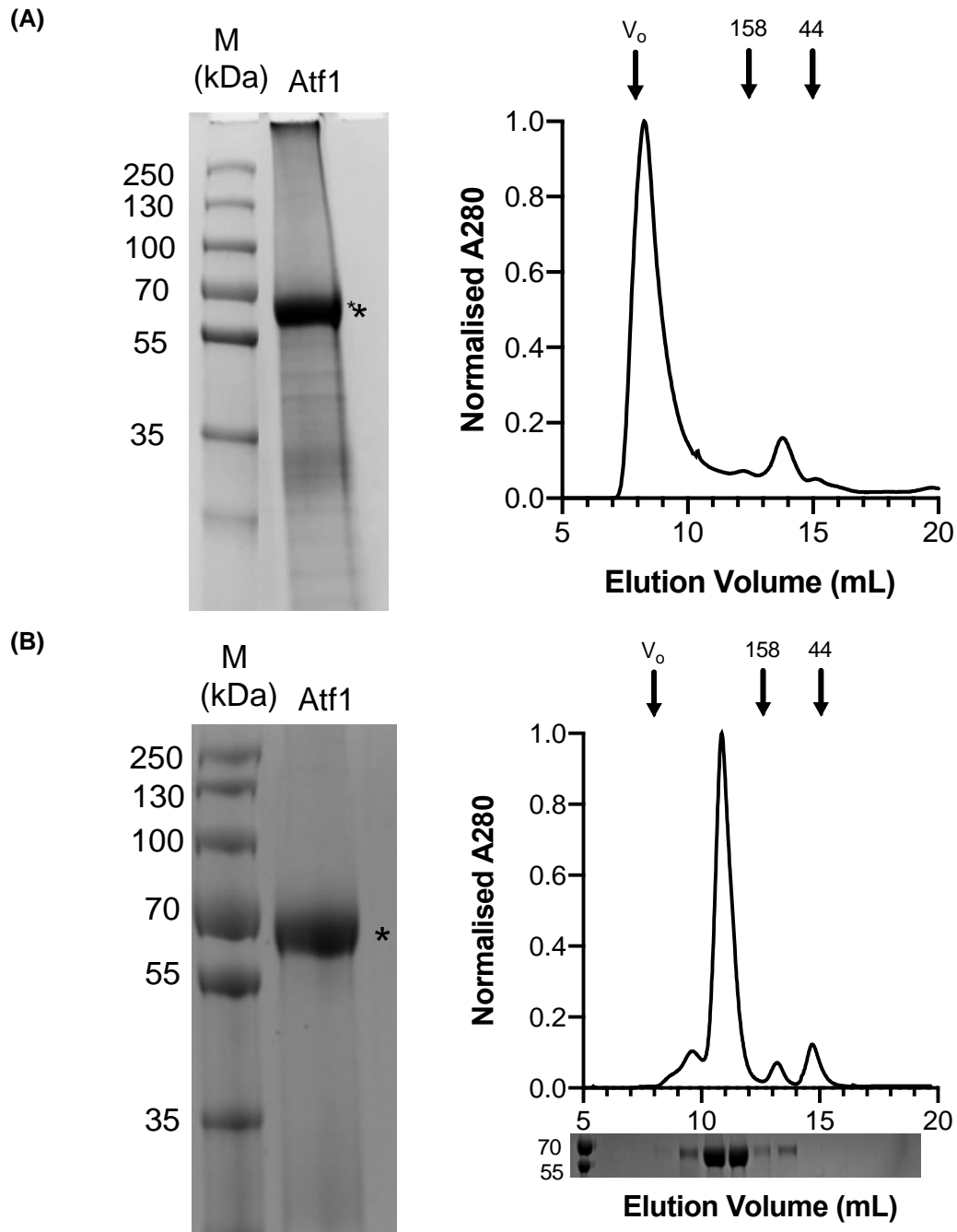


Figure 9.11. Large scale expression of Atf1 in Cymal-5 and FC-14.

(A) Coomassie-stained SDS-PAGE of 20 μ g purified Atf1 in Cymal-5. The right panel shows SEC of Atf1 purified in Cymal-5. **(B)** Coomassie-stained SDS-PAGE gel of 20 μ g purified Atf1 in FC-14. The right panel shows SEC of Atf1 purified in FC-14. V_0 indicates the void volume.

9.2.2.3. Biophysical characterisation and activity assays of Atf1 protein

SEC-MALS

Figure 9.12 shows the SEC-MALS trace of purified Atf1 in FC-14. This data can be analysed to provide a shape-independent estimate of molecular weight. The estimated molecular weight of Atf1 in FC-14 based on the three-detector method of Slotboom *et al.*³⁷¹ was 43 kDa. The theoretical molecular weight of Atf1 is around 65 kDa, so the MALS analysis implies that Atf1 is a monomer in FC-14. Other SEC-MALS results elsewhere in this thesis suggest that MALS analysis consistently slightly underestimates molecular weight for unknown reasons.

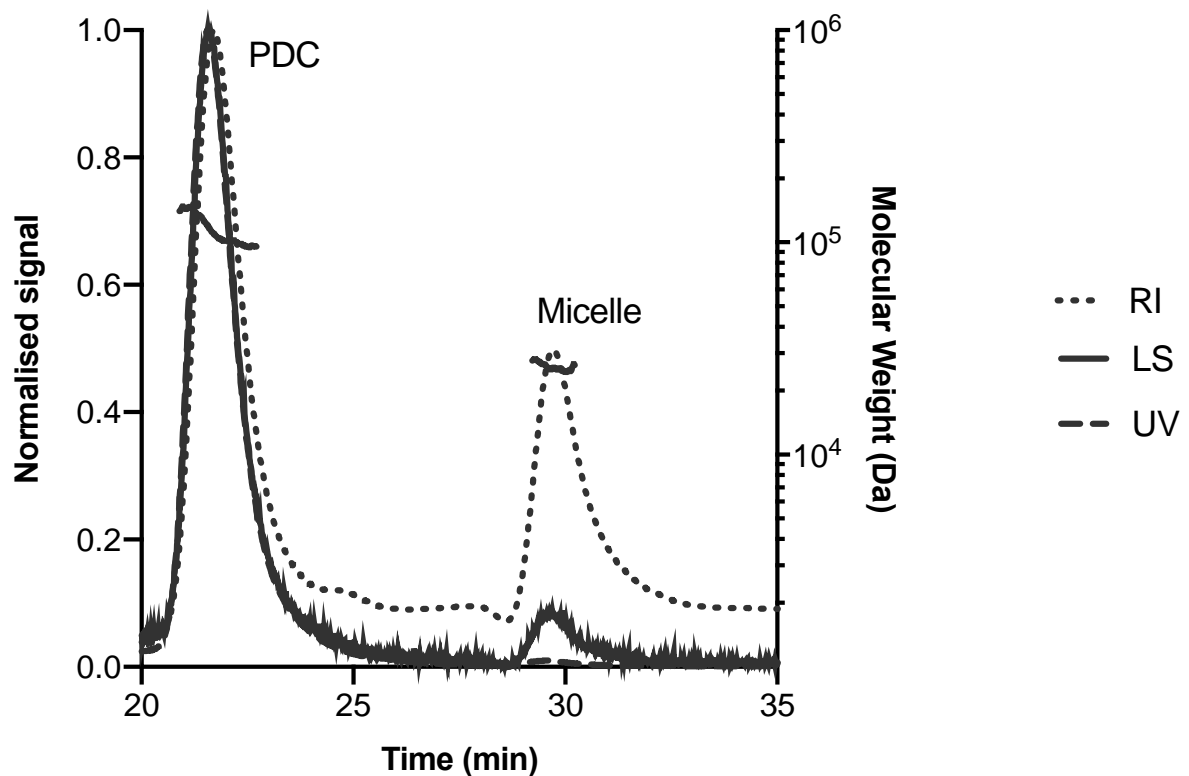


Figure 9.12. SEC-MALS analysis of the Atf1-FC-14 protein-detergent complex.

The chromatogram shows the normalised readings of light scattering (LS), absorbance at 280 nm (UV) and refractive index (RI) between the elution time of 20 and 35 minute. Black lines represent the molecular mass calculations across elution peaks.

Circular dichroism

The secondary structure and thermal stability of Atf1 in FC-14 were determined by circular dichroism (CD) (Figure 9.13). UV-CD spectra (Figure 9.13A) showed this was similar to published data recorded in Thesis³⁶⁸. Regarding thermal stability, it seems that Atf1 lost secondary structure with increasing temperature (Figure 9.13B). However, this was not via the cooperative transition that is the hallmark of well-folded proteins.

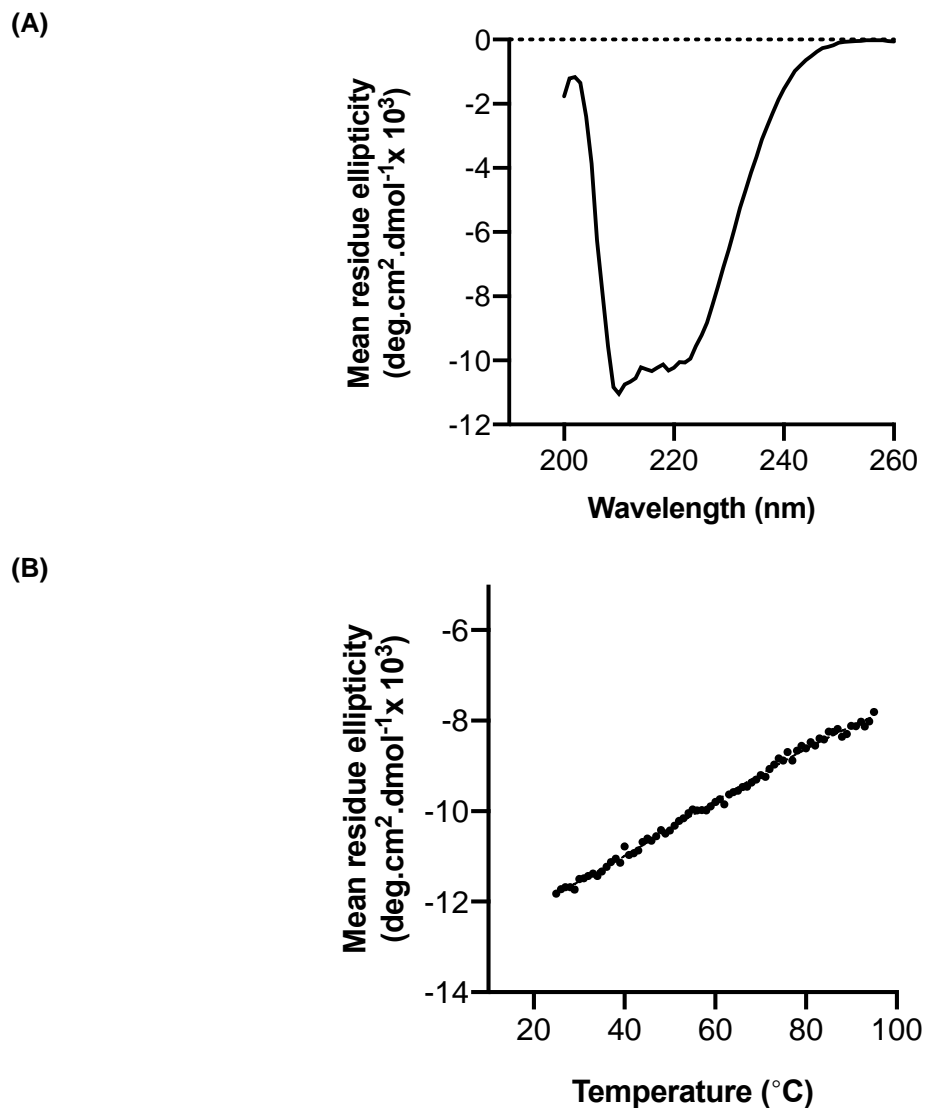


Figure 9.13. Circular dichroism spectrum of purified Atf1.

(A) CD spectrum of Atf1 in FC-14 detergent. Data which values of HT were above 700 V were excluded.

(B) Thermal stability of Atf1. Data were collected from 25°C to 95°C with 1°C increments.

9.2.2.4. Thioesterase and biochemical acyltransferase activity

Two activity assays were performed to test whether Atf1 was functional in FC-14. The first assay used *p*-nitrophenylthioacetate (*p*-NPTA) to test the previously observed thioesterase activity of Atf1 directly. Cleavage of this thioacetate substrate by Atf1 produces the coloured product *p*-nitrophenol. The results of this assay are shown in Figure 9.14A. Recombinant Atf1 solubilised in FC-14 was able to hydrolyse *p*-NPTA at a slightly lower specific activity ($7 \mu\text{M}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$) than the published value for Atf1 solubilised in Thesit ($10 \mu\text{M}\cdot\text{min}^{-1}\cdot\text{mg}^{-1}$)³⁶⁸. In the second functional assay, Atf1 solubilised in FC-14 was used in the coupled acyltransferase assay described above. Surprisingly, purified Atf1 did not display any activity against octanoyl-CoA in this assay (Figure 9.14B). Thus, it is not clear that Atf1 is functional after purification in FC-14.

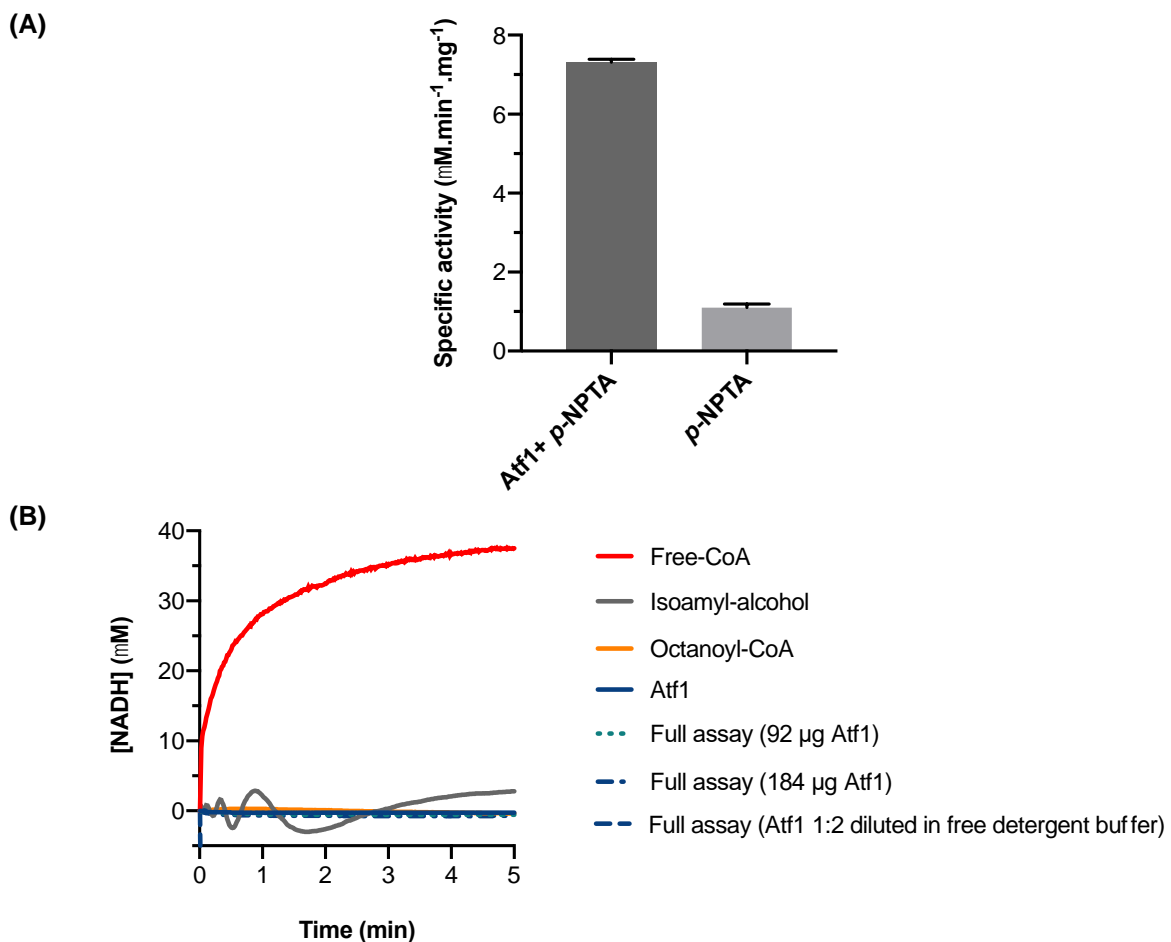


Figure 9.14. Thioesterase and biochemical acyltransferase activity assay of Atf1 in FC-14.

(A) Atf1 was incubated with *p*-NPTA to produce *p*-nitrophenol. A control reaction was performed with only *p*-NPTA. Data were collected from triplicate experiments. Error bars represent the standard deviation. **(B)** Atf1p against octanoyl-CoA in a coupled assay did not show activity. Different concentrations of Atf1 protein were used, and experiments included diluting the protein into a detergent-free buffer to try and minimise the impact of FC-14 on the protein.

9.2.3. Discussion

The results here exploited a detergent screen to establish a new, reliable and consistent method for purifying Atf1. This uses the detergent FC-14, a member of the fos-choline family of zwitterionic detergents considered powerful solubilising agents¹⁰⁵. These detergents can be denaturing and do not always support the purification of folded and functional proteins. Nonetheless, there are published studies in which membrane proteins have been successfully purified in fos-choline detergents, and crystal structures obtained^{544,545}.

SEC studies showed that Atf1 is homogenous and monodisperse in FC-14 with an apparent molecular weight of around 283 kDa, which is higher than expected since Atf1 has a molecular weight of 65 kDa and the proposed FC-14 micelle size is about 44 kDa¹⁰⁹. However, the elution volume of a membrane protein should not be used to estimate the oligomeric state or stoichiometry. Three primary considerations should be considered. These are that firstly the calibration curve was based on soluble proteins; secondly, the amount of detergent bound is unknown and can vary dramatically depending on both the protein and the nature of the detergent³⁷¹; and finally, that protein-detergent complexes can deviate markedly from being spherical, and so show unpredictable behaviour in hydrodynamic methods. Because of these drawbacks, SEC-MALS is the method of choice for determining the oligomeric state of membrane proteins in complex with detergent. SEC-MALS confirms that Atf1 is a stable monomer in FC-14.

CD data showed that the secondary structure of Atf1 is similar to the CD spectra published by Nancolas *et al.*³⁶⁸. However, melt data displayed weak cooperativity, which is often interpreted as resulting from the protein being in a partially unfolded or molten globule state.

Regarding biochemical activity, a thioesterase assay showed that FC-14-purified Atf1 was able to hydrolyse the acyl-CoA thioester bond of a generic esterase substrate, albeit with lower activity than published for Atf1 in Thesit. Atf1 in FC-14 did not exhibit acyltransferase activity against octanoyl-CoA. In short, the results are consistent with Atf1 being partially unfolded in FC-14, leading to a loss of activity which could be because some Atf1 was partially inactive. This is consistent with recent concerns that FC detergents are generally unsuitable for biophysical studies⁵⁴⁶. Alternatively, it is known that fos-cholines can efficiently solubilise unfolded membrane proteins¹⁰⁹. Whatever the case here, the results of this chapter show that FC14 is not a good candidate for further work on protein structure and function. Further work should pursue refolding studies that attempt to remove the detergent and refold Atf1 while still

preserving solubility. Brief efforts were made here to remove detergent via dilution (Figure 9.14), but a more thorough attempt should be made to exchange FC-14 for renaturing detergents rather than trying to remove detergent altogether.

9.3. Heterologous expression of fruit AATases in *S. cerevisiae*

9.3.1. Introduction and aim

One important class of chemicals responsible for fruit aroma profile is the esters, made in nature by various enzymes known as acyl-CoA:alcohol *O*-acyltransferases (AATases). These enzymes can be found in agricultural fruit plants such as banana, strawberry, papaya, melon, and kiwifruit, with functional similarity to Atf1 from *S. cerevisiae*. However, it is not clear how these enzymes work. In particular, the basis for substrate specificity is poorly understood. If plant AATases could be recombinantly expressed, it should be possible to bring the 'toolkit' developed to study Atf1 to bear on this family of proteins for the first time. Understanding the molecular biochemistry of plant AATases may lead to improvements in fruit crops, fruit flavour and shelf life.

The structure and function of AATases from papaya (VpAAT1), melon (MAAT), and strawberry (FaAAT1) AATases have been studied before by comparative computational modelling^{533–535}. Papaya and melon AATases have been expressed in yeast^{502,503}, and strawberry AATases in *E. coli*⁵²⁹. To date, no biophysical characterisation has been reported for these enzymes, and the quality and purity of the recombinant proteins are unclear (often not being reported).

This part of the thesis aims to express selected AATases in *S. cerevisiae* and purify them using affinity chromatography. Ultimately, this would lead to the first *in vitro* characterisation of these enzymes. To achieve this objective, we hypothesised that AATases could be successfully expressed and purified using the well-established protocols used for Atf1³⁶⁸.

9.3.2. Results

9.3.2.1. Crystallizability analysis of selected fruit AATases

The crystallizability of proteins can be predicted by using XtalPred web server³⁹⁶. According to the prediction, the protein might belong to one of the five crystallisation groups; easy to crystallise corresponds to a score of 1, while particularly challenging is a score of 5³⁹⁶. Table 9.3 displays the XtalPred results for the fruit AATases considered.

The scores were generally high, indicating that these fruit AATases will be a challenging family to crystallise. Melon and strawberry were chosen as the best scoring AATases for future research. Furthermore, VpAAT1 was chosen as the best candidate among the remaining since it is the most well-studied AATase to date.

Table 9.3. Crystallisation group of AATase from papaya (VpAAT1), cultivated strawberry (FaAAT1), kiwi (AcAT16 and AcAT9), apple (MpAAT1) and melon (MAAT).

Protein	Accession number	Crystallisation group
VpAAT1	FJ548611	4
FaAAT1	AAG13130	3
AcAT16	KJ626344	4
AcAT9	HO772640	4
MpAAT1	AAU14879	4
MAAT	EU431334	3

Synthetic genes corresponding to each of these proteins were obtained for recombinant expression, including C-terminal V5 and His tags for immunodetection and purification, respectively. These genes were cloned into the yeast plasmid pYES2. Constructs were confirmed by sequencing (Figure 9.15) and transformed into *S. cerevisiae* FGY217 for expression.

FaAAT1-V5-His₁₀ tag					
MEKIEVSINS	KHTIKPSTSS	TPLQPYKLT	LDQLTPPAYV	PIVFFYPITD 50	
HDFNLPQTLA	DLRQALSETL	TLYYPLSGRV	KNNLYIDDFE	EGVPYLEARV 100	
NCDMTDFLRL	RKIECLNEFV	PIKPFSMEAI	SDERYPLLGV	QVNVFDSGIA 150	
IGVSVSHKLI	DGGTADCFLK	SWGAVFRGCR	ENIIHPSLSE	AALLFPPRDD 200	
LPEKYVDQME	ALWFAGKKVA	TRRFVFGVKA	ISSIQDEAKS	ESVPKPSRVH 250	
AVTGFLWKHL	IAASRALTS	TTSTRLSIAA	QAVNLRTRMN	METVLDNATG 300	
NLFWWAQAIL	ELSHTTPEIS	DLKLCDLVNL	LNGSVKQCNG	DYFETFKGKE 350	
GYGRMCEYLD	FQRTMSSMEP	APDIYLFSSW	TNFFNPLDFG	WGRTSWIGVA 400	
GKIESASCKF	IILVPTQCGS	GIEAWVNLEE	EKMAMLEQDP	HFLALASPKT 450	
LISRGPFEGK	PIP NPLLGLD	STR TGHHHHH	HHHHH *	486	
VpAAT1-V5-His₁₀ tag					
MAEKASSLMF	NVRRHEPELI	TPAKPTPREI	KLLSDIDDQD	GLRFQVPIIQ 50	
FYKNNSSMQG	KNPAKIIKSA	LAETLVHYYP	LAGRLREGFG	RKLMVECTGE 100	
GILFIEADAD	VTLHEFGDDL	PPFPCLVEL	LYDVPGSSGI	IDTPLLIIQV 150	
TRLKCGGFIF	ALRLNHTMSD	ASGLVQFMTA	VGEMARGQRS	LSIQPVWERH 200	
LLNARDPPRV	THIHHEYDDL	EDTKGTIIP	DDMVHRSFFF	GPSEMAAIRR 250	
LVPAHFHRST	TSEVLTAYLW	RCYTIALQPD	PEEEMRVICV	VNSRTKLNPP 300	
LPTGFYNGI	AFPAAISQAK	KICENPFGYT	LQLVKQTKVD	VTEEYMRSAA 350	
DLMAMKGRPH	FTVVRRYMVS	DVTRAGFGLV	DFGWGRPEPV	YGGPAKGGVG 400	
PIPGVTSFFV	PFKNRKGEKG	IVVPTCLPTP	AMERFAKLMN	EILQNQLLVS 450	
AEENKSVFIV	SAISRGPFEG	K IPNPLLGL	D STRTGHHHH	HHHHHH *	497
MAAT-V5-His₁₀ tag					
METMQTIDFS	FQVRKCQPEL	IAPANPTPYE	FKQLSDVDDQ	QSLRFQLPLV 50	
NIYHHNPSLE	GRDPVKVIKE	AIKALVFY	PLAGRLREGP	GRKLFVECTG 100	
EGILFIEADA	DVSLEQFRDT	LPYSLSSMEN	NIIHNSLNSD	GVLNSPLLLI 150	
QVTRLKCGGF	IFGIHFDHTM	ADGFGIAQFM	KAI AEIARGA	FAPSILPVWQ 200	
RALLTARDPP	RITVRHYEYD	QVVDTKSTLI	PANNMIDRLF	FFTQRQISTL 250	
RQTLPAHLHD	CSSFVLAAY	VWRLRTIAFQ	LKPEEEVRF	CVVNLRSKID 300	
IPLGFYGNAI	VFPTVITVA	KLCGNPLGYA	VDLIRKAKAK	ATKEYIKSMV 350	
DFMVIKGRPR	FTEIGPFMMS	DITRIGFENV	DFGWGKAIFG	GPIIGGCGII 400	
RGMISYSIAF	MNRNGEKGIV	VPLCLPPPAM	ERFRANVHAS	LQVIQVLDKV 450	
DRDMQTILSA	LSRGPFEGK	IPNPLLGLD	TRTGHHHHHH	HHHH *	495

Figure 9.15. Protein sequences of FaAAT1, VpAAT1 and MAAT used in this work.

The V5 epitope is highlighted in red while His₁₀ tag in blue.

9.3.2.2. Expression and purification of fruit AATases in *S. cerevisiae*

The expression of MAAT, VpAAT1 and FaAAT1 was performed as described in Chapter 8. Cellular localisation of these AATases was confirmed by Western blotting of all cellular fractions (lysate, cytosol and membranes). Figure 9.16 showed that proteins were expressed by *S. cerevisiae*, although they were not easily detectable by Coomassie staining. However, specific bands were evident on Western blots close to the theoretical molecular weights. MAAT appeared to be associated with the membrane fraction, and VpAAT1 appeared to be degraded, showing multiple bands at a similar molecular weight. FaAAT1 was more encouraging, with higher levels of expression (greater intensity of blot bands) and protein distributed between the cytosol and membrane fractions.

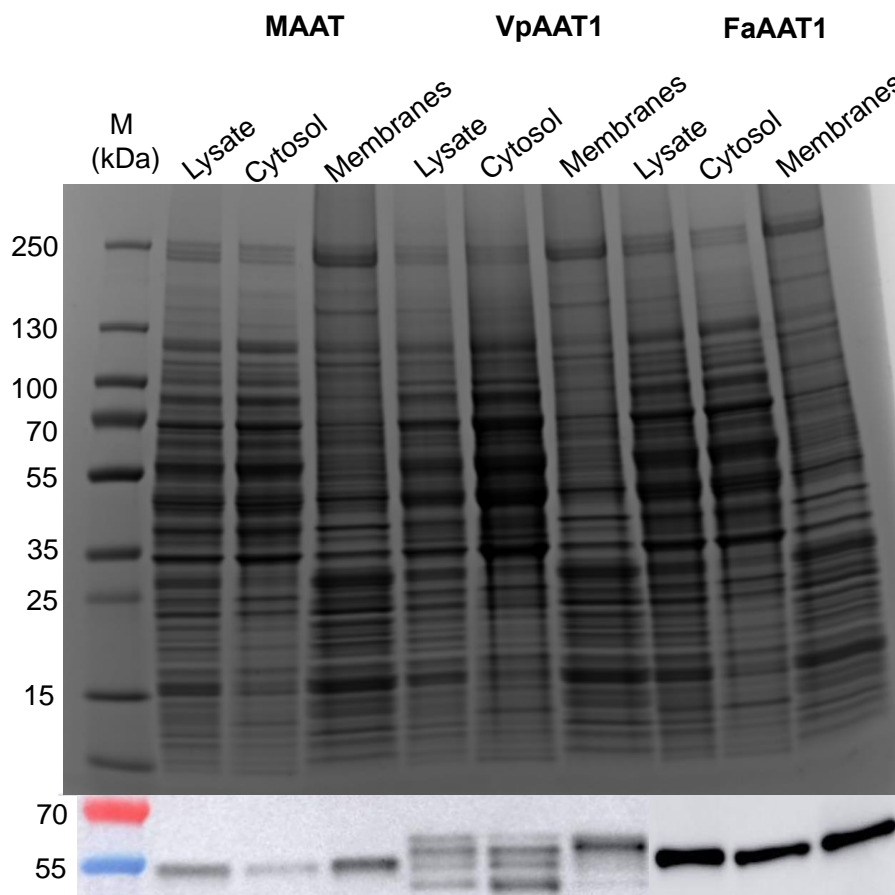


Figure 9.16. Expression of MAAT, VpAAT1 and FaAAT1 in *S. cerevisiae*.

(A) Coomassie-stained SDS-PAGE gel of cell fractions (lysate, cytosol, membranes). 50 μg of protein were loaded for all fractions. **(B)** Western blotting of cellular fractions showing the presence of protein in either cytosol and membrane (30 μg of each fraction were loaded). MAAT and FaAAT1 proteins ran at the expected molecular weights of 56 and 55 kDa, respectively.

The membrane fraction of MAAT and both protein-containing fractions of FaAAT1 (cytosol and membranes) were taken forward for purification. 2% Thesit was used to solubilise membranes, and purification was performed according to the protocol developed for Atf1. However, in this case, Thesit was not successful in solubilising MAAT membranes since no band was visualised in soluble membranes, and thus no protein was eluted from IMAC (Figure 9.17C). A similar result was observed for FaAAT1 membranes (Figure 9.17B). Additionally, purification of the cytosol fraction of FaAAT1 was unsuccessful, with the protein being eluted in the column wash step, and no visible band was observed on an SDS-PAGE gel (Figure 9.17A).

Table 9.4 compares the recombinant expression of AATases from fruit (FaAAt1, MAAT and VpAAT1) against Atf1 protein from *S. cerevisiae*. None of the fruit AATases appears to be suitable candidates for further characterisation.

Table 9.4. Results of AATase expression in *S. cerevisiae*.

Purified protein	Expressed?	Fraction purified	Yield (mg/ L culture)	Bands on SDS-PAGE?
Atf1	+++	Membranes	2.04*	Visible band is observed
FaAAT1	+	Cytosol and membranes	No protein detected in eluate	No bands observed in eluate fraction
MAAT	+	Membranes	No protein detected in eluate	No bands observed in eluate fraction
VpAAT1	Protein was degraded	-	-	-

* value obtained from Atf1 purification described in Section 9.1.

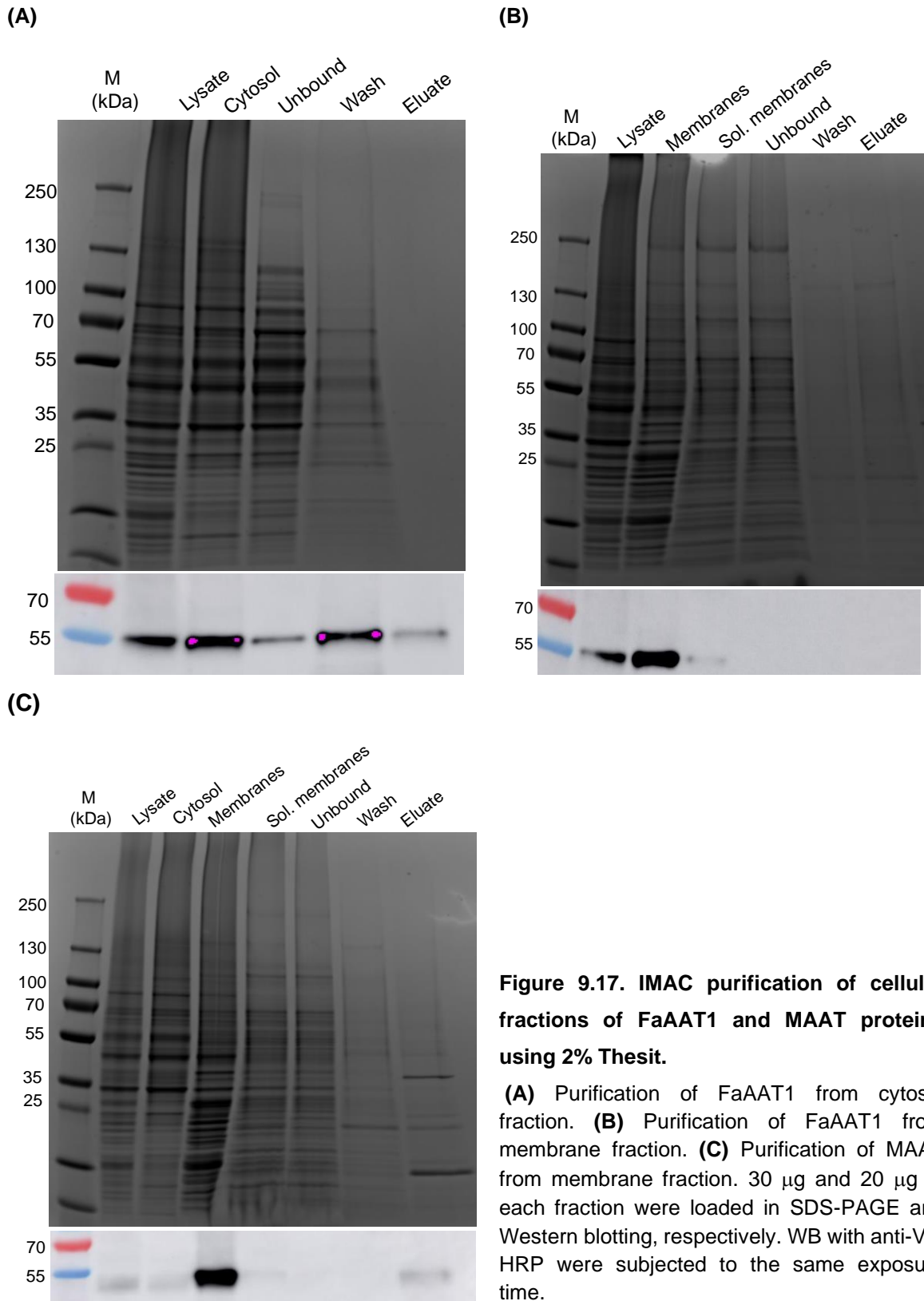


Figure 9.17. IMAC purification of cellular fractions of FaAAT1 and MAAT proteins using 2% Thesit.

(A) Purification of FaAAT1 from cytosol fraction. **(B)** Purification of FaAAT1 from membrane fraction. **(C)** Purification of MAAT from membrane fraction. 30 μ g and 20 μ g of each fraction were loaded in SDS-PAGE and Western blotting, respectively. WB with anti-V5-HRP were subjected to the same exposure time.

9.3.3. Discussion

The main objective of this section was to study the structure and function of selected AATases from commercial fruit. The genes from mountain papaya (*VpAAT1*), cultivated strawberry (*FaAAT1*) and melon (*MAAT*) were successfully cloned into the pYES2 vector for expression in *S. cerevisiae*.

When MAAT and FaAAT1 were expressed at a large scale, both proteins were found to be in either cytosol and/or membrane fraction, but the AATase from papaya was expressed in a degraded state. Further purification of cytosol or membrane fractions did not yield any protein, which corroborated the absence of bands in SDS-PAGE gels. Bands on Western blotting were visualised due to the high sensitivity of this technique, but this did not correspond to significant protein expression.

These findings significantly differ from previous results reported in the literature. Balbontín *et al.* and Morales-Quintana *et al.* claimed the expression of VpAAT1 protein in the yeast cell line INVSc1^{502,532}. Other studies have shown that melon AATases (CmAAT1-4) were expressed in the same yeast cell line^{503,522,523}. Moreover, the expression and purification of those AATases were only proven by activity assays. The yeast cell line used here, FGY217, differs only in the deletion of a major vacuolar protease that should improve expression yields. Even though published AATases purification procedures did not mention the use of any solubilising agent, our results suggested that either FaAAT1 or MAAT might be membrane-bound proteins and thus need to be extracted from membranes. Additionally, results indicated that Thesit was not successful in solubilising membrane fractions from FaAAT1 and MAAT. This result corroborates Pérez's findings, which showed partial purification of a strawberry AATase from crude *F. ananassa x Duchessne* var. Chandler extracts using Triton X-100⁵¹³. Therefore, future work might focus on the solubilisation of yeast membranes using another detergent or high-throughput screening of multiple AATase homologues to try and identify enzymes that are more amenable to expression.

In conclusion, the approach used for *S. cerevisiae* Atf1 was not applicable for the purification of other alcohol acyltransferases. Our results call into question the reliability of previous studies that show robust expression of these plant enzymes in yeast cell lines.

9.4. Using pOPIN vectors to optimise protein expression

9.4.1. Introduction and aim

Considering the results shown above, neither Atf1 nor the fruit AATases tested are readily expressible in heterologous recombinant cell lines, prohibiting their further study. Although Atf1 was successfully expressed in *S. cerevisiae* (Section 9.1), its heterogeneous size exclusion profile in detergent complexes suggested that this would not be a good candidate for crystallography. This section of the thesis sought to test whether transcriptional fusion tags could improve the expression profile and solubility of Atf1 and plant AATases. Several commercial plasmids with fusion protein tags, such as Maltose Binding Protein (MBP) or N-utilisation substance A (NusA), have been used to increase protein expression and improve solubility⁵⁴⁷. Particularly, MBP has demonstrated chaperone-like behaviour to support the folding of fusion proteins⁵⁴⁸. Here, MBP and NusA were fused to Atf1 and fruit AATases and small scale tests performed to evaluate their recombinant expression and solubility.

9.4.2. Results

9.4.2.1. Expression and purification of VpAAT1, FaAAT1, MAAT and Atf1 fusion proteins

Synthetic genes of AATases (*VpAAT1* (papaya), *FaAAT1* (strawberry) and *MAAT* (melon)), as well as the *atf1* gene, were cloned into pOPINF, pOPINM and pOPINNUSA vectors by ligation-independent approach⁵⁴⁹ as described in Chapter 8. For expression, the sequenced plasmids with the confirmed insert were transformed into *E. coli* BL21(DE3) strain. Each of these constructs had a 6-residue polyhistidine tag (His₆) at the N-terminal and a C-terminal V5 epitope to allow immunoblotting. Constructs in the pOPINM and pOPINNUSA vectors additionally have the protein fusion tags MBP and NusA, respectively, at the N-terminal.

All protein sequences are described in subsequent figures (from Figure 9.18 to Figure 9.21).

His₆tag – Atf1 – V5

MA HHHHHH SS	GLEVLFGQPM	NEIDEKNQAP	VQOECLKEMI	QNGHARRMGS	50
VEDLYVALNR	QNLRYNFCTY	GELSDYCTRD	QLTLALREIC	LKNPTLLHIV	100
LPTRWPNHEN	YYSRSEYYSR	PHPVHDYISV	LQELKLSGVV	LNEQPEYSAV	150
MKQILEEFKN	SKGSYTAKIF	KLTTTLTIPY	FGPTGPSWRL	ICLPEEHTEK	200
WKKFIFVSNH	CMSDGRSSIH	FFHDLRDELN	NIKTTPKKLD	YIFKYEEDYQ	250
LLRKLPEPIE	KVIDFRPPYL	FIPKSLLSGF	IYNHLRFSSK	GVCMRMDDVE	300
KTDDVVTEII	NISPTEFQAI	KANIKSNIQG	KCTITPFLHV	CWFVSLHKWG	350
KFFKPLNFEW	LTDIFIPADC	RSQLPDDDEM	RQMYRYGANV	GFIDFTPWIS	400
EFDMDNKEN	FWPLIEHYHE	VISEALRNKK	HLHGLGFNIQ	GFVQKYVNID	450
KVMCDRAIGK	RRGGTLLSNV	GLFNQLEEPD	AKYSICDLAF	GQFQGSWHQA	500
FSLGVCSTNV	KGMNIVVAST	KNVVGSQESL	EELCSIYKAL	LLGPSRGPFE	550
GKPIPNLLG	LDST*	564			

His₆tag – MBP- Atf1 – V5

MA HHHHHH SS	GMKIEEGKLV	IWINGDKGYN	GLAEVGGKFE	KDTGIKVTVE	50
HPDKLEEKFP	QVAATGDGPD	IIFWAHDREF	GYAQSGLLAE	ITPDKAFQDK	100
LYPFTWDAVR	YNGKLIAYPI	AVEALSIIYN	KDLLPNPPKT	WEEIPALDKE	150
LKAKGKSALM	FNLQEPYFTW	PLIAADGGYA	FKYENGKYDI	KDVGVDNAGA	200
KAGLTFVLVDL	IKNKHMNADT	DYSIAEAAFN	KGETAMTING	PWAWSNIDTS	250
KVNYGVTVLP	TFKGQPSKPF	VGVLSAGINA	ASPNKELAKE	FLENYLLTDE	300
GLEAVNKDKP	LGAVALKSYE	EELAKDPRIA	ATMENAQKGE	IMPNI PQMSA	350
FWYAVRTAVI	NAASGRQTV	EALKDAQTSS	GLEVLFGQPM	NEIDEKNQAP	400
VQOECLKEMI	QNGHARRMGS	VEDLYVALNR	QNLRYNFCTY	GELSDYCTRD	450
QLTLALREIC	LKNPTLLHIV	LPTRWPNHEN	YYSRSEYYSR	PHPVHDYISV	500
LQELKLSGVV	LNEQPEYSAV	MKQILEEFKN	SKGSYTAKIF	KLTTTLTIPY	550
FGPTGPSWRL	ICLPEEHTEK	WKKFIFVSNH	CMSDGRSSIH	FFHDLRDELN	600
NIKTTPKKLD	YIFKYEEDYQ	LLRKLPEPIE	KVIDFRPPYL	FIPKSLLSGF	650
IYNHLRFSSK	GVCMRMDDVE	KTDDVVTEII	NISPTEFQAI	KANIKSNIQG	700
KCTITPFLHV	CWFVSLHKWG	KFFKPLNFEW	LTDIFIPADC	RSQLPDDDEM	750
RQMYRYGANV	GFIDFTPWIS	EFDMDNKEN	FWPLIEHYHE	VISEALRNKK	800
HLHGLGFNIQ	GFVQKYVNID	KVMCDRAIGK	RRGGTLLSNV	GLFNQLEEPD	850
AKYSICDLAF	GQFQGSWHQA	FSLGVCSTNV	KGMNIVVAST	KNVVGSQESL	900
EELCSIYKAL	LLGPSRGPFE	GKPIPNLLG	LDST*	935	

His₆tag – NusA - Atf1 – V5

MA HHHHHH HNK	EILAVVEAVS	NEKALPREKI	FEALESALAT	ATKKKYEQEI	50
DVRVQIDRKS	GDFDTFRRWL	VVDEVTQPTK	EITLEAARYE	DESLNLGDYV	100
EDQIESVTFD	RITTQTAKQV	IVQKVREAER	AMVVDQFREQ	EGEITGVVK	150
KVNRDNISLD	LGNNAEAVIL	REDMLPRENF	RPGDRVRGVL	YSVRPEARGA	200
QLFVTRSKPE	MLIELFRIEV	PEIGEEVIEI	KAAARDPGSR	AKIAVKTNDK	250
RIDPVGACVG	MARGARVQTVS	TELGGERIDI	VLWDDNPAQF	VINAMAPADV	300
ASIVVDEDKH	TMDIAVEAGN	LAQAIGRNGQ	NVRLASQLSG	WELNVMTVDD	350
LQAKHQAEAH	AAIDTFTKYL	DIDEDFATVL	VEEGFSTLEE	LAYVPMKELL	400
EIEGLDEPTV	EALRERAKNA	LATIAQAQEE	SLGDNKPADD	LLNLEGVDRD	450
LAFKLAARGV	CTLEDLAEQG	IDDLADIEGL	TDEKAGALIM	AARNICWFGD	500
EASSGLEVLV	QGPMNEIDEK	NQAPVQOECL	KEMIQNGHAR	RMGSVEDLYV	550
ALNRQNLRYN	FCTYGELSDY	CTRDQLTLAL	REICLKNPTL	LHIVLPTRWP	600
NHENYYRSSE	YYSRPHPVHD	YISVLQELKL	SGVVLNEQPE	YSAVMKQILE	650
EFKNSKGSYT	AKIFKLTTTL	TIPYFGPTGP	SWRLICLPEE	HTEKWKKFIF	700
VSNHCMSDGR	SSIHFFHDLR	DELNNIKTPP	KKLDYIFKYE	EDYQLLRKLP	750
EPIEKVIDFR	PPYLFIPKSL	LSGFIYNHLR	FSSKGVCMRM	DDVEKTDDVV	800
TEIINISPTE	FPAIKANIKS	NIQKCTITP	FLHVCWFVSL	HKWGKFFKPL	850
NFEWLTDIFI	PADCRSQLPD	DDEMROMYRY	GANVGFIDFT	PWISEFDMND	900
NKENFWPLIE	HYHEVISEAL	RNKKHLHGLG	FNIQGFVQKY	VNIDKVMCDR	950
AIGKRRGGTL	LSNVGLFNQL	EEDAKYSIC	DLAFGQFQGS	WHQAFSLGVC	1000
STNVKGMNIV	VASTKNVVG	QESLEELCSI	YKALLLGP	SRGPFE GKPIPN	1050
PLGLDST*	1059				

Figure 9.18. Protein sequence of non-fusion Atf1 and fused with MBP and NusA protein tags.V5 epitope is highlighted in red, His₆ tag in blue and fusion protein tags in grey.

His₆tag – FaAAT1 – V5

MA HHHHHH SS	GLEVLFGQPM	EKIEVSINSK	HTIKPSTSST	PLQPYKLTL	50
DQLTPPAYVP	IVFFYPITDH	DFNLPQTLAD	LRQALSETLT	LYYPLSGRVK	100
NNLYIDDFEE	GVPYLEARVN	CDMTDFLRLR	KIECLNEFVP	IKPFSMEAIS	150
DERYPLLGVQ	VNVFDSGIAI	GVSVSHKLID	GGTADCFLKS	WGAVFRGCRE	200
NI IHPSLSEA	ALLFPPRDDL	PEKYVDQMEA	LWFAGKKVAT	RRFVFGVKAI	250
SSIQDEAKSE	SVPKPSRVHA	VTGFLWKHLI	AASRALTSQT	TSTRLSIAAQ	300
AVNLRTRMNM	ETVLDNATGN	LFWWAQAI	LSHTTPEISD	LKLCDLVNLL	350
NGSVKQCNGD	YFETFKGKEG	YGRMCEYLD	QRTMSSMEPA	PDIYLFSSWT	400
NFFNPLDFGW	GRTSWIGVAG	KIESASCKFI	ILVPTQCGSG	IEAWVNLEEE	450
KMAMLEQDPH	FLALASPCTL	ISRGPFEG KP	IPNPLLGLDS	T*	492

His₆tag – MBP- FaAAT1 – V5

MA HHHHHH SS	GMKIEEGKLV	IWINGDKGYN	GLAEVGGKFE	KDTGIKVTVE	50
HPDKLEEKFP	QVAATGDGPD	IIFWAHDRFG	GYAQSGLLAE	ITPDKAFQDK	100
LYPFTWDAVR	YNGKLIAYPI	AVEALSLIYN	KDLLPNPPKT	WEEIPALDKE	150
LKAKGKSALM	FNLQEPYFTW	PLIAADGGYA	FKYENGKYDI	KDVGVDNAGA	200
KAGLTFVLVDL	IKNKHMNADT	DYSIAEAAFN	KGETAMTING	PWAWSNIDTS	250
KVNYGVTVLP	TFKGQPSKPF	VGVL SAGINA	ASPKNELAKE	FLENYLLTDE	300
GLEAVNKDKP	LGAVALKSYE	EELAKDPRIA	ATMENAQKGE	IMPNI PQMSA	350
FWYAVRTAVI	NAASGRQTV	EALKDAQTSS	GLEVLFGQPM	EKIEVSINSK	400
HTIKPSTSST	PLQPYKLTL	DQLTPPAYVP	IVFFYPITDH	DFNLPQTLAD	450
LRQALSETLT	LYYPLSGRVK	NNLYIDDFEE	GVPYLEARVN	CDMTDFLRLR	500
KIECLNEFVP	IKPFSMEAIS	DERYPLLGVQ	VNVFDSGIAI	GVSVSHKLID	550
GGTADCFLKS	WGAVFRGCRE	NI IHPSLSEA	ALLFPPRDDL	PEKYVDQMEA	600
LWFAGKKVAT	RRFVFGVKAI	SSIQDEAKSE	SVPKPSRVHA	VTGFLWKHLI	650
AASRALTSQT	TSTRLSIAAQ	AVNLRTRMNM	ETVLDNATGN	LFWWAQAI	700
LSHTTPEISD	LKLCDLVNLL	NGSVKQCNGD	YFETFKGKEG	YGRMCEYLD	750
QRTMSSMEPA	PDIYLFSSWT	NFFNPLDFGW	GRTSWIGVAG	KIESASCKFI	800
ILVPTQCGSG	IEAWVNLEEE	KMAMLEQDPH	FLALASPCTL	ISRGPFEG KP	850
IPNPLLGLDS	T*				861

His₆tag – NusA – FaAAT1 – V5

MA HHHHHH HNK	EILAVVEAVS	NEKALPREKI	FEALESALAT	ATKKKYEQEI	50
DVRVQIDRKS	GDFDTFRRWL	VVDEVTOPTK	EITLEAARYE	DESLNLGDYV	100
EDQIESVTFD	RITTQTAKQV	IVQKVREAER	AMVVDQFREQ	EGEITGVVK	150
KVNRDNISLD	LGNNAAEAVIL	REDMLPRENF	RPGDRVRGVL	YSVRPEARGA	200
QLFVTRSKPE	MLIELFRIEV	PEIGEEVIEI	KAARDPGSR	AKIAVKTNDK	250
RIDPVGACVG	MARGARQTVS	TELGGERIDI	VLWDDNPAQF	VINAMAPADV	300
ASIVVDEDKH	TMDIAVEAGN	LAQAIGRNGQ	NVRLASQLSG	WELNVMTVDD	350
LQAKHQAEAH	AAIDTFTKYL	DIDEDFATVL	VEEGFSTLEE	LAYVPMKELL	400
EIEGLDEPTV	EALRERAKNA	LATIAQAQEE	SLGDNKPADD	LLNLEGVDRD	450
LAFKLAARGV	CTLEDLAEQG	IDDLADIEGL	TDEKAGALIM	AARNICWFGD	500
EASSGLEVLV	QGPMEKIEVS	INSKHTIKPS	TSSTPLQPYK	LTLLDQLTPP	550
AYVPIVFFYP	ITDHDFNLPO	TLADLRQALS	ETLTLYYPLS	GRVKNNLYID	600
DFEEGVPYLE	ARVNCDMTDF	LRLRKIECLN	EFVPIKPFMS	EASDERYPL	650
LGVQVNVFDS	GIAIGVSVSH	KLIDGGTADC	FLKSWGAVFR	GCRENI IHPS	700
LSEAALLFPP	RDDLPEKYVD	QMEALWFAGK	KVATRRFVFG	VKAISSIQDE	750
AKSESVPKPS	RVHAVTGFLW	KHLIAASRAL	TSGT TSTRLS	IAAQAVNLRT	800
RMNMETVLDN	ATGNLFWWAQ	AILELSHTTP	EISDLKLCDL	VNLLNGSVKQ	850
CNGDYFETFK	GKEGYGRMCE	YLD FQRTMSS	MEPAPDIYLF	SSWTNFFNPL	900
DFGWGRTSWI	GVAGKIESAS	CKFIILVPTQ	CGSGIEAWVN	LEEEKMAMLE	950
QDPHFLALAS	PKTLISRGPF	EGKPIPNPLL	GLDST*		986

Figure 9.19. Protein sequence of non-fusion FaAAT1 and fused with MBP and NusA protein tags.V5 epitope is highlighted in red, His₆ tag in blue and fusion protein tags in grey.

His₆tag – MAAT – V5

MA HHHHHH SS	GLEVLFGQPM	ETMQTIDFSF	QVRKCQPELI	APANPTPYEF	50
KQLSDVDDQQ	SLRFQLPLVN	IYHHNPSLEG	RDPVKVIKEA	IAKALVFYYP	100
LAGRLREGPG	RKLFVECTGE	GILFIEADAD	VSLEQFRDTL	PYSLSSMENN	150
IIHNSLNSDG	VLNSPLLLIQ	VTRLKCGGFI	FGIHFHDHTMA	DGFGIAQFMK	200
AIAEIARGAF	APSILPVWQR	ALLTARDPPR	ITVRHYEYDQ	VVDTKSTLIP	250
ANNMIDRLFF	FTQRQISTLR	QTLPAHLHDC	SSFEVLAAYV	WRLRTIAFQL	300
KPEEEVRFLC	VVNLRSKIDI	PLGFYGNNAIV	FPTVITTVAK	LCGNPLGYAV	350
DLIRKAKAKA	TKEYIKSMVD	FMVIKGRPRF	TEIGPFMMSD	ITRIGFENVD	400
FGWGKAIFGG	PIIGGCGIIR	GMISYSIAFM	NRNGEKGIVV	PLCLPPPAME	450
RFRANVHASL	QVIQVLDKVD	RDMQTILSAL	SRGPFE GKPI	PNPLLLGLDST	500

* 501

His₆tag – MBP- MAAT – V5

MA HHHHHH SS	GMKIEEGKLV	IWINGDKGYN	GLAEVGKKFE	KDTGIKVTVE	50
HPDKLEEKFP	QVAATGDGPD	IIFWAHDRFG	GYAQSGLLAE	ITPKAFQDK	100
LYPFTWDAVR	YNGKLIAYPI	AVEALSIIYN	KDLLPNPPKT	WEEIPALDKE	150
LKAKGKSALM	FNLQEPYFTW	PLIAADGGYA	FKYENGKYDI	KDVGVDNAGA	200
KAGLTFLVDL	IKNKHMNADT	DYSIAEAAFN	KGETAMTING	PWAWSNIDTS	250
KVNYGVTVLP	TFKGQPSKPF	VGVL SAGINA	ASPKNELAKE	FLENYLLTDE	300
GLEAVNKDKP	LGAVALKSYE	EELAKDPRIA	ATMENAQKGE	IMPNI PQMSA	350
FWYAVRTAVI	NAASGRQTV	EALKDAQTSS	GLEVLFGQPM	ETMQTIDFSF	400
QVRKCQPELI	APANPTPYEF	KQLSDVDDQQ	SLRFQLPLVN	IYHHNPSLEG	450
RDPVKVIKEA	IAKALVFYYP	LAGRLREGPG	RKLFVECTGE	GILFIEADAD	500
VSLEQFRDTL	PYSLSSMENN	IIHNSLNSDG	VLNSPLLLIQ	VTRLKCGGFI	550
FGIHFHDHTMA	DGFGIAQFMK	AIAEIARGAF	APSILPVWQR	ALLTARDPPR	600
ITVRHYEYDQ	VVDTKSTLIP	ANNMIDRLFF	FTQRQISTLR	QTLPAHLHDC	650
SSFEVLAAYV	WRLRTIAFQL	KPEEEVRFLC	VVNLRSKIDI	PLGFYGNNAIV	700
FPTVITTVAK	LCGNPLGYAV	DLIRKAKAKA	TKEYIKSMVD	FMVIKGRPRF	750
TEIGPFMMSD	ITRIGFENVD	FGWGKAIFGG	PIIGGCGIIR	GMISYSIAFM	800
NRNGEKGIVV	PLCLPPPAME	RFRANVHASL	QVIQVLDKVD	RDMQTILSAL	850
SRGPFE GKPI	PNPLLLGLDST	* 871			

His₆tag – NusA - MAAT – V5

MA HHHHHH HNK	EILAVVEAVS	NEKALPREKI	FEALESALAT	ATKKKYEQEI	50
DVRVQIDRKS	GDFDTFRRWL	VVDEVTQPTK	EITTEAARYE	DESLNLGDYV	100
EDQIESVTFD	RITTQTAKQV	IVQKVREAER	AMVVDQFREQ	EGEIIITGVVK	150
KVNRDNISLD	LGNNAEAVIL	REDMLPRENF	RPGDRVRGVL	YSVRPEARGA	200
QLFVTRSKPE	MLIELFRIEV	PEIGEEVIEI	KAAARDPGSR	AKIAVKTNDK	250
RIDPVGACVG	MARGARQTVS	TELGGERIDI	VLWDDNPAQF	VINAMAPADV	300
ASIVVDEDKH	TMDIAVEAGN	LAQAIGRNGQ	NVRLASQLSG	WELNVMTVDD	350
LQAKHQAEAH	AAIDTFTKYL	DIDEDFATVL	VEEGFSTLEE	LAYVPMKELL	400
EIEGLDEPTV	EALRERAKNA	LATIAQAQEE	SLGDNKPADD	LLNLEGVDRD	450
LAFKLAARGV	CTLEDLAEQG	IDDLADIEGL	TDEKAGALIM	AARNICWFGD	500
EASSGLEVLV	QGPMETMQTI	DFSFQVRKCQ	PELIAPANPT	PYEFKQLSDV	550
DDQQLRFQL	PLVNIYHHNP	SLEGRDPVKV	IKEAIAKALV	FYYPLAGRLR	600
EGPGRKLFVE	CTGEGILFIE	ADADVSLEQF	RDTLPYSLSS	MENNI IHNSL	650
NSDGLVNSPL	LLIQVTRLKC	GGFIFGIHFD	HTMADGFGIA	QFMKAIAEIA	700
RGAFAPSILP	VWQRALLTAR	DPPRITVRHY	EYDQVVDTKS	TLIPANNMID	750
RLFFFTQRQI	STLRQTLPAH	LHDCSSFEVL	AAVWRLRTI	AFQLKPEEEV	800
RFLCVVNLRS	KIDIPLGFYV	NAIVFPTVIT	TVAKLCGNPL	GYAVDLIRKA	850
KAKATKEYIK	SMVDFMVIK	RPRFTEIGPF	MMSDITRIGF	ENVDFGWGKA	900
IFGGPIIGGC	GIIRGMISYS	IAFMNRNGEK	GIVVPLCLPP	PAMERFRANV	950
HASLQVIQVL	DKVDRDMQTI	LSALS RGPFE	GKPIPNPLLG	LDST	* 995

Figure 9.20. Protein sequence of non-fusion MAAT and fused with MBP and NusA protein tags.V5 epitope is highlighted in red, His₆ tag in blue and fusion protein tags in grey.

His₆tag – VpAAT1 – V5

MA HHHHHH SS	GLEVLFGQPM	AEKASSLMFN	VRRHEPELIT	PAKPTPREIK	50
LLSDIDDQDG	LRFQVPIIQF	YKNNSSMQGK	NPAKIIKSAL	AETLVHYYPL	100
AGRLREGFGR	KLMVECTGEG	ILFIEADADV	TLHEFGDDL	PPFPCLVELL	150
YDVPGSSGII	DTPLLLIQVT	RLKCGGFIFA	LRLNHTMSDA	SGLVQFMTAV	200
GEMARGQRSL	SIQPVWERHL	LNARDPPRVT	HIHHEYDDLE	DTKGTIIPLD	250
DMVHRSFFFG	PSEMAAIRRL	VPAHFHRSTT	SEVLTAYLWR	CYTIALQDPD	300
EEEMRVICVV	NSRTKLNPPPL	PTGFYGNIGIA	FPAAISQAKK	ICENPFGYTL	350
QLVKQTKVDV	TEEYMRSAAD	LMAMKGRPHF	TVVRRYMVSD	VTRAGFGLVD	400
FGWGRPEPVY	GGPAKGGVGP	IPGVTSFFVVP	FKNRKGEKGI	VVPTCLPTPA	450
MERFAKLMNE	ILQNQLLVSA	EENKSVFIVS	AISRGPFE GK	PIP NPL LGLD	500

ST* 503

His₆tag – MBP- VpAAT1 – V5

MA HHHHHH SS	GMKIEEGKLV	IWINGDKGYN	GLAEVGGKFE	KDTGIKVTVE	50
HPDKLEEKFP	QVAATGDGPD	IIFWAHDRFG	GYAQSGLLAE	ITPDKAFQDK	100
LYPFTWDAVR	YNGKLIAYPI	AVEALSIIYN	KDLLPNPPKT	WEEIPALDKE	150
LKAKGKSALM	FNLQEPYFTW	PLIAADGGYA	FKYENKDYDI	KDVGVDNAGA	200
KAGLTFVLVDL	IKNKHMNADT	DYSIAEAAFN	KGETAMTING	PWAWSNIDTS	250
KVNYGVTVLP	TFKGQPSKPF	VGVLASAGINA	ASPKNELAKE	FLENYLLTDE	300
GLEAVNKDKP	LGAVALKSYE	EELAKDPRIA	ATMENAQKGE	IMPNIPOMSA	350
FWYAVRTAVI	NAASGRQTV	EALKDAQTSS	GLEVLFGQPM	AEKASSLMFN	400
VRRHEPELIT	PAKPTPREIK	LLSDIDDQDG	LRFQVPIIQF	YKNNSSMQGK	450
NPAKIIKSAL	AETLVHYYPL	AGRLREGFGR	KLMVECTGEG	ILFIEADADV	500
TLHEFGDDL	PPFPCLVELL	YDVPGSSGII	DTPLLLIQVT	RLKCGGFIFA	550
LRLNHTMSDA	SGLVQFMTAV	GEMARGQRSL	SIQPVWERHL	LNARDPPRVT	600
HIHHEYDDLE	DTKGTIIPLD	DMVHRSFFFG	PSEMAAIRRL	VPAHFHRSTT	650
SEVLTAYLWR	CYTIALQDPD	EEEMRVICVV	NSRTKLNPPPL	PTGFYGNIGIA	700
FPAAISQAKK	ICENPFGYTL	QLVKQTKVDV	TEEYMRSAAD	LMAMKGRPHF	750
TVVRRYMVSD	VTRAGFGLVD	FGWGRPEPVY	GGPAKGGVGP	IPGVTSFFVVP	800
FKNRKGEKGI	VVPTCLPTPA	MERFAKLMNE	ILQNQLLVSA	EENKSVFIVS	850
AISRGPFE GK	PIP NPL LGLD	ST* 873			

His₆tag – NusA – VpAAT1 – V5

MA HHHHHH HNK	EILAVVEAVS	NEKALPREKI	FEALESALAT	ATKKKYEQEI	50
DVRVQIDRKS	GDFDTFRRWL	VVDEVTQPTK	EITLEAARYE	DESLNLDYV	100
EDQIESVTFD	RITTQTAKQV	IVQKVREAER	AMVVDQFREQ	EGEIIITGVVK	150
KVNRDNISLD	LGNNAAEAVIL	REDMLPRENF	RPGDRVRGVL	YSVRPEARGA	200
QLFVTRSKPE	MLIELFRIEV	PEIGEEVIEI	KAAARDPGSR	AKIAVKTNDK	250
RIDPVGACVG	MARGARVQTVS	TELGGERIDI	VLWDDNPAQF	VINAMAPADV	300
ASIVVDEDKH	TMDIAVEAGN	LAQAIGRNGQ	NVRLASQLSG	WELNVMTVDD	350
LQAKHQAEAH	AAIDTFTKYL	DIDEDFATVL	VEEGFSTLEE	LAYVPMKELL	400
EIEGLDEPTV	EALRERAKNA	LATIAQAQEE	SLGDNKPADD	LLNLEGVDRD	450
LAFKLAARGV	CTLEDLAEQG	IDDLADIEGL	TDEKAGALIM	AARNICWFGD	500
EASSGLEVLV	QGPMAEKASS	LMFNVRREHEP	ELITPAKPTP	REIKLLSDID	550
DQDGLRFQVP	IIQFYKNNSS	MQGKNPAKII	KSALAEVLVH	YYPLAGRLRE	600
GFGKRLMVEC	TGEGILFIEA	DADVTLHEFG	DDLPPFPCL	VELLYDVPGS	650
SGIIDTPLLL	IQVTRLKCGG	FIFALRLNHT	MSDASGLVQF	MTAVGEMARG	700
QRSLSIQPVW	ERHLLNARDP	PRVTHIHHEY	DDLEDTKGTI	IPLDDMVHRS	750
FFFGPSEMAA	IRRLVPAHFH	RSTTSEVLTA	YLWRCYTIAL	QDPPEEMRV	800
ICVNSRTKL	NPPLPTGFY	NGIAFPAAIS	QAKKICENPF	GYTLQLVKQT	850
KVDVTEEYMR	SAADLMAMKG	RPHFTVRRY	MVSDVTRAGF	GLVDFGWGRP	900
EPVYGGPAKG	GVGPIPGVTS	FFVPFKNRKG	EKGIVVPTCL	PTPAMERFAK	950
LMNEILQNQL	LVSAAENKSV	FIVSAISRGP	FE GKPIP NPL	LGLDST* 997	

Figure 9.21. Protein sequence of non-fusion VpAAT1 and fused with MBP and NusA protein tags.V5 epitope is highlighted in red, His₆ tag in blue and fusion protein tags in grey.

To assess if AATase proteins were being expressed and their degree of solubility, small scale expression trials were carried out. These initial trials systematically screened each fusion protein against different media (LB, 2YT or SB) and two induction systems (IPTG or auto-induction). After cells were harvested and lysed using the BugBuster reagent, a centrifugation step was performed to separate the soluble cell lysate (supernatant) from the insoluble pellet, which was resuspended in 6 M urea for analysis. Table 9.5 reports the expression of each recombinant protein (Atf1, FaAAT1, MAAT and VpAAT1) under each condition. The most promising data presented in the table are described in the subsequent text.

Table 9.5. Screening expression constructs against cell media and expression conditions for each protein of interest (FaAAT1, MAAT, VpAAT1 and Atf1).

‘-’ means that protein was not expressed and ‘+’ means protein was expressed, as determined by western blotting. ND, not determined. IB, inclusion bodies (protein was detected in the insoluble fraction after cell lysis using BugBuster reagent).

Protein	Vector	Fusion protein	Small scale expression				Large scale expression
			LB _{IPTG} ^a	LB _{AIM} ^b	2YT _{AIM} ^b	SB _{AIM} ^b	Best candidate
FaAAT1	pOPINF	-	IB	-	-	-	-
	pOPINM	MBP	+	-	++	++	SB _{AIM}
	pOPINNUSA	NusA	-	-	-	-	-
MAAT	pOPINF	-	IB	-	-	-	-
	pOPINM	MBP	+	-	-	-	-
	pOPINNUSA	NusA	-	-	-	-	-
VpAAT1	pOPINF	-	-	-	-	-	-
	pOPINM	MBP	-	-	-	-	-
	pOPINNUSA	NusA	-	-	-	-	-
Atf1	pOPINF	-	ND	+++	+++	++	LB _{AIM}
	pOPINM	MBP	ND	+++	+++	++	LB _{AIM}
	pOPINNUSA	NusA	ND	-	-	-	-

^a Induced with 1 mM IPTG and incubated overnight (~20 hour) at 20°C. ^b Cultures were incubated for 24 hour at 37°C. All expression conditions are described in Section 8.7

Concerning small scale expression trials of Atf1, no expression in LB_{IPTG} was tested since it was already known that Atf1 forms inclusion bodies under such conditions. Protein electrophoresis shown in Figure 9.22 revealed that Atf1 was successfully expressed in all auto-induction culture media (LB, 2YT or SB) in the absence of a fusion protein or when fused with MBP. Therefore, these two Atf1 constructs were chosen for further purification. Large scale expression revealed that Atf1 was better expressed in LB_{AIM} when compared to other AIM, and the highest expression level was observed for non-fusion Atf1 (Figure 9.23A and B).

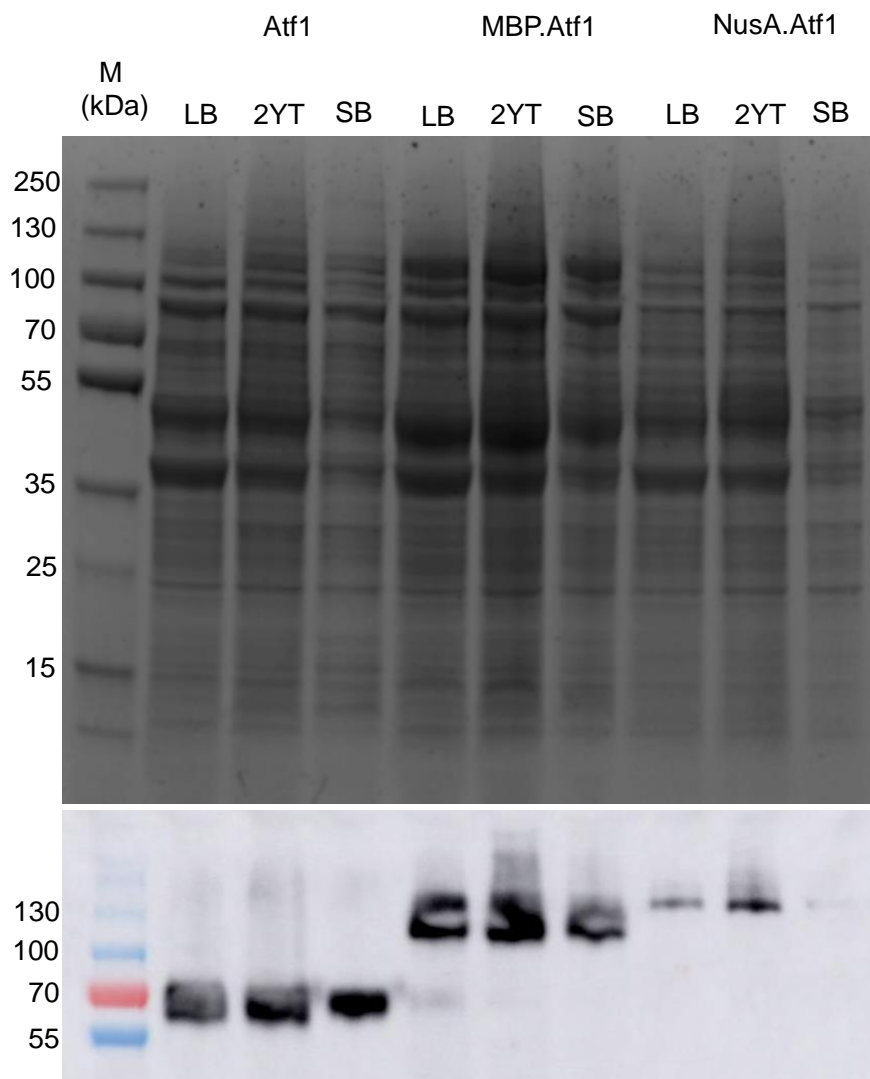


Figure 9.22. Small expression of non-fusion Atf1 or fused with protein tags.

Coomassie-stained SDS-PAGE gel (top panel) and Western blot (bottom panel) of whole cell lysates of LB, 2YT and SB auto-induction media cultures used for expression of non-fusion Atf1 and Atf1 fused with MBP or NusA. Non-fusion Atf1 and MBP-Atf1 ran at the expected molecular weight of 65 kDa and 106 kDa, respectively. 10 μ g of each fraction were loaded in SDS-PAGE and WB, respectively.

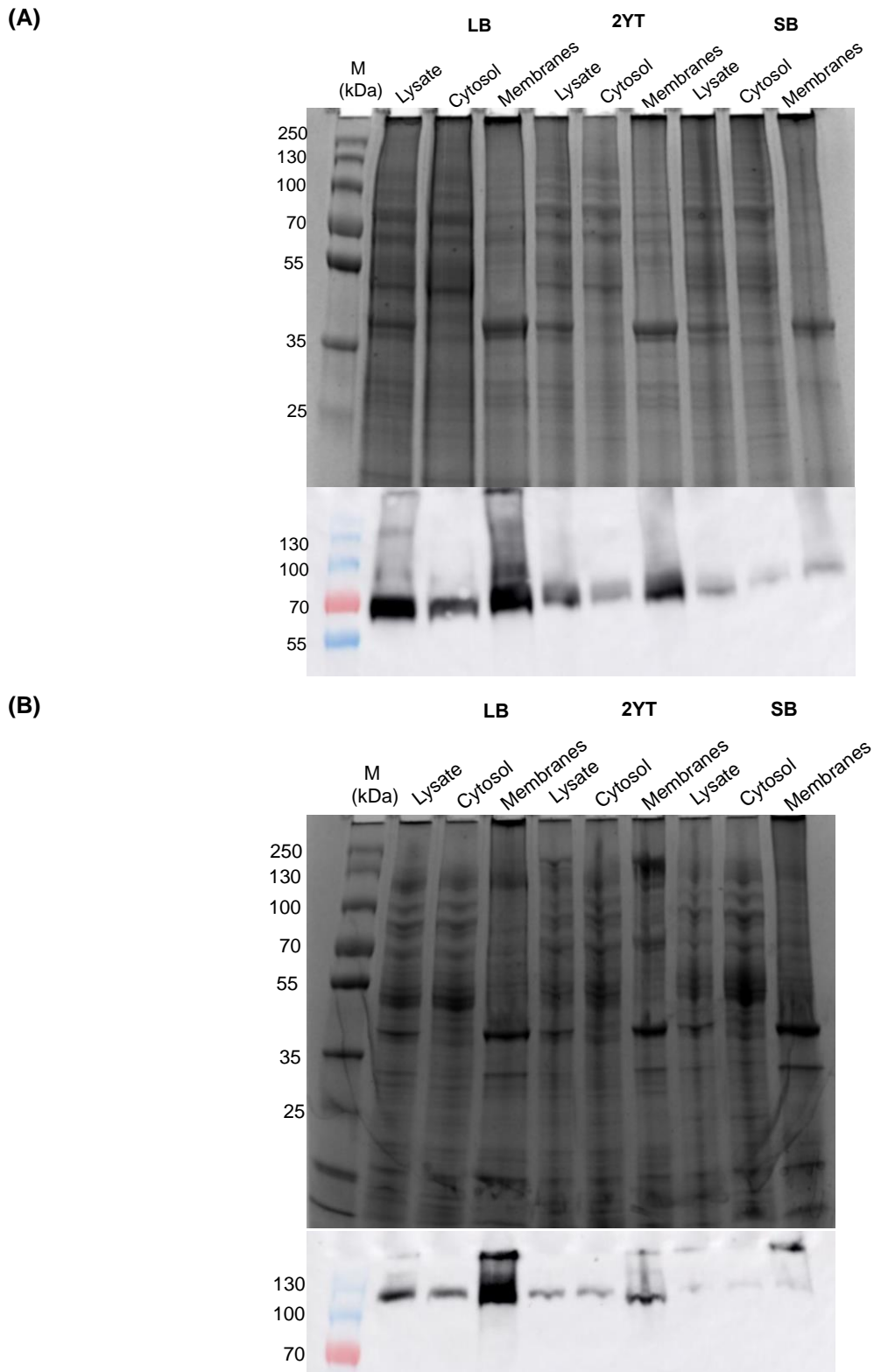
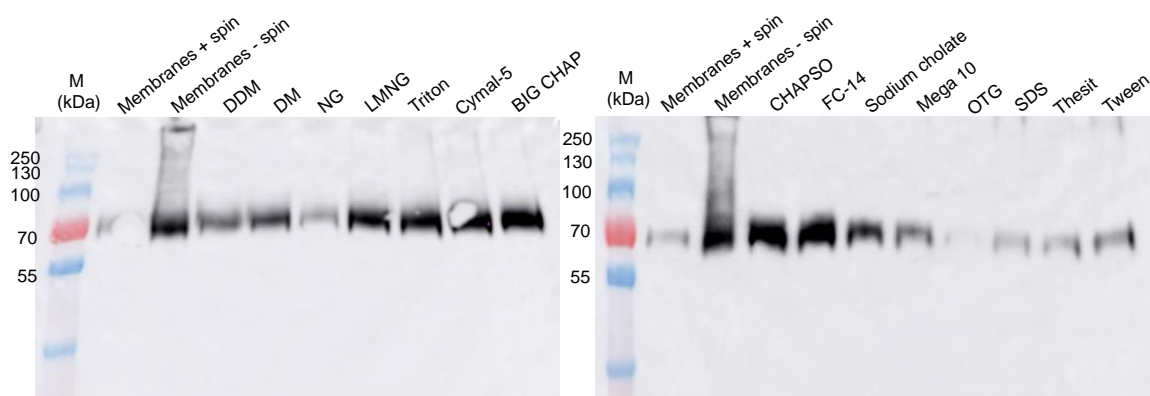


Figure 9.23. Large scale expression of Atf1 and MBP-Atf1.

Coomassie-stained SDS-PAGE gel (top panel) and Western blot (bottom panel) of lysate, cytosol and membranes of LB, 2YT and SB auto-induction media cultures used for expression of **(A)** non-fusion Atf1 protein and **(B)** Atf1 fused with MBP. Atf1 and MBP-Atf1 ran at the expected molecular weight of 65 kDa and 106 kDa, respectively. 10 μ g of each fraction were loaded for both Coomassie staining and Western blotting.

Purification of Atf1 was attempted from cytosol and membrane fractions by IMAC. No protein was purified from the cytosol. The membrane fraction was subjected to a detergent screening before total solubilisation (Figure 9.24A). A number of detergents were identified as supporting the solubilisation of Atf1, including LMNG, Triton X-100, Cymal-5, BIG CHAP, CHAPSO and FC-14. One of these 'hits', Cymal-5, was used in purification. Although solubilisation with Cymal-5 yielded 0.29 mg/L of culture, the SEC chromatogram showed a highly heterogenous purification with no visible band on the SDS-PAGE gel (Figure 9.24B).

(A)



(B)

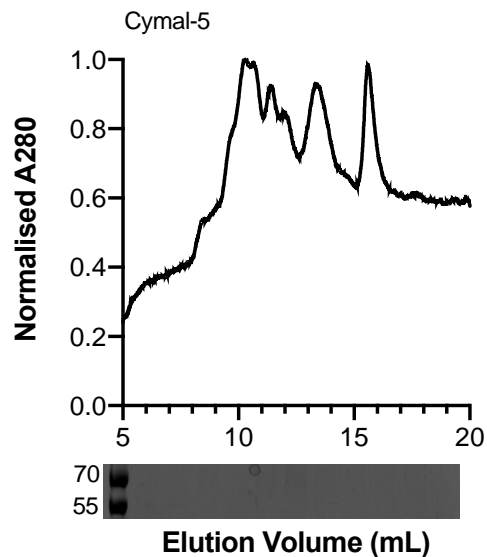


Figure 9.24. Detergent screening of Atf1 expressed in *E. coli* BL21(DE3) cells cloned in pOPINF and growth in LB_{AIM}.

(A) Western blotting of solubilised membrane fraction of Atf1. Aliquots of membrane fraction were treated as shown, and supernatants were collected for analysis. 10 μ g were loaded. *Membrane + spin*, supernatant from the centrifugation of untreated membranes. *Membrane - spin*, untreated membranes prior to centrifugation. **(B)** Size exclusion chromatography of Atf1 solubilised in Cymal-5. SDS-PAGE gel showed no visible band.

The most interesting results obtained with fruit AATase results are shown in Table 9.5 and described as follows. FaAAT1 can be successfully expressed by *E. coli* after induction with IPTG. However, cell fractionation showed that FaAAT1 was found in the insoluble fraction (pellet, Figure 9.25A) and so probably forms inclusion bodies during expression. Similar results were observed with MAAT1 (Figure 9.25B). Inclusion bodies were treated with different chemicals, such as SDS, urea or GndHCl and supernatants were collected for analysis. The presence of protein in supernatants was analysed by western blotting, which revealed no protein in those fractions. Thus, neither FaAAT1 nor MAAT1 could be recovered from inclusion bodies.

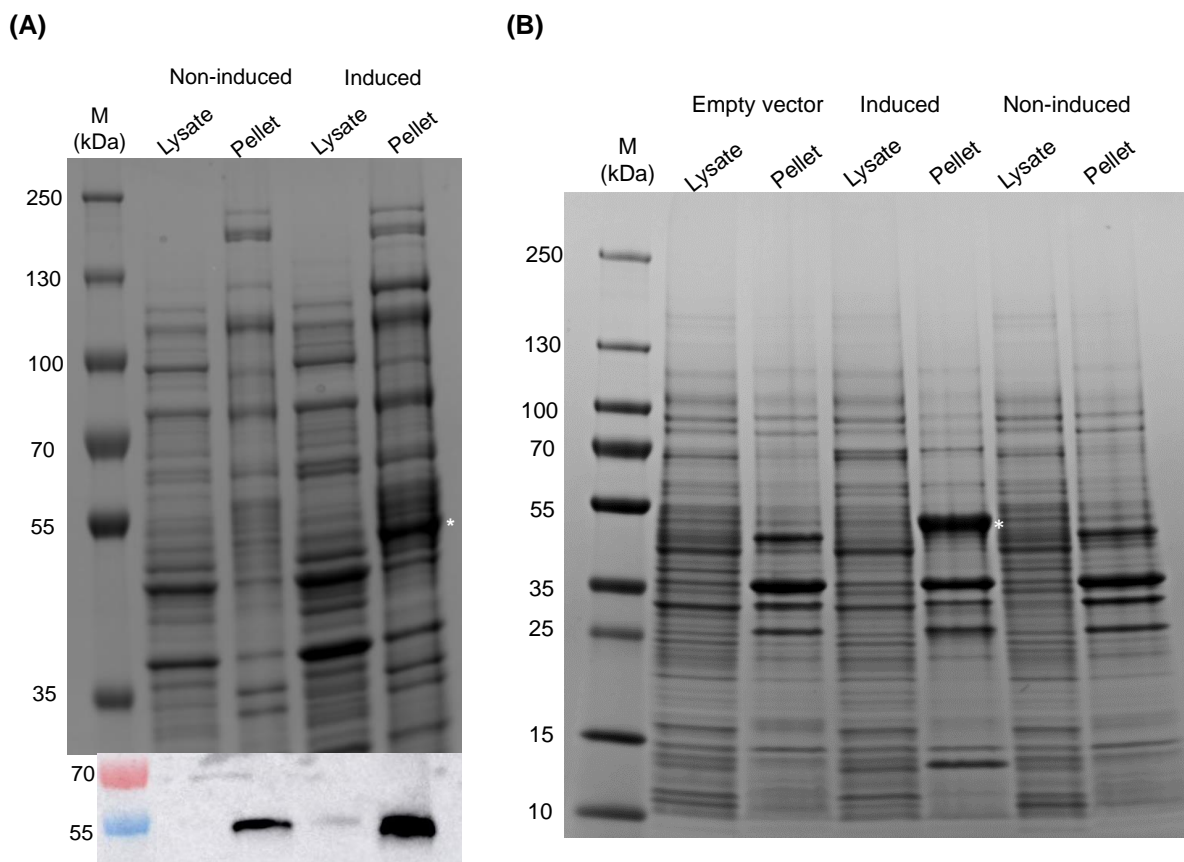


Figure 9.25. Coomassie-stained SDS-PAGE gels showing non-fusion FaAAT1 and MAAT expressed in the insoluble fraction.

(A) FaAAT1 protein ran at the expected molecular weight of 55 kDa (*), and **(B)** MAAT protein ran at the expected molecular weight of 56 kDa. *Empty vector* means popinF vector without an insert. Lysate (supernatant) and pellet (resuspended in 6 M urea) were from non-induced cells and induced cells at 1 mM of IPTG. 10 μ g of each fraction were loaded.

Surprisingly, the VpAAT1 enzyme was not overexpressed in either pOPINF, pOPINM or pOPINNUSA vectors (Figure 9.26).

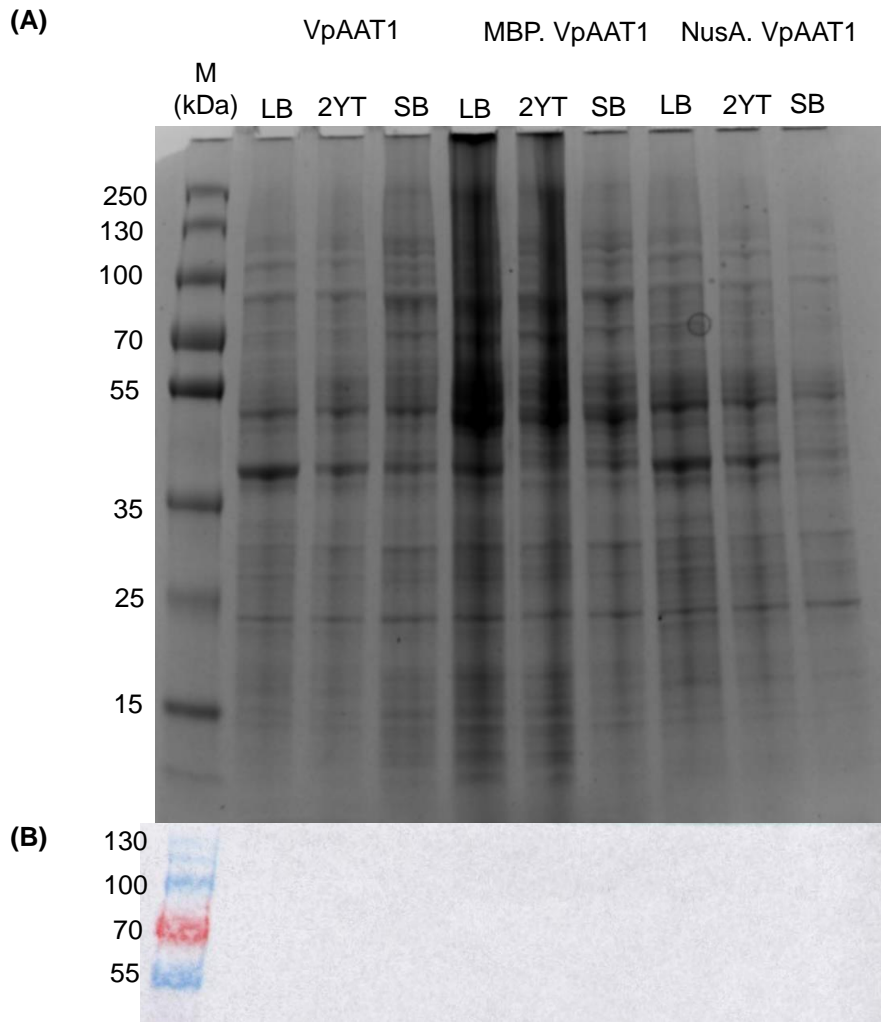


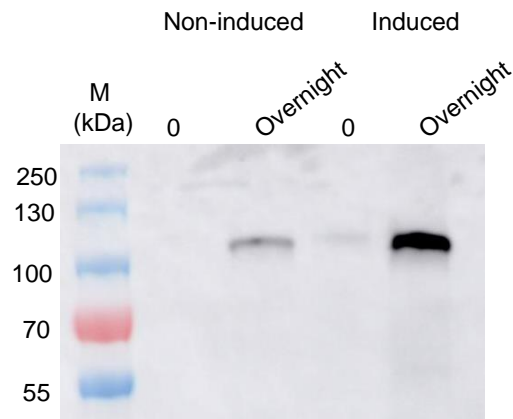
Figure 9.26. Small scale expression of non-fusion VpAAT1 or fused with protein tags.

(A) Coomassie-stained SDS-PAGE gel and **(B)** Western blot of whole cell lysates of LB, 2YT and SB auto-induction cultures used for expression of non-fusion VpAAT1 and VpAAT1 fused with MBP or NusA. No protein expression was observed in any condition. 20 μ g of each fraction were loaded in SDS-PAGE and WB, respectively.

Regarding MAAT expression using LB media and 1 mM IPTG, the protein was not expressed using the NusA fusion tag, but a certain amount of expression was observed when the MBP fusion tag was present; some basal expression was also observed (Figure 9.27A). To complement this study, small scale expression trials of MAAT fused with MBP or NusA in auto-induction media (LB, 2YT or SB) were carried out. The results showed that no protein was being expressed, and aggregates were observed at the top of the Coomassie-stained SDS-

PAGE gel (Figure 9.27B). The fact that these experiments were performed at 37°C for 24 hour may impact the formation of protein aggregates.

(A)



(B)

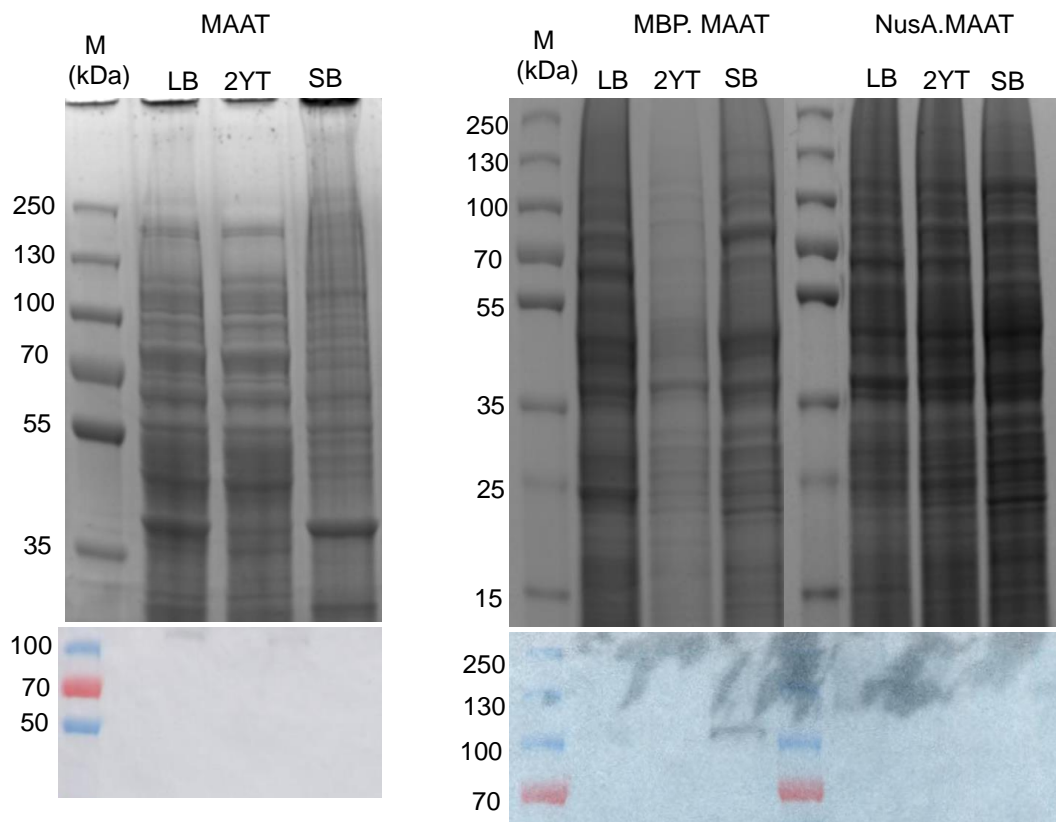


Figure 9.27. Small scale expression of non-fusion MAAT or fused with protein tags.

(A) Western blotting of MAAT fused with MBP confirmed that the recombinant protein was expressed in LB media induced with IPTG. Non-induced cell cultures were used as control. 10 µg of each fraction were loaded. MAAT fused with MBP ran at the expected molecular weight of 97 kDa. (B) MAAT was not expressed in either condition tested. Coomassie-stained SDS-PAGE gel (top panel) and Western blot (bottom panel) of whole-cell lysates of LB, 2YT and SB auto-induction cultures used for expression of non-fusion MAAT (left panel) and MAAT fused with MBP or NusA (right panel). 20 µg of each fraction were loaded in SDS-PAGE and WB, respectively.

Figure 9.28A and B shows the expression of MBP-FaAAT1 and NusA-FaAAT1 in LB media with 1 mM IPTG, respectively. Some expression of the MBP fusion was observed after overnight induction, but levels of the NusA construct were similar to basal expression levels. However, when large scale expression of FaAAT1 fused with MBP was attempted, it revealed that protein had a toxic effect on the host strain, inhibiting growth (Figure 9.28C and D).

To complement these results, auto-induction media were used for protein expression tests. Figure 9.29 shows the results of expressing FaAAT1 constructs in auto-induction media. Western blotting demonstrates that MBP-FaAAT1 is expressed in 2YT and SB, while NusA-FaAAT1 is not expressed under the conditions tested. Collectively, these results suggested that the MBP fusion of FaAAT1 was the most promising candidate for further purification.

A large scale expression of MBP-FaAAT1 in super broth auto-induction media was attempted. Figure 9.30A shows that FaAAT1 fused with MBP was expressed as expected but appeared to localise to the sedimenting membrane fraction, which was unexpected since the MBP fusion should render the protein soluble so that it resides in the cytosol⁵⁴⁹. Presumably, this represents protein aggregates that sediment under centrifugation but are not equivalent to classical inclusion bodies (which are removed before this spin).

A detergent screening for the best component to solubilise this membrane fraction was performed. After treatment with each detergent, the remaining insoluble component was pelleted, and supernatants were collected for analysis. Western blotting (Figure 9.30B) revealed that SDS, which is a harsh anionic detergent used for membrane disruption, was the best solubilising agent. Other, milder detergents were ineffective. Based on this result, membranes were treated with SDS to try and facilitate purification, but no protein could be obtained by nickel affinity chromatography. Additionally, 1% DDM was used to solubilise membranes; However, no protein was visible either in SDS-PAGE or Western Blot (Figure 9.31).

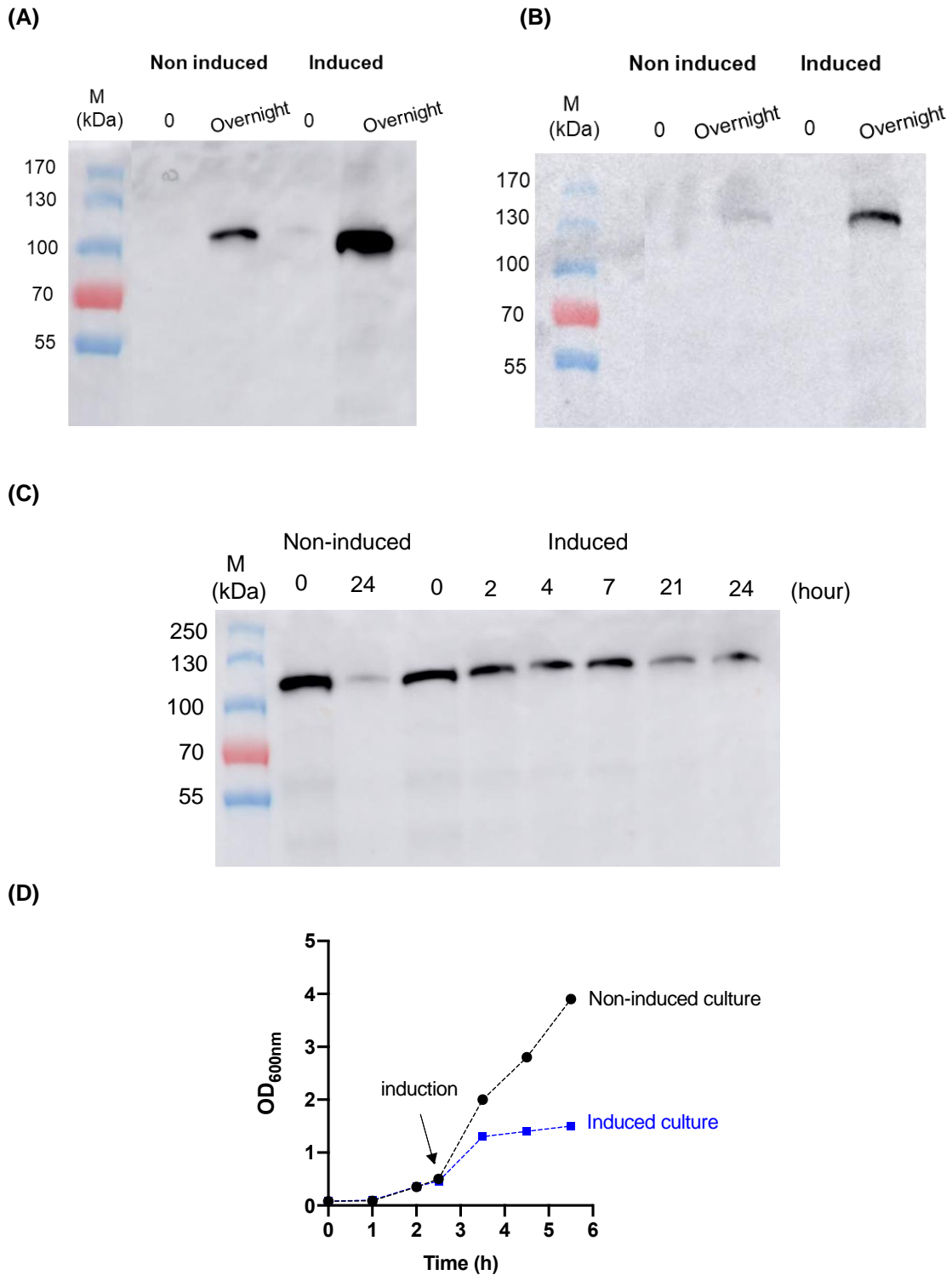


Figure 9.28. Small expression of FaAAT1 fused with protein tags in LB cultures induced with IPTG.

Western blotting of FaAAT1 fused with **(A)** MBP and **(B)** NusA confirmed that this recombinant protein was expressed upon IPTG induction. Non-induced cell cultures were used as control. **(C)** Western blotting of time points samples from large scale expression of FaAAT1 fused with MBP. The initial time corresponds to the point of inoculation of LB cultures. **(D)** OD₆₀₀ values plotted against time of expression (hour).

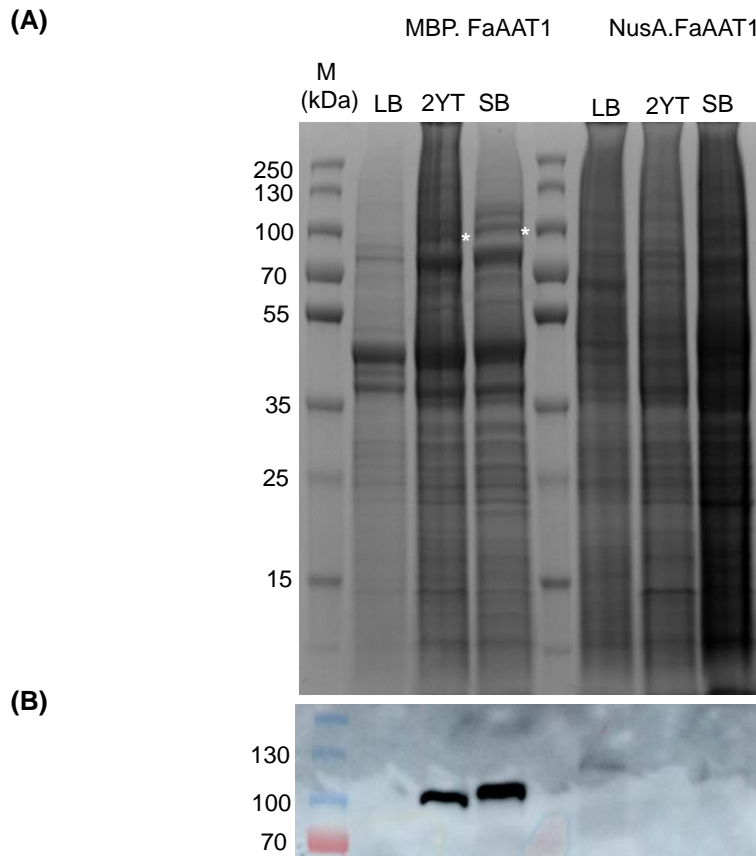


Figure 9.29. Small scale expression trials of FaAAT1 fused with protein tags.

(A) Coomassie-stained SDS-PAGE gel and (B) Western blotting of whole-cell lysates of LB, 2YT and SB auto-induction media cultures used for expression of FaAAT1 fused with MBP or NusA. FaAAT1 fused with MBP was expressed and ran at the expected molecular weight of 96 kDa.

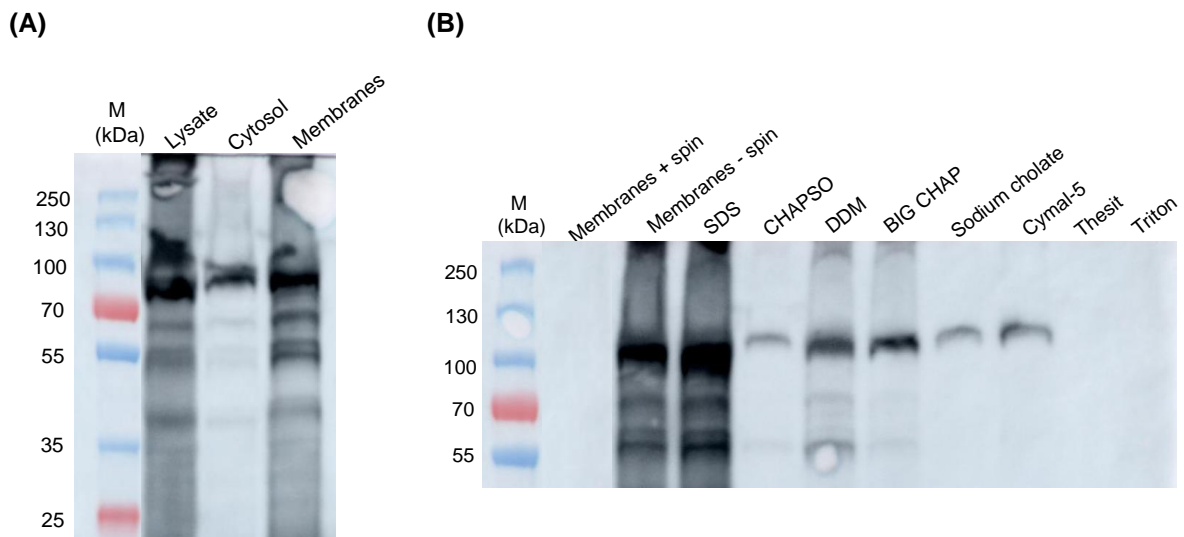


Figure 9.30. MBP-FaAAT1 was localised to sedimenting membranes.

(A) Large scale expression of FaAAT1 fused with MBP showed protein in the membrane fraction on Western blots. (B) Aliquots of membrane fraction were treated as shown, and supernatants were collected for analysis. 20 μ g were loaded. *Membrane + spin*, supernatant from the centrifugation of untreated membranes. *Membrane - spin*, untreated membranes prior to centrifugation.

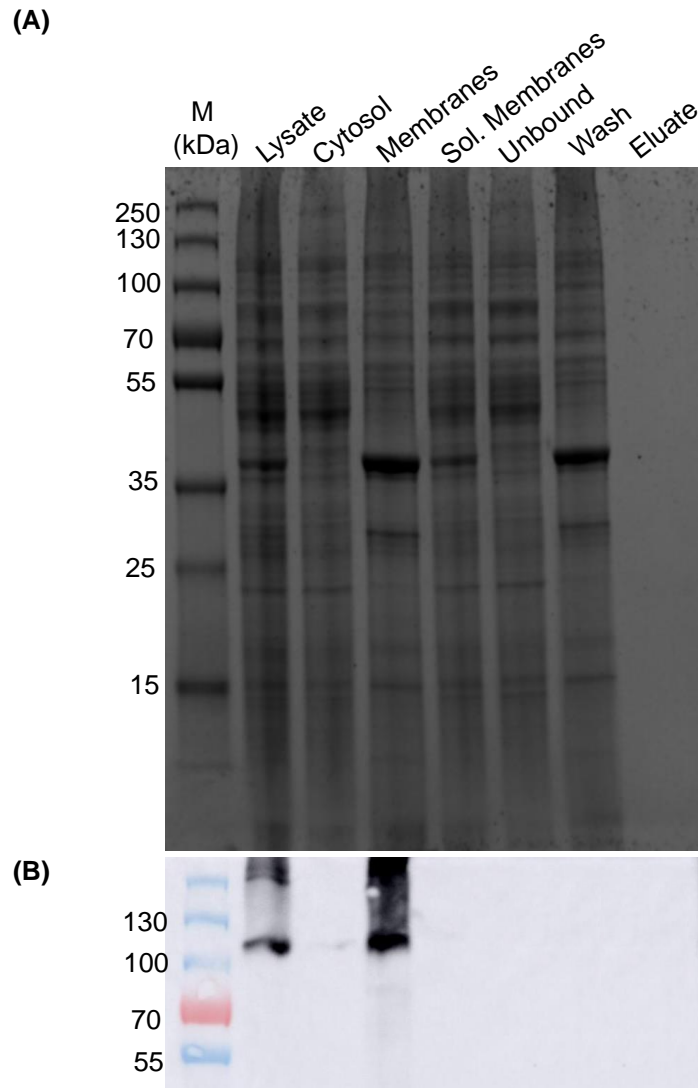


Figure 9.31. Large scale expression and purification of FaAAT1 fused with MBP solubilised with 1% DDM.

(A) Coomassie-stained SDS-PAGE gel of cell fractions and purified fractions from Ni²⁺ column. Lane headings are relevant to cell fractionation (lysate, cytosol, membranes, soluble membranes) or affinity purification (unbound, wash and purified protein). FaAAT1 fused with MBP ran at the expected molecular weight of 96 kDa. 20 µg of protein were loaded for all fractions. **(B)** Western blotting of cellular fractions confirmed the membrane localisation and lack of effective solubilisation (10 µg of protein were loaded).

9.4.3. Discussion

This section of the thesis aimed to clone and recombinantly express plant AATase genes in *E. coli*. Three genes from mountain papaya (*VpAAT1*), cultivated strawberry (*FaAAT1*) and melon (*MAAT*), as well as the functionally-related yeast *atf1* gene, were successfully cloned into the pOPINF, pOPINM and pOPINNUSA vectors. Ultimately, these constructs would allow the study of the structure and function of those AATases.

Small scale expression trials suggested that the production of fruit AATases was challenging since, in most cases, the construct was not expressed. Despite exploring numerous optimisation protocols and different *E. coli* strains (BL21(DE3) and BL21-AI), expression of plant AATases was generally poor, and the low levels of protein that were produced were unable to be purified. Atf1 was better expressed in a non-fusion construct or fused with MBP; However, Atf1 was still detected in the sedimenting membranes fraction, from which it and could not be successfully solubilised or purified. Therefore, Atf1 expressed in *S. cerevisiae* still the best expressing system.

Challenges in expressing the strawberry AATases have also been published in the literature. Ollas *et al.* reported the partial purification of FaAAT1 (or SAAT) from crude fruit extracts, where cells were resuspended in a buffer containing 0.1% Triton X-100. The same authors observed that without this detergent, only low levels of AATase activity were detected. Moreover, AATase activity was assessed in three subcellular locations, being only detected in the microsomal fraction. These *in vivo* observations suggested that strawberry AATase could be a membrane-bound enzyme since Triton X-100 is a mild non-ionic detergent^{512, 513}. However, the detergent screening performed here reported that Triton X-100 could not solubilise recombinant FaAAT1. In 2004, Beekwilder *et al.* reported the expression of two AATases from strawberry (*FaAAT1* and *FvAAT1*) in *Escherichia coli* (BL21 Gold (DE3) strain)⁴⁹¹. Surprisingly, both recombinant proteins were expressed in the soluble fraction and purified using Ni-NTA spin column without detergent. Nonetheless, no evidence of protein yield or stability was provided; there were no SDS-PAGE gels, Western Blots or size exclusion profiles in their study. Our results challenge their conclusions by presenting a very different picture, where FaAAT1 is not amendable to recombinant expression in *E. coli*.

Concerning *VpAAT1* and *MAAT*, no *E. coli* expression has previously been published in the literature. According to our findings, these proteins might prove very challenging to express. This agrees with the study from Yahyaoui *et al.*⁵⁰⁶, who showed that two melon AATases

(CmAAT1 or CmAAT2) were inactive when expressed in *E. coli*. Those authors suggested that plant AATases might need posttranslational modifications to support their expression. While this remains speculative, the need for such modifications would also be a reason for the absence of expression observed here.

It is interesting to note that many of these fruit AATases are strain-dependent, meaning that specific protein is only expressed in a specific *E. coli* host. For example, an apple AATase (MpAAT1) was only found in the soluble fraction when expressed in *E. coli* C43(DE3) cells⁵²⁸. Also, MpAAT1 was purified from the soluble fraction using IMAC, which afterwards also revealed the presence of the *E. coli* GroEL chaperone by peptide electrospray MS-MS analysis⁵²⁸. This unintentional co-expression of GroEL with MpAAT1 presumably kept MpAAT1 soluble during the process. Activity assays showed that MpAAT1 was active in cell extracts, presumably producing volatile esters using endogenous *E. coli* acetyl-CoA⁵²⁸. Another apple AATase, MdAAT2, was expressed in BL21(DE3)pLysS strain and purified by Ni-NTA spin column⁵⁵⁰. The authors used immunoblotting to show that MdAAT2 is mainly expressed in the fruit peel.

Chapter 10. Conclusions on alcohol acyltransferase work

The overarching aim of this section of the thesis was to derive new structural and functional information on alcohol acyltransferases from yeast and fruit. These enzymes are of significant interest in the food and drink industry since they produce a range of volatile esters with distinct aromas and flavours. These are implicated in both the taste and ripening processes of commercial crops as well as the taste of fermented alcoholic drinks such as beer and wine. The products formed by these enzymes – short- and medium-chain fatty acid esters – are also widely applied in the dairy, cosmetic and pharmaceutical sectors. As a result, in the last decade, industrial synthesis of these compounds has focused on metabolic engineering endeavour^{445,536}. Understanding the detailed catalytic mechanism of these enzymes would provide fundamental insight into an uncharacterised protein family and support the development of cellular ‘factories’ dedicated to the production of these chemicals.

A robust system for the recombinant expression and purification of alcohol acyltransferases (Atf1, VpAAT1, MAAT and FaAAT1) was required to achieve this goal. However, the expression and purification of those alcohol acyltransferases were found to be very challenging.

Concerning *S. cerevisiae* Atf1, the protocols developed and optimised here did not improve either solubility or expression yields beyond those already published³⁶⁸. Truncated Atf1 variants, in which terminal hydrophobic domains were removed, did not lead to the expected changes in cellular localisation and did not improve the size exclusion chromatography profile of Atf1. The use of common fusion tags (MBP, NusA) made no difference to these results. A notable outcome of the work described here was the identification of FC-14 as a solubilising detergent that could maintain Atf1 as a monodisperse monomer – the first time that this has been reported. Nonetheless, FC-14-solubilised protein was neither correctly folded nor fully functional in biochemical assays. While future work may be able to develop our findings a refold Atf1 from FC-14, it appears that Atf1 remains a challenging target for biophysical studies.

In the other topic covered in this chapter, selected fruit AATases were expressed in yeast and *E. coli*. This again was also unsuccessful, despite efforts to find appropriate strains, media conditions and fusion partners that could support expression. This difficulty in expressing AATases was unexpected given the numerous reports in the literature that claim to have

produced such enzymes recombinantly. Our work appears to call those studies into question, although we did not attempt to reproduce these experiments directly; for example, we used different expression vectors. However, the data here strongly suggest that recombinant yeast and bacterial cells are generally poor expression hosts for plant AATases. Future work on these enzymes should focus on construct optimisation, homologue screening, or the screening of completely different expression platforms from the ones already tested, such as baculovirus insect cells or plant cell lines.

Chapter 11. Overall conclusions

Membrane proteins perform a wide range of biological processes vital for cellular functions, and a deeper understanding of these proteins remains of intense interest. They are naturally found embedded in the heterogenous environment of the lipid membrane and have a reputation as being difficult to study. This thesis exemplifies many of these challenges and shows that, despite recent progress, there are still many hurdles that restrict the use of biochemical approaches for analysing membrane proteins.

This project presented the study of two different class of membrane proteins, an integral membrane protein, LicB choline permease, and a group of peripheral membranes, known as acyl-CoA:alcohol *O*-acyltransferase. In both cases, two critical and time-consuming steps were experienced. One of these was the choice of expression conditions – for example, the most suitable host strain or culture media to support successful, high-level expression. Secondly, the screening of different solubilising agents that could yield a sufficient amount of membrane proteins for subsequent functional and biophysical studies.

Although this thesis does provide indications of some productive routes forward, clearly, both *H influenzae* LicB and the acyltransferases are inherently difficult targets for characterisation. In both cases, it may be that extensive homologue screening to identify more tractable sequences. That is the only way to make significant progress towards answering the research questions posed here.

Concerning LicB permease, it was shown to be purified to homogeneity and conduct the first biophysical studies of this protein. The results confirm that LicB is an α -helical membrane protein, with the extent of helicity consistent with bioinformatic predictions. A novel fluorescence assay for ligand binding *in vitro* was established, allowing us to determine the ligand binding affinity and specificity of LicB. Alongside, the AATases study suggests that their expression and solubility are highly dependent on the host microorganism. Thus, a broader screening of different host strains would be an advantage.

The number of membrane protein structures is rapidly rising, if not quite at the exponential rate that was predicted in 2005 (<https://blanco.biomol.uci.edu/mpstruc/>). This increase is mainly due to recent technologies aimed at precisely solving the problems outlined in this

thesis. Novel expression systems, automated high-throughput screening and new solubilising agents might eventually identify conditions under which these proteins could be successfully crystallised. Without such developments, it seems likely that the fine details of these proteins will remain enigmatic for some time.

Chapter 12. References

1. Luckey, M. *Membrane Structural Biology: With Biochemical and Biophysical Foundations*. (Cambridge University Press, 2014).
2. Alberts, B. *et al.* Membrane structure. in *Molecular Biology of the Cell* 565–596 (W. W. Norton & Company, 2014).
3. Rietveld, A. & Simons, K. The differential miscibility of lipids as the basis for the formation of functional membrane rafts. *Biochim. Biophys. Acta - Biomembr.* **1376**, 467–479 (1998).
4. Singer, S. J. & Nicolson, G. L. The fluid mosaic model of the structure of cell membranes. *Science (80-)*. **175**, 720–731 (1972).
5. Nelson, D. & Cox, M. *Lehninger Principles of Biochemistry*. (W. H. Freeman and Company, 2008).
6. van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: Where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9**, 112–124 (2008).
7. Watson, H. Biological membranes. *Essays Biochem.* **59**, 43–69 (2015).
8. Brown, D. A. & London, E. Functions of lipid rafts in biological membranes. *Annu. Rev. Cell Dev. Biol.* **14**, 111–136 (1998).
9. Brown, R. E. Sphingolipid organization in biomembranes: What physical studies of model membranes reveal. *J. Cell Sci.* **111**, 1–9 (1998).
10. Becher, A. & McIlhinney, R. A. J. Consequences of lipid raft association on G-protein-coupled receptor function. *Biochem. Soc. Symp.* **72**, 151–164 (2005).
11. Simons, K. & Gerl, M. J. Revitalizing membrane rafts: New tools and insights. *Nat. Rev. Mol. Cell Biol.* **11**, 688–699 (2010).
12. Engelman, D. M. Membranes are more mosaic than fluid. *Nature* **438**, 578–580 (2005).
13. Whited, A. M. & Johs, A. The interactions of peripheral membrane proteins with biological membranes. *Chem. Phys. Lipids* **192**, 51–59 (2015).
14. Mouritsen, O. G. Lipids, curvature, and nano-medicine. *Eur. J. Lipid Sci. Technol.* **113**, 1174–1187 (2011).
15. Von Heijne, G. The membrane protein universe: What's out there and why bother? *J. Intern. Med.* **261**, 543–557 (2007).
16. Wallin, E. & Heijne, G. Von. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7**, 1029–1038 (1998).
17. Curnow, P. Membrane proteins in nanotechnology. *Biochem. Soc. Trans.* **37**, 643–652 (2009).

18. Abbott, G. Molecular mechanisms of cardiac voltage-gated potassium channelopathies. *Curr. Pharm. Des.* **12**, 3631–3644 (2006).
19. Nelson, M., Todorovic, S. & Perez-Reyes, E. The role of T-type calcium channels in epilepsy and pain. *Curr. Pharm. Des.* **12**, 2189–2197 (2006).
20. Riordan, J. R. Assembly of functional CFTR chloride channels. *Annu. Rev. Physiol.* **67**, 701–718 (2005).
21. Suzuki, Y. *et al.* Genetic and molecular bases of peroxisome biogenesis disorders. *Genet. Med.* **3**, 372–376 (2001).
22. Yin, H. & Flynn, A. D. Drugging membrane protein interactions. *Annu. Rev. Biomed. Eng.* **18**, 51–76 (2016).
23. Thomik, T., Wittig, I., Choe, J., Boles, E. & Oreb, M. An artificial transport metabolon facilitates improved substrate utilization in yeast. *Nat. Chem. Biol.* **13**, 1158–1163 (2017).
24. Bayoumi, M., Bayley, H., Maglia, G. & Sapra, K. T. Multi-compartment encapsulation of communicating droplets and droplet networks in hydrogel as a model for artificial cells. *Sci. Rep.* **7**, 45167 (2017).
25. Booth, M. J., Schild, V. R., Graham, A. D., Olof, S. N. & Bayley, H. Light-activated communication in synthetic tissues. *Sci. Adv.* **2**, e1600056 (2016).
26. Leister, D. Genetic engineering, synthetic biology and the light reactions of photosynthesis. *Plant Physiol.* **179**, 778–793 (2019).
27. Lalaurie, C. J. *et al.* The *de novo* design of a biocompatible and functional integral membrane protein using minimal sequence complexity. *Sci. Rep.* **8**, 14564 (2018).
28. Curnow, P. Designing minimalist membrane proteins. *Biochem. Soc. Trans.* **47**, 1233–1245 (2019).
29. Curnow, P. *et al.* Small-residue packing motifs modulate the structure and function of a minimal *de novo* membrane protein. *Sci. Rep.* **10**, 15203 (2020).
30. Niggli, V. Structural properties of lipid-binding sites in cytoskeletal proteins. *Trends Biochem. Sci.* **26**, 604–611 (2001).
31. Ben-Tal, N., Honig, B., Miller, C. & McLaughlin, S. Electrostatic binding of proteins to membranes. Theoretical predictions and experimental results with charybdotoxin and phospholipid vesicles. *Biophys. J.* **73**, 1717–1727 (1997).
32. Rytömaa, M. & Kinnunen, P. K. J. Reversibility of the binding of cytochrome C to liposomes. Implications for lipid-protein interactions. *J. Biol. Chem.* **270**, 3197–3202 (1995).
33. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1666**, 62–87 (2004).
34. von Heijne, G. & Manoil, C. Membrane proteins: From sequence to structure. *Protein*

- Eng.* **4**, 109–112 (1990).
35. Johnson, J. E. & Cornell, R. B. Amphitropic proteins: Regulation by reversible membrane interactions (review). *Mol. Membr. Biol.* **16**, 217–235 (1999).
 36. Murray, D. *et al.* Electrostatics and the membrane association of Src: Theory and experiment. *Biochemistry* **37**, 2145–2159 (1998).
 37. Newton, A. C. Protein kinase C: Structure, function, and regulation. *J. Biol. Chem.* **270**, 28495–28498 (1995).
 38. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459**, 347–355 (2009).
 39. Law, C. J., Maloney, P. C. & Wang, D.-N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **62**, 289–305 (2008).
 40. Saier, M. H. A functional-phylogenetic classification system for transmembrane solute transporters. *Microbiol. Mol. Biol. Rev.* **64**, 354–411 (2000).
 41. Morth, J. P. *et al.* A structural overview of the plasma membrane Na⁺, K⁺ -ATPase and H⁺ -ATPase ion pumps. *Nat. Rev. Mol. Cell Biol.* **12**, 60–70 (2011).
 42. Lolkema, J. S. & Slotboom, D.-J. Estimation of structural similarity of membrane proteins by hydropathy profile alignment. *Mol. Membr. Biol.* **15**, 33–42 (1998).
 43. Lolkema, J. S. & Slotboom, D.-J. Classification of 29 families of secondary transport proteins into a single structural class using hydropathy profile analysis. *J. Mol. Biol.* **327**, 901–909 (2003).
 44. Saier, M. H. *et al.* The transporter classification database (TCDB): Recent advances. *Nucleic Acids Res.* **44**, D372–D379 (2016).
 45. Saier, M. H., Reddy, V. S., Tamang, D. G. & Västermark, Å. The transporter classification database. *Nucleic Acids Res.* **42**, D251–D258 (2014).
 46. Lam, V. H. *et al.* Pathways of transport protein evolution: Recent advances. *Biol. Chem.* **392**, (2011).
 47. Jack, D. L., Yang, N. M. & Saier, M. H. The drug/metabolite transporter superfamily. *Eur. J. Biochem.* **268**, 3620–39 (2001).
 48. Yan, A., Guan, Z. & Raetz, C. R. H. An undecaprenyl phosphate-aminoarabinose flippase required for polymyxin resistance in *Escherichia coli*. *J. Biol. Chem.* **282**, 36077–36089 (2007).
 49. White, S. H. & Wimley, W. C. Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 319–365 (1999).
 50. Almén, M., Nordström, K. J., Fredriksson, R. & Schiöth, H. B. Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* **7**, 50 (2009).
 51. Oberai, A., Ihm, Y., Kim, S. & Bowie, J. U. A limited universe of membrane protein

- families and folds. *Protein Sci.* **15**, 1723–1734 (2006).
52. Imai, K., Gromiha, M. M. & Horton, P. Mitochondrial β -barrel proteins, an exclusive club? *Cell* **135**, 1158–1159 (2008).
 53. Elofsson, A. & Heijne, G. von. Membrane protein structure: Prediction versus reality. *Annu. Rev. Biochem.* **76**, 125–140 (2007).
 54. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: Resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
 55. Hedin, L. E., Illergård, K. & Elofsson, A. An introduction to membrane proteins. *J. Proteome Res.* **10**, 3324–3331 (2011).
 56. McKay, M. J., Afrose, F., Koeppe, R. E. & Greathouse, D. V. Helix formation and stability in membranes. *Biochim. Biophys. Acta - Biomembr.* **1860**, 2108–2117 (2018).
 57. Stansfeld, P. J. *et al.* MemProtMD: Automated insertion of membrane protein structures into explicit lipid membranes. *Structure* **23**, 1350–1361 (2015).
 58. Granseth, E., von Heijne, G. & Elofsson, A. A study of the membrane–water interface region of membrane proteins. *J. Mol. Biol.* **346**, 377–385 (2005).
 59. von Heijne, G. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. *J. Mol. Biol.* **225**, 487–494 (1992).
 60. Lins, L. & Brasseur, R. The hydrophobic effect in protein folding. *FASEB J.* **9**, 535–540 (1995).
 61. Eilers, M., Shekar, S. C., Shieh, T., Smith, S. O. & Fleming, P. J. Internal packing of helical membrane proteins. *PNAS* **97**, 5796–5801 (2000).
 62. Walters, R. F. S. & DeGrado, W. F. Helix-packing motifs in membrane proteins. *PNAS* **103**, 13658–13663 (2006).
 63. Kim, S. *et al.* Transmembrane glycine zippers: Physiological and pathological roles in membrane proteins. *PNAS* **102**, 14278–14283 (2005).
 64. Kauko, A., Illergård, K. & Elofsson, A. Coils in the membrane core are conserved and functionally important. *J. Mol. Biol.* **380**, 170–180 (2008).
 65. Almeida, J. G., Preto, A. J., Koukos, P. I., Bonvin, A. M. J. J. & Moreira, I. S. Membrane proteins structures: A review on computational modeling tools. *Biochim. Biophys. Acta - Biomembr.* **1859**, 2021–2039 (2017).
 66. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).
 67. Claros, M. G. & Heijne, G. von. TopPred II: An improved software for membrane protein structure predictions. *Bioinformatics* **10**, 685–686 (1994).
 68. Viklund, H. & Elofsson, A. Best α -helical transmembrane protein topology predictions are achieved using hidden Markov models and evolutionary information. *Protein Sci.*

- 13**, 1908–1917 (2004).
69. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580 (2001).
 70. Tusnády, G. E. & Simon, I. The HMMTOP transmembrane topology prediction server. *Bioinformatics* **17**, 849–850 (2001).
 71. Sonnhammer, E. L., von Heijne, G. & Krogh, A. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–82 (1998).
 72. Tusnády, G. E. & Simon, I. Principles governing amino acid composition of integral membrane proteins: Application to topology prediction. *J. Mol. Biol.* **283**, 489–506 (1998).
 73. Bernsel, A. *et al.* Prediction of membrane-protein topology from first principles. *PNAS* **105**, 7177–7181 (2008).
 74. Viklund, H., Bernsel, A., Skwark, M. & Elofsson, A. SPOCTOPUS: A combined predictor of signal peptides and membrane protein topology. *Bioinformatics* **24**, 2928–2929 (2008).
 75. Jones, D. T. Improving the accuracy of transmembrane protein topology prediction using evolutionary information. *Bioinformatics* **23**, 538–544 (2007).
 76. Reynolds, S. M., Käll, L., Riffle, M. E., Bilmes, J. A. & Noble, W. S. Transmembrane topology and signal peptide prediction using dynamic Bayesian networks. *PLoS Comput. Biol.* **4**, (2008).
 77. Liang, J., Naveed, H., Jimenez-Morales, D., Adamian, L. & Lin, M. Computational studies of membrane proteins: Models and predictions for biological understanding. *Biochim. Biophys. Acta - Biomembr.* **1818**, 927–941 (2012).
 78. Nugent, T. & Jones, D. T. Membrane protein structural bioinformatics. *J. Struct. Biol.* **179**, 327–337 (2012).
 79. Bernsel, A., Viklund, H., Hennerdal, A. & Elofsson, A. TOPCONS: Consensus prediction of membrane protein topology. *Nucleic Acids Res.* **37**, W465–W468 (2009).
 80. Käll, L., Krogh, A. & Sonnhammer, E. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **338**, 1027–1036 (2004).
 81. Nugent, T. & Jones, D. T. Transmembrane protein topology prediction using support vector machines. *BMC Bioinformatics* **10**, 159 (2009).
 82. Koehler Leman, J., Ulmschneider, M. B. & Gray, J. J. Computational modeling of membrane proteins. *Proteins Struct. Funct. Bioinforma.* **83**, 1–24 (2015).
 83. Viklund, H. & Elofsson, A. OCTOPUS: Improving topology prediction by two-track ANN-based preference scores and an extended topological grammar. *Bioinformatics* **24**,

- 1662–1668 (2008).
84. Fagerberg, L., Jonasson, K., von Heijne, G., Uhlén, M. & Berglund, L. Prediction of the human membrane proteome. *Proteomics* **10**, 1141–1149 (2010).
 85. Venko, K., Roy Choudhury, A. & Novič, M. Computational approaches for revealing the structure of membrane transporters: Case study on bilitranslocase. *Comput. Struct. Biotechnol. J.* **15**, 232–242 (2017).
 86. Peters, C., Tsirigos, K. D., Shu, N. & Elofsson, A. Improved topology prediction using the terminal hydrophobic helices rule. *Bioinformatics* **32**, 1158–1162 (2016).
 87. Kall, L., Krogh, A. & Sonnhammer, E. L. L. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* **21**, i251–i257 (2005).
 88. Tsirigos, K. D., Peters, C., Shu, N., Käll, L. & Elofsson, A. The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. *Nucleic Acids Res.* **43**, W401–W407 (2015).
 89. Drew, D., Fröderberg, L., Baars, L. & De Gier, J. W. L. Assembly and overexpression of membrane proteins in *Escherichia coli*. *Biochim. Biophys. Acta - Biomembr.* **1610**, 3–10 (2003).
 90. Bernaudat, F. *et al.* Heterologous expression of membrane proteins: Choosing the appropriate host. *PLoS One* **6**, e29191 (2011).
 91. Wagner, S. *et al.* Consequences of membrane protein overexpression in *Escherichia coli*. *Mol. Cell. Proteomics* **6**, 1527–1550 (2007).
 92. Booth, P. J. & Curnow, P. Membrane proteins shape up: Understanding *in vitro* folding. *Curr. Opin. Struct. Biol.* **16**, 480–488 (2006).
 93. Miroux, B. & Walker, J. E. Over-production of proteins in *Escherichia coli*: Mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289–298 (1996).
 94. Wagner, S. *et al.* Tuning *Escherichia coli* for membrane protein overexpression. *PNAS* **105**, 14371–14376 (2008).
 95. Kunji, E., Slotboom, D.-J. & Poolman, B. *Lactococcus lactis* as host for overproduction of functional membrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1610**, 97–108 (2003).
 96. Frelet-Barrand, A. *et al.* *Lactococcus lactis*, an alternative system for functional expression of peripheral and intrinsic *Arabidopsis* membrane proteins. *PLoS One* **5**, e8746 (2010).
 97. Wiener, M. A pedestrian guide to membrane protein crystallization. *Methods* **34**, 364–372 (2004).
 98. Shao, S., Brown, A., Santhanam, B. & Hegde, R. S. Structure and assembly pathway

- of the ribosome quality control complex. *Mol. Cell* **57**, 433–44 (2015).
99. Kost, T. A., Condeary, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23**, 567–575 (2005).
 100. Rosenbusch, J., Lustig, A., Grabo, M., Zulauf, M. & Regenass, M. Approaches to determining membrane protein structures to high resolution: Do selections of subpopulations occur? *Micron* **32**, 75–90 (2001).
 101. Gutmann, D. A. P. *et al.* A high-throughput method for membrane protein solubility screening: The ultracentrifugation dispersity sedimentation assay. *Protein Sci.* **16**, 1422–1428 (2007).
 102. Cross, T. A., Murray, D. T. & Watts, A. Helical membrane protein conformations and their environment. *Eur. Biophys. J.* **42**, 731–755 (2013).
 103. Choy, B. C., Cater, R. J., Mancina, F. & Pryor, E. E. A 10-year meta-analysis of membrane protein structural biology: Detergents, membrane mimetics, and structure determination techniques. *Biochim. Biophys. Acta - Biomembr.* **1863**, 183533 (2021).
 104. Mineev, K. S. & Nadezhdin, K. D. Membrane mimetics for solution NMR studies of membrane proteins. *Nanotechnol. Rev.* **6**, 15–32 (2017).
 105. Garavito, R. M. & Ferguson-Miller, S. Detergents as tools in membrane biochemistry. *J. Biol. Chem.* **276**, 32403–32406 (2001).
 106. Cockley, D. Detergents and their uses in membrane protein science. *Anatrace Man.* 1–17 (2007).
 107. Privé, G. G. Detergents for the stabilization and crystallization of membrane proteins. *Methods* **41**, 388–397 (2007).
 108. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination — The next generation. *Biochim. Biophys. Acta - Biomembr.* **1838**, 78–87 (2014).
 109. Kotov, V. *et al.* High-throughput stability screening for detergent-solubilized membrane proteins. *Sci. Rep.* **9**, 10379 (2019).
 110. Mandon, E. D., Agez, M., Pellegrin, R., Igonet, S. & Jawhari, A. Novel systematic detergent screening method for membrane proteins solubilization. *Anal. Biochem.* **517**, 40–49 (2017).
 111. Lantéz, V., Nikolaidis, I., Rechenmann, M., Vernet, T. & Noirclerc-Savoye, M. Rapid automated detergent screening for the solubilization and purification of membrane proteins and complexes. *Eng. Life Sci.* **15**, 39–50 (2015).
 112. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).
 113. Newstead, S., Kim, H., von Heijne, G., Iwata, S. & Drew, D. High-throughput

- fluorescent-based optimization of eukaryotic membrane protein overexpression and purification in *Saccharomyces cerevisiae*. *PNAS* **104**, 13936–13941 (2007).
114. Drew, D., Lerch, M., Kunji, E., Slotboom, D.-J. & de Gier, J.-W. Optimization of membrane protein overexpression and purification using GFP fusions. *Nat. Methods* **3**, 303–313 (2006).
 115. Drew, D. *et al.* A scalable, GFP-based pipeline for membrane protein overexpression screening and purification. *Protein Sci.* **14**, 2011–2017 (2005).
 116. Valiyaveetil, F. I., Zhou, Y. & MacKinnon, R. Lipids in the structure, folding, and function of the KcsA K⁺ Channel. *Biochemistry* **41**, 10771–10777 (2002).
 117. Lemieux, M. J. *et al.* Three-dimensional crystallization of the *Escherichia coli* glycerol-3-phosphate transporter: A member of the major facilitator superfamily. *Protein Sci.* **12**, 2748–2756 (2003).
 118. Gupta, K. *et al.* The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* **541**, 421–424 (2017).
 119. Banerjee, P., Joo, J. B., Buse, J. T. & Dawson, G. Differential solubilization of lipids along with membrane proteins by different classes of detergents. *Chem. Phys. Lipids* **77**, 65–78 (1995).
 120. Ding, Y., Fujimoto, L. M., Yao, Y., Plano, G. V. & Marassi, F. M. Influence of the lipid membrane environment on structure and activity of the outer membrane protein Ail from *Yersinia pestis*. *Biochim. Biophys. Acta - Biomembr.* **1848**, 712–720 (2015).
 121. Zhang, M., Huang, R., Im, S.-C., Waskell, L. & Ramamoorthy, A. Effects of membrane mimetics on cytochrome P450-Cytochrome b₅ interactions characterized by NMR spectroscopy. *J. Biol. Chem.* **290**, 12705–12718 (2015).
 122. Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172–175 (2014).
 123. Frey, L., Lakomek, N.-A., Riek, R. & Bibow, S. Micelles, bicelles and nanodiscs: Comparing the impact of membrane mimetics on membrane protein backbone dynamics. *Angew. Chemie Int. Ed.* **56**, 380–383 (2017).
 124. Warschawski, D. E. *et al.* Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1808**, 1957–1974 (2011).
 125. le Maire, M., Champeil, P. & Møller, J. V. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim. Biophys. Acta - Biomembr.* **1508**, 86–111 (2000).
 126. Privé, G. G. Lipopeptide detergents for membrane protein studies. *Curr. Opin. Struct. Biol.* **19**, 379–385 (2009).
 127. Kunji, E. R. S., Harding, M., Butler, P. J. G. & Akamine, P. Determination of the molecular mass and dimensions of membrane proteins by size exclusion

- chromatography. *Methods* **46**, 62–72 (2008).
128. Singh, S. K. & Sigworth, F. J. Cryo-EM: Spinning the micelles away. *Structure* **23**, 1561 (2015).
 129. Zhou, X., Kini, R. M. & Sivaraman, J. Application of isothermal titration calorimetry and column chromatography for identification of biomolecular targets. *Nat. Protoc.* **6**, 158–165 (2011).
 130. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. Biophys. Acta - Biomembr.* **1666**, 105–117 (2004).
 131. Linke, D. Detergents: An overview. *Methods Enzymol.* **463**, 603–617 (2009).
 132. Wu, S. *et al.* Determining the critical micelle concentration of surfactants by a simple and fast titration method. *Anal. Chem.* **92**, 4259–4265 (2020).
 133. Lin, S.-H. & Guidotti, G. Purification of membrane proteins. *Methods Enzymol.* **463**, 619–629 (2009).
 134. Stetsenko, A. & Guskov, A. An overview of the top ten detergents used for membrane protein crystallization. *Crystals* **7**, 197 (2017).
 135. Dilworth, M. V. *et al.* Microbial expression systems for membrane proteins. *Methods* **147**, 3–39 (2018).
 136. Sanders, C. R. & Prosser, R. S. Bicelles: A model membrane system for all seasons? *Structure* **6**, 1227–1234 (1998).
 137. Gayen, A., Goswami, S. K. & Mukhopadhyay, C. NMR evidence of GM1-induced conformational change of substance P using isotropic bicelles. *Biochim. Biophys. Acta - Biomembr.* **1808**, 127–139 (2011).
 138. Yamaguchi, T., Uno, T., Uekusa, Y., Yagi-Utsumi, M. & Kato, K. Ganglioside-embedding small bicelles for probing membrane-landing processes of intrinsically disordered proteins. *Chem. Commun.* **49**, 1235 (2013).
 139. Bae, H. E. *et al.* Tandem neopentyl glycol maltosides (TNMs) for membrane protein stabilisation. *Chem. Commun.* **52**, 12104–12107 (2016).
 140. Matar-Merheb, R. *et al.* Structuring detergents for extracting and stabilizing functional membrane proteins. *PLoS One* **6**, e18036 (2011).
 141. Zhang, Q. *et al.* Designing facial amphiphiles for the stabilization of integral membrane proteins. *Angew. Chemie Int. Ed.* **46**, 7023–7025 (2007).
 142. Lee, S. C. *et al.* Steroid-based facial amphiphiles for stabilization and crystallization of membrane proteins. *PNAS* **110**, E1203–E1211 (2013).
 143. He, X. *et al.* Structure of a cation-bound multidrug and toxic compound extrusion transporter. *Nature* **467**, 991–994 (2010).
 144. Ciancaglini, P. *et al.* Proteoliposomes in nanobiotechnology. *Biophys. Rev.* **4**, 67–81 (2012).

145. Rigaud, J. L., Bluzat, A. & Buschlen, S. Incorporation of bacteriorhodopsin into large unilamellar liposomes by reverse phase evaporation. *Biochem. Biophys. Res. Commun.* **111**, 373–382 (1983).
146. Seigneuret, M. & Rigaud, J. L. Analysis of passive and light-driven ion movements in large bacteriorhodopsin liposomes reconstituted by reverse-phase evaporation. 2. Influence of passive permeability and back-pressure effects upon light-induced proton uptake. *Biochemistry* **25**, 6723–6730 (1986).
147. Seigneuret, M. & Rigaud, J.-L. Use of the fluorescent pH probe pyranine to detect heterogeneous directions of proton movement in bacteriorhodopsin reconstituted large liposomes. *FEBS Lett.* **188**, 101–106 (1985).
148. Wheeler, T. J. & Hinkle, P. C. Kinetic properties of the reconstituted glucose transporter from human erythrocytes. *J. Biol. Chem.* **256**, 8907–14 (1981).
149. Kasahara, M. & Hinkle, P. C. Reconstitution and purification of the D-glucose transport protein from human erythrocytes. *J. Biol. Chem.* **252**, 7384–7390 (1977).
150. Rigaud, J.-L., Pitard, B. & Levy, D. Reconstitution of membrane proteins into liposomes: Application to energy-transducing membrane proteins. *Biochim. Biophys. Acta - Bioenerg.* **1231**, 223–246 (1995).
151. Bayburt, T. H., Grinkova, Y. V. & Sligar, S. G. Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins. *Nano Lett.* **2**, 853–856 (2002).
152. Denisov, I. G. & Sligar, S. G. Nanodiscs in membrane biochemistry and biophysics. *Chem. Rev.* **117**, 4669–4713 (2017).
153. Frauenfeld, J. *et al.* A saposin-lipoprotein nanoparticle system for membrane proteins. *Nat. Methods* **13**, 345–351 (2016).
154. Autzen, H. E., Julius, D. & Cheng, Y. Membrane mimetic systems in CryoEM: Keeping membrane proteins in their native environment. *Curr. Opin. Struct. Biol.* **58**, 259–268 (2019).
155. Denisov, I. G. & Sligar, S. G. Nanodiscs for structural and functional studies of membrane proteins. *Nat. Struct. Mol. Biol.* **23**, 481–486 (2016).
156. Akkaladevi, N. *et al.* Following nature's lead: On the construction of membrane-inserted toxins in lipid bilayer nanodiscs. *J. Membr. Biol.* **248**, 595–607 (2015).
157. Efremov, R. G., Leitner, A., Aebersold, R. & Raunser, S. Architecture and conformational switch mechanism of the ryanodine receptor. *Nature* **517**, 39–43 (2015).
158. Nikolaev, M. *et al.* Integral membrane proteins can be crystallized directly from nanodiscs. *Cryst. Growth Des.* **17**, 945–948 (2017).
159. Rasmussen, S. G. F. *et al.* Structure of a nanobody-stabilized active state of the $\beta 2$ adrenoceptor. *Nature* **469**, 175–180 (2011).
160. Faas, R. *et al.* Time-course and degradation rate of membrane scaffold protein

- (MSP1D1) during recombinant production. *Biotechnol. Reports* **17**, 45–48 (2018).
161. Hagn, F., Nasr, M. L. & Wagner, G. Assembly of phospholipid nanodiscs of controlled size for structural studies of membrane proteins by NMR. *Nat. Protoc.* **13**, 79–98 (2018).
 162. Stroud, Z., Hall, S. C. L. & Dafforn, T. R. Purification of membrane proteins free from conventional detergents: SMA, new polymers, new opportunities and new insights. *Methods* **147**, 106–117 (2018).
 163. Lee, S. C. *et al.* A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat. Protoc.* **11**, 1149–1162 (2016).
 164. Dörr, J. M. *et al.* The styrene–maleic acid copolymer: A versatile tool in membrane research. *Eur. Biophys. J.* **45**, 3–21 (2016).
 165. Scheidelaar, S. *et al.* Molecular model for the solubilization of membranes into nanodisks by styrene maleic acid copolymers. *Biophys. J.* **108**, 279–290 (2015).
 166. Dörr, J. M. *et al.* Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K⁺ channel: The power of native nanodiscs. *PNAS* **111**, 18607–18612 (2014).
 167. Swainsbury, D. J. K., Scheidelaar, S., van Grondelle, R., Killian, J. A. & Jones, M. R. Bacterial reaction centers purified with styrene maleic acid copolymer retain native membrane functional properties and display enhanced stability. *Angew. Chemie Int. Ed.* **53**, 11803–11807 (2014).
 168. Broecker, J., Eger, B. T. & Ernst, O. P. Crystallography of membrane proteins mediated by polymer-bounded lipid nanodiscs. *Structure* **25**, 384–392 (2017).
 169. Postis, V. *et al.* The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochim. Biophys. Acta - Biomembr.* **1848**, 496–501 (2015).
 170. Bersch, B., Dörr, J. M., Hessel, A., Killian, J. A. & Schanda, P. Proton-detected solid-state NMR spectroscopy of a zinc diffusion facilitator protein in native nanodiscs. *Angew. Chemie Int. Ed.* **56**, 2508–2512 (2017).
 171. Gulati, S. *et al.* Detergent-free purification of ABC (ATP-binding-cassette) transporters. *Biochem. J.* **461**, 269–278 (2014).
 172. Paulin, S. *et al.* Surfactant-free purification of membrane protein complexes from bacteria: Application to the staphylococcal penicillin-binding protein complex PBP2/PBP2a. *Nanotechnology* **25**, 285101 (2014).
 173. Kopf, A. H. *et al.* Factors influencing the solubilization of membrane proteins from *Escherichia coli* membranes by styrene–maleic acid copolymers. *Biochim. Biophys. Acta - Biomembr.* **1862**, 183125 (2020).
 174. Puthenveetil, R., Nguyen, K. & Vinogradova, O. Nanodiscs and solution NMR: Preparation, application and challenges. *Nanotechnol. Rev.* **6**, (2017).

175. Ravula, T., Hardin, N. Z., Ramadugu, S. K., Cox, S. J. & Ramamoorthy, A. Formation of pH-resistant monodispersed polymer-lipid nanodiscs. *Angew. Chemie Int. Ed.* **57**, 1342–1345 (2018).
176. Fiori, M. C., Jiang, Y., Altenberg, G. A. & Liang, H. Polymer-encased nanodiscs with improved buffer compatibility. *Sci. Rep.* **7**, 7432 (2017).
177. Popot, J.-L. Amphipols: Where from? Where to? *J. Membr. Biol.* **247**, 755–757 (2014).
178. Zoonens, M. & Popot, J.-L. Amphipols for each season. *J. Membr. Biol.* **247**, 759–796 (2014).
179. Picard, M. *et al.* Protective and inhibitory effects of various types of amphipols on the Ca²⁺-ATPase from sarcoplasmic reticulum: A comparative study. *Biochemistry* **45**, 1861–1869 (2006).
180. Diab, C., Tribet, C., Gohon, Y., Popot, J.-L. & Winnik, F. M. Complexation of integral membrane proteins by phosphorylcholine-based amphipols. *Biochim. Biophys. Acta - Biomembr.* **1768**, 2737–2747 (2007).
181. Champeil, P., Menguy, T., Tribet, C., Popot, J.-L. & le Maire, M. Interaction of amphipols with sarcoplasmic reticulum Ca²⁺-ATPase. *J. Biol. Chem.* **275**, 18623–18637 (2000).
182. Picard, M., Duval-Terrié, C., Dé, E. & Champeil, P. Stabilization of membranes upon interaction of amphipathic polymers with membrane proteins. *Protein Sci.* **13**, 3056–3058 (2004).
183. Popot, J.-L. *et al.* Amphipols: Polymeric surfactants for membrane biology research. *Cell. Mol. Life Sci.* **60**, 1559–1574 (2003).
184. Tribet, C., Audebert, R. & Popot, J.-L. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci.* **93**, 15047–15050 (1996).
185. Popot, J.-L. Amphipols, nanodiscs and fluorinated surfactants: Three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu. Rev. Biochem.* **79**, 737–775 (2010).
186. Pocanschi, C. L., Popot, J.-L. & Kleinschmidt, J. H. Folding and stability of outer membrane protein A (OmpA) from *Escherichia coli* in an amphipathic polymer, amphipol A8-35. *Eur. Biophys. J.* **42**, 103–118 (2013).
187. Pocanschi, C. L. *et al.* Amphipathic polymers: Tools to fold integral membrane proteins to their active form. *Biochemistry* **45**, 13954–13961 (2006).
188. Kiley, P. *et al.* Self-assembling peptide detergents stabilize isolated photosystem ion a dry surface for an extended time. *PLoS Biol.* **3**, e230 (2005).
189. Larsen, A. N. *et al.* Dimeric peptides with three different linkers self-assemble with phospholipids to form peptide nanodiscs that stabilize membrane proteins. *Soft Matter* **12**, 5937–5949 (2016).
190. Tao, H. *et al.* Engineered nanostructured β -sheet peptides protect membrane proteins.

- Nat. Methods* **10**, 759–761 (2013).
191. McGregor, C.-L. *et al.* Lipopeptide detergents designed for the structural study of membrane proteins. *Nat. Biotechnol.* **21**, 171–176 (2003).
 192. Schafmeister, C., Miercke, L. & Stroud, R. Structure at 2.5 Å of a designed peptide that maintains solubility of membrane proteins. *Science (80-.)*. **262**, 734–738 (1993).
 193. Zhao, X. *et al.* Designer short peptide surfactants stabilize G protein-coupled receptor bovine rhodopsin. *PNAS* **103**, 17707–17712 (2006).
 194. Zhang, S. Lipid-like self-assembling peptides. *Acc. Chem. Res.* **45**, 2142–2150 (2012).
 195. Corin, K. *et al.* Designer lipid-like peptides: A class of detergents for studying functional olfactory receptors using commercial cell-free systems. *PLoS One* **6**, e25067 (2011).
 196. Yeh, J. I., Du, S., Tortajada, A., Paulo, J. & Zhang, S. Peptergents: Peptide detergents that improve stability and functionality of a membrane protein, glycerol-3-phosphate dehydrogenase. *Biochemistry* **44**, 16912–16919 (2005).
 197. Kumpitsch, C., Koskinen, K., Schöpf, V. & Moissl-Eichinger, C. The microbiome of the upper respiratory tract in health and disease. *BMC Biol.* **17**, 87 (2019).
 198. Khan, S., Priti, S. & Ankit, S. Bacteria etiological agents causing lower respiratory tract infections and their resistance patterns. *Iran. Biomed. J.* **19**, 240–6 (2015).
 199. Watson, K. *et al.* Upper respiratory tract bacterial carriage in aboriginal and non-aboriginal children in a semi-arid area of western Australia. *Pediatr. Infect. Dis. J.* **25**, 782–790 (2006).
 200. Murphy, T. F., Bakaletz, L. O. & Smeesters, P. R. Microbial interactions in the respiratory tract. *Pediatr. Infect. Dis. J.* **28**, S121–S126 (2009).
 201. Man, W. H., de Steenhuijsen Piters, W. A. A. & Bogaert, D. The microbiota of the respiratory tract: Gatekeeper to respiratory health. *Nat. Rev. Microbiol.* **15**, 259–270 (2017).
 202. Prinzi, A. Normal respiratory microbiota in health and disease. *American Society for Microbiology* (2020). Available at: <https://asm.org/Articles/2020/February/Normal-Respiratory-Microbiota-in-Health-and-Diseas>. (Accessed: 20th January 2021)
 203. Bosch, A. A. T. M., Biesbroek, G., Trzcinski, K., Sanders, E. A. M. & Bogaert, D. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog.* **9**, e1003057 (2013).
 204. Siegel, S. J. & Weiser, J. N. Mechanisms of bacterial colonization of the respiratory tract. *Annu. Rev. Microbiol.* **69**, 425–444 (2015).
 205. Clark, S. E. & Weiser, J. N. Microbial modulation of host immunity with the small molecule phosphorylcholine. *Infect. Immun.* **81**, 392–401 (2013).
 206. Mathers, C. D. & Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* **3**, e442 (2006).

207. Foresight. Infectious diseases: Preparing for the future. *UK Office of Science and Innovation* (2006).
208. Burki, T. K. The economic cost of respiratory disease in the UK. *Lancet Respir. Med.* **5**, 381 (2017).
209. Chonmaitree, T. *et al.* Viral upper respiratory tract infection and otitis media complication in young children. *Clin. Infect. Dis.* **46**, 815–823 (2008).
210. Ayala, O. D. *et al.* Characterization of bacteria causing acute otitis media using Raman microspectroscopy. *Anal. Methods* **9**, 1864–1871 (2017).
211. José, R. J. Respiratory infections: A global burden. *Ann. Res. Hosp.* **2**, 12–12 (2018).
212. World Health Organization. Pneumonia the forgotten killer of children. *World Heal. Organ.* **1**, 44 (2006).
213. Cappelletty, D. Microbiology of bacterial respiratory infections. *Pediatr. Infect. Dis. J.* **17**, S55–S61 (1998).
214. Paul, S. P., Wilkinson, R. & Routley, C. Management of respiratory tract infections in children. *Nurs. Res. Rev.* **4**, 135–148 (2014).
215. Brook, I. Microbiology of sinusitis. *Proc. Am. Thorac. Soc.* **8**, 90–100 (2011).
216. King, P. *Haemophilus influenzae* and the lung. *Clin. Transl. Med.* **1**, (2012).
217. Van Eldere, J., Slack, M. P. E., Ladhani, S. & Cripps, A. W. Non-typeable *Haemophilus influenzae*, an under-recognised pathogen. *Lancet Infect. Dis.* **14**, 1281–1292 (2014).
218. Blaser, M. J. & Falkow, S. What are the consequences of the disappearing human microbiota? *Nat. Rev. Microbiol.* **7**, 887–894 (2009).
219. Adam, H. J. *et al.* Changing epidemiology of invasive *Haemophilus influenzae* in Ontario, Canada: Evidence for herd effects and strain replacement due to Hib vaccination. *Vaccine* **28**, 4073–4078 (2010).
220. Erwin, A. L. & Smith, A. L. Nontypeable *Haemophilus influenzae*: Understanding virulence and commensal behavior. *Trends Microbiol.* **15**, 355–362 (2007).
221. Agrawal, A. & Murphy, T. F. *Haemophilus influenzae* infections in the *H. influenzae* type b conjugate vaccine era. *J. Clin. Microbiol.* **49**, 3728–3732 (2011).
222. Ladhani, S. *et al.* Invasive *Haemophilus influenzae* disease, Europe, 1996–2006. *Emerg. Infect. Dis.* **16**, 455–463 (2010).
223. Bondy, J. *et al.* *Haemophilus influenzae* LicB contributes to lung damage in an aged mice co-infection model. *Microb. Pathog.* **90**, 1–6 (2016).
224. Murphy, T. V., Pastor, P., Medley, F., Osterholm, M. T. & Granoff, D. M. Decreased *Haemophilus* colonization in children vaccinated with *Haemophilus influenzae* type b conjugate vaccine. *J. Pediatr.* **122**, 517–523 (1993).
225. Ulanova, M. & Tsang, R. S. W. Invasive *Haemophilus influenzae* disease: Changing epidemiology and host–parasite interactions in the 21st century. *Infect. Genet. Evol.* **9**,

- 594–605 (2009).
226. Murphy, T. F. Respiratory infections caused by non-typeable *Haemophilus influenzae*. *Curr. Opin. Infect. Dis.* **16**, 129–134 (2003).
 227. Morey, P. *et al.* Evidence for a non-replicative intracellular stage of nontypable *Haemophilus influenzae* in epithelial cells. *Microbiology* **157**, 234–250 (2011).
 228. Shukla, S. D., Sohal, S. S., O'Toole, R. F., Eapen, M. S. & Walters, E. H. Platelet activating factor receptor: Gateway for bacterial chronic airway infection in chronic obstructive pulmonary disease and potential therapeutic target. *Expert Rev. Respir. Med.* **9**, 473–85 (2015).
 229. Malanovic, N. & Lohner, K. Gram-positive bacterial cell envelopes: The impact on the activity of antimicrobial peptides. *Biochim. Biophys. Acta - Biomembr.* **1858**, 936–946 (2016).
 230. Beveridge, T. J. Structures of Gram-negative cell walls and their derived membrane vesicles. *J. Bacteriol.* **181**, 4725–4733 (1999).
 231. Epand, R. M., Walker, C., Epand, R. F. & Magarvey, N. A. Molecular mechanisms of membrane targeting antibiotics. *Biochim. Biophys. Acta - Biomembr.* **1858**, 980–987 (2016).
 232. Brown, L., Wolf, J. M., Prados-Rosales, R. & Casadevall, A. Through the wall: Extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi. *Nat. Rev. Microbiol.* **13**, 620–630 (2015).
 233. Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of *Streptococcus pneumoniae* virulence factors in host respiratory colonization and disease. *Nat. Rev. Microbiol.* **6**, 288–301 (2008).
 234. Galanos, C. & Freudenberg, M. A. Bacterial endotoxins: Biological properties and mechanisms of action. *Mediators Inflamm.* **2**, S11–S16 (1993).
 235. Bryant, C. E., Spring, D. R., Gangloff, M. & Gay, N. J. The molecular basis of the host response to lipopolysaccharide. *Nat. Rev. Microbiol.* **8**, 8–14 (2010).
 236. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71**, 635–700 (2002).
 237. Park, B. S. *et al.* The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. *Nature* **458**, 1191–1195 (2009).
 238. Preston, A., Mandrell, R. E., Gibson, B. W. & Apicella, M. A. The lipooligosaccharides of pathogenic Gram-negative bacteria. *Crit. Rev. Microbiol.* **22**, 139–180 (1996).
 239. Weiser, J. N., Shchepetov, M. & Chong, S. T. Decoration of lipopolysaccharide with phosphorylcholine: A phase-variable characteristic of *Haemophilus influenzae*. *Infect. Immun.* **65**, 943–50 (1997).
 240. Schweda, E., Richards, J., Hood, D. & Moxon, E. Expression and structural diversity of

- the lipopolysaccharide of *Haemophilus influenzae*: Implication in virulence. *Int. J. Med. Microbiol.* **297**, 297–306 (2007).
241. Virji, M. Pathogenic *Neisseriae*: Surface modulation, pathogenesis and infection control. *Nat. Rev. Microbiol.* **7**, 274–286 (2009).
 242. Weiser, J. N., Lindberg, A. A., Manning, E. J., Hansen, E. J. & Moxon, E. R. Identification of a chromosomal locus for expression of lipopolysaccharide epitopes in *Haemophilus influenzae*. *Infect. Immun.* **57**, 3045–52 (1989).
 243. Weiser, J. N., Love, J. M. & Moxon, E. R. The molecular mechanism of phase variation of *H. influenzae* lipopolysaccharide. *Cell* **59**, 657–665 (1989).
 244. High, N. J., Deadman, M. E. & Moxon, E. R. The role of a repetitive DNA motif (5'-CAAT-3') in the variable expression of the *Haemophilus influenzae* lipopolysaccharide epitope alpha Gal(1-4)beta Gal. *Mol. Microbiol.* **9**, 1275–1282 (1993).
 245. High, N. J., Jennings, M. P. & Moxon, E. R. Tandem repeats of the tetramer 5'-CAAT-3' present in *lic2A* are required for phase variation but not lipopolysaccharide biosynthesis in *Haemophilus influenzae*. *Mol. Microbiol.* **20**, 165–174 (1996).
 246. Weiser, J. N. Relationship between colony morphology and the life cycle of *Haemophilus influenzae*: The contribution of lipopolysaccharide phase variation to pathogenesis. *J. Infect. Dis.* **168**, 672–680 (1993).
 247. Weiser, J. N., Maskell, D. J., Butler, P. D., Lindberg, A. A. & Moxon, E. R. Characterization of repetitive sequences controlling phase variation of *Haemophilus influenzae* lipopolysaccharide. *J. Bacteriol.* **172**, 3304–3309 (1990).
 248. Hood, D. W. *et al.* DNA repeats identify novel virulence genes in *Haemophilus influenzae*. *PNAS* **93**, 11121–11125 (1996).
 249. Jarosik, G. P. & Hansen, E. J. Identification of a new locus involved in expression of *Haemophilus influenzae* type b lipooligosaccharide. *Infect. Immun.* **62**, 4861–7 (1994).
 250. Martí-Llitas, P. *et al.* Nontypable *Haemophilus influenzae* displays a prevalent surface structure molecular pattern in clinical isolates. *PLoS One* **6**, e21133 (2011).
 251. Weiser, J. N. & Pan, N. Adaptation of *Haemophilus influenzae* to acquired and innate humoral immunity based on phase variation of lipopolysaccharide. *Mol. Microbiol.* **30**, 767–775 (1998).
 252. Mandrell, R. E. *et al.* Lipooligosaccharides (LOS) of some *Haemophilus* species mimic human glycosphingolipids, and some LOS are sialylated. *Infect. Immun.* **60**, 1322–8 (1992).
 253. Hood, D. W. *et al.* Identification of a lipopolysaccharide alpha-2,3-sialyltransferase from *Haemophilus influenzae*. *Mol. Microbiol.* **39**, 341–351 (2001).
 254. Månsson, M. *et al.* A new structural type for *Haemophilus influenzae* lipopolysaccharide. Structural analysis of the lipopolysaccharide from nontypeable

- Haemophilus influenzae* strain 486. *Eur. J. Biochem.* **268**, 2148–2159 (2001).
255. Schweda, E., Li, J., Moxon, E. & Richards, J. Structural analysis of lipopolysaccharide oligosaccharide epitopes expressed by non-typeable *Haemophilus influenzae* strain 176. *Carbohydr. Res.* **337**, 409–420 (2002).
256. Engskog, M. K. R., Deadman, M., Li, J., Hood, D. W. & Schweda, E. K. H. Detailed structural features of lipopolysaccharide glycoforms in nontypeable *Haemophilus influenzae* strain 2019. *Carbohydr. Res.* **346**, 1241–1249 (2011).
257. Lysenko, E. *et al.* The position of phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* affects binding and sensitivity to C-reactive protein-mediated killing. *Mol. Microbiol.* **35**, 234–245 (2000).
258. Fischer, W., Behr, T., Hartmann, R., Peter-Katalinic, J. & Egge, H. Teichoic acid and lipoteichoic acid of *Streptococcus pneumoniae* possess identical chain structures. A reinvestigation of teichoic acid (C polysaccharide). *Eur. J. Biochem.* **215**, 851–857 (1993).
259. Gillespie, S. H., Ainscough, S., Dickens, A. & Lewin, J. Phosphorylcholine-containing antigens in bacteria from the mouth and respiratory tract. *J. Med. Microbiol.* **44**, 35–40 (1996).
260. Weiser, J. N., Goldberg, J. B., Pan, N., Wilson, L. & Virji, M. The phosphorylcholine epitope undergoes phase variation on a 43-kilodalton protein in *Pseudomonas aeruginosa* and on pili of *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *Infect. Immun.* **66**, 4263–7 (1998).
261. McCrea, K. W., Xie, J., Marrs, C. F. & Gilsdorf, J. R. Prevalence of genetic differences in phosphorylcholine expression between nontypeable *Haemophilus influenzae* and *Haemophilus haemolyticus*. *BMC Microbiol.* **10**, 286 (2010).
262. Serino, L. & Virji, M. Genetic and functional analysis of the phosphorylcholine moiety of commensal *Neisseria* lipopolysaccharide. *Mol. Microbiol.* **43**, 437–448 (2002).
263. Serino, L. & Virji, M. Phosphorylcholine decoration of lipopolysaccharide differentiates commensal *Neisseriae* from pathogenic strains: Identification of *licA*-type genes in commensal *Neisseriae*. *Mol. Microbiol.* **35**, 1550–1559 (2002).
264. Ben-Menachem, G., Zähringer, U. & Rottem, S. The phosphocholine motif in membranes of *Mycoplasma fermentans* strains. *FEMS Microbiol. Lett.* **199**, 137–141 (2001).
265. Young, N. M. *et al.* Structural characterization and MHCII-dependent immunological properties of the zwitterionic O-chain antigen of *Morganella morganii*. *Glycobiology* **21**, 1266–1276 (2011).
266. Humphries, H. E. & High, N. J. The role of *licA* phase variation in the pathogenesis of invasive disease by *Haemophilus influenzae* type b. *FEMS Immunol. Med. Microbiol.*

- 34**, 221–230 (2002).
267. Weiser, J. N., Williams, A. & Moxon, E. R. Phase-variable lipopolysaccharide structures enhance the invasive capacity of *Haemophilus influenzae*. *Infect. Immun.* **58**, 3455–3457 (1990).
268. Henderson, I. R., Owen, P. & Nataro, J. P. Molecular switches - the ON and OFF of bacterial phase variation. *Mol. Microbiol.* **33**, 919–932 (1999).
269. Weiser, J. N. *et al.* Phosphorylcholine on the lipopolysaccharide of *Haemophilus influenzae* contributes to persistence in the respiratory tract and sensitivity to serum killing mediated by C-reactive protein. *J. Exp. Med.* **187**, 631–640 (1998).
270. Cundell, D. R., Gerard, N. P., Gerard, C., Idanpaan-Heikkila, I. & Tuomanen, E. I. *Streptococcus pneumoniae* anchor to activated human cells by the receptor for platelet-activating factor. *Nature* **377**, 435–438 (1995).
271. Swords, W. E. *et al.* Binding of the non-typeable *Haemophilus influenzae* lipooligosaccharide to the PAF receptor initiates host cell signalling. *Cell. Microbiol.* **3**, 525–536 (2001).
272. Swords, W. E. *et al.* Non-typeable *Haemophilus influenzae* adhere to and invade human bronchial epithelial cells via an interaction of lipooligosaccharide with the PAF receptor. *Mol. Microbiol.* **37**, 13–27 (2000).
273. Ishii, S. & Shimizu, T. Platelet-activating factor (PAF) receptor and genetically engineered PAF receptor mutant mice. *Prog. Lipid Res.* **39**, 41–82 (2000).
274. Mariano, F. *et al.* Platelet-activating factor synthesis by neutrophils, monocytes, and endothelial cells is modulated by nitric oxide production. *Shock* **19**, 339–344 (2003).
275. Bulger, E. M., Arbabi, S., Garcia, I. & Maier, R. V. The macrophage response to endotoxin requires platelet activating factor. *Shock* **17**, 173–179 (2002).
276. Schenkein, H. A., Barbour, S. E., Berry, C. R., Kipps, B. & Tew, J. G. Invasion of human vascular endothelial cells by *Actinobacillus actinomycetemcomitans* via the receptor for platelet-activating factor. *Infect. Immun.* **68**, 5416–5419 (2000).
277. Branger, J. *et al.* Platelet-activating factor receptor-deficient mice show an unaltered clearance of nontypeable *Haemophilus influenzae* from their respiratory tract. *Shock* **22**, 543–547 (2004).
278. Clementi, C. F. & Murphy, T. F. Non-typeable *Haemophilus influenzae* invasion and persistence in the human respiratory tract. *Front. Cell. Infect. Microbiol.* **1**, (2011).
279. Tong, H. H., Blue, L. E., James, M. A., Chen, Y. P. & DeMaria, T. F. Evaluation of phase variation of nontypeable *Haemophilus influenzae* lipooligosaccharide during nasopharyngeal colonization and development of otitis media in the chinchilla model. *Infect. Immun.* **68**, 4593–4597 (2000).
280. Fujita, K., Hirano, T., Kodama, S. & Suzuki, M. Prognostic impact of phosphorylcholine

- expression in nontypeable *Haemophilus influenzae* in otitis media with effusion. *Acta Otolaryngol.* **129**, 832–838 (2009).
281. Pang, B. *et al.* Lipooligosaccharides containing phosphorylcholine delay pulmonary clearance of nontypeable *Haemophilus influenzae*. *Infect. Immun.* **76**, 2037–2043 (2008).
 282. Elswaifi, S. F., Scarratt, W. K. & Inzana, T. J. The role of lipooligosaccharide phosphorylcholine in colonization and pathogenesis of *Histophilus somni* in cattle. *Vet. Res.* **43**, 49 (2012).
 283. Kharat, A. S. & Tomasz, A. Drastic reduction in the virulence of *Streptococcus pneumoniae* expressing type 2 capsular polysaccharide but lacking choline residues in the cell wall. *Mol. Microbiol.* **60**, 93–107 (2006).
 284. Lysenko, E. S., Gould, J., Bals, R., Wilson, J. M. & Weiser, J. N. Bacterial phosphorylcholine decreases susceptibility to the antimicrobial peptide LL-37/hCAP18 expressed in the upper respiratory tract. *Infect. Immun.* **68**, 1664–1671 (2000).
 285. Johnson, R. W., McGillivray, G., Denoël, P., Poolman, J. & Bakaletz, L. O. Abrogation of nontypeable *Haemophilus influenzae* protein D function reduces phosphorylcholine decoration, adherence to airway epithelial cells, and fitness in a chinchilla model of otitis media. *Vaccine* **29**, 1211–1221 (2011).
 286. Goldenberg, H. B., McCool, T. L. & Weiser, J. N. Cross-reactivity of human immunoglobulin G2 recognizing phosphorylcholine and evidence for protection against major bacterial pathogens of the human respiratory tract. *J. Infect. Dis.* **190**, 1254–1263 (2004).
 287. Casey, R., Newcombe, J., McFadden, J. & Bodman-Smith, K. B. The acute-phase reactant C-reactive protein binds to phosphorylcholine-expressing *Neisseria meningitidis* and increases uptake by human phagocytes. *Infect. Immun.* **76**, 1298–1304 (2008).
 288. Clark, S. E., Snow, J., Li, J., Zola, T. A. & Weiser, J. N. Phosphorylcholine allows for evasion of bactericidal antibody by *Haemophilus influenzae*. *PLoS Pathog.* **8**, e1002521 (2012).
 289. West-Barnette, S., Rockel, A. & Swords, W. E. Biofilm growth increases phosphorylcholine content and decreases potency of nontypeable *Haemophilus influenzae* endotoxins. *Infect. Immun.* **74**, 1828–1836 (2006).
 290. Hong, W. *et al.* Phosphorylcholine decreases early inflammation and promotes the establishment of stable biofilm communities of nontypeable *Haemophilus influenzae* strain 86-028NP in a chinchilla model of otitis media. *Infect. Immun.* **75**, 958–965 (2007).
 291. Hong, W., Pang, B., West-Barnette, S. & Swords, W. E. Phosphorylcholine expression

- by nontypeable *Haemophilus influenzae* correlates with maturation of biofilm communities *in vitro* and *in vivo*. *J. Bacteriol.* **189**, 8300–8307 (2007).
292. Mold, C., Du Clos, T. W., Nakayama, S., Edwards, K. M. & Gewurz, H. C-reactive protein reactivity with complement and effects on phagocytosis. *Ann. N. Y. Acad. Sci.* **389**, 251–262 (1982).
 293. Volanakis, J. E. & Kaplan, M. H. Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Exp. Biol. Med.* **136**, 612–614 (1971).
 294. Gould, J. M. & Weiser, J. N. Expression of C-reactive protein in the human respiratory tract. *Infect. Immun.* **69**, 1747–1754 (2001).
 295. Erwin, A. L. *et al.* Role of lipopolysaccharide phase variation in susceptibility of *Haemophilus influenzae* to bactericidal immunoglobulin M antibodies in rabbit sera. *Infect. Immun.* **68**, 2804–2807 (2000).
 296. Gehre, F. *et al.* Role of teichoic acid choline moieties in the virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **77**, 2824–2831 (2009).
 297. Gould, J. M. & Weiser, J. N. The inhibitory effect of C-reactive protein on bacterial phosphorylcholine platelet-activating factor receptor-mediated adherence is blocked by surfactant. *J. Infect. Dis.* **186**, 361–371 (2002).
 298. Baumgarth, N., Tung, J. W. & Herzenberg, L. A. Inherent specificities in natural antibodies: A key to immune defense against pathogen invasion. *Springer Semin. Immunopathol.* **26**, 347–362 (2005).
 299. Lieberman, R., Potter, M., Mushinski, E. B., Humphrey, W. & Rudikoff, S. Genetics of a new IgVH (T15 idiotype) marker in the mouse regulating natural antibody to phosphorylcholine. *J. Exp. Med.* **139**, 983–1001 (1974).
 300. Kenny, J. J. *et al.* Antigen binding and idiotype analysis of antibodies obtained after electroporation of heavy and light chain genes encoding phosphocholine-specific antibodies: A model for T15-idiotype dominance. *J. Exp. Med.* **176**, 1637–1643 (1992).
 301. Briles, D. E. *et al.* Antiphosphocholine antibodies found in normal mouse serum are protective against intravenous infection with type 3 *Streptococcus pneumoniae*. *J. Exp. Med.* **153**, 694–705 (1981).
 302. Briles, D. E., Latham Claflin, J., Schroer, K. & Forman, C. Mouse IgG3 antibodies are highly protective against infection with *Streptococcus pneumoniae*. *Nature* **294**, 88–90 (1981).
 303. Briles, D. E., Forman, C., Hudak, S. & Claflin, J. L. Anti-phosphorylcholine antibodies of the T15 idiotype are optimally protective against *Streptococcus pneumoniae*. *J. Exp. Med.* **156**, 1177–1185 (1982).
 304. Mi, Q.-S. *et al.* Highly reduced protection against *Streptococcus pneumoniae* after deletion of a single heavy chain gene in mouse. *Proc. Natl. Acad. Sci.* **97**, 6031–6036

- (2000).
305. Gray, B. M., Dillon, H. C. & Briles, D. E. Epidemiological studies of *Streptococcus pneumoniae* in infants: Development of antibody to phosphocholine. *J. Clin. Microbiol.* **18**, 1102–1107 (1983).
 306. Shaw, P. X. *et al.* Natural antibodies with the T15 idiotype may act in atherosclerosis, apoptotic clearance, and protective immunity. *J. Clin. Invest.* **105**, 1731–1740 (2000).
 307. Frostegård, J. Low level natural antibodies against phosphorylcholine: A novel risk marker and potential mechanism in atherosclerosis and cardiovascular disease. *Clin. Immunol.* **134**, 47–54 (2010).
 308. Thiagarajan, D. *et al.* Antibodies against phosphorylcholine and malondialdehyde during the first two years of life. *J. Immunol.* **205**, 2109–2116 (2020).
 309. Goodridge, H. S. *et al.* Immunomodulation via novel use of TLR4 by the filarial nematode phosphorylcholine-containing secreted product, ES-62. *J. Immunol.* **174**, 284–293 (2005).
 310. Marshall, F. A., Grierson, A. M., Garside, P., Harnett, W. & Harnett, M. M. ES-62, an immunomodulator secreted by filarial nematodes, suppresses clonal expansion and modifies effector function of heterologous antigen-specific T cells *in vivo*. *J. Immunol.* **175**, 5817–5826 (2005).
 311. Hage, F. G. & Szalai, A. J. C-reactive protein gene polymorphisms, C-reactive protein blood levels, and cardiovascular disease risk. *J. Am. Coll. Cardiol.* **50**, 1115–1122 (2007).
 312. Munson, R. S. & Sasaki, K. Protein D, a putative immunoglobulin D-binding protein produced by *Haemophilus influenzae*, is glycerophosphodiester phosphodiesterase. *J. Bacteriol.* **175**, 4569–4571 (1993).
 313. Oguoma, V. M. *et al.* 10-Valent pneumococcal non-typeable *H. influenzae* protein D conjugate vaccine (PHiD-CV10) versus 13-valent pneumococcal conjugate vaccine (PCV13) as a booster dose to broaden and strengthen protection from otitis media (PREVIX_BOOST) in Australian . *BMJ Open* **10**, e033511 (2020).
 314. Toropainen, M. *et al.* Pneumococcal *Haemophilus influenzae* Protein D conjugate vaccine induces antibodies that inhibit glycerophosphodiester phosphodiesterase activity of Protein D. *Infect. Immun.* **76**, 4546–4553 (2008).
 315. Prymula, R. *et al.* Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both *Streptococcus pneumoniae* and non-typable *Haemophilus influenzae*: A randomised double-blind efficacy study. *Lancet* **367**, 740–748 (2006).
 316. Novotny, L. *et al.* Passive immunization with human anti-protein D antibodies induced by polysaccharide protein D conjugates protects chinchillas against otitis media after

- intranasal challenge with *Haemophilus influenzae*. *Vaccine* **24**, 4804–4811 (2006).
317. Shiri, T., McCarthy, N. D. & Petrou, S. The impact of childhood pneumococcal vaccination on hospital admissions in England: A whole population observational study. *BMC Infect. Dis.* **19**, 510 (2019).
318. Bay, S. *et al.* Phosphorylcholine–carbohydrate– protein conjugates efficiently induce Hhpten-specific antibodies which recognize both *Streptococcus pneumoniae* and *Neisseria meningitidis*: A potential multitarget vaccine against respiratory infections. *J. Med. Chem.* **47**, 3916–3919 (2004).
319. Fischer, R. T., Longo, D. L. & Kenny, J. J. A novel phosphocholine antigen protects both normal and X-linked immune deficient mice against *Streptococcus pneumoniae*. Comparison of the 6-O-phosphocholine hydroxyhexanoate-conjugate with other phosphocholine-containing vaccines. *J. Immunol.* **154**, 3373–82 (1995).
320. Kenny, J. J., Guelde, G., Fischer, R. T. & Longo, D. L. Induction of phosphocholine-specific antibodies in X-linked immune deficient mice: *In vivo* protection against a *Streptococcus pneumoniae* challenge. *Int. Immunol.* **6**, 561–568 (1994).
321. Ohori, J., Iuchi, H., Maseda, Y. & Kurono, Y. Phosphorylcholine intranasal immunization with a 13-valent pneumococcal conjugate vaccine can boost immune response against *Streptococcus pneumoniae*. *Vaccine* **38**, 699–704 (2020).
322. Zhang, J.-R., Idanpaan-Heikkila, I., Fischer, W. & Tuomanen, E. I. Pneumococcal *licD2* gene is involved in phosphorylcholine metabolism. *Mol. Microbiol.* **31**, 1477–1488 (1999).
323. Kimura, A. & Hansen, E. J. Antigenic and phenotypic variations of *Haemophilus influenzae* type b lipopolysaccharide and their relationship to virulence. *Infect. Immun.* **51**, 69–79 (1986).
324. Waldow, F. *et al.* Attachment of phosphorylcholine residues to pneumococcal teichoic acids and modification of substitution patterns by the phosphorylcholine esterase. *J. Biol. Chem.* **293**, 10620–10629 (2018).
325. Young, N. M., Foote, S. J. & Wakarchuk, W. W. Review of phosphocholine substituents on bacterial pathogen glycans: Synthesis, structures and interactions with host proteins. *Mol. Immunol.* **56**, 563–573 (2013).
326. Alrousan, E. & Fan, X. Glycerophosphorylcholine regulates *Haemophilus influenzae* *glpQ* gene expression. *FEMS Microbiol. Lett.* **362**, (2015).
327. Fan, X., Goldfine, H., Lysenko, E. & Weiser, J. N. The transfer of choline from the host to the bacterial cell surface requires *glpQ* in *Haemophilus influenzae*. *Mol. Microbiol.* **41**, 1029–1036 (2001).
328. Song, X. M., Forsgren, A. & Janson, H. The gene encoding protein D (*hpd*) is highly conserved among *Haemophilus influenzae* type b and nontypeable strains. *Infect.*

- Immun.* **63**, 696–9 (1995).
329. Janson, H., Hedén, L. O. & Forsgren, A. Protein D, the immunoglobulin D-binding protein of *Haemophilus Influenzae*, is a lipoprotein. *Infect. Immun.* **60**, 1336-1342 (1992).
330. Whiting, G. Investigation of a choline phosphate synthesis pathway in *Streptococcus pneumoniae*: Evidence for choline phosphate cytidylyltransferase activity. *FEMS Microbiol. Lett.* **143**, 279–284 (1996).
331. Wang, L., Jiang, Y.-L., Zhang, J.-R., Zhou, C.-Z. & Chen, Y. Structural and enzymatic characterization of the choline kinase LicA from *Streptococcus pneumoniae*. *PLoS One* **10**, e0120467 (2015).
332. Elswaifi, S. F. *et al.* Molecular characterization of phosphorylcholine expression on the lipooligosaccharide of *Histophilus somni*. *Microb. Pathog.* **47**, 223–230 (2009).
333. Sohlenkamp, C., López-Lara, I. M. & Geiger, O. Biosynthesis of phosphatidylcholine in bacteria. *Prog. Lipid Res.* **42**, 115–162 (2003).
334. Ziegler, C., Bremer, E. & Krämer, R. The BCCT family of carriers: From physiology to crystal structure. *Mol. Microbiol.* **78**, 13–34 (2010).
335. Saier Lab Bioinformatics Group. Transporter Classification Database. Available at: <http://www.tcdb.org/search/result.php?tc=2.A.7>. (Accessed: 1st February 2021)
336. Rock, C. O., Heath, R. J., Park, H.-W. & Jackowski, S. The *licC* gene of *Streptococcus pneumoniae* encodes a CTP:phosphocholine cytidylyltransferase. *J. Bacteriol.* **183**, 4927–4931 (2001).
337. Campbell, H. A. & Kent, C. The CTP:phosphocholine cytidylyltransferase encoded by the *licC* gene of *Streptococcus pneumoniae*: Cloning, expression, purification, and characterization. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1534**, 85–95 (2001).
338. Kwak, B.-Y. *et al.* Structure and mechanism of CTP:phosphocholine cytidylyltransferase (LicC) from *Streptococcus pneumoniae*. *J. Biol. Chem.* **277**, 4343–4350 (2002).
339. Fox, K. L. *et al.* Duplicate copies of *lic1* direct the addition of multiple phosphocholine residues in the lipopolysaccharide of *Haemophilus influenzae*. *Infect. Immun.* **76**, 588–600 (2008).
340. Morona, J. K., Morona, R. & Paton, J. C. Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. *Mol. Microbiol.* **23**, 751–763 (1997).
341. Morona, J. K., Morona, R. & Paton, J. C. Analysis of the 5' portion of the type 19A capsule locus identifies two classes of *cpsC*, *cpsD*, and *cpsE* genes in *Streptococcus pneumoniae*. *J. Bacteriol.* **181**, 3599–3605 (1999).
342. Qu, J. *et al.* Proteomic expression profiling of *Haemophilus influenzae* grown in pooled

- human sputum from adults with chronic obstructive pulmonary disease reveal antioxidant and stress responses. *BMC Microbiol.* **10**, 162 (2010).
343. Lamark, T. *et al.* DNA sequence and analysis of the *bet* genes encoding the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. *Mol. Microbiol.* **5**, 1049–1064 (1991).
 344. Sleator, R. D., Gahan, C. G. M., Abee, T. & Hill, C. Identification and disruption of BetL, a secondary glycine betaine transport system linked to the salt tolerance of *Listeria monocytogenes* LO28. *Appl. Environ. Microbiol.* **65**, 2078–2083 (1999).
 345. Eichler, K., Bourgis, F., Buchet, A., Kleber, H.-P. & Mandrand-Berthelot, M.-A. Molecular characterization of the *cai* operon necessary for carnitine metabolism in *Escherichia coli*. *Mol. Microbiol.* **13**, 775–786 (1994).
 346. Peter, H., Burkovski, A. & Krämer, R. Isolation, characterization, and expression of the *Corynebacterium glutamicum betP* gene, encoding the transport system for the compatible solute glycine betaine. *J. Bacteriol.* **178**, 5229–5234 (1996).
 347. Weinand, M., Krämer, R. & Morbach, S. Characterization of compatible solute transporter multiplicity in *Corynebacterium glutamicum*. *Appl. Microbiol. Biotechnol.* **76**, 701–708 (2007).
 348. Kappes, R. M., Kempf, B. & Bremer, E. Three transport systems for the osmoprotectant glycine betaine operate in *Bacillus subtilis*: Characterization of OpuD. *J. Bacteriol.* **178**, 5071–5079 (1996).
 349. Chen, C. & Beattie, G. A. *Pseudomonas syringae* BetT is a low-affinity choline transporter that is responsible for superior osmoprotection by choline over glycine betaine. *J. Bacteriol.* **190**, 2717–2725 (2008).
 350. Peter, H., Burkovski, A. & Krämer, R. Osmo-sensing by N- and C-terminal extensions of the glycine betaine uptake system BetP of *Corynebacterium glutamicum*. *J. Biol. Chem.* **273**, 2567–2574 (1998).
 351. Tøndervik, A. & Strøm, A. R. Membrane topology and mutational analysis of the osmotically activated BetT choline transporter of *Escherichia coli*. *Microbiology* **153**, 803–813 (2007).
 352. Jung, H. *et al.* CaiT of *Escherichia coli*, a new transporter catalyzing L-carnitine/ γ -butyrobetaine exchange. *J. Biol. Chem.* **277**, 39251–39258 (2002).
 353. Farwick, M., Siewe, R. M. & Krämer, R. Glycine betaine uptake after hyperosmotic shift in *Corynebacterium glutamicum*. *J. Bacteriol.* **177**, 4690–4695 (1995).
 354. Ressler, S., Terwisscha van Scheltinga, A. C., Vornrhein, C., Ott, V. & Ziegler, C. Molecular basis of transport and regulation in the Na⁺/betaine symporter BetP. *Nature* **458**, 47–52 (2009).
 355. Tsai, C.-J. *et al.* Structural asymmetry in a trimeric Na⁺/betaine symporter, BetP, from

- Corynebacterium glutamicum*. *J. Mol. Biol.* **407**, 368–381 (2011).
356. Tang, L., Bai, L., Wang, W. & Jiang, T. Crystal structure of the carnitine transporter and insights into the antiport mechanism. *Nat. Struct. Mol. Biol.* **17**, 492–496 (2010).
357. Schulze, S., Köster, S., Geldmacher, U., Terwisscha van Scheltinga, A. C. & Kühlbrandt, W. Structural basis of Na⁺-independent and cooperative substrate/product antiport in CaiT. *Nature* **467**, 233–236 (2010).
358. Perez, C., Koshy, C., Yildiz, Ö. & Ziegler, C. Alternating-access mechanism in conformationally asymmetric trimers of the betaine transporter BetP. *Nature* **490**, 126–130 (2012).
359. Oswald, C. *et al.* Crystal structures of the choline/acetylcholine substrate-binding protein ChoX from *Sinorhizobium meliloti* in the liganded and unliganded-closed states. *J. Biol. Chem.* **283**, 32848–32859 (2008).
360. Lamark, T., Røkenes, T. P., McDougall, J. & Strøm, A. R. The complex bet promoters of *Escherichia coli*: Regulation by oxygen (ArcA), choline (BetI), and osmotic stress. *J. Bacteriol.* **178**, 1655–1662 (1996).
361. Fan, X., Pericone, C. D., Lysenko, E., Goldfine, H. & Weiser, J. N. Multiple mechanisms for choline transport and utilization in *Haemophilus influenzae*. *Mol. Microbiol.* **50**, 537–548 (2003).
362. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, 931 (2017).
363. Eberhardt, A., Wu, L. J., Errington, J., Vollmer, W. & Veening, J.-W. Cellular localization of choline-utilization proteins in *Streptococcus pneumoniae* using novel fluorescent reporter systems. *Mol. Microbiol.* **74**, 395–408 (2009).
364. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
365. Ebejer, J. P., Hill, J. R., Kelm, S., Shi, J. & Deane, C. M. Memoir: Template-based structure prediction for membrane proteins. *Nucleic Acids Res.* **41**, 379–383 (2013).
366. Zimmermann, L. *et al.* A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core. *J. Mol. Biol.* **430**, 2237–2243 (2018).
367. Kelm, S., Shi, J. & Deane, C. M. MEDELLER: Homology-based coordinate generation for membrane proteins. *Bioinformatics* **26**, 2833–2840 (2010).
368. Nancolas, B., Bull, I. D., Stenner, R., Dufour, V. & Curnow, P. *Saccharomyces cerevisiae* Atf1p is an alcohol acetyltransferase and a thioesterase *in vitro*. *Yeast* **34**, 239–251 (2017).
369. Ma, J. & Xia, D. The use of blue native PAGE in the evaluation of membrane protein aggregation states for crystallization. *J. Appl. Crystallogr.* **41**, 1150–1160 (2008).
370. Calvopiña, K. *et al.* Structural/mechanistic insights into the efficacy of nonclassical β-

- lactamase inhibitors against extensively drug resistant *Stenotrophomonas maltophilia* clinical isolates. *Mol. Microbiol.* **106**, 492–504 (2017).
371. Slotboom, D. J., Duurkens, R. H., Olieman, K. & Erkens, G. B. Static light scattering to characterize membrane proteins in detergent solution. *Methods* **46**, 73–82 (2008).
 372. Zhao, H., Brown, P. H. & Schuck, P. On the distribution of protein refractive index increments. *Biophys. J.* **100**, 2309–2317 (2011).
 373. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119–139 (2005).
 374. MacRaild, C. A., Hatters, D. M., Howlett, G. J. & Gooley, P. R. NMR structure of human apolipoprotein C-II in the presence of sodium dodecyl sulfate. *Biochemistry* **40**, 5414–5421 (2001).
 375. Curnow, P. *et al.* Expression, purification, and reconstitution of a diatom silicon transporter. *Biochemistry* **51**, 3776–3785 (2012).
 376. Weber, G. *et al.* The use of a cholinergic fluorescent probe for the study of the receptor proteolipid. *Mol. Pharmacol.* **7**, 530–7 (1971).
 377. Yung-Chi, C. & Prusoff, W. H. Relationship between the inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **22**, 3099–3108 (1973).
 378. Uzdavinyis, P. *et al.* Dissecting the proton transport pathway in electrogenic Na^+/H^+ antiporters. *PNAS* **114**, E1101–E1110 (2017).
 379. Parker, J. L., Mindell, J. A. & Newstead, S. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. *Elife* **3**, (2014).
 380. Mathieu, K. *et al.* Functionality of membrane proteins overexpressed and purified from *E. coli* is highly dependent upon the strain. *Sci. Rep.* **9**, 2654 (2019).
 381. Gopal, G. J. & Kumar, A. Strategies for the production of recombinant protein in *Escherichia coli*. *Protein J.* **32**, 419–425 (2013).
 382. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
 383. Grabski, A., Mehler, M. & Drott, D. The overnight express autoinduction system: High-density cell growth and protein expression while you sleep. *Nat. Methods* **2**, 233–235 (2005).
 384. Crowley, E. L. & Rafferty, S. P. Review of lactose-driven auto-induction expression of isotope-labelled proteins. *Protein Expr. Purif.* **157**, 70–85 (2019).
 385. Rath, A., Glibowicka, M., Nadeau, V. G., Chen, G. & Deber, C. M. Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *PNAS* **106**, 1760–1765 (2009).
 386. Reynolds, J. A. & Tanford, C. Binding of dodecyl sulfate to proteins at high binding

- ratios. Possible implications for the state of proteins in biological membranes. *PNAS* **66**, 1002–1007 (1970).
387. Miyake, J., Ochiai-Yanagi, S., Kasumi, T. & Takagi, T. Isolation of a membrane protein from *R. rubrum* chromatophores and its abnormal behavior in SDS-polyacrylamide gel electrophoresis due to a high binding capacity for SDS. *J. Biochem.* **83**, 1679–1686 (1978).
388. Pollock, N. L. *et al.* SMA-PAGE: A new method to examine complexes of membrane proteins using SMALP nano-encapsulation and native gel electrophoresis. *Biochim. Biophys. Acta - Biomembr.* **1861**, 1437–1445 (2019).
389. le Maire, M. *et al.* Gel chromatography and analytical ultracentrifugation to determine the extent of detergent binding and aggregation, and Stokes radius of membrane proteins using sarcoplasmic reticulum Ca²⁺-ATPase as an example. *Nat. Protoc.* **3**, 1782–1795 (2008).
390. Chaptal, V. *et al.* Quantification of detergents complexed with membrane proteins. *Sci. Rep.* **7**, 41751 (2017).
391. Gimpl, K., Klement, J. & Keller, S. Characterising protein/detergent complexes by triple-detection size-exclusion chromatography. *Biol. Proced. Online* **18**, 4 (2016).
392. Miles, A. J. & Wallace, B. A. Circular dichroism spectroscopy of membrane proteins. *Chem. Soc. Rev.* **45**, 4859–4872 (2016).
393. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89**, 392–400 (2008).
394. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
395. Hofmann, K. & Stoffel, W. TMBASE-A database of membrane spanning protein segments. *Biol Chem Hoppe Seyler* **347**, (1993).
396. Slabinski, L. *et al.* XtalPred: A web server for prediction of protein crystallizability. *Bioinformatics* **23**, 3403–3405 (2007).
397. Benkert, P., Biasini, M. & Schwede, T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27**, 343–350 (2011).
398. Tsuchiya, H. *et al.* Structural basis for amino acid export by DMT superfamily transporter YddG. *Nature* **534**, 417–420 (2016).
399. Nji, E., Gulati, A., Qureshi, A. A., Coincon, M. & Drew, D. Structural basis for the delivery of activated sialic acid into Golgi for sialylation. *Nat. Struct. Mol. Biol.* **26**, 415–423 (2019).
400. Rahman, M., Ismat, F., McPherson, M. J. J. & Baldwin, S. A. Topology-informed strategies for the overexpression and purification of membrane proteins. *Mol. Membr.*

- Biol.* **24**, 407–418 (2007).
401. Eshaghi, S. An efficient strategy for high-throughput expression screening of recombinant integral membrane proteins. *Protein Sci.* **14**, 676–683 (2005).
 402. Kwon, S.-K., Kim, S. K., Lee, D.-H. & Kim, J. F. Comparative genomics and experimental evolution of *Escherichia coli* BL21(DE3) strains reveal the landscape of toxicity escape from membrane protein overproduction. *Sci. Rep.* **5**, 16076 (2015).
 403. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. The toxicity of recombinant proteins in *Escherichia coli*: A comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expr. Purif.* **37**, 203–206 (2004).
 404. Deacon, S. E. *et al.* Reliable scale-up of membrane protein over-expression by bacterial auto-induction: From microwell plates to pilot scale fermentations. *Mol. Membr. Biol.* **25**, 588–598 (2008).
 405. Jin, T. A new fluorometric method for the detection of the neurotransmitter acetylcholine in water using a dansylcholine complex with p-sulfonated calix[8]arene. *J. Incl. Phenom.* 195–201 (2003).
 406. Yakimchuk, K. Protein receptor-ligand interaction/binding assays. *Mater. Methods* **1**, (2011).
 407. Jares-Erijman, E. A. & Jovin, T. M. FRET imaging. *Nat. Biotechnol.* **21**, 1387–1395 (2003).
 408. Ghisaidoobe, A. & Chung, S. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: A focus on Förster resonance energy transfer techniques. *Int. J. Mol. Sci.* **15**, 22518–22538 (2014).
 409. Gustiananda, M., Liggins, J. R., Cummins, P. L. & Gready, J. E. Conformation of prion protein repeat peptides probed by FRET measurements and molecular dynamics simulations. *Biophys. J.* **86**, 2467–2483 (2004).
 410. Bian, X. & Lockless, S. W. Preparation to minimize buffer mismatch in isothermal titration calorimetry experiments. *Anal. Chem.* **88**, 5549–5553 (2016).
 411. Courtois, J. & Berret, J.-F. Probing oppositely charged surfactant and copolymer interactions by isothermal titration microcalorimetry. *Langmuir* **26**, 11750–11758 (2010).
 412. Velazquez-Campoy, A., Leavitt, S. A. & Freire, E. Characterization of protein-protein interactions by isothermal titration calorimetry. in *Methods Mol Biol.* **1278**, 183–204 (2015).
 413. Ladbury, J. E. Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochem. Soc. Trans.* **38**, 888–893 (2010).
 414. Rajarathnam, K. & Rösgen, J. Isothermal titration calorimetry of membrane proteins — Progress and challenges. *Biochim. Biophys. Acta - Biomembr.* **1838**, 69–77 (2014).

415. Doyle, M. L. Characterization of binding interactions by isothermal titration calorimetry. *Curr. Opin. Biotechnol.* **8**, 31–35 (1997).
416. Boudker, O. & Oh, S. Isothermal titration calorimetry of ion-coupled membrane transporters. *Methods* **76**, 171–182 (2015).
417. Nie, Y., Smirnova, I., Kasho, V. & Kaback, H. R. Energetics of ligand-induced conformational flexibility in the lactose permease of *Escherichia coli*. *J. Biol. Chem.* **281**, 35779–35784 (2006).
418. Guan, L. & Kaback, H. R. Lessons from lactose permease. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 67–91 (2006).
419. Broecker, J., Vargas, C. & Keller, S. Revisiting the optimal *c* value for isothermal titration calorimetry. *Anal. Biochem.* **418**, 307–309 (2011).
420. Nimigean, C. M. A radioactive uptake assay to measure ion transport across ion channel-containing liposomes. *Nat. Protoc.* **1**, 1207–1212 (2006).
421. Schuldiner, S., Kerwar, G. K., Kaback, H. R. & Weil, R. Energy-dependent binding of dansylgalactosides to the beta-galactoside carrier protein. *J. Biol. Chem.* **250**, 1361–1370 (1975).
422. Lee, C. *et al.* A two-domain elevator mechanism for sodium/proton antiport. *Nature* **501**, 573–577 (2013).
423. Minhas, G. S. & Newstead, S. Structural basis for prodrug recognition by the SLC15 family of proton-coupled peptide transporters. *PNAS* **116**, 804–809 (2019).
424. Krämer, R. & Morbach, S. BetP of *Corynebacterium glutamicum*, a transporter with three different functions: Betaine transport, osmosensing, and osmoregulation. *Biochim. Biophys. Acta - Bioenerg.* **1658**, 31–36 (2004).
425. Tsai, M.-F. & Miller, C. Substrate selectivity in arginine-dependent acid resistance in enteric bacteria. *PNAS* **110**, 5893–5897 (2013).
426. Spooner, P. J. R., Friesen, R. H. E., Knol, J., Poolman, B. & Watts, A. Rotational mobility and orientational stability of a transport protein in lipid membranes. *Biophys. J.* **79**, 756–766 (2000).
427. Pitard, B., Richard, P., Dunarach, M., Girault, G. & Rigaud, J.-L. ATP synthesis by the F₀F₁ ATP synthase from *Thermophilic bacillus* PS₃ reconstituted into liposomes with bacteriorhodopsin. 1. Factors defining the optimal reconstitution of ATP synthases with bacteriorhodopsin. *Eur. J. Biochem.* **235**, 769–778 (1996).
428. le Coutre, J., Narasimhan, L. R., Patel, C. K. N. & Kaback, H. R. The lipid bilayer determines helical tilt angle and function in lactose permease of *Escherichia coli*. *PNAS* **94**, 10167–10171 (1997).
429. Richard, P., Rigaud, J.-L. & Graber, P. Reconstitution of CF₀F₁ into liposomes using a new reconstitution procedure. *Eur. J. Biochem.* **193**, 921–925 (1990).

430. de Kruijff, B. *et al.* Lipid polymorphism and membrane function. in *The Enzymes of Biological Membranes* **1**, 131–204 (Springer US, 1985).
431. Seto-Young, D., Chen, C.-C. & Wilson, T. H. Effect of different phospholipids on the reconstitution of two functions of the lactose carrier of *Escherichia coli*. *J. Membr. Biol.* **84**, 259–267 (1985).
432. Paradies, G., Petrosillo, G. & Ruggiero, F. M. Cardiolipin-dependent decrease of cytochrome *c* oxidase activity in heart mitochondria from hypothyroid rats. *Biochim. Biophys. Acta - Bioenerg.* **1319**, 5–8 (1997).
433. Warren, G. B., Houslay, M. D., Metcalfe, J. C. & Birdsall, N. J. M. Cholesterol is excluded from the phospholipid annulus surrounding an active calcium transport protein. *Nature* **255**, 684–687 (1975).
434. Knol, J. *et al.* Unidirectional reconstitution into detergent-destabilized liposomes of the purified lactose transport system of *Streptococcus thermophilus*. *J. Biol. Chem.* **271**, 15358–15366 (1996).
435. Knol, J., Sjollem, K. & Poolman, B. Detergent-mediated reconstitution of membrane proteins. *Biochemistry* **37**, 16410–16415 (1998).
436. Mynarcik, D. C., Yu, G. Q. & Whittaker, J. Alanine-scanning mutagenesis of a C-terminal ligand binding domain of the insulin receptor α subunit. *J. Biol. Chem.* **271**, 2439–2442 (1996).
437. Ashkenazy, H. *et al.* ConSurf 2016: An improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
438. Saerens, S. M. G., Delvaux, F. R., Verstrepen, K. J. & Thevelein, J. M. Production and biological function of volatile esters in *Saccharomyces cerevisiae*. *Microb. Biotechnol.* **3**, 165–177 (2010).
439. Mason, A. B. & Dufour, J.-P. Alcohol acetyltransferases and the significance of ester synthesis in yeast. *Yeast* **16**, 1287–1298 (2000).
440. El Hadi, M., Zhang, F.-J., Wu, F.-F., Zhou, C.-H. & Tao, J. Advances in fruit aroma volatile research. *Molecules* **18**, 8200–8229 (2013).
441. Robinson, A. L. *et al.* Origins of grape and wine aroma. Part 1. Chemical components and viticultural impacts. *Am. J. Enol. Vitic.* **65**, 1–24 (2014).
442. Pérez, A. G., Sanz, C., Olías, R., Ríos, J. J. & Olías, J. M. Evolution of strawberry alcohol acyltransferase activity during fruit development and storage. *J. Agric. Food Chem.* **44**, 3286–3290 (1996).
443. Runguphan, W. & Keasling, J. D. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab. Eng.* **21**, 103–113 (2014).

444. de Jong, B. W., Shi, S., Siewers, V. & Nielsen, J. Improved production of fatty acid ethyl esters in *Saccharomyces cerevisiae* through up-regulation of the ethanol degradation pathway and expression of the heterologous phosphoketolase pathway. *Microb. Cell Fact.* **13**, 39 (2014).
445. Rodriguez, G. M., Tashiro, Y. & Atsumi, S. Expanding ester biosynthesis in *Escherichia coli*. *Nat. Chem. Biol.* **10**, 259–265 (2014).
446. Spence, C. Just how much of what we taste derives from the sense of smell? *Flavour* **4**, 30 (2015).
447. Guichard, E. Interactions between flavor compounds and food ingredients and their influence on flavor perception. *Food Rev. Int.* **18**, 49–70 (2002).
448. Holt, S., Miks, M. H., de Carvalho, B. T., Foulquié-Moreno, M. R. & Thevelein, J. M. The molecular biology of fruity and floral aromas in beer and other alcoholic beverages. *FEMS Microbiol. Rev.* **43**, 193–222 (2019).
449. Engan, S. Esters in beer. *Brew Dig.* **49**, 40–48 (1974).
450. Meilgaard, M. C. Flavour chemistry of beer. Part II: Flavor and threshold of 239 aroma volatiles. *MBAA. Technol Q* **12**, 151–168 (1975).
451. Song, J. & Forney, C. F. Flavour volatile production and regulation in fruit. *Can. J. Plant Sci.* **88**, 537–550 (2008).
452. FAOSTAT - Food and Agriculture Organization of the United Nations. Available at: <http://www.fao.org/faostat/en/#data/QC>.
453. Plotto, A., Bai, J. & Baldwin, E. Fruits. in *Springer Handbook of Odor* 27–28 (Springer, Cham, 2017).
454. Nordström, K. Formation of ethyl acetate in fermentation with brewer's yeast. III. Participation of coenzyme A. *J. Inst. Brew.* **68**, 398–407 (1962).
455. Lewendon, A., Murray, I. A., Shaw, W. V., Gibbs, M. R. & Leslie, A. G. W. Replacement of catalytic histidine-195 of chloramphenicol acetyltransferase: Evidence for a general base role for glutamate. *Biochemistry* **33**, 1944–1950 (1994).
456. Wu, D. *et al.* Structure of human carnitine acetyltransferase. *J. Biol. Chem.* **278**, 13159–13165 (2003).
457. Cai, Y. *et al.* Choline acetyltransferase structure reveals distribution of mutations that cause motor disorders. *EMBO J.* **23**, 2047–2058 (2004).
458. Verstrepen, K. J. *et al.* The *Saccharomyces cerevisiae* alcohol acetyl transferase gene *ATF1* is a target of the cAMP/PKA and FGM nutrient-signalling pathways. *FEMS Yeast Res.* **4**, 285–96 (2003).
459. Pires, E. J., Teixeira, J. A., Brányik, T. & Vicente, A. A. Yeast: The soul of beer's aroma - a review of flavour-active esters and higher alcohols produced by the brewing yeast. *Appl. Microbiol. Biotechnol.* **98**, 1937–1949 (2014).

460. Gallone, B. *et al.* Domestication and divergence of *Saccharomyces cerevisiae* beer yeasts. *Cell* **166**, 1397-1410.e16 (2016).
461. Kruis, A. J. *et al.* Contribution of Eat1 and other alcohol acyltransferases to ester production in *Saccharomyces cerevisiae*. *Front. Microbiol.* **9**, (2018).
462. Rodríguez-Vargas, S., Sánchez-García, A., Martínez-Rivas, J. M., Prieto, J. A. & Randez-Gil, F. Fluidization of membrane lipids enhances the tolerance of *Saccharomyces cerevisiae* to freezing and salt stress. *Appl. Environ. Microbiol.* **73**, 110–116 (2007).
463. Ding, B.-J. *et al.* The yeast ATF1 acetyltransferase efficiently acetylates insect pheromone alcohols: Implications for the biological production of moth pheromones. *Lipids* **51**, 469–475 (2016).
464. Christiaens, J. F. *et al.* The fungal aroma gene *ATF1* promotes dispersal of yeast cells through insect vectors. *Cell Rep.* **9**, 425–432 (2014).
465. Yoshimoto, H. *et al.* Characterization of the *ATF1* and *Lg-ATF1* genes encoding alcohol acetyltransferases in the bottom fermenting yeast *Saccharomyces pastorianus*. *J. Ferment. Bioeng.* **86**, 15–20 (1998).
466. Malcorps, P. & Dufour, J. P. Short-chain and medium-chain aliphatic-ester synthesis in *Saccharomyces cerevisiae*. *Eur. J. Biochem.* **210**, 1015–22 (1992).
467. Tamai, Y. Alcohol acetyl transferase genes and ester formation in brewer's yeast. in *Biotechnology for Improved Foods and Flavour* 196–205 (Society, American Chemical, 1996).
468. Yoshimoto, H. *et al.* Isolation and characterization of the *ATF2* gene encoding alcohol acetyltransferase II in the bottom fermenting yeast *Saccharomyces pastorianus*. *Yeast* **15**, 409–417 (1999).
469. Fujii T, Yoshimoto H, T. Y. Acetate ester production by *Saccharomyces cerevisiae* lacking the *ATF1* gene encoding the alcohol acetyltransferase. *J. Ferment. Bioeng.* **81**, 538–542 (1996).
470. Lilly, M., Lambrechts, M. G. & Pretorius, I. S. Effect of increased yeast alcohol acetyltransferase activity on flavor profiles of wine and distillates. *Appl. Environ. Microbiol.* **66**, 744–53 (2000).
471. Verstrepen, K. J. *et al.* Expression levels of the yeast alcohol acetyltransferase genes *ATF1*, *Lg-ATF1*, and *ATF2* control the formation of a broad range of volatile esters. *Appl. Environ. Microbiol.* **69**, 5228–37 (2003).
472. Fujii, T. *et al.* Molecular cloning, sequence analysis, and expression of the yeast alcohol acetyltransferase gene. *Appl. Environ. Microbiol.* **60**, 2786–92 (1994).
473. Nagasawa, N., Bogaki, T., Iwamatsu, A., Hamachi, M. & Kumagai, C. Cloning and nucleotide sequence of the alcohol acetyltransferase II gene (*ATF2*) from

- Saccharomyces cerevisiae* Kyokai No. 7. *Biosci. Biotechnol. Biochem.* **62**, 1852–1857 (1998).
474. Lilly, M. *et al.* The effect of increased yeast alcohol acetyltransferase and esterase activity on the flavour profiles of wine and distillates. *Yeast* **23**, 641–659 (2006).
475. Fujii, T., Kobayashi, O., Yoshimoto, H., Furukawa, S. & Tamai, Y. Effect of aeration and unsaturated fatty acids on expression of the *Saccharomyces cerevisiae* alcohol acetyltransferase gene. *Appl. Environ. Microbiol.* **63**, 910–5 (1997).
476. Yoshioka, K. & Hashimoto, N. Ester formation by alcohol acetyltransferase from brewers' yeast. *Agric. Biol. Chem.* **45**, 2183–2190 (1981).
477. Fujiwara, D., Yoshimoto, H., Sone, H., Harashima, S. & Tamai, Y. Transcriptional co-regulation of *Saccharomyces cerevisiae* alcohol acetyltransferase gene, *ATF1* and Δ -9 fatty acid desaturase gene, *OLE1* by unsaturated fatty acids. *Yeast* **14**, 711–721 (1998).
478. Fujiwara, D., Kobayashi, O., Yoshimoto, H., Harashima, S. & Tamai, Y. Molecular mechanism of the multiple regulation of the *Saccharomyces cerevisiae* *ATF1* gene encoding alcohol acetyltransferase. *Yeast* **15**, 1183–97 (1999).
479. Zitomer, R. S., Carrico, P. & Deckert, J. Regulation of hypoxic gene expression in yeast. *Kidney Int.* **51**, 507–513 (1997).
480. Vasconcelles, M. J. *et al.* Identification and characterization of a low oxygen response element involved in the hypoxic induction of a family of *Saccharomyces cerevisiae* genes. Implications for the conservation of oxygen sensing in eukaryotes. *J. Biol. Chem.* **276**, 14374–14384 (2001).
481. Saerens, S. M. G. *et al.* Parameters affecting ethyl ester production by *Saccharomyces cerevisiae* during fermentation. *Appl. Environ. Microbiol.* **74**, 454–461 (2008).
482. Verstrepen, K. J. *et al.* Flavor-active esters: Adding fruitiness to beer. *J. Biosci. Bioeng.* **96**, 110–8 (2003).
483. Urit, T., Li, M., Bley, T. & Löser, C. Growth of *Kluyveromyces marxianus* and formation of ethyl acetate depending on temperature. *Appl. Microbiol. Biotechnol.* **97**, 10359–10371 (2013).
484. Verstrepen, K. J. *et al.* The *Saccharomyces cerevisiae* alcohol acetyl transferase Atf1p is localized in lipid particles. *Yeast* **21**, 367–377 (2004).
485. Lin, J.-L. & Wheeldon, I. Dual N- and C-terminal helices are required for endoplasmic reticulum and lipid droplet association of alcohol acetyltransferases in *Saccharomyces cerevisiae*. *PLoS One* **9**, e104141 (2014).
486. Tiwari, R., Köffel, R. & Schneider, R. An acetylation/deacetylation cycle controls the export of sterols and steroids from *S. cerevisiae*. *EMBO J.* **26**, 5109–5119 (2007).
487. Minetoki, T., Bogaki, T., Iwamatsu, A., Fujii, T. & Hamachi, M. The purification,

- properties and internal peptide sequences of alcohol acetyltransferase isolated from *Saccharomyces cerevisiae* Kyokai No. 7. *Biosci. Biotechnol. Biochem.* **57**, 2094–2098 (1993).
488. Akita, O., Suzuki, S., Obata, T, Hara, S. Purification and some properties of alcohol acetyltransferase from sake yeast. *Agric. Biol. Chem.* **54**, 1485–1490 (1990).
489. Zhu, J., Lin, J.-L., Palomec, L. & Wheeldon, I. Microbial host selection affects intracellular localization and activity of alcohol-O-acetyltransferase. *Microb. Cell Fact.* **14**, 35 (2015).
490. D'Auria, J. C. Acyltransferases in plants: A good time to be BAHD. *Curr. Opin. Plant Biol.* **9**, 331–340 (2006).
491. Beekwilder, J. Functional characterization of enzymes forming volatile esters from strawberry and banana. *PLANT Physiol.* **135**, 1865–1878 (2004).
492. Fellman, J. K., Miller, T. W., Mattinson, D. S. & Mattheis, J. P. Factors that influence biosynthesis of volatile flavor compounds in apple fruits. *HortScience* **35**, 1026–1033 (2000).
493. Pérez, A. G., Olías, R., Luaces, P. & Sanz, C. Biosynthesis of strawberry aroma compounds through amino acid metabolism. *J. Agric. Food Chem.* **50**, 4037–4042 (2002).
494. Wyllie, S. G., Leach, D. N., Wang, Y. & Shewfelt, R. L. Key aroma compounds in melons. in *Fruit Flavors* 248–257 (ACS Symposium Series, 1995).
495. Pérez, A. G., Rios, J. J., Sanz, C. & Olías, J. M. Aroma components and free amino acids in strawberry variety Chandler during ripening. *J. Agric. Food Chem.* **40**, 2232–2235 (1992).
496. Yu, M. H., Salunkhe, D. K. & Olson, L. E. Production of 3-methylbutanal from L-leucine by tomato extract. *Plant Cell Physiol.* **9**, 633–638 (1968).
497. Tressl, R. & Drawert, F. Biogenesis of banana volatiles. *J. Agric. Food Chem.* **21**, 560–565 (1973).
498. Myers, M. J., Issenberg, P. & Wick, E. L. L-leucine as a precursor of isoamyl alcohol and isoamyl acetate, volatile aroma constituents of banana fruit discs. *Phytochemistry* **9**, 1693–1700 (1970).
499. Wang, C., Xing, J., Chin, C.-K., Ho, C.-T. & Martin, C. E. Modification of fatty acids changes the flavor volatiles in tomato leaves. *Phytochemistry* **58**, 227–232 (2001).
500. Manríquez, D. *et al.* Two highly divergent alcohol dehydrogenases of melon exhibit fruit ripening-specific expression and distinct biochemical characteristics. *Plant Mol. Biol.* **61**, 675–685 (2006).
501. Yu, X.-H., Gou, J.-Y. & Liu, C.-J. BAHD superfamily of acyl-CoA dependent acyltransferases in *Populus* and *Arabidopsis*: Bioinformatics and gene expression.

- Plant Mol. Biol.* **70**, 421–442 (2009).
502. Balbontín, C. *et al.* *VpAAT1*, a gene encoding an alcohol acyltransferase, is involved in ester biosynthesis during ripening of mountain papaya fruit. *J. Agric. Food Chem.* **58**, 5114–5121 (2010).
 503. El-Sharkawy, I. *et al.* Functional characterization of a melon alcohol acyltransferase gene family involved in the biosynthesis of ester volatiles. Identification of the crucial role of a threonine residue for enzyme activity. *Plant Mol. Biol.* **59**, 345–362 (2005).
 504. González, M. *et al.* Aroma development during ripening of *Fragaria chiloensis* fruit and participation of an alcohol acyltransferase (*FcAAT1*) gene. *J. Agric. Food Chem.* **57**, 9123–9132 (2009).
 505. Bayer, A., Ma, X. & Stöckigt, J. Acetyltransfer in natural product biosynthesis - Functional cloning and molecular analysis of vinorine synthase. *Bioorg. Med. Chem.* **12**, 2787–2795 (2004).
 506. Yahyaoui, F. E. L. *et al.* Molecular and biochemical characteristics of a gene encoding an alcohol acyl-transferase involved in the generation of aroma volatile esters during melon ripening. *Eur. J. Biochem.* **269**, 2359–66 (2002).
 507. Ma, X., Koepke, J., Panjekar, S., Fritsch, G. & Stöckigt, J. Crystal structure of vinorine synthase, the first representative of the BAHD superfamily. *J. Biol. Chem.* **280**, 13576–13583 (2005).
 508. Yamashita, I., Nemoto, Y. & Yoshikawa, S. Studies on flavor development in strawberries. I. Formation of volatile esters in strawberries. *Agric. Biol. Chem.* **39**, 2303–2307 (1975).
 509. Idstein, H., Keller, T. & Schreier, P. Volatile constituents of mountain papaya (*Carica candamarcensis*, *syn. C. pubescens* *Lenne et Koch*) fruit. *J. Agric. Food Chem.* **33**, 663–666 (1985).
 510. Morales, A. L. & Duque, C. Aroma constituents of the fruit of the mountain papaya (*Carica pubescens*) from Colombia. *J. Agric. Food Chem.* **35**, 538–540 (1987).
 511. Wang, M. Y., MacRae, E., Wohlers, M. & Marsh, K. Changes in volatile production and sensory quality of kiwifruit during fruit maturation in *Actinidia deliciosa* 'Hayward' and *A. chinensis* 'Hort16A'. *Postharvest Biol. Technol.* **59**, 16–24 (2011).
 512. Pérez, A. G., Sanz, C. & Olías, J. M. Partial purification and some properties of alcohol acyltransferase from strawberry fruits. *J. Agric. Food Chem.* **41**, 1462–1466 (1993).
 513. Olías, R., Pérez, A. G. & Sanz, C. Catalytic properties of alcohol acyltransferase in different strawberry species and cultivars. *J. Agric. Food Chem.* **50**, 4031–6 (2002).
 514. Harada, M., Ueda, Y. & Iwata, T. Purification and some properties of alcohol acetyltransferase from banana fruit. *Plant Cell Physiol.* **26**, 1067–1074 (1985).
 515. Olías, J. M., Sanz, C., Rios, J. J. & Pérez, A. G. Substrate specificity of alcohol

- acyltransferase from strawberry and banana fruits. in *Fruit Flavors* 134–141 (ACS Symposium Series, 1995).
516. Ueda, Y., Tsuda, A., Bai, J.-H., Fujishita, N. & Chachin, K. Characteristic pattern of aroma ester formation from banana, melon, and strawberry with reference to the substrate specificity of ester synthetase and alcohol contents in pulp. *Japanese Soc. Food Sci. Technol.* **39**, 183–187 (1992).
517. Salas, J. J. Characterization of alcohol acyltransferase from olive fruit. *J. Agric. Food Chem.* **52**, 3155–3158 (2004).
518. Fellman, J. K., Mattinson, D. S., Bostick, B. C., Mattheis, J. P. & Patterson, M. E. Ester biosynthesis in “Rome” apples subjected to low-oxygen atmospheres. *Postharvest Biol. Technol.* **3**, 201–214 (1993).
519. Botondi, R., De Sanctis, F., Bartoloni, S. & Mencarelli, F. Simultaneous application of ethylene and 1-MCP affects banana ripening features during storage. *J. Sci. Food Agric.* **94**, 2170–2178 (2014).
520. Balbontín, C., Gaete-Eastman, C., Vergara, M., Herrera, R. & Moya-León, M. A. Treatment with 1-MCP and the role of ethylene in aroma development of mountain papaya fruit. *Postharvest Biol. Technol.* **43**, 67–77 (2007).
521. Li, D.-P. *et al.* Salicylic acid, ethephon, and methyl jasmonate enhance ester regeneration in 1-MCP-treated apple fruit after long-term cold storage. *J. Agric. Food Chem.* **54**, 3887–3895 (2006).
522. Lucchetta, L. *et al.* Biochemical and catalytic properties of three recombinant alcohol acyltransferases of melon. Sulfur-containing ester formation, regulatory role of CoA-SH in activity, and sequence elements conferring substrate preference. *J. Agric. Food Chem.* **55**, 5213–5220 (2007).
523. Flores, F. B., Manriquez, D., El-Sharkawy, I., Latché, A. & Pech, J. C. Characterization of genes involved in the formation of aroma volatiles in Charentais melon fruit. *Acta Hortic.* **682**, 673–680 (2005).
524. Günther, C. S., Chervin, C., Marsh, K. B., Newcomb, R. D. & Souleyre, E. J. F. Characterisation of two alcohol acyltransferases from kiwifruit (*Actinidia* spp.) reveals distinct substrate preferences. *Phytochemistry* **72**, 700–710 (2011).
525. Souleyre, E. J. F., Günther, C. S., Wang, M. Y., Newcomb, R. D. & Marsh, K. B. Ester biosynthesis in kiwifruit - From genes to enzymes to pathways. *Acta Hortic.* **913**, 205–211 (2011).
526. Garcia, C. V., Stevenson, R. J., Atkinson, R. G., Winz, R. A. & Quek, S.-Y. Changes in the bound aroma profiles of ‘Hayward’ and ‘Hort16A’ kiwifruit (*Actinidia* spp.) during ripening and GC-olfactometry analysis. *Food Chem.* **137**, 45–54 (2013).
527. Cumplido-Laso, G. *et al.* The fruit ripening-related gene *FaAAT2* encodes an acyl

- transferase involved in strawberry aroma biogenesis. *J. Exp. Bot.* **63**, 4275–4290 (2012).
528. Souleyre, E. J. F., Greenwood, D. R., Friel, E. N., Karunairetnam, S. & Newcomb, R. D. An alcohol acyl transferase from apple (cv. Royal Gala), MpAAT1, produces esters involved in apple fruit flavor. *FEBS J.* **272**, 3132–3144 (2005).
529. Aharoni, A. *et al.* Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. *Plant Cell* **12**, 647–62 (2000).
530. Morales-Quintana, L., Fuentes, L., Gaete-Eastman, C., Herrera, R. & Moya-León, M. A. Structural characterization and substrate specificity of VpAAT1 protein related to ester biosynthesis in mountain papaya fruit. *J. Mol. Graph. Model.* **29**, 635–642 (2011).
531. Morales-Quintana, L., Moya-León, M. A. & Herrera, R. Molecular docking simulation analysis of alcohol acyltransferases from two related fruit species explains their different substrate selectivities. *Mol. Simul.* **38**, 912–921 (2012).
532. Morales-Quintana, L., Nuñez-Tobar, M. X., Moya-León, M. A. & Herrera, R. Molecular dynamics simulation and site-directed mutagenesis of alcohol acyltransferase: A proposed mechanism of catalysis. *J. Chem. Inf. Model.* **53**, 2689–2700 (2013).
533. Morales-Quintana, L., Moya-León, M. A. & Herrera, R. Computational study enlightens the structural role of the alcohol acyltransferase DFGWG motif. *J. Mol. Model.* **21**, 216 (2015).
534. Navarro-Retamal, C., Gaete-Eastman, C., Herrera, R., Caballero, J. & Alzate-Morales, J. H. Structural and affinity determinants in the interaction between alcohol acyltransferase from *F. x ananassa* and several alcohol substrates: A computational study. *PLoS One* **11**, e0153057 (2016).
535. Galaz, S., Morales-Quintana, L., Moya-León, M. A. & Herrera, R. Structural analysis of the alcohol acyltransferase protein family from *Cucumis melo* shows that enzyme activity depends on an essential solvent channel. *FEBS J.* **280**, 1344–1357 (2013).
536. Tashiro, Y., Desai, S. H. & Atsumi, S. Two-dimensional isobutyl acetate production pathways to improve carbon yield. *Nat. Commun.* **6**, 7488 (2015).
537. Bohnenkamp, A. C. *et al.* Multilevel optimisation of anaerobic ethyl acetate production in engineered *Escherichia coli*. *Biotechnol. Biofuels* **13**, 65 (2020).
538. Huang, A. *et al.* BioBits™ Explorer: A modular synthetic biology education kit. *Sci. Adv.* **4**, eaat5105 (2018).
539. Kota, J., Gilstring, C. F. & Ljungdahl, P. O. Membrane chaperone Shr3 assists in folding amino acid permeases preventing precocious ERAD. *J. Cell Biol.* **176**, 617–628 (2007).
540. Gietz, R. D. & Schiestl, R. H. Frozen competent yeast cells that can be transformed with high efficiency using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 1–4 (2007).

541. Knight, M. J., Bull, I. D. & Curnow, P. The yeast enzyme Eht1 is an octanoyl-CoA:ethanol acyltransferase that also functions as a thioesterase. *Yeast* **31**, 463–474 (2014).
542. Dunn, B. J., Cane, D. E. & Khosla, C. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Biochemistry* **52**, 1839–1841 (2013).
543. Tai, Y.-S., Xiong, M. & Zhang, K. Engineered biosynthesis of medium-chain esters in *Escherichia coli*. *Metab. Eng.* **27**, 20–28 (2015).
544. Kefala, G. *et al.* Structures of the OmpF porin crystallized in the presence of foscholine-12. *Protein Sci.* **19**, 1117–1125 (2010).
545. Bass, R. B. Crystal structure of *Escherichia coli* MscS, a voltage-modulated and mechanosensitive channel. *Science (80-.)*. **298**, 1582–1587 (2002).
546. Chipot, C. *et al.* Perturbations of native membrane protein structure in alkyl phosphocholine detergents: A critical assessment of NMR and biophysical studies. *Chem. Rev.* **118**, 3559–3607 (2018).
547. Shih, Y.-P. *et al.* High-throughput screening of soluble recombinant proteins. *Protein Sci.* **11**, 1714–1719 (2009).
548. Kapust, R. B. & Waugh, D. S. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci.* **8**, 1668–1674 (1999).
549. Berrow, N. S. *et al.* A versatile ligation-independent cloning method suitable for high-throughput expression screening applications. *Nucleic Acids Res.* **35**, e45 (2007).
550. Li, D. *et al.* Molecular cloning and expression of a gene encoding alcohol acyltransferase (*MdAAT2*) from apple (cv. Golden Delicious). *Phytochemistry* **67**, 658–667 (2006).