

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



## Developing novel methods for co-translational studies of membrane protein folding

Pellowe, Grant

*Awarding institution:*  
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

### END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Developing Novel Methods for Co-translational Studies of Membrane Protein Folding

Grant A. Pellowe

A thesis submitted for the degree of *Doctor of Philosophy*

2021

Under the supervision of Professor Paula Booth

Department of Chemistry

King's College London

## Abstract

---

Membrane proteins are a diverse range of macromolecules and when misfolded, disease phenotypes often occur. Membrane protein folding studies ascertain how a linear polypeptide forms its final functional three-dimensional structure. Current methods for studying membrane protein folding rely on sophisticated biophysical methods to probe the folded state of a purified protein in detergent micelles or an artificial lipid mimetic. However, most proteins *in vivo* fold in a co-translational process, from N- to C- terminal as the peptide emerges from the ribosome. This thesis develops new methods to expand the toolbox of co-translational folding studies for membrane proteins. Firstly, I develop a new method for purification of a 6 TM Rhomboid protease in a partially translated state while still attached to the ribosome to form a ribosome-bound nascent chain complex (RNC). This RNC purification method utilises a novel polymer-based system to not only capture the partially expressed protein, but also its surrounding lipids from its native *E. coli* membrane allowing a more physiological snapshot of the co-translational folding process, without the need to first purify using detergents as is common in the field. I also show evidence that the SecYEG bacterial translocon can co-purify with the RNC complex which opens many avenues of structural and biochemical research projects, of which these RNC samples are highly amenable to. The final chapter of this thesis focuses on the development of an atomic-force microscopy (AFM)-based protein unfolding experiment for a fully translated and reconstituted membrane protein. This approach can be adapted to further study the effects of lipid environment, and intrinsic stability of membrane protein RNCs at varying positions through their translation cycle which should revolutionise methods for studying co-translational membrane protein folding.

## Acknowledgements

---

Many people must be thanked for making this thesis possible.

Firstly, to my supervisor Paula, without who I would have never had the opportunity to carry out this research - you've not only been an excellent scientist and mentor, but also exceptionally understanding and more than willing to help with any bumps that happen along the road, and it has been a pleasure to work with you.

Secondly, the entire 208 office, past and present, for all the fun times in the lab and the many hours in the pub after long days of unsuccessful experiments. Special thanks to Drs Heather Findlay, Nicola Harris and Eamonn Reading - you guys have taught me so much, and I am forever grateful to you each of you.

Thanks also to the technical, admin and security teams at Britannia House who always ensure experiments and lab life can run as smoothly as possible.

Also, to Yong Jian Wang - Jim, thank you for your patience and persistence when training me in AFM and for teaching me physics - it really would not have been possible without you. I am also grateful to the rest of the Garcia-Manyes group for making me feel welcome in those long days in the basement, and finally to Sergi for allowing us AFM time to begin with.

Finally, I need to also thank my friends and family for their continued support, not sure I could have finished this PhD without you all!



## Publications

---

My first author publications in *BBA Biomembranes* and *Biochemistry*, can be found in **Appendices 9 and 10**

### Published

- Jia R, Martens C, Shekar M, Pant S, **Pellowe GA**, Lau AM, Findlay HE, Harris NJ, Tajkorshid E, Booth PJ, Politis A. Hydrogen-deuterium exchange mass spectrometry captures distinct dynamics upon substrate and inhibitor binding to a transporter. *Nature Communications*. 2020  
doi: 10.1038/s41467-020-20032-3
- **Pellowe GA**, Findlay HE, Lee K, Gemeinhardt TM, Blackholly LR, Reading E, Booth PJ. Capturing Membrane Protein Ribosome Nascent Chain Complexes in a Native-like Environment for Co-translational Studies. *Biochemistry*. 2020  
doi: 10.1021/acs.biochem.0c00423
- Harris NJ, **Pellowe GA**, Booth PJ. Cell-free expression tools to study co-translational folding of alpha helical membrane transporters. *Scientific Reports*. 2020  
doi: 10.1038/s41598-020-66097-4
- **Pellowe GA**, Booth PJ. Structural insight into co-translational membrane protein folding. *Biochim Biophys Acta – Biomembranes*. 2020  
doi: 10.1016/j.bbamem.2019.07.007

### In preparation

- Harris NJ, **Pellowe GA**, Blackholly LR, Gulaidi-Breen S, Findlay HE, Booth PJ. TBD – membrane protein folding review article, *OpenBiology*. 2020/21

# Table of Contents

---

Abstract.....	2
Acknowledgements .....	3
Publications .....	4
Table of Contents .....	5
List of Figures.....	12
List of Tables .....	15
List of Equations.....	15
List of Abbreviations .....	16
1. Introduction.....	18
1.1. Membranes and membrane proteins .....	19
1.1.1. Membranes and lipids.....	19
1.1.2. Membrane proteins.....	23
1.1.3. Studying membrane protein folding is important.....	24
1.2. Membrane mimetics for protein folding studies .....	27
1.2.1. The problem with detergents .....	28
1.2.2. Mixed micelles and bicelles.....	29
1.2.3. Synthetic liposomes and vesicles.....	30
1.2.4. MSP-based nanodiscs .....	30
1.2.5. Polymer-based nanodiscs.....	31
1.3. <i>In vivo</i> and co-translational folding studies of membrane proteins .....	34
1.3.1. Introduction to co-translational folding.....	34
1.3.2. Biochemical methods to study co-translational folding .....	36
1.3.3. Time-resolved studies of co-translational folding .....	40
1.4. <i>In vitro</i> folding studies of membrane proteins .....	43
1.4.1. <i>In vitro</i> folding of $\alpha$ -helical membrane proteins.....	43
1.4.2. Artificial denaturation.....	44
1.4.3. Single-molecule mechanical unfolding from membranes .....	46
1.5. Ribosome-bound nascent chain complexes for co-translational studies .....	48
1.5.1. RNCs as a tool to study protein folding.....	48
1.5.2. An overview of stalling sequences used for RNC production.....	49
1.5.2.1. A brief note on the 70S ribosome .....	49
1.5.2.2. SecM peptide stalling .....	51
1.5.2.3. TnaC peptide stalling .....	53

1.5.2.4.	Other stalling methods.....	54
1.5.3.	Current studies using soluble protein RNCs .....	54
1.5.3.1.	Biochemical studies .....	55
1.5.3.2.	Fluorescence spectroscopy.....	56
1.5.3.3.	Structural studies .....	57
1.5.3.4.	Force spectroscopy .....	58
1.5.4.	Membrane protein RNCs.....	60
1.6.	Thesis aims and layout.....	65
2.	Materials and methods.....	67
2.1.	Reagents.....	68
2.1.1.	Common biochemical reagents.....	68
2.1.2.	DIBMA stock preparation.....	68
2.2.	Molecular biology for expression constructs .....	68
2.2.1.	Materials and plasmids .....	68
2.2.2.	Bacterial strains and buffers.....	69
2.2.3.	Construct design and PCR.....	70
2.2.3.1.	Cloning of GlpG truncations and (enhanced) SecM.....	70
2.2.3.2.	Addition of Avi-Tag to RNC constructs.....	70
2.2.3.3.	Cloning of SecYEG and RNC into RSF-Duet 1.....	71
2.2.3.4.	Poly-glycine linker addition to Xyle .....	72
2.2.3.5.	Addition of HaloTag and TMH to Xyle .....	72
2.3.	SDS-Page and Western blotting.....	72
2.3.1.	Isolated protein SDS-PAGE .....	72
2.3.2.	RNC sample SDS-Page.....	73
2.3.3.	Western blot .....	73
2.4.	Ribosome nascent chain expression and purification .....	74
2.4.1.	Growth and expression of RNCs.....	74
2.4.1.1.	Minimal media components.....	74
2.4.1.2.	Growth and expression of non-biotinylated RNCs .....	75
2.4.1.3.	Growth and expression of Biotinylated RNCs.....	75
2.4.2.	Buffers.....	76
2.4.3.	Purification of RNCs in DDM .....	76
2.4.4.	Purification of RNCs in DIBMA.....	77
2.5.	RNC sucrose gradient purifications.....	78
2.6.	Biotinylated RNC-avidin binding .....	78

2.7.	Thin-layer chromatography lipid analysis for DIBMA samples .....	78
2.8.	Negative staining of RNCs in DIBMA and DDM .....	79
2.8.1.	Sample preparation for negative stain visualisation .....	79
2.8.2.	Image processing.....	79
2.9.	Isolated protein expression and purification .....	80
2.9.1.	Expression and purification of GlpG and associated mutants.....	80
2.9.1.1.	Growth and expression of GlpG and mutants.....	80
2.9.1.2.	Buffers.....	80
2.9.1.3.	Purification of GlpG and mutants.....	80
2.9.2.	Expression and purification of XyleE and associated mutants .....	81
2.9.2.1.	Growth and expression of XyleE and mutants .....	81
2.9.2.2.	Buffers.....	81
2.9.2.3.	Purification of XyleE and mutants.....	82
2.9.3.	Expression and purification of separate XyleE half domains.....	82
2.10.	Reconstitution methods .....	82
2.10.1.	Preparation of lipid films .....	82
2.10.2.	Reconstitution of FL protein into synthetic lipids.....	83
2.10.3.	Solubilisation of proteoliposomes with DIBMA .....	83
2.10.4.	Markwell-Lowry assay for reconstitution efficiency .....	83
2.10.5.	Carbohydrate assay for OG and DDM in proteoliposomes.....	84
2.11.	General biophysical methods for protein analysis.....	84
2.11.1.	Isothermal titration calorimetry (ITC) .....	84
2.11.2.	Circular dichroism (CD) and synchrotron-radiation CD (SRCD).....	84
2.11.3.	Dynamic light scattering (DLS) for particle size .....	85
2.12.	Biochemical assay methods for protein analysis .....	85
2.12.1.	EnzCheck protease assay for GlpG activity .....	85
2.12.2.	Xylose dehydrogenase assay .....	85
2.12.3.	HaloLink Resin binding to Halo-TMH-XyleE .....	87
2.13.	Atomic force microscopy (AFM) methods.....	87
2.13.1.	Sample deposition .....	87
2.13.2.	Calibration of cantilever in fluid.....	88
2.13.3.	AFM imaging .....	90
2.13.4.	Force/extension (FX) mode .....	90
2.13.4.1.	Data collection .....	90
2.13.4.2.	Data analysis.....	90

2.13.5.	Force/clamp (FC) mode .....	91
2.13.5.1.	Data collection .....	91
2.13.5.2.	Data analysis.....	91
3.	A strategy for the generation of membrane protein RNCs.....	92
3.1.	Introduction.....	93
3.1.1.	Rhomboid protease - GlpG .....	93
3.1.1.1.	Structure and dynamics .....	93
3.1.1.2.	Folding studies.....	95
3.1.1.3.	Purity .....	96
3.1.1.4.	Activity .....	97
3.1.2.	Considerations for membrane protein RNC production .....	98
3.1.3.	Adapting GlpG for RNC study .....	98
3.1.4.	Aims of chapter .....	99
3.2.	Results.....	100
3.2.1.	N-His GlpG expression and purification in DDM.....	100
3.2.2.	N-His GlpG is active and folded in DDM .....	101
3.2.3.	RNC stalling sequence design rational .....	103
3.2.4.	Preparation of $\alpha$ -helical membrane protein RNCs within DDM .....	104
3.2.5.	Negative stain EM of RNCs in DDM.....	108
3.2.6.	Co-expression with BirA for efficient <i>In vivo</i> biotinylation of RNCs.....	109
3.2.7.	An enhanced SecM arrest sequence to improve RNC stability .....	113
3.2.8.	RNC overexpression with SecYEG for structural studies .....	115
3.3.	Discussion.....	118
3.3.1.	GlpG preparation and activity .....	118
3.3.2.	Designing RNCs with membranes in mind .....	119
3.3.3.	Advantages of <i>in vivo</i> generated MP RNCs purified in DDM.....	121
3.3.4.	Polytopic $\alpha$ -helical membrane protein RNC purification in DDM.....	122
3.3.5.	<i>In vivo</i> biotinylation for purification and structural experimentation .....	126
3.4.	Chapter summary.....	128
4.	Capturing RNCs in a native-like environment for co-translational studies .....	129
4.1.	Introduction.....	130
4.1.1.	Native nanodiscs for <i>in vivo</i> -like <i>in vitro</i> study .....	130
4.1.1.1.	Stability and functional studies .....	131
4.1.1.2.	Structure and dynamics of proteins in SMALPs.....	133
4.1.2.	GlpG in native nanodiscs .....	136

4.1.3.	DIBMA and its suitability for RNC capture .....	136
4.1.4.	Non-native nanodiscs and membrane protein RNCs.....	138
4.1.5.	Aims of chapter .....	139
4.2.	Part A: DIBMA - Moving towards an <i>in vivo</i> environment.....	140
4.2.1.	Preparation of DIBMA .....	140
4.2.2.	DIBMA solubilises synthetic lipids.....	142
4.2.3.	DIBMA captures reconstituted protein with the correct lipid composition....	143
4.3.	Part B: GlpG RNCs in DIBMALPs .....	146
4.3.1.	Capturing membrane protein RNCs within DIBMA nanodiscs .....	146
4.3.2.	Evidence of endogenous SecYEG translocon association with GlpG RNCs ....	151
4.3.3.	DIBMA captures membrane protein RNCs in a native lipid environment.....	152
4.3.4.	Negative stain EM of RNCs produced in DIBMA.....	155
4.3.5.	Arrest-enhanced SecM improves DIBMA purified RNC quality .....	156
4.4.	Discussion.....	158
4.4.1.	Preparation and functionality of DIBMA.....	158
4.4.2.	Purification of RNCs into DIBMA native nanodiscs .....	158
4.4.3.	Capture of native lipids and endogenous insertion machinery.....	161
4.5.	Chapter summary.....	162
5.	Expression, preparation, and characterisation of Xyle .....	163
5.1.	Introduction.....	164
5.1.1.	D-xylose/H <sup>+</sup> symporter - Xyle .....	164
5.1.1.1.	Structure and stability.....	164
5.1.1.2.	Binding and transport.....	165
5.1.2.	Aims of chapter .....	169
5.1.2.1.	Purification, purity, and reconstitution.....	169
5.1.2.2.	Determining purified Xyle functionality .....	169
5.2.	Results.....	170
5.2.1.	Xyle expression and purification in DDM .....	170
5.2.2.	Expression and purification of separate Xyle domains .....	172
5.2.3.	WT Xyle and Xyle-pGly28 are folded DDM .....	174
5.2.4.	WT Xyle and Xyle-pGly28 bind ligand in DDM.....	175
5.2.5.	Protein reconstitution into synthetic lipid vesicles.....	177
5.2.6.	A novel NAD <sup>+</sup> -linked Xyle functional assay .....	179
5.2.6.1.	Preparation of xylose dehydrogenase (XDH).....	179
5.2.6.2.	Encapsulation of XDH, and Xyle activity .....	181

5.2.7.	Modifying Xyle for single-molecule experiments .....	182
5.3.	Discussion.....	185
5.3.1.	Xyle preparation and activity .....	185
5.3.2.	Xyle separate domain preparation and relevance .....	188
5.3.3.	Tagging Xyle for downstream single-molecule work.....	189
5.4.	Chapter summary.....	190
6.	Single-molecule mechanical unfolding of Xyle.....	191
6.1.	Single-molecule spectroscopy techniques for membrane proteins.....	192
6.1.1.	Past AFM studies on membrane proteins.....	192
6.1.2.	Force-extension spectroscopy for membrane proteins .....	197
6.1.2.1.	Collecting and analysing force-extension data.....	197
6.1.2.2.	Fitting of force-extension data.....	198
6.1.2.3.	Extracting energetics from AFM experiments .....	199
6.1.3.	Force-Clamp spectroscopy .....	200
6.1.4.	Magnetic tweezer spectroscopy .....	201
6.1.5.	Xyle as a model candidate for SMFS .....	203
6.2.	Aims of chapter .....	205
6.3.	Part A: Bilayers, reconstitution, and discs for AFM.....	206
6.3.1.	Determining mica deposition conditions using liposomes.....	206
6.3.2.	Imaging Xyle in 40:40:20 liposomes .....	207
6.3.3.	Improving AFM protein-pickup rate .....	209
6.3.4.	Polymer nanodiscs for AFM .....	210
6.4.	Part B: Unfolding Xyle .....	211
6.4.1.	Mapping the expected unfolding pathway of Xyle .....	211
6.4.2.	Defining the Xyle-tip anchor point using FX SMFS.....	214
6.4.3.	Xyle and Halo-TMH-Xyle in DIBMA nanodiscs .....	217
6.4.4.	Unfolding Xyle from the C-terminal .....	219
6.4.5.	Unfolding Xyle from the N-terminal.....	223
6.4.6.	D-xylose binding does not affect Xyle unfolding pathways.....	225
6.4.7.	Force-clamp unfolding of WT Xyle.....	226
6.5.	Discussion.....	229
6.5.1.	Xyle unfolding trajectory and force regimes .....	229
6.5.2.	Limitations of this study .....	232
6.5.3.	Moving towards magnetic tweezer spectroscopy .....	233
6.6.	Chapter summary.....	234

7.	Summary and future directions.....	236
7.1.	Summary of results.....	237
7.2.	Visual summary of RNC samples generated.....	238
7.3.	A note on the Coronavirus lab shutdown.....	239
7.4.	Significance and future work.....	239
7.4.1.	Generation of GlpG RNCs captured within in DDM micelles.....	239
7.4.2.	Capture of GlpG RNCs in native nanodiscs with endogenous SecYEG.....	240
7.4.3.	Towards in-depth biophysical studies of folding using AFM and MT.....	244
7.5.	Closing remarks.....	245
	References.....	247
	Appendices.....	259
1.	N-His GlpG for pET28a.....	259
2.	WT Xyle for pET28a.....	259
3.	Xyle mechanical unfolding sequences and constructs.....	260
4.	Primers and PCR conditions for GlpG and RNC related mutants.....	261
5.	Primers and PCR conditions for Xyle and associated mutants.....	261
6.	RNC protein sequences and constructs.....	262
7.	Bell-Evans-Richie energy landscape example calculation.....	263
8.	Theoretical contour lengths for mapped Xyle PBD 4GBY.....	265
9.	<i>BBA Biomembranes</i> : Mini review.....	266
10.	<i>Biochemistry</i> : Capturing membrane protein RNCs article.....	279



## List of Figures

---

Figure 1-1: Hypothetical protein folding funnel for $\alpha$ -helical membrane protein.....	26
Figure 1-2: Membrane mimics for membrane protein folding studies.....	28
Figure 1-3: In vivo translocon assisted co-translational folding .....	35
Figure 1-4: SEIRAS set-up and methodology .....	41
Figure 1-5: Possible models for co-translational folding in the absence of translocon from SEIRAS .....	42
Figure 1-6: Denaturant unfolding of membrane proteins.....	45
Figure 1-7: Structure of the 70S ribosome .....	51
Figure 1-8: Summary of RNC use to study co-translational membrane protein folding, targeting and translocon dynamics .....	65
Figure 2-1: Example agarose gel for Infusion cloning of RNCs into MCS1 of RSF-Duet1-SecYEGavi.....	71
Figure 3-1: Structure of the GlpG Rhomboid protease .....	94
Figure 3-2: Gel-filtration chromatogram and SDS-PAGE analysis of N-his GlpG.....	101
Figure 3-3: N-His GlpG is active and folded in DDM.....	102
Figure 3-4: Construct design of GlpG RNCs at different co-translational ‘intermediates’ ....	104
Figure 3-5: Quality control for DDM purified RNCs .....	105
Figure 3-6: Confirmation of purified, stalled membrane protein RNCs.....	106
Figure 3-7: Sucrose preparation of DDM RNCs as an alternative to SEC purification.....	107
Figure 3-8: Negative staining of 4 TM GlpG in DDM .....	109
Figure 3-9: In vivo biotinylation of RNCs to prevent non-specific Ni-NTA binding .....	110
Figure 3-10: Biotinylated 4 TM GlpG RNC binds to streptavidin Sepharose beads .....	111
Figure 3-11: Low pH RNC binding to streptavidin and neutravidin coated beads .....	112
Figure 3-12: Monomeric avidin binding and eluting of biotinylated 4 TM GlpG RNCs .....	113
Figure 3-13: Arrest-enhanced SecM yields RNCs with a greater stability, occupancy, and purity in DDM.....	114
Figure 3-14: RNC occupancy calculation using GlpG standards .....	115
Figure 3-15: Endogenous SecYEG co-purifies with RNC in DDM detergent.....	116
Figure 3-16: Overexpression of biotinylated SecYEG and 4 TM GlpG RNCs .....	117
Figure 3-17: Basic summary of DDM RNC preparation steps .....	122
Figure 4-1: DIBMA preparation and IR quality control .....	142

Figure 4-2: DIBMA solubilisation trials of synthetic liposomes .....	143
Figure 4-3: Preparation of Xyle DIBMA nanodiscs in a synthetic lipid mix .....	144
Figure 4-4: DIBMA solubilisation of 40:40:20 synthetic lipids with Xyle.....	145
Figure 4-5: Schematic of DIBMA RNC preparation steps .....	146
Figure 4-6: Solubilisation and gel-filtration purification of GlpG RNCs in DIBMA.....	148
Figure 4-7: Preparation of DIBMA RNCs using SEC and sucrose gradient.....	150
Figure 4-8: Endogenous SecYEG is captured by DIBMA .....	151
Figure 4-9: GlpG RNCs prepared in DIBMA nanodiscs contain native lipid compositions....	153
Figure 4-10: HPTLC plate staining and quantification of DIBMA nanodisc RNCs prepared from membranes extracted from cells grown in Luria-Bertani media.....	154
Figure 4-11: Negative stains of 2 TM and 4 TM GlpG in DIBMA native nanodiscs .....	155
Figure 4-12: Negative stain of 6 TM GlpG RNC in DIBMA native nanodiscs .....	156
Figure 4-13: GlpG RNCs are stable in DIBMA nanodiscs and tolerant to freeze thaw cycles	157
Figure 5-1: MFS alternate access transport mechanism for Xyle .....	166
Figure 5-2: Growth, expression, and purification of WT Xyle and Xyle-pGly28.....	171
Figure 5-3: Purification of the Xyle N-domain .....	173
Figure 5-4: Purification of the Xyle C domain .....	174
Figure 5-5: CD and OCD of WT Xyle and Xyle-pGly28 in DDM.....	175
Figure 5-6: ITC for WT Xyle and Xyle-pGly28 in DDM.....	176
Figure 5-7: Markwell-Lowry and OG detergent standard curves .....	178
Figure 5-8: Schematic for novel XDH assay for Xyle transport.....	179
Figure 5-9: Preparation and calculating the activity of XDH .....	180
Figure 5-10: Transport of D-xylose across a bilayer assayed using XDH.....	181
Figure 5-11: Schematic of doubly tagged Xyle for perpendicular AFM or MT pulling .....	183
Figure 5-12: DDM preparation of Halo-TMH-Xyle construct for AFM .....	184
Figure 6-1: AFM SMFS to (un)fold membrane proteins .....	193
Figure 6-2: Unfolding energy barrier for a membrane protein .....	199
Figure 6-3: MT setup to unfold membrane proteins .....	202
Figure 6-4: PyMol structure of Xyle showing protein dimensions.....	204
Figure 6-5: Optimisation of liposome deposition conditions onto freshly cleaved mica .....	207
Figure 6-6: High-resolution imaging of single Xyle proteins in bilayers .....	208
Figure 6-7: AFM high-resolution image of WT Xyle reconstituted in DIBMA nanodiscs .....	211
Figure 6-8: Theoretical unfolding of WT Xyle from C-terminal cantilever binding.....	213
Figure 6-9: Example traces from WT Xyle corresponding to N- or C- terminal binding .....	214

Figure 6-10: Summary of trajectory assignment for C-terminal Xyle binding using Xyle-pGly28.....	215
Figure 6-11: Empty and WT Xyle reconstituted DIBMA trace comparison .....	217
Figure 6-12: Assaying the tethering of HaloTag to halo-ligand of DIBMA native nanodisc Halo-TMH-Xyle .....	218
Figure 6-13: 2D histogram summary of Xyle-pGly28 C-terminal-binding force peaks.....	220
Figure 6-14: 2D histogram summary of WT Xyle C-terminal-binding force peaks .....	221
Figure 6-15: 2D histogram summary of Xyle and Xyle-pGly28 N-terminal-binding force peaks .....	223
Figure 6-16: N-terminal binding traces showing $L_{C_4}$ intermediate.....	224
Figure 6-17: WT Xyle binding D-xylose plotting rupture force against $L_C$ .....	225
Figure 6-18: Selection of clamp traces for WT Xyle at 65 pN .....	227
Figure 6-19: Example FC trace showing intermediate 'hopping' at 65 pN.....	228
Figure 6-20: Summary of FX SMFS anchor points for WT Xyle.....	230

## List of Tables

---

Table 1-1: Lipid composition variations between organisms and organelles .....	22
Table 1-2: Overview of polymers used for native nanodisc production .....	33
Table 2-1: Antibodies and dilutions for Western Blotting .....	73
Table 2-2: MDG media components .....	74
Table 2-3: EM9 media components .....	74
Table 2-4: 1000x Trace metal stocks, pH 7.0.....	75
Table 3-1: Methods summary to obtain homogenous RNC preparation in DDM micelles.....	123
Table 4-1: Structures of membrane proteins in SMA polymer nanodiscs .....	135
Table 5-1: Summary and literature comparisons of biochemical values obtained for WT XylE and XylE-pGly28 .....	186
Table 6-1: Energetics extracted from AFM based SMFS on membrane proteins .....	195
Table 6-2: Comparison of major peak $L_C$ between C-terminally bound WT and pGly28 XylE .....	216
Table 6-3: Central $L_C$ values and anchor points for C-terminal binding XylE-pGly28 and WT XylE.....	222
Table 6-4: Summary of $L_C$ values and anchor points for N-terminal extraction of XylE.....	224
Table 6-5: Step-Sizes for XylE unfolding at 65 pN.....	228

## List of Equations

---

A = Appendix

Equation 1-1: Folding pathway for T4 lysozyme determined using optical tweezer SMFS ....	59
Equation 2-1: Mean residue ellipticity (MRE) calculation .....	85
Equation 2-2: Beer-Lambert equation.....	86
Equation 2-3: Hooke's law .....	88
Equation 2-4: Simple harmonic motion with one degree of freedom for AFM cantilever.....	89
Equation 2-5: $k_s$ calculation from power spectrum .....	89
Equation 4-1: Molar refractive index increment for DIBMA .....	140
Equation 6-1: WLC model of polymer elasticity.....	198
Equation A-1: Loading rate varies with pulling velocity.....	263
Equation A-2: Bell-Evans-Richie model .....	263
Equation A-3: Arrhenius equation.....	263

## List of Abbreviations

---

$\Delta G_u$	Gibbs free energy of unfolding	MFS	Major Facilitator superfamily
$\Delta x$	Distance to transition state	MP	Membrane protein
$\alpha$ NPG	<i>p</i> -nitrophenyl- $\alpha$ - <i>D</i> -galactopyranoside	MPEX	Membrane protein explorer
AA	Amino acid	MRE	Mean-residue ellipticity
AAP	Arginine attenuator peptide	mRNA	Messenger ribonucleic acid
AE1	Arrest Enhanced SecM	MRW	Mean-residue weight
AFM	Atomic Force microscopy	MSP	Membrane-scaffold protein
AMS	4-Acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid disodium salt	MST	Microscale Thermophoresis
AP	Arrest peptide	MT	Magnetic tweezers
APAA	Alkyl polyacrylic	MWCO	Molecular weight cut-off
APC	Amino Acid-Polyamine-Organocation protein family	NAD <sup>+</sup>	Nicotinamide adenine dinucleotide
AQP1	Aquaporin-1	NADH	Reduced nicotinamide adenine dinucleotide
ATP	Adenosine-triphosphate	NBD	N-(7-Nitrobenz-2-Oxa-1,3-Diazol-4-yl)
BirA	Biotin ligase	NC	Nascent chain
$\beta$ ME	$\beta$ -mercaptoethanol	Ni-NTA	Nickel-nitrilotriacetic acid
BODIPY	4,4-difluoro-4-bora-3a,4a-diaza-s-indacene	NMR	Nuclear magnetic resonance
BSA	Bovine serum albumin	OCD	Oriented circular dichroism
CD	Circular dichroism	OG	$\beta$ -D-octyl glucoside
CFP	Cyan-fluorescent protein	OPM	Orientations of Proteins in Membranes database
CFTR	Cystic fibrosis transmembrane conductance regulator	OST	Oligosaccharyltransferase
CHAPSO	3-[(3-cholamidopropyl) dimethylammonio]-2-hydroxyl-1-propanesulfonate	<i>P. aeruginosa</i>	<i>Pseudomonas aeruginosa</i>
CL	Cardiolipin	PBS	Phosphate-buffered saline
CV	Column volume	PBST	PBS-tween-20
DAGK	Diacylglycerol kinase	PCR	Polymerase chain reaction
DDM	Dodecyl- $\beta$ -D-maltoside	PDB	Protein data bank
DEPC	Diethyl pyrocarbonate	PET	Photoinduced electron transfer
DFP	Diisopropyl fluorophosphate	pGly28	Poly-glycine linker of 28 amino acids
DFS	Dynamic force spectroscopy	PMA	Poly(methyl acrylate)
DHFR	Dihydrofolate reductase	PMF	Proton-motive force
DHPC	Dihexanoyl-phosphatidylcholine	PMSF	Phenylmethylsulphonyl fluoride
DIBMA	Diisobutylene-maleic acid copolymer	PR	Proteorhodopsin
DLS	Dynamic light scattering	PTC	Peptidyl-transferase centre
DMPC	Dimyristoyl-phosphatidylcholine	PVDF	Polyvinylidene fluoride
DMPG	Dimyristoyl-phosphatidylglycerol	QNM	Quantitative nanomechanical property mapping
dNTPs	Deoxyribonucleotide triphosphates	RF	Retention-factor
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine	RNC	Ribosome bound nascent chain complex

DOPE	1,2-dioleoyl-sn-glycero-3-phosphoethanolamine	ROS	Rod outer segment
DOPG	1,2-dioleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)	RRL	Rabbit reticulocyte lysate
<i>E. coli</i>	<i>Escherichia coli</i>	rRNA	Ribosomal ribonucleic acid
ECH	Extra-cytoplasmic helix	rSAP	Shrimp-alkaline phosphatase
EDTA	Ethylenediaminetetraacetic acid	RT	Room temperature
EM9	Enhanced Mg media	SAS	Signal-anchor-sequence
ER	Endoplasmic reticulum	SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
FC	Force-clamp spectroscopy	SEC	Size-exclusion chromatography
FCS	Fluorescence correlation spectroscopy	SecM	Secretion monitor peptide
FRET	Förster resonance energy transfer	SEIRAS	Surface-enhanced infrared spectroscopy
FTIR	Fourier transform Infra-red spectroscopy	SLC	Solute carrier
FX	Force-extension spectroscopy	SM	Sphingomyelin
g	Centrifugal force	SMA	Styrene-maleic acid copolymer
GFP	Green fluorescent protein	SMALP	Styrene-maleic acid lipid particle
GPCR	G-Protein coupled receptor	SMFS	Single-molecule force spectroscopy
GuHCl	Guanidine hydrochloride	SMI	Imidized poly(styrene-co-maleic anhydride)
<i>H. salinarum</i>	<i>Halobacterium salinarum</i>	SRCD	Synchrotron radiation circular dichroism
HDX-MS	Hydrogen/deuterium exchange mass spectrometry	SRP	Signal recognition particle
HEPES	2-[4-(2-hydroxyethyl)piperazin-1-yl]ethane sulfonic acid	TCA	Trichloroacetic acid
HPTLC	High performance thin layer chromatography	TEM	Transmission electron microscopy
HRP	Horseradish-peroxidase	TEV	Tobacco Etch Virus protease
Hsp	Heat-shock protein	TM	Transmembrane
HTL	Holotranslocon	TMD	Transmembrane domain
ICH	Intra-cytoplasmic helix	TMH	Transmembrane helix
IMAC	Immobilised affinity chromatography	<i>tna</i>	Tryptophanase
IPTG	$\beta$ -D-1-thiogalactopyranoside	TRAM	Translocon-associated membrane protein
ITC	Isothermal titration calorimetry	Trap	Translocon-associated protein
IVTT	<i>In vitro</i> transcription/translation	TRiC	Tailless complex polypeptide 1 ring complex
$k_D$	Dissociation constant	tRNA	Transfer ribonucleic acid
KLD	Kinase/ligase/DpnI	UAA	Unnatural amino acid
LB	Luria-Bertani	UV	Ultraviolet
$L_c$	Contour length	WLC	Worm-like chain
LILBID-MS	Laser Induced Liquid Bead Ion Desorption Mass spectrometry	WT	Wild type
MD	Molecular dynamics	XDH	Xylose dehydrogenase
MDG	Phosphate(M):Aspartate(D):Glucose(G) media	XMR	Xylose mutarotase

# 1. Introduction

---

**Section 1.3. has been published as a mini-review article (appendix 9):**

Structural insight into co-translational membrane protein folding

**Pellowe GA, Booth PJ. *Biochim Biophys Acta – Biomembranes*, 2020**

doi: 10.1016/j.bbamem.2019.07.007

**Some of chapter 1 is in preparation for a review article:**

*OpenBiology* 2020/21

## **1.1. Membranes and membrane proteins**

### **1.1.1. Membranes and lipids**

Biological membranes comprise of lipid molecules and protein and surround the cell and its organelles. At its most basic function, the cell membrane acts as a physical barrier which compartmentalises different regions of the cell. The lipid membrane is selectively permeable which prevents leakage of the contents of the compartment, while allowing generation of chemical and electrical concentration gradients across the membrane. These gradients are often used to drive the uptake of essential nutrients into the cell, and the decoration of the lipid bilayer with integral membrane proteins is often a prerequisite for this. Cell membranes and their associated proteins play an integral role in osmotic homeostasis, cell-cell signalling, cell-cell adhesion and cell motility (Findlay and Booth, 2006, Harayama and Riezman, 2018).

The basic lipid bilayer is formed spontaneously in water and is driven by the characteristic amphipathic structure of lipids; the most common *Escherichia coli* (*E. coli*) lipid family is the phospholipid, which contains a phosphate containing polar headgroup, connected to two acyl chains via a glycerol moiety. Bilayer self-assembly is driven by the apolar acyl chains packing together with hydrophobic interactions to shield themselves from the surrounding aqueous environment, and the polar headgroups of the lipid face outwards. The chemical make-up of the lipid mix can direct the self-assembly of the lipid molecules into different phases; micellar, lamellar, hexagonal, and cubic phases can all be formed depending on the variety of lipids used (Findlay and Booth, 2006, Corin and Bowie, 2020). However only a fluid lamellar phase; produced with lipids like phosphatidylcholine (PC), phosphatidylserine (PS) and sphingomyelin (SM), can form a biological, non-permeable lipid bilayer.

Lipid identity, and therefore chemistry can vary massively between membranes in cell types, as well as various compartments within single cells, and the relative mix of lipids can greatly affect global bilayer properties which alter folding dynamics of proteins. The polar headgroup size and type, as well as the length and saturation and asymmetry of the acyl chains can greatly affect the fluidity, lateral pressure, elasticity, and curvature of the lipid bilayer. Lipids contributing to the global properties of the bilayer can be found extensively reviewed in Corin and Bowie (2020). Briefly, unsaturated acyl chains produce fluid membranes, whereas highly saturated chain lipids such as phosphatidylethanolamine (PE) produce kinks in the lipid tails, which increase the number of possible configurations of the chain and hence increase rotational freedom and disorder in the hydrophobic region, all leading to a more ordered solid gel phase of membrane which is stabilised by the many packed Van der Waals interactions



between the tails (van Meer et al., 2008). Fluidity can be decreased with the addition of cholesterol, or like-molecules which sequester between the acyl chains, this further restricts lipid motion. Additionally, the length of the acyl chain region can alter hydrophobic thickness, which can alter the trafficking of certain eukaryotic membrane proteins based on the length of their helices and the corresponding thickness of the bilayer at varying parts of the secretory pathway (Findlay and Booth, 2006). Headgroup chemistries can also alter global bilayer properties. Smaller headgroups can increase negative curvature of the bilayer, which in turn increases lateral pressure in the chain region as the hydrophobic effect compensates, the opposite is true with larger headgroup sizes. Certain headgroups also contain charge, such as phosphatidylglycerol (PG) molecules, which increase electrostatic interactions with other membranes and proteins residing within the bilayer (Corin and Bowie, 2020).

The structure of the biological membrane with lipid and protein (see **section 1.1.2**) was proposed first as the ‘fluid mosaic model’ (Singer and Nicolson, 1972) where the proteins spanning the bilayer can diffuse laterally along the plane of the membrane. However, this has been shown to be an oversimplification, with far more proteins embedded in the bilayer, and with lipids residing in ‘rafts’ rather than as a uniform distribution, resulting in various bilayer thicknesses, lateral pressures, and lipid types in an incredibly complex environment (Engelman, 2005, Marinko et al., 2019). Lipid environments can vary dramatically between different organelles within a cell, and between membranes of different organisms. Differences between lipid compositions in *E. coli* inner membranes, *Acholeplasma laidlawii* (*A. laidlawii*) and organelles of nucleated mammalian cells are shown in **Table 1-1**.

Certain organisms have been shown to adapt their lipid complexity at a genomic level in response to their surrounding of pH, temperature, or pressure environment. This response was observed in *A. laidlawii* varying ratios between two dominating glycolipids in the bacterial membrane. Diglucoyldiglyceride (DGlcDAG) and monoglucoyldiglyceride (MGlcDAG) are bilayer and non-bilayer forming respectively, and altered ratios of these can alter the fluidity of the bilayer (Wieslander et al., 2002). In humans, rather than being affected by environment, acyl chain composition is affected by diet (Corin and Bowie, 2020). Differences here largely affect the lateral pressure profile throughout the bilayer in a similar way to PE, which may alter pathways of protein folding. Regions of high pressure in the acyl chain region (low headgroup pressure) promote TM helix association with lipid headgroups, whereas lower acyl pressure may suggest TM insertion across the bilayer is favoured. This phenomenon was characterised *in vitro* using MscL, by titrating PE lipids with unsaturated acyl chains into

synthetic liposomes to increase bilayer pressure (Harris et al., 2018), as well as refolding experiments of MscL from a chemically denatured state (Miller et al., 2016). Much of these systems have used fully formed and reconstituted proteins refolding into lipids, reviewed in Harris et al., (2008), but these experiments cannot accurately depict what occurs co-translationally *in vivo*.

Another study of the effect of lipid properties on proteins has been explored using a PE deficient *E. coli* strain which resulted in the inverted topology of the 6 transmembrane (TM) helix N-terminal domain of the MFS LacY transporter. This saw helix 7, which has an overall lower hydrophathy profile to the rest of the helices, no longer residing in the membrane but instead associated with lipid headgroups. Only the final 5 TM helices therefore remained correctly oriented. Domain flipping was able to rescue on addition of PE *in vitro*, and has been suggested that the effect likely results from the increased PG and negative charge at the cytoplasmic lipid interaction regions of the protein, which was sufficient enough to induce domain flipping (Dowhan and Bogdanov, 2011).

The differences in lipid compositions can also vary between organelles. For example, in the Golgi and ER there are much lower quantities of sphingolipids and cholesterol found compared to the plasma membrane (**Table 1-1**), possibly resulting in similar effects on bilayer pressures mentioned above. Proteins produced in these organelles which are then trafficked towards the plasma membrane therefore experience vast changes in the lipid environment and must have adapted specific mechanisms to remain stable both in and between membranes with altered global properties. Unfortunately the specific effect of the energetics of membrane protein folding and stability in differing lipid systems between organelles is still largely unexplored (Marinko et al., 2019), and much of the work has been explored in terms of protein function; cholesterol being required for GPCR function, a negatively charged lipid for correct KcsA function, and cardiolipin for protein oligomerisation, as examples (Marinko et al., 2019, Corin and Bowie, 2020). Overall, explorations into further lipid species governing folding, and the effect of this changing environment is largely understudied, and this remains an area of great interest for the studies of membrane protein folding.

**Table 1-1: Lipid composition variations between organisms and organelles**

Lipid compositions as percentages of total lipid content from *E. coli*, *A. laidlawii* and nucleated mammalian cells. Table adapted from Corin and Bowie 2020. Abbreviations: PE: phosphatidylethanolamine, PG: phosphatidylglycerol, PC: phosphatidylcholine, PS: phosphatidylserine, PI: phosphatidylinositol, CL: cardiolipin, Chol: cholesterol, SM: sphingomyelin, M- and DGlcDAG: mono- and di-glucosyldiacylglycerol and BmP: bis(monoacylglycerol)phosphate.

Adapted from Corin and Bowie 2020	PE	PG	PC	PS	PI	CL	Chol	SM	MGlcDAG	DGlcDAG	DPG	BmP
<i>A. laidlawii</i>		10-20							20-43	31-56	9-17	
<i>E. coli</i>	70-75	20-25				5-10						
Plasma membrane	11		23	8			34	17				
Golgi	21		36	6	12		18	6				
ER	20		54		11		8					
Mitochondria	31		37		6	22						
Endosomes/lysosomes	11		30		7		30	15				7

### 1.1.2. Membrane proteins

Membrane proteins have a diverse range of structure and functions and are essential for cell viability. They make up between 20-80 % of the membrane by weight across most kingdoms, and can either be peripherally associated with the membrane, or transverse the membrane entirely. In Human cells, around 30 % of the total protein in the cell consists of integral membrane proteins, and of these around 50 % are druggable (Corin and Bowie, 2020, Marinko et al., 2019). Membrane proteins can provide a selective permeability of molecules across the membrane, using channels or pumps, or act as membrane receptors which transmit a signal across cells or tissues. Additionally, membrane proteins are involved in the most important energy generation reactions across all kingdoms, during oxidative phosphorylation and photosynthesis. Many human integral membrane proteins are druggable, and are often associated with disease phenotypes when mutated, or their oligomers incorrectly assembled - leading to misfolding or an absence of the protein in the proteome. Perhaps the most renowned example of this manifests itself as a single amino acid deletion,  $\Delta F508$ , of the cystic fibrosis transmembrane conductance regulator (CFTR) gene. This produces a misfolded CFTR protein which cannot exit the endoplasmic reticulum (ER) and is digested by the proteasome, resulting in a build-up of chloride ions in the cytoplasm of epithelial cells, which an otherwise function CFTR channel would remove, leading to a sticky mucus forming in the airway (Bartoszewski et al., 2010) and a subsequent onset of Cystic Fibrosis symptoms.

Additionally, the solute carrier (SLC) gene family, and major-facilitator superfamily (MFS) secondary transporters in particular, have varied modes of the facilitation of ion, sugars, small peptides, vitamins and a host of other biomolecules across to lipid bilayer in response to an electrochemical gradient (Hediger et al., 2013, Sun et al., 2012, Yan, 2015). These proteins act to maintain normal cellular, and whole-body homeostasis. It is therefore no surprise that when these proteins have a limited function, a host of disease phenotypes may arise. Proteins of the human glucose transporter family, namely GLUTs1-4, have been extensively studied in relation to metabolic and glucose storage diseases, with mutated GLUT1 resulting in De Vivo disease, or GLUT1 deficiency syndrome (Deng et al., 2014). This manifests itself in a broad range of inimical symptoms including infantile-onset seizures (Brockmann, 2009).

Other protein misfolding phenotypes, such as aggregates of globular proteins; Tau and  $\alpha$ -synuclein, promote Alzheimer's and Parkinson's disease phenotypes, respectively. This highlights the importance of studying the (mis)folding pathways of proteins to prevent disease consequences down the line. The proteostasis network attempts to prevent misfolding by

chaperoning folding, and removing misfolded aggregates, due to an intricate, complex network of proteins. Unfortunately, mistakes can still happen and lead to unfavourable disease phenotypes. In depth reviews of soluble and membrane protein folding in health and disease can be found in reviews (Balchin et al., 2016, Marinko et al., 2019).

### **1.1.3. Studying membrane protein folding is important**

Membrane proteins are complex structures which reside in a chemically complex environment at the interface between a hydrophobic bilayer, and an exposed aqueous environment. They interact closely with both lipid and water in their environment however they still require synthesis by the ribosome and trafficking to the correct destination in the cell.

Most proteins fold in a vectorial fashion from N- to C- terminus as they are translated by the ribosome. The primary sequence of the protein contains all the information for the correct folding of the protein, which can fold into more complex secondary structure motifs, which are further complicated by long range tertiary structure formation. The final folded protein structure is crucial for correct protein activity, and for membrane proteins in particular – the hydrophobic bilayer is crucial to promote formation of this correct structure. Although the primary sequence of the protein does dictate the final structure of a protein, as described by Anfinsen in 1973 (Anfinsen, 1973), this has been shown to be an oversimplification, and folding is effected by many factors, including the environment, as well as folding assistance by chaperoning proteins to prevent protein aggregation.

The classic view of soluble protein folding supports a folding funnel, depicted as the energy of the protein against the topological arrangement of atoms (Dill and Chan, 1997). The unfolded primary sequence starts at the top of the funnel and topples down through the kinetic traps and energy wells as it folds, finally landing at the lowest energy conformation in its final folded state. The rugged funnel environment can vary greatly between different types of protein and for membrane proteins, it is not as simple as depicted in the folding funnel model, due to the additional complexity of the lipid bilayer solvent for the membrane protein, which can be difficult to depict in such models.  $\alpha$ -helical proteins readily collapse through the conformational space to form the hydrophobic core which can immediately form local interactions and gain tertiary structure. Mixed structure  $\alpha/\beta$  proteins initially collapse to form the core, before assembling into ordered structures where  $\beta$ -sheet formation seems to be the rate limiting step for folding (Schultz, 2000). Some folding pathways therefore include a stable intermediate which can have its own energy well, which must then be overcome to promote

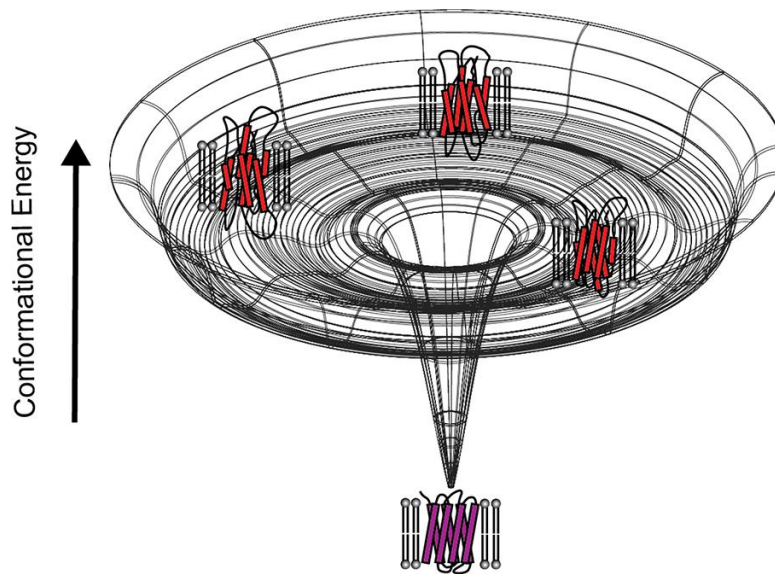
native protein folding. Overcoming this energy barrier is often made easier by folding chaperones and even the ribosome itself *in vivo*.

Reimagining the classic folding funnel is for membrane proteins and co-translational folding is incredibly difficult. The lipid solvent which as explored in **section 1.1.1** can vary massively, and local interactions between the lipids and the protein will change the surface of the free energy landscape (vertical axis), and the global properties of the bilayer may alter the depth of the funnel (horizontal axis), as helix insertion becomes more, or less stable as governed by intrinsic bilayer mechanics, in effect reducing the configurational entropy of the funnel.

Assuming the folding funnel schematic is a correct, albeit incomplete, depiction of protein folding; an  $\alpha$ -helical protein must sample a variety of pre-formed helices before eventually collapsing into a native topology, perhaps suggesting a relatively flat but wide upper funnel region as helices and pre-formed from random-coils, before all helices are formed and a significant collapse into final native structure residing in a deep well occurs. This has been proposed for bacteriorhodopsin (bR) refolding into bicelles, reviewed in Marinko et al, (2019)

**Figure 1-1.** The collapse into native structure may happen very slowly, as helices are searching conformational space to form interhelical contacts and may be sped up by reducing the size of the bicelle. This would result in a thinner, but deeper funnel as fewer conformations can be searched in a confined space. In the case of bR however it has been shown experimentally (Lu and Booth, 2000) that parallel folding pathways which sample different intermediates are observed with different kinetic and thermodynamic parameters (Allen et al., 2004, Curran et al., 1999) when folding is studied in differing lipid compositions. These specific lipid interactions will alter the shape of the funnel to produce a more ‘rugged’ energy landscape with additional energy wells in which the bR intermediates must sample before finding its native state. The extent of any changes of the funnel, however, would need to be characterised by extensive experimental and molecular dynamics studies.

bR experiments highlight a small piece of the incredibly complex effect on protein folding by the lipid environment, with extensive energetics of its folding pathways defined. For co-translational folding however, this is an incredibly complex process with many contributing factors affecting folding rates and preventing off-pathway misfolding with the specific binding of lipids, and action of chaperoning and insertion machinery to significantly reduce energy barriers and configurational entropy which likely cannot be simply depicted using a folding funnel diagram.



**Figure 1-1: Hypothetical protein folding funnel for  $\alpha$ -helical membrane protein**

*In the classic folding funnel depicted by Dill and Chan (1997), proteins start at the top of the funnel in an unfolded state. As they start to fold, they make their way down through the funnel to one of the many minima of a folded protein. The fully folded native protein is found at the very bottom of the funnel in the lowest energy state. Chaperones are required to move a misfolded protein over its energy barrier and back into its correctly folded state. Here, the funnel has been reimagined for bR folding into a bicelle from a chemically denatured state. The upper rim represents the random coiled state which is unlikely to be sampled in a membrane environment. the formation of helices occurs in the secondary basin and it is here helices explore topological configuration before the native structure is achieved in the deep energy well, suggesting stable protein is formed. Figure taken from Marinko et al., (2019) and based on Dill and Chan (1997).*

The energetic cost of membrane protein partitioning into a bilayer is high, and only two structural motifs for integral membrane proteins are observed -  $\alpha$ -helical bundles and  $\beta$ -barrels, where the complete hydrogen bonding in the secondary structure compensates for this higher energetic barrier (White and Wimley, 1999). This energetic cost is perhaps significantly reduced, particularly in  $\alpha$ -helical proteins with the aid of the heterotrimeric SecYEG translocon in bacteria (Sec61 $\alpha\beta\gamma$  in eukaryotes) (Komar et al., 2016, Gold et al., 2007) and other insertase proteins which capture the helical nascent chain from the SecY lateral gate, such as YidC (Serek et al., 2004, Wang and Dalbey, 2011, Komar et al., 2016) (see **section 1.3.1**). This allows the partitioning of the  $\alpha$ -helices across the bilayer, with the aid of the SecA ATPase to lower the energy barrier of insertion during membrane protein biogenesis in the cell, allowing correct folding on a biologically relevant timescale.

The complex lipid environment poses challenging demands on membrane proteins with respect to folding, translocation, and stability, and compared with like studies for soluble protein, membrane protein studies lag behind due to the difficulty in providing a physiologically relevant environment. The field is still some way from understanding the

molecular details of membrane protein co-translational folding, with molecular and mechanistic knowledge currently deriving from refolding studies of full length isolated nascent chains (Marinko et al., 2019), and new methods are still needed for mechanistic and structural detail on folding of membrane proteins. This thesis aims to develop a major advance in the membrane protein folding field; a method to directly observe the co-translational folding and insertion of a polytopic  $\alpha$ -helical protein into a native lipid environment.

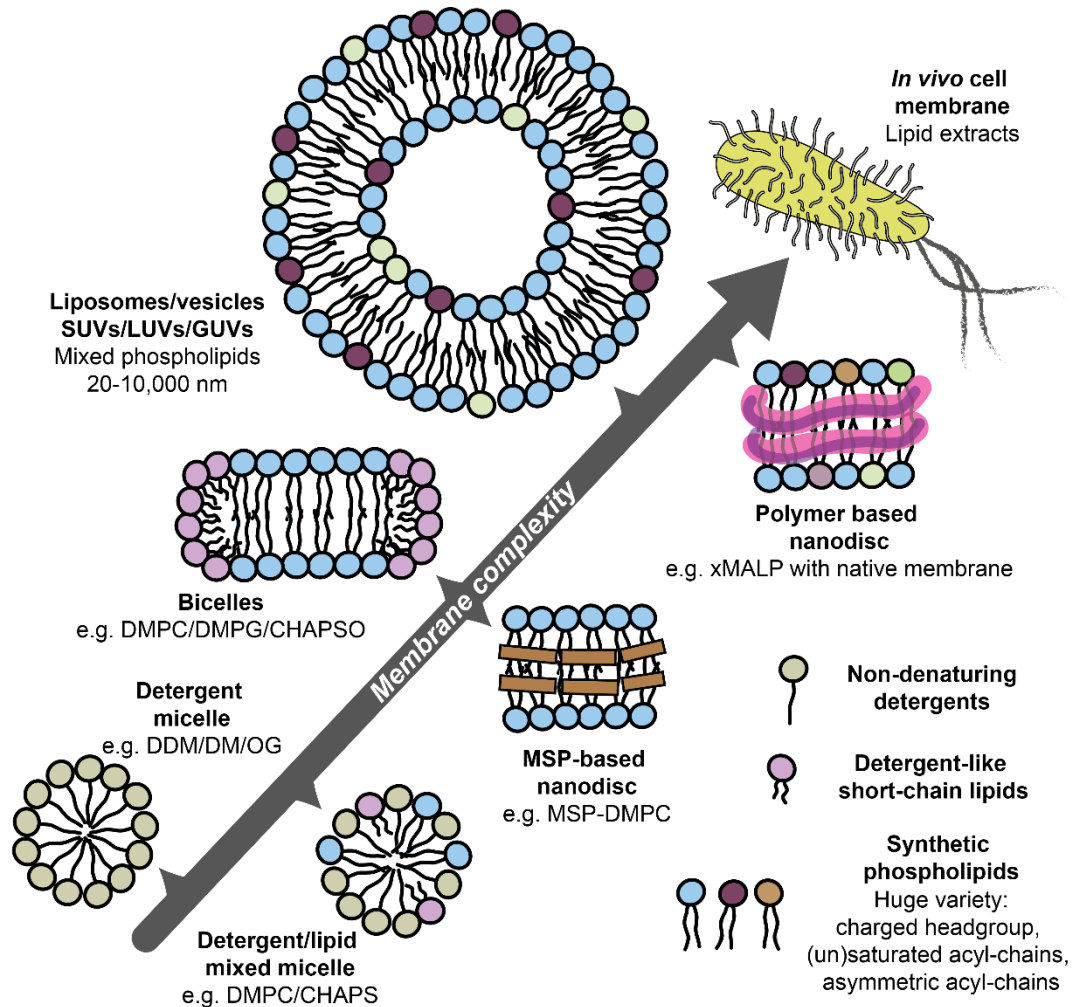
## 1.2. Membrane mimetics for protein folding studies

Unlike globular proteins, membrane proteins have specific hydrophobic moieties which span the lipid bilayer and polar regions which are exposed to the sides of the membrane as well as line the solute channel. Mimicking the physiological conditions of this complex membrane can be tricky; it is common practice to purify such proteins into detergent micelles which act as a temporary membrane mimic, which are used to carry out both functional assays and *in vitro* folding studies. Although good for purification and handling, detergents are destabilising and can cause protein activity to diminish over time (Seddon et al., 2004) and reconstitution into higher order lipid mimics is often necessary for stability as well as providing a more physiological environment for *in vitro* studies.

*In vitro* membrane mimics come in varying complexities **Figure 1-2**, each with advantages and disadvantages in terms of physiological relevance and suitability for the mode of study – the decision of which mimic is best is often a compromise between their intrinsic properties, and their suitability for an experiment. For instance, smaller and easy to handle detergents are good for biochemical studies of protein secondary structure using circular dichroism (CD)\*, but do not accurately portray the *in vivo* lipid environment. Alternatively, larger bilayer forming liposomes mixed with correct synthetic lipids are a close portrayal to the correct *in vivo* environment, but it can be difficult to obtain high resolution structural or biophysical data due to their size, and consequent light scattering by the lipids leading to low signal-to-noise.

\*CD measures the differential absorption of right-handed and left-handed circularly polarised light. This, when measured over various wavelengths allows the examination of chiral structures, much like those formed by protein secondary structure –  $\alpha$ -helices and  $\beta$ -sheets which have their own characteristic spectra and can therefore be used to follow unfolding of the secondary structure when subject to heat, or chemical denaturants.





**Figure 1-2: Membrane mimics for membrane protein folding studies**

Synthetic lipids and detergents are commonly mixed to provide a desired membrane environment based on the lipid properties. The relative mixes, and lipid/detergent types can affect the protein in a variety of ways. The complexity of the system can be increased from detergent micelles, to an *in vivo* like environment by altering the lipids, or directly extracting the membrane environment from the host system by using entire lipid extracts. Each mimic has its own advantages and disadvantages, micelles are the least membrane like, bicelles and nanodiscs (brown blocks are indicative of membrane scaffold protein (MSP)) form bilayers. Liposomes, and small, large, and giant unilamellar vesicles (SUV/LUV/GUV) can also form bilayers but are compartmentalised which useful to assay transport. Novel co-polymer (purple rings) systems can be used to extract membrane proteins with the native host lipid environment, or whole lipid extracts can be extracted and moulded into the most convenient environment for biophysical analysis, a bicelle for magnetic tweezer unfolding, or vesicles for cell-free study, as examples.

### 1.2.1. The problem with detergents

The use of detergent for the study of membrane proteins is very common and has many uses which have allowed some incredibly detailed folding studies of membrane proteins. Detergents often act as a first avenue for the *in vitro* study of membrane proteins, unfortunately however, there are significant drawbacks when using detergent to solubilise, extract and isolate a membrane protein from its native environment for further investigation.

The importance of lipid-protein interactions should not be understated, and there is concern that not all membrane protein structures can be considered biologically relevant as they lack associated lipids that may be structurally significant (Guo, 2020, Swiecicki et al., 2020). Using this classical method of purifying membrane proteins results in unfavourable effects being introduced whereby protein integrity is disrupted through the process of stripping away the potentially stabilising, physiologically relevant lipid environment (Seddon et al., 2004, Ravula et al., 2019). Detergents may also remove annular lipids which have a direct effect on protein function (Charalambous et al., 2008). To counter these issues, specific detergents must be chosen for the membrane protein of interest, and this is usually achieved through trial and error. The denaturing nature of detergents on membrane proteins was improved slightly using amphiphilic polymers called amphipols (Tribet et al., 1996), however these were still moderately denaturing to sensitive membrane proteins. Recent advances in cryo-electron microscopy however have since allowed for high resolution structure acquisition in lipid environments directly (Liao et al., 2013, Tascon et al., 2020, Lee et al., 2019, Sun et al., 2018). A full review of the uses on detergents for membrane protein study can be found in Seddon et al., (2004).

### **1.2.2. Mixed micelles and bicelles**

Mixed micelles and lipid bicelles are spontaneously formed when lipids and low concentrations of detergents are mixed in the correct ratio and at a certain temperature. The discoidal structures that form can closer mimic the *in vivo* bilayer environment to a higher quality than detergent micelle alone and show some bilayer properties (Seddon et al., 2004). Early protein studies in bicelle systems saw the mixing of long chain lipids such as dimyristoyl-phosphatidylcholine (DMPC) with shorter chain lipids like dihexanoyl-phosphatidylcholine (DHPC) or detergents like 3-[(3-cholamidopropyl) dimethylammonio]-2-hydroxyl-1-propanesulfonate (CHAPSO). This mix allows the DMPC to form a planar bilayer, usually with a depth similar to the length of a transmembrane helix, whereas the shorter chain DHPC lipids or zwitterionic detergent CHAPSO, forms the rim of the bicelle (Sanders and Prosser, 1998, Seddon et al., 2004). Bicelles are perfect for protein structural study once protein is reconstituted into the disc, and are particularly useful for NMR due to their magnetic susceptibility allowing uniform alignment of each reconstituted disc in a sample (Sanders and Prosser, 1998). Further study into whether bicelles were a good membrane mimic for membrane proteins saw the reconstitution of diacylglycerol kinase (DAGK) into various bicelle systems (Czerski and Sanders, 2000) and the successful functionality of the protein was

recorded. Although DAGK activity showed a preference for mixed micelles, or vesicles formed with lipids used for bicelles, significant activity was recorded within each bicelle and thus were shown to exhibit some bilayer properties.

### **1.2.3. Synthetic liposomes and vesicles**

Liposomes are bilayer forming mixes of lipid which are compartmentalised. The size of these particles can be tuned by extrusion, or sonication, and can have an array of different properties based on the chemical make-up of the lipid mix and the resulting mix can have differing effects on properties of the MFS protein, LacY (Findlay and Booth, 2017). Mixes of three lipids were used to obtain different bilayer properties. DOPC, a lamellar neutral lipid, was mixed with DOPE, a non-lamellar lipid which increases lateral pressure, and bilayer curvature, or DOPG - a negatively charged lipid which provides a net charge in the global bilayer. An extensive study of reconstitution efficiency, thermostability as assessed using CD or SRCD, topology using cysteine loop-labelling, and function with active o-NPG transport across the bilayer was assayed in varying lipid conditions. The increase in DOPG increases reconstitution efficiency, but has detrimental effect on folding, promoting an inverted topology hence abolished function at mole fractions above 0.4. Topology experiments here were consistent with early studies of LacY in PE deficient cell strains where LacY was shown to exhibit an inverted structure of the N-terminal domain. This was later found to be induced by the increased negative charge at the membrane (Bogdanov and Dowhan, 2012, Bogdanov et al., 2014, Dowhan and Bogdanov, 2011), and mutagenesis of the proteins charged residues corrected this topological shift and the flipping was attributed to the charge balance rule (Wang et al., 2002).

When DOPE was increased in the tertiary lipid mix, it was shown to be required for stability, correct topology, and active transport, as well as refolding from a urea denatured state, but when too high, PE limits folding and reconstitution yields (Findlay and Booth, 2017). Earlier work showed a similar effect of the DOPE lipid on GalP, another MFS transporter (Findlay et al., 2010). The highest refolding yields for GalP and LacY in (DOPE/DOPC bilayers) occurred at DOPE molar fractions of 0.5-0.6 and 0.2-0.4 for GalP and LacY respectively. This provided further evidence for the notion that increased lateral chain pressure increase protein stability and testing this hypothesis can be achieved using synthetic liposome mixes.

### **1.2.4. MSP-based nanodiscs**

A synthetic lipid bilayer surrounded by a ring of amphipol or membrane scaffold protein (MSP) were the first membrane nanodiscs used for protein study (Denisov et al., 2004). Many

complex proteins such as ion channels (Xu et al., 2015) and cytochrome P450 were reconstituted into these discs (Luthra et al., 2013), and have been particularly useful for structural studies of membrane proteins using high resolution techniques such as NMR, CryoEM and X-ray crystallography, of each protein in a functionally relevant environment (Seddon et al., 2004). The major downside of using MSP-based nanodiscs is their high spectroscopic absorption, which is similar to the protein of study due to the proteinaceous nature of MSP, thus hindering UV or visible light spectra acquisition for structure, or functional assays of the protein. Reconstitution of protein into MSP-nanodiscs still requires the need for detergent and synthetic lipids, with DMPC as the most used lipid. DMPC however is not found in *E. coli*. Synthetic lipid mixes can be used to form MSP nanodiscs, however these do still not recapitulate the correct chemistry of the *in vivo* environment but do allow experimentation into the effects of certain lipid properties (charge, tail saturation and size) on the protein in this environment.

#### **1.2.5. Polymer-based nanodiscs**

More recently, new developments into the use of styrene-maleic acid (SMA) co-polymers have been able to offer a solution to overcome many of the issues highlighted above, acting as a platform for solubilising membrane proteins while enabling the membrane protein to be extracted from its native lipid environment to form a 'native nanodisc' from the cells' intrinsic lipid mix (Reading, 2018, Postis et al., 2015, Reading et al., 2017, Lee et al., 2016, Knowles et al., 2009, Haffke et al., 2020, Lemieux and Overduin, 2021). This process retains any associated lipids that may play pivotal roles in stabilisation, or in protein functionality. These co-polymers form stable SMA lipid particles (SMALPs), encapsulating a protein of interest in a near native environment, allowing for further investigation and to help bridge the gap between studies of membrane proteins *in vitro* and *in vivo* (Hesketh et al., 2020, Kopf et al., 2020, Swiecicki et al., 2020) and exhibiting an environment of improved stability for membrane proteins to reside in, and do not require excess detergent which can cause difficulties when performing structural studies (Ganapathy et al., 2020, Pollock et al., 2018).

The original SMA polymer consists of styrene, the hydrophobic residue, and maleic acid as the hydrophilic residue which act to solubilise the membrane through partitioning into the bilayer spontaneously, forming the discoidal lipid particle which is stabilised by the polymer (Orwick et al., 2012). The ratio and order of styrene to maleic acid can also vary, resulting in different sized nanodiscs, with the common being SMA(2:1) and SMA(3:1) (S:MA). The size of the discs

produced can also be varied from 10-30 nm by altering the polymer to lipid ratio (Craig et al., 2016).

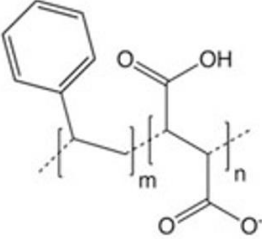
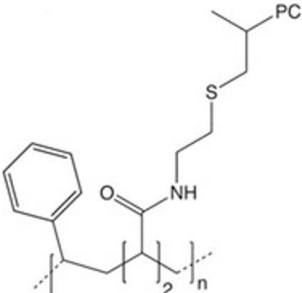
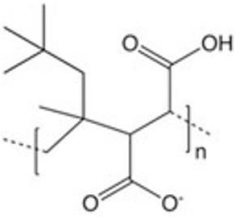
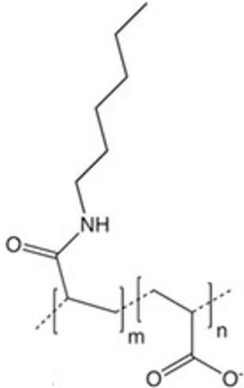
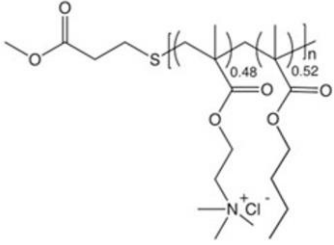
Although investigations into optimal conditions have been performed, SMA is not an ideal polymer for every protein. SMA is unstable at low pH and can precipitate in the presence of high divalent cations (Lee et al., 2016), which can make this polymer unsuitable for proteins which require these specific co-factors or environment for their structure or function. Alternative co-polymers such as the diisobutylene-maleic acid (DIBMA) (Oluwole et al., 2017a), and polymethacrylate (PMA) (Parmar et al., 2016) are shown to solubilise membranes similarly to SMA, but with a milder effect on acyl chain order resulting in a more physiological environment, and both lack the styrene moiety and therefore do not hinder spectroscopic experiments and are also much more tolerant to the presence of divalent cations (Barniol-Xicota and Verhelst, 2018). Additionally, SMA derivatives have been developed to increase stability in low pH, or to tune the size of nanodiscs (Ravula et al., 2018). However, a solution that suits all proteins and all applications has not yet been developed (Pollock et al., 2018, Oluwole et al., 2017b, Lloris-Garcera et al., 2020, Kopf et al., 2020, Ganapathy et al., 2020).

An overview of common polymers and associated properties which have been used to isolate membrane proteins from their native lipid environment is shown in **Table 1-2** with information taken from (Overduin and Esmaili, 2019) and [www.smalp.net](http://www.smalp.net).

More detail on polymer based nanodiscs and the use of DIBMA, can be found in **Chapter 4**.

**Table 1-2: Overview of polymers used for native nanodisc production**

\* SMA-EA: ethanolamine addition, broader range of pH, temperature and salts tolerated; SMA-ED: ethylenediamine addition, solubilises in pH <5 and >7; SMA-SH: Amino-mercapto-ethane addition for dye or tag labelling; SMAd-A: dehydrated (maleimide) SMA-ED, solubilised in pH <6; SMA-QA: addition of quaternary ammonium (positively charged SMA), solubilises at pH 2-10 and tolerance of high metal cations; SMI: positively charged SM, similar to SMA-QA.

Polymer	SMA Styrene-maleic acid						zSMA Zwitterionic co-polymer	DIBMA Diisobutylene-maleic acid	APAA Alkyl polyacrylic	PMA Poly(methyl acrylate)
Structure										
Subunit ratio	3:1, 2:1, 1:4:1, 2:3:1						N/A	1:1	1:1	1:1.1
Modifications from SMA	SMA-EA*	SMA-ED*	SMA-SH*	SMAd-A*	SMA-QA*	SMI*	SMA, lacking maleic acid	Replace styrene with diisobutylene	Replace styrene with Butyl, pentyl, or hexyl sidechains	Replace styrene with hydrophobic, cationic chains
References	(Knowles et al., 2009, Overduin and Esmaili, 2019)						(Overduin and Esmaili, 2019, Fiori et al., 2017)	(Oluwole et al., 2017a)	(Hardin et al., 2019)	(Parmar et al., 2016)

### **1.3. *In vivo* and co-translational folding studies of membrane proteins**

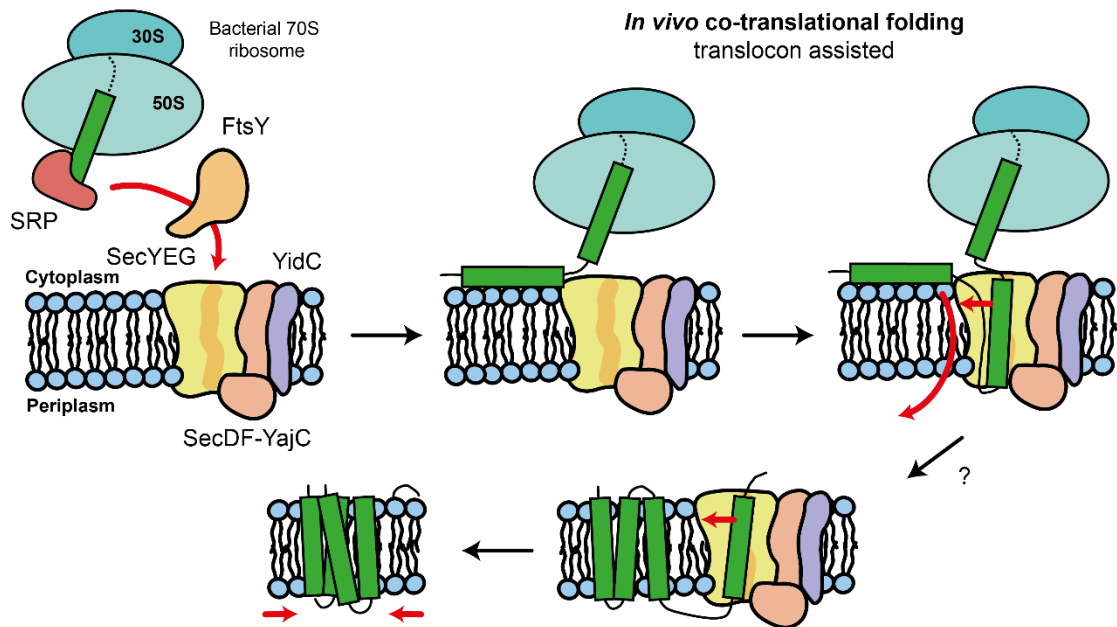
Much of this section (pages 31-40, and **Figure 1-4** and **Figure 1-5**) was previously published as a mini-review article in *BBA Biomembranes*. doi: [10.1016/j.bbamem.2019.07.007](https://doi.org/10.1016/j.bbamem.2019.07.007), of which I am the first author and have the rights to reproduce here. The full paper can be found in **Appendix 9**.

#### **1.3.1. Introduction to co-translational folding**

The majority of *E. coli*  $\alpha$ -helical nascent chains follow the Sec-dependent pathway of insertion, and are captured by the signal recognition particle (SRP) and delivered towards its receptor FtsY for the binding of the ribosome nascent chain (RNC) to the SecYEG insertion apparatus (Gold et al., 2007). SecYEG with associated chaperones **Figure 1-3** is often termed the holotranslocon (HTL); this larger complex is composed of SecYEG-SecDFYajC-YidC (Komar et al., 2016) and acts as a protein channel, translocase and insertase, and protects the nascent chain from aggregation as it begins to fold on the cytoplasmic side of the membrane (Dowhan and Bogdanov, 2011, Harris et al., 2017b). In eukaryotic organisms, a hetero-trimeric complex Sec61, of three subunits ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), is essential for both co-translational insertion of membrane proteins and translocation of secretory proteins into, and across the endoplasmic reticulum membrane for cellular trafficking. Sec62 and Sec63 bind during a post-translational event in yeast, and most likely other eukaryotes, to ratchet peptides through the Sec61 channel with the aid of BiP, an ATPase of the Hsp70 family (Rapoport, 2007). This mechanism is similar to protein secretion mechanisms across the bacterial inner membrane, however SecA pushes the peptide through the channel in an ATP-dependent manner (Gold et al., 2007). Sec61 also associates itself with chaperoning proteins in higher-order complexes like translocon-associated membrane protein (TRAM), translocon-associated protein (Trap) and oligosaccharyltransferase (OST), which together are required for the folding of glycosylated proteins in eukaryotes (Rapoport, 2007).

At the *E. coli* membrane, in addition to the function of the HTL **Figure 1-3**, lipid interactions influence membrane insertion (Harris et al., 2018, Findlay and Booth, 2017), with lipid headgroup charge and interactions with the nascent chain affecting TM insertion efficiency and topology (Bogdanov et al., 2014). The bulk properties of the lipid bilayer, such as its lateral pressure and hydrophobic thickness, also likely impact on membrane protein insertion and folding (Harris et al., 2018). Thus, the cell membrane ‘arena’ required for membrane protein

folding adds increasing complexity to the already intricate co-translational folding pathway of proteins and therefore this native environment should also be studied when exploring membrane protein folding.



**Figure 1-3: In vivo translocon assisted co-translational folding**

Membrane protein folding is much more complex in vivo where several insertase proteins and chaperoning machinery aid in the folding of polytopic  $\alpha$ -helical proteins. A translating ribosome with emerging nascent chain is first captured by SRP and with the aid of FtsY, is delivered the holo-translocon, comprising of SecYEG, SecDF-YajC, and insertase YidC. It is generally accepted that helices formed in the ribosome exit tunnel first bind to the lipid headgroups, before the next emerging helix enters the SecY channel, where it then exits the lateral gate and partitions through the lipid bilayer. The mechanistic steps of this process are largely unknown, but a concert of chaperoning machinery and lipid dynamics act together to fold the protein into its final structure.

TMs of polytopic membrane proteins are generally assumed to insert into the membrane sequentially, however more complex scenarios suggest that insertion can also occur as two-helix hairpins which are assembled close to the membrane or even in the vicinity of the Sec61/SecY channels before their coordinated release into the membrane (Hermansson and von Heijne, 2003, Pitonzo and Skach, 2006, Lu et al., 2000, Foster et al., 2000). The helical structure can form in the ribosome exit tunnel, as shown by Förster resonance energy transfer (FRET), gel-shift assays and structural studies (Fedyukina and Cavagnero, 2011). A long standing model of translocon function proposes that transmembrane (TM) helices insert into the bilayer via a channel and lateral gate in the translocon (Cymer et al., 2015b). This model is supported by chemical-crosslinking studies whereby stalled nascent chains were found to crosslink to both the translocon and surrounding lipids, in isolated ER-microsomes (Martoglio et al., 1995). A recent hypothesis suggests that TM helices do not necessarily enter the SecYEG



channel, but slide down the outside of the translocon, inserting via lipid head groups before the hydrophobic interior of the membrane provides the driving for insertion of the TMs (Cymer et al., 2015b). More polar regions of the protein may use the translocon channel to insert into or cross the bilayer. The chemistry of the lipid head groups, packing of hydrophobic tails and overall lipid bilayer properties are important for the insertion of nascent TM helices in either model.

Smaller and less complex proteins have been observed to insert with only YidC-type insertases in a 'Sec-independent' manner. Substrates for YidC, or Oxa1 of mitochondria or Alb3/Alb4 of chloroplasts, are likely to favour smaller proteins with one or two helices (Dalbey et al., 2014). The precise function of the YidC-like proteins is currently unknown, with only a direct catalytic activity observed thus far when reconstituted *in vitro* (Serek et al., 2004). Mitochondria do not contain the Sec machinery and only contain Oxa1, suggesting that direct contact with lipids is essential for MP insertion and spontaneous folding can occur with the insertase partitioning charged/highly polar residues across the membrane (Cymer et al., 2015b). However, Oxa1 may cooperate with the TIM23 machinery and also Mba1 during membrane protein insertion, forming a larger insertion complex like HTL in *E. coli*, therefore it remains unclear whether YidC/Oxa1 act as standalone insertases, or remain as a small unknown part of a whole (Wang and Dalbey, 2011)

Membrane protein co-translational folding and insertion has largely been studied using *in vitro* transcription/translation (IVTT). This *in vitro* approach enables control of protein translation within simplified cell and membrane extracts, or purified components such as the PURExpress® system (Kuruma and Ueda, 2015) and synthetic lipid mixtures. Assorted investigations of the inserted state have been undertaken using biochemical assays such as epitope binding, limited proteolysis, and chemical cross-linking (Ismail et al., 2015, Ismail et al., 2012, Dale et al., 2000, Dale and Krebs, 1999, Harris et al., 2017b, Ataka et al., 2013). Moreover, structure formation during co-translational folding has been probed by infra-red spectroscopy (Harris et al., 2017b). These studies have led to exceptional insight into the co-translational insertion of membrane proteins; however, these translation systems are often low yielding and expensive.

### **1.3.2. Biochemical methods to study co-translational folding**

Early biochemical techniques to ascertain a time-courses for the order of polytopic TM helix insertion *in vivo*, used single-cysteine mutants amenable to labelling and were observed by a

phosphorescent gel shift. Membrane impermeable 4-Acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid disodium salt (AMS) was used alongside <sup>35</sup>S-Methionine pulse-chase labelling to determine the insertion rates of radiolabelled bacterioopsin (bO) (bacteriorhodopsin without retinal co-factor), a well-studied  $\alpha$ -helical archaic proton pump of *Halobacterium salinarum* with each helix labelled A-G. A cysteine located near the N-terminus showed a translocation rate much slower than the translation of isolated protein, as determined by AMS derivatisation at certain time points. This suggests that the insertion of the first TM helix occurs co-translationally (Dale and Krebs, 1999) and acts as a folding 'nucleus' for the rest of the protein. This was later advanced, and each of the extracellular loops in bO were also labelled and pulse chased. The BC and FG loops of bO are translocated in sequence order after the first helix (Dale et al., 2000, Dale and Krebs, 1999), but the FG loop, unlike the BC, completes its translocation after full length protein elongation is complete, and therefore must occur post-translationally. These results suggested a sequential order of co-translational TM insertion for bO in *H. salinarum*. Direct structure formation, however, cannot be characterised using this approach.

Co-translational folding of eukaryotic polytopic proteins has been studied using protease digestion to assess proteins regions outside the membrane in the cytosol. Truncated polypeptides of the 6 TM human Aquaporin-1 (AQP1) water channel (Lu et al., 2000) were fused into protease sensitive reporter constructs at TM helix-connecting loops via epitopes, which allowed determination of topology for the truncated proteins at the endoplasmic reticulum in canine pancreas rough microsomal membranes with rabbit-reticulocyte lysate (RRL). The location of protease reporter in the cytosol or ER lumen indicated the orientation of TM helices. It was determined that 4 of 6 helices initially transverse the membrane, with helices 2 and 4 binding the lipid interfacial region. A late-stage maturation process appeared to flip TM helix 3 from an N-out orientation 180° through the membrane to N-in orientation, in turn pulling helices 2 and 4 into the bilayer. This is a co-translational process and rotation of TM 3 increases as relative C-terminal helices are released from the ribosome (Lu et al., 2000). Other AQP1 TM segments appear to transiently insert into the bilayer, these regions may remain in the translocon before integration adding yet another variable to the folding pathway. The pathway for AQP1 is much more complex than the sequential insertion studies of bO. However, truncated protein chains were used to ascertain topology in AQP1, which may have given the NC extra time to alter its folded conformation before measurements have taken place. In addition, bO studies were time resolved and therefore offer a more direct

measurement of co-translational folding and topology that epitope labelling can. Yet, neither study can assess structure formation during co-translational insertion nor actual folding of the protein. In other work, lipid composition has been shown to influence topology with altered *E. coli* lipids (Bogdanov et al., 2002). Notably lack of the main lipid phosphatidylethanolamine (PE), causes an inversion of a 6 TM domain of the 12 TM lactose permease in the membrane, which can be rescued on addition of PE.

Epitope protease assays have also been used to show that the helices themselves contain topological information in the form of signal-anchoring sequences, which open the translocon allowing translocation of loops into the ER lumen and result in an N-out helix orientation, and stop-transfer sequences, which close the translocon keeping the connecting loops on the cytosolic side of the ER membrane (Moss et al., 1998). The strength of stop-transfer activity, as a result of hydrophilic/phobic residues in helices, and loop length, can vary resulting in extensive topology and diversity among a range of proteins. The 6 TM helix AQP4 protein regulates the co-translational folding of its final structure in this way unlike AQP1 (Foster et al., 2000), where there is no co-translational helix rotation.

As NCs are elongated, the environments of the ribosome exit tunnel and the cytosol influence folding. Incorporating fluorescent or photo-crosslinking labels into the nascent chain, changes in the peptide conformation as it is produced can be monitored. Unnatural amino acid (UAA) technology uses modified aminoacyl-tRNAs to incorporate a non-natural amino acid directly into the nascent chain during translation (Johnson, 2005). FRET can occur between a donor and acceptor dye when in close proximity, to produce an observable fluorescence when fluorophore based UAAs are used in translation. The efficiency of this energy transfer is proportional to the distance between the two fluorophores. This can give spatial information between nascent chain positions, and when photo-crosslinking agents are introduced, interactions between the UAA and translocon components are observed (Johnson, 2005, Woolhead et al., 2004).

Single membrane-spanning TM peptides appear to fold near the ribosome PTC to produce compact  $\alpha$ -helices. This is shown by a high FRET efficiency. The compact nature is lost as the NC is released into the exit-tunnel, characterised by a reduced FRET efficiency. Together, this shows that the ribosome itself can play a role in the initial folding strategy for TM spanning proteins. This pathway was proven to be orchestrated by specific ordered ribosome/TM interactions using photo-crosslinking reagents. The TM segment was shown to crosslink with three eukaryotic ribosomal proteins (Woolhead et al., 2004), whereas non-TM helix sequences

crosslinked to only one. This suggests that the ribosome contains specific sequence selectors for TM helices, with the first acting as an initial sensor for TM segments, to provide a hydrophobic nucleation point for folding (Woolhead et al., 2004), indicating that folding pathways are dependent on sequence, and protein identity. These interactions further show that specific sequences in the TM-helix primary structure can control the entry of the helix into the translocon much like the epitope labelling work, however UAA has only been used to study single-membrane spanning proteins with the overall picture of insertion of polytopic proteins still to be uncovered. Certain fluorescence dyes coupled with UAA technology can also give insightful information into its hydrophobic microenvironment due to a red or blue shift in its emission maxima. This was nicely used to uncover minor structural alterations in the TIM23 mitochondrial channel when subject to a proton motive force (PMF) using wheat-germ extract, an alternative cell-free expression system sometimes used for eukaryotic protein expression (Alder et al., 2008).

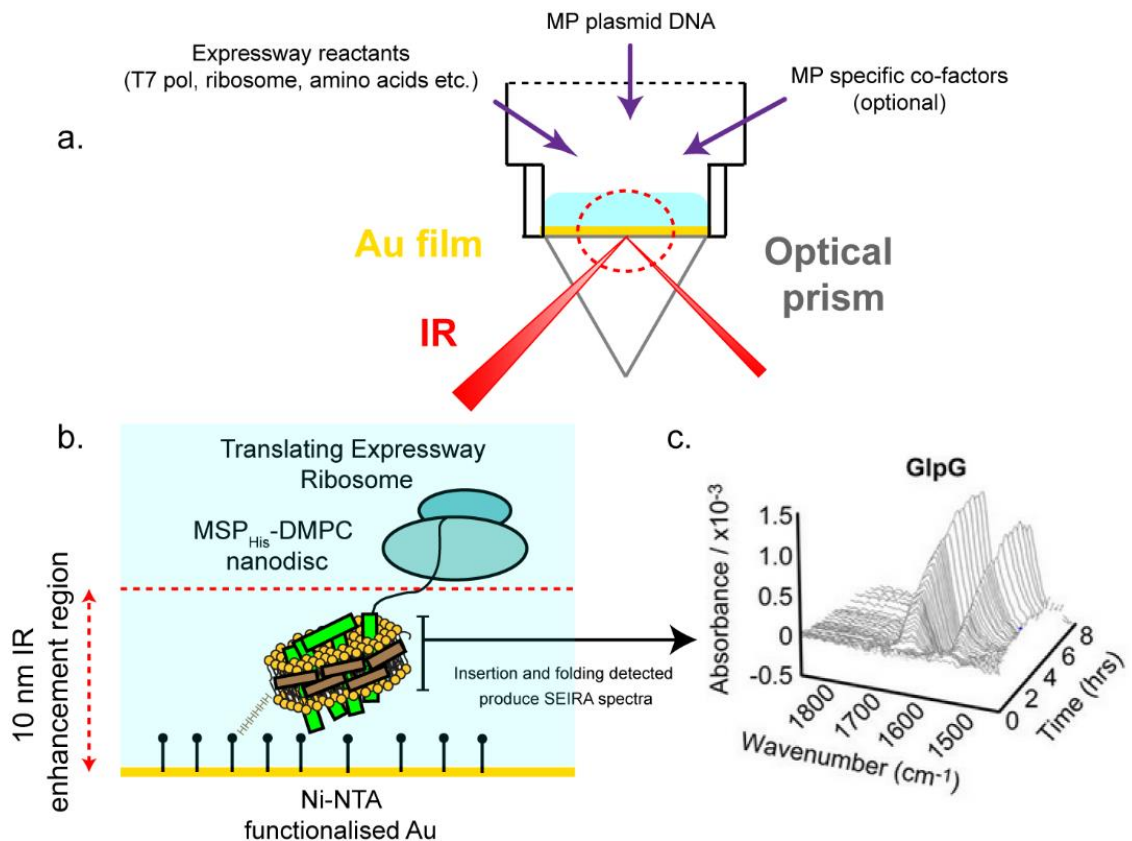
Use of translational arrest peptides (APs) can be applied as transplantable *in vivo* force sensors to measure forces acting upon a nascent polypeptide chain during translation and to study the kinetics of insertion and folding. The SecM arrest-peptide (see **section 1.5.2.2**) has been used as an *in vivo* force-sensor by the von Heijne group to give insight into forces acting on the co-translationally inserting and folding segments of the NC (Niesen et al., 2018, Goldman et al., 2015). The force of NC release is related to a measurement of the fraction of truncated protein per fully translated protein ( $f_{FL}$ ), and plotted against helix number, and a force-profile for the particular protein is determined. This methodology seeks to describe the force (or  $f_{FL}$ ) acting on a hydrophobic segment at varying distance from the AP, most likely describing interactions arising from NC insertion into the translocation machinery and TM helix partitioning into the membrane.

It was hypothesised that the detected force experienced by hydrophobic domains increase when the upstream TM helices fold and partition into the membrane, possibly signifying interactions between helices in the co-translational folding process (Ismail et al., 2012). Polytopic membrane proteins; CaiT, NhaA, EmrD, BtuC and GlpT, were cloned into the AP construct (Cymer and von Heijne, 2013). The apparent free-energies of protein insertion ( $\Delta G_{app}$ ) were predicted using  $\Delta G$ -predictor (Hessa et al., 2007) and of each candidate, one TM with a positive predicted  $\Delta G_{app}$  was selected suggesting unfavourable helix insertion. The interactions between the helix and the upstream N-terminal helices were quantified by an  $f_{FL}$  value, and show that an increased hydrophobicity subsequently increased insertion, as did the

presence of previously synthesised upstream helices. The shorter the loop length between the helix and the prior helices also increased  $f_{FL}$  signifying the reason for the short loop length in polytopic membrane proteins. The exact identity each of these interactions themselves however cannot be determined, nor can they give us a time resolved sequence of events of folding and tertiary structure formation on the growing polypeptide.

### **1.3.3. Time-resolved studies of co-translational folding**

A recent approach to directly observe temporal formation of NC structure, during co-translational folding of MPs into membranes is surface enhanced IR spectroscopy (SEIRAS). The SEIRA setup **Figure 1-4** allows for IR spectra to be collected in the field of IR enhancement, resulting from plasmonic resonance at a thin gold surface deposited on a silicon prism. This gives a 10-100-fold increase in sensitivity compared to conventional IR spectroscopy (Ataka et al., 2013). There is a 10 nm enhancement region, which allows spectra to be obtained only of uniformly oriented lipid-nanodiscs and protein within the nanodisc, but not the translating ribosomes nor anything outside the membrane. The IR amide I (C=O stretching vibration) and II band (C-N stretching and N-H bending) intensities give fingerprints of protein structure, and particularly the position and width of the amide I band indicate helix formation packing during co-translational insertion.



**Figure 1-4: SEIRAS set-up and methodology**

a) Reaction chamber and contents. The chamber sits on silicon prism with a thin layer of gold. A 10 nm enhancement area above the gold surface results from plasmon resonance, due a reflection of the IR beam in the gold-prism interface. Plasmid DNA of the MP of interest is added, along with the cell-free kit containing ribosomes, T7 polymerase and ‘feed’ mixture with amino acids and energy replenishment components. Here, any co-factors required for MP folding (such as retinal for bR) are added. b) The IR beam illuminates a region up to 10 nm where a uniform layer of MSP-DMPC poly-histidine tethered nanodiscs sit on the Ni-NTA functionalised gold. The ribosomes translate the plasmid DNA producing NCs, which spontaneously insert into the nanodisc allowing in situ measurements of co-translational folding. c) Schematic SEIRA spectra output showing amide I and amide II peak intensity, which increase over time, and yielding kinetic information on folding.

The first MP SEIRAS co-translational study was carried out on bacteriorhodopsin (bR), using a commercial cell-free kit with bR DNA, retinal cofactor, and DMPC-based MSP-nanodiscs. Both amide I and II bands increased as a function of time after the reaction started until a significant shift in peaks suggested that membrane insertion had occurred. In the late stages of folding, tertiary structure formation was observed with a sharpening of the amide I band, characteristic of  $\alpha$ -helical bundling (Baumann et al., 2016). No structure formation was observed when retinal was absent **Figure 1-5**.

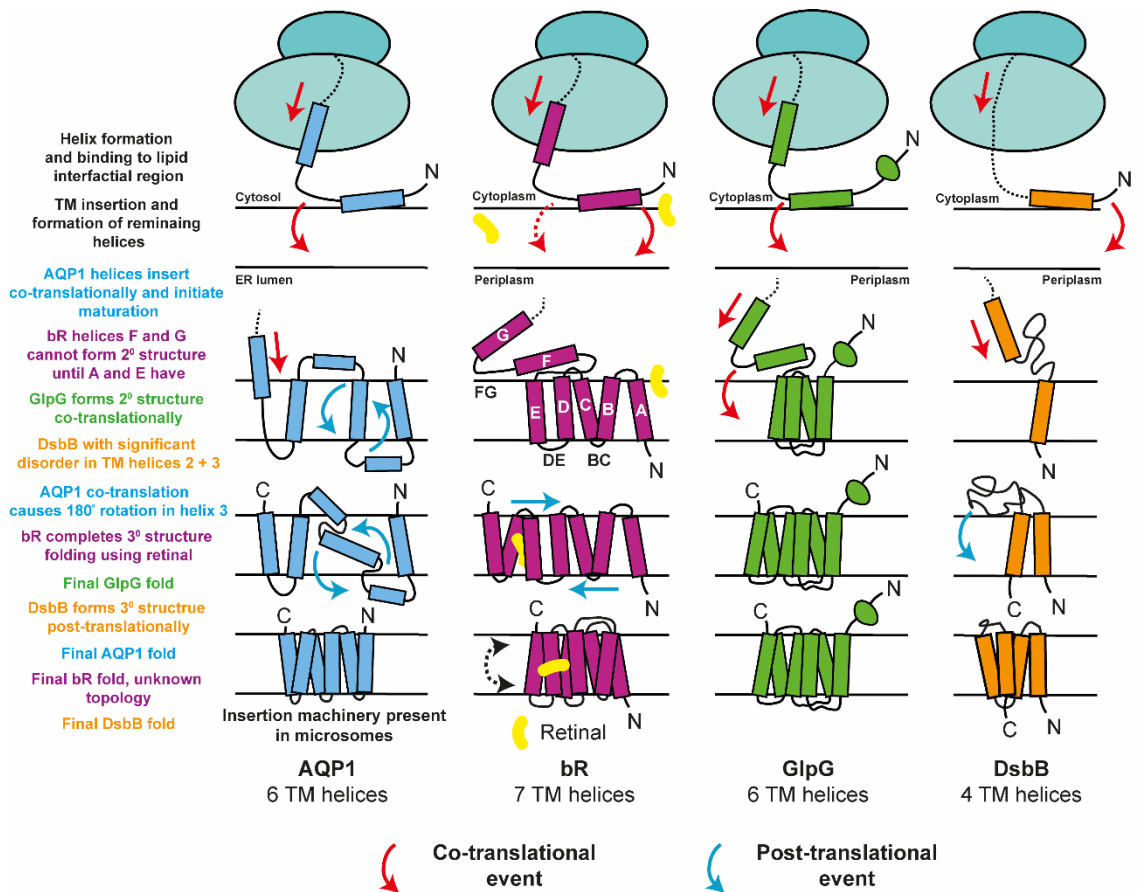


Figure 1-5: Possible models for co-translational folding in the absence of translocon from SEIRAS

Summary of possible models for protein folding and insertion using SEIRAS data for bR, and SEIRAS with crystal structure and MPEX predictions for GlpG and DsbB adapted from (Harris et al., 2017b). AQP1 comparison from protease reporter work in rough-ER microsomes (Lu and Booth, 2000). Topology for bR in this cell-free system is unknown, the topologies of GlpG and DsbB depicted were confirmed by AMS labelling in liposomes. For each protein, helices are co-translationally (red arrows) produced and bind to the membrane interfacial region. AQP1 undergoes a co-translational maturation step where TM<sub>3</sub> flips in the membrane 180° to produce the final folded structure. Insertion machinery may aid this process in the system used. Truncated proteins were used in the AQP1 study, this may have given the NC more time to rearrange their folding pathway. bR forms helices A-E co-translationally but F and G cannot fold until A and E are correctly inserted (loops labelled for **section 1.3.2**). A post-translational (blue arrows) event then forms the final 3D structure of bR using retinal included in the reaction. GlpG forms secondary and tertiary structure co-translationally. DsbB Regions corresponding to TMs 2 and 3 in DsbB are not predicted to insert suggesting that the unordered regions in SEIRA correspond to TMs 2 and 3, and they insert only insert post-translationally after TMs 1 and 4. Most likely, helices 2 and 3 favour a disordered state at the DMPC interface rather than partitioning through the bilayer in the absence of any translocon machinery.

More recently, cell-free expression of the *E. coli* membrane proteins rhomboid protease GlpG and disulphide reductase DsbB were followed by SEIRAS. As with bO, no translocon was present and thus the observed folding was spontaneous and efficient, unassisted by translation apparatus (Harris et al., 2017b). For GlpG, the earliest detected IR bands corresponded to  $\alpha$ -helices and  $\beta$ -structure, and disordered structure. The latter then decreased as more  $\alpha$ -helical structure was produced.  $\alpha$ -helical bundling between TM helices occurred (Baumann et al.,

2016, Harris et al., 2017b) as GlpG folds into its polytopic structure. These structural changes all occurred within the time required to make the full-length protein, strongly indicating that all folding occurred co-translationally.

DsbB co-translational folding was slower with more disordered structure initially, and helix formation and packing occurring both co- and post-translationally (Harris et al., 2017b). The observed structure formation during folding of both GlpG and DsbB was consistent with hydropathy plots; all 6 helices of GlpG are predicted to be stable by hydrophobicity scales for insertion into the lipid headgroups as well as via the translocon (Snider et al., 2009). In contrast, TMs 2 and 3 of the 4 TM DsbB are predicted to be unstable, and thus could initially form disordered structure and require TMs 1 and 4 already in the membrane for the remaining helices to insert and fold to give the final protein structure. SEIRAS itself, without labelling of the protein, cannot identify which regions of the protein are folding.

Each protein tested using SERIAS so far has inserted into the bilayer in the absence of any translocation machinery suggesting, and confirming previous work (Booth et al., 2001, Curnow and Booth, 2007, Lu and Booth, 2000, Bowie, 2004), that the lipids themselves possess the necessary mechanical energy to modulate lipid binding independent of the translocon. The co-translational folding of GlpG for example was enhanced by certain lipids, including PE which alters the mechanical properties and PG that introduces negative headgroup charge. There is a degree of folded and misfolded structure for GlpG, DsbB and bR in their early stages, and it may be that addition of a translocon would reduce the amount of misfolded protein.

#### **1.4. *In vitro* folding studies of membrane proteins**

Studying co-translational folding pathways of membrane proteins is a relatively new field. Much of the existing research into membrane protein folding stems from *in vitro* folding work, where proteins are purified and reconstituted into a membrane mimetic and forced to unfold or re-fold in an artificial system to generate thermodynamic and kinetic properties of folding. A summary of methods previously used to do this are outlined in this section.

##### **1.4.1. *In vitro* folding of $\alpha$ -helical membrane proteins**

Early models of membrane protein folding were thought about from a thermodynamic perspective, as a two-stage process (Popot and Engelman, 1990), with the first stage describing the formation of  $\alpha$ -helical secondary structure and subsequent transitioning of the  $\alpha$ -helix



across the bilayer, and the second stage describing the interaction of these helices inside the bilayer to form the final folded structure.

The folding of membrane proteins into the bilayer environment is a complex process, thought to be driven by relative contributions of dominant forces. These forces arise from hydrogen bonding in the secondary and tertiary structure of the protein and transient interactions of van der Waals forces between elements of secondary structure, and any prosthetic groups which may be necessary for folding, but also the hydrophobic effect, which causes the collapse of the secondary structures into their final three-dimensional structure as water molecules expel the non-polar helices (White and Wimley, 1999).

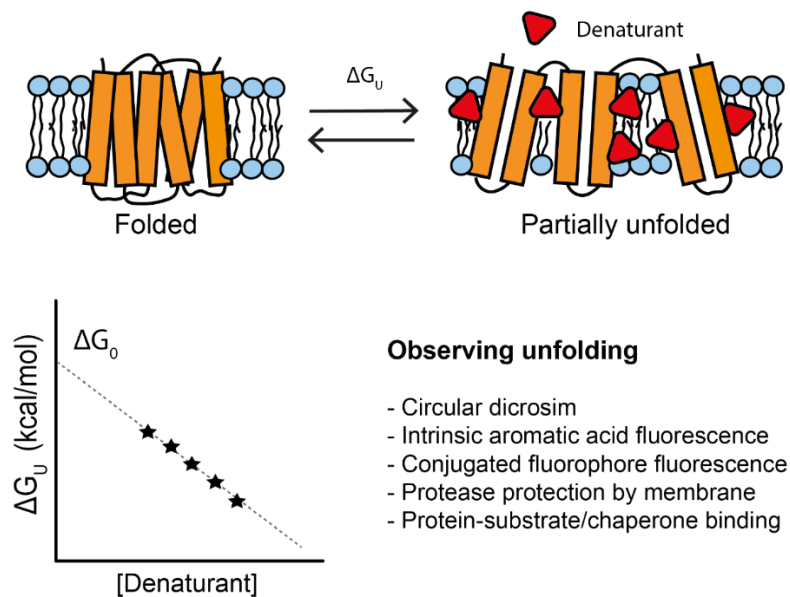
Hydrophobic collapse is perhaps the most dominant driving force for membrane protein collapse when studying *in vitro*, but of course *in vivo*, this process is more complex with co-translational folding and chaperone assistance, which the two-stage model does not capture. One example of this is the initial formation of  $\alpha$ -helical secondary structure, which usually occurs within the ribosome exit tunnel *in vivo*. Therefore, the two-stage model, although elegant, is overly simplistic model of membrane protein insertion and folding, and many mechanistic details are lost.

#### 1.4.2. Artificial denaturation

*In vitro* bulk-ensemble unfolding of membrane proteins in synthetic membrane mimics were initially used to study the folded, and unfolded states of non-denaturing detergent purified and reconstituted samples, or through membrane extracts containing overexpressed protein. These purified proteins can be reconstituted into mixed detergent lipid micelles or bicelles which provide the hydrophobic TM spanning region for the protein. These proteins can be forced to unfold with increasing concentrations of chaotropic denaturants **Figure 1-6** like urea, guanidine hydrochloride (GuHCl) or sodium-dodecyl sulphate (SDS), or thermal denaturation. The unfolding of these proteins can be followed using intrinsic tryptophan fluorescence (Booth et al., 1995), retinal fluorescence (in the case of bacteriorhodopsin) (Booth and Farooq, 1997), or circular dichroism (Riley et al., 1997) to assess the degree of secondary structure loss. The ratio of folded to unfolded protein at each denaturant concentration can be converted into free energy of unfolding ( $\Delta G_U$ ) by extrapolation to a zero denaturant concentration due to the linear dependence on denaturant to unfolding, assuming a two-state process, folded to unfolded through no stable intermediates, is followed, as shown in **Figure 1-6**. This value gives

an indication into the stability of the protein and can be used as a comparison between different mutants, environments, and even different proteins.

SDS appears to be the favoured detergent for membrane transporters experimentally, as the detergent micelle maintains helical structure (Bowie, 2004). Subsequent removal of the chaotrope may allow the protein to re-fold to its native state. The refolding can also be fitted to a two-state process with linear dependence on the free energy of the denaturant and equilibrium unfolding constants along the reaction coordinate can be calculated from the fraction of folded to unfolded protein. This process, however, is complex with many folding intermediates (Lu and Booth, 2000, Curnow and Booth, 2007) which can be difficult to assess by bulk unfolding. These methods have been used extensively to measure the thermodynamics of the protein inserting into the lipid bilayer; protein folding studies of bR for example started with mixed micelles of DMPC/CHAPS and titrations of SDS to reversibly unfold the protein (Booth et al., 2001) and this allowed the determination of the thermodynamic driving forces of bR folding and also kinetics of binding, and role of the retinal co-factor in bR folding.



**Figure 1-6: Denaturant unfolding of membrane proteins**

*Biochemical unfolding: Reversible unfolding of membrane proteins is necessary for the extraction of unfolding kinetics and thermodynamics. Denaturant folding studies titrate a denaturant such as urea or guanidine hydrochloride and follow the unfolding from a reconstituted fully folded protein, to a partially unfolded state in the membrane mimic. Unfolding is observed using a variety of techniques from assaying secondary structure with circular dichroism, or proteolysis assays which assay folded, from unfolded protein.*

More recently chemical denaturation has been used to determine the thermodynamic folding principles and the effect of lipid chemistry on thermodynamic parameters of LeuT, a complex

knotted leucine transporter protein from the hyper-thermophilic *Aquifex aeolicus* (Penmatsa and Gouaux, 2014). Unfolding LeuT using urea in synthetic liposomes showed a significant loss (35 %) of secondary structure was observed using CD when denatured with 8 M urea, this was subsequently refolded by dilution of urea, and approximately 95 % of structure was recovered, and LeuT transport activity was reinstated. Both DOPE and DOPG increases were directly related to an increased thermodynamic stability in liposomes (Sanders et al., 2018).

To further probe unfolding intermediates, bulk unfolding methods such as phi-value analysis can be used to describe the position of a certain amino acid at its transition state to the fully folded protein. This provides another powerful tool to study the kinetics of unfolding, thermodynamics, and energetics of the protein folding. Individual amino acids are mutated to alanine and the above denaturing experiments are performed. The phi-value itself relates the change in activation energy (as the protein folds through a particular transition) to the overall free-energy of folding. A ratio closer to 1 suggests the mutant position is closer to the native structure of the protein (Curnow and Booth, 2009, Curnow et al., 2011).

#### **1.4.3. Single-molecule mechanical unfolding from membranes**

Single-molecule force spectroscopy (SMFS) investigations of membrane proteins have been dominated by atomic force microscopy (AFM) because of its versatility, allowing researchers to image bilayers, individual proteins, and mechanically unfold individual membrane proteins in both synthetic bilayers and native membranes. In these experiments, an AFM tip is adsorbed to the protein and is unfolded stepwise, corresponding to stable sequential segments of the protein unfolding as the cantilever is retracted from the surface (Jefferson et al., 2018). A major advantage to single molecule work is that much lower concentrations can be used, which can prevent aggregation of the sample, and allow hidden-intermediates which may be overlooked in bulk-ensemble methods explored in **section 1.4.2**.

Much of the work previously carried out on membrane proteins observes the small intermolecular forces between stable structural segments between helices or domains and provides a basis for the mechanical studies under proteins in differing biological environments using AFM. A substantial amount of work has been carried out on LacY (Serdiuk et al., 2019, Serdiuk et al., 2015, Serdiuk et al., 2017, Serdiuk et al., 2014, Serdiuk et al., 2016) and NhaA (Kedrov et al., 2004, Kedrov et al., 2007, Kedrov et al., 2006b, Kedrov et al., 2008, Kedrov et al., 2005, Kedrov et al., 2006a) of the MFS family of transporters. Many rhodopsins (Kawamura et al., 2010, Kawamura et al., 2013), including bR (Preiner et al., 2007, Oesterhelt

et al., 2000), have also been studied in their native membranes as an extract, as well as reconstituted nanodiscs and reconstituted synthetic liposomes (Zocher et al., 2012a), where the resulting data can give insight into the refolding/insertion both in the absence and presence of chaperones (Serdiuk et al., 2016, Serdiuk et al., 2017, Serdiuk et al., 2019), ligand (Serdiuk et al., 2014, Bippes et al., 2013), and the effects of surrounding lipid composition on mechanical unfolding and topology (Serdiuk et al., 2015). SMFS can be used to determine key kinetic and energetic parameters that provide insight into the underlying (un)folding energy landscape of the protein of interest.

Single proteins cycle through various intermediates on unfolding, whose position, and corresponding force can change depending on many factors, such as temperature (Janovjak et al., 2003), pH (Damaghi et al., 2010), loading rate (unfolding rate), the presence of ligand (Whited and Park, 2014), oligomerisation (Sapra et al., 2006), bilayer composition, and the vector in which force is applied (Muller et al., 2006, Serdiuk et al., 2015). Therefore, many single trajectories must be averaged for portion of protein unfolded, and the corresponding rupture forces, allowing the most probable unfolding pathways can be mapped onto the structure of the protein.

High-force AFM extraction can assess both stages of two-stage folding, but in a complex way which is difficult to deconvolute. More recently, magnetic tweezers (MT) have been used to study proteins immersed inside a membrane environment, often bicelles. Magnetic tweezers assess only the second stage of folding, while the protein remains in an unfolded state in a bilayer, much like artificial denaturation studies outlined above. The differences in energy based on the two different folding stages can be highlighted using bR studies. In AFM experiments, where helices are pulled from a hydrophobic environment using AFM into an aqueous environment, the unfolding free energy is calculated to be 290.5 kcal/mol (Preiner et al., 2007). This is much higher than the  $\Delta G_U$  of 11.2-12.3 kcal/mol (Chang and Bowie, 2014) when studied with denaturant unfolding in bicelles with the latter corresponding only to second-stage folding, suggesting that the bilayer itself is the major player in defining the folding energies of a transmembrane protein.

An MT set-up can access much lower force intermediates, as well as directing the pulling force in the plane of the membrane which is much more physiologically relevant than extracting a protein out of the membrane under high force. The protein of interest in a lipid bilayer is tethered at both ends, one to the stage, and the other to a magnetic bead, often using long DNA handles as a tether. A magnet is used to carefully manipulate the bead to gradually unfold the

protein, which highlights distinct unfolding intermediates within the bilayer. To date, GlpG and CIC have been studied using bicelles (Min et al., 2015, Min et al., 2018), and more recently GlpG and the Human  $\beta_2$ -adrenergic receptor in liposomes to determine whether the boundaries of the disc influence the forces of unfolding, and to determine the direction of unfolding (Choi et al., 2019).

Further detail into mechanically unfolding membrane proteins using AFM and MT spectroscopy can be found in the introduction to **chapter 6**.

## **1.5. Ribosome-bound nascent chain complexes for co-translational studies**

*In vitro* studies of folding are incredibly useful to understand the driving forces underpinning folding in the membrane. Unfortunately, these systems rely on the reconstitution of a fully translated protein reconstituted into an artificial mimetic. RNC technology can bring together the mechanistic details of co-translational protein folding, explored in **section 1.3**, with the powerful *in vitro* methods to obtain thermodynamic properties of the process explored in **section 1.4**, enabling researchers to overcome the overly simplistic two-stage model by incorporating snapshots of the co-translational process to fully decipher folding pathways during co-translational protein biogenesis.

Further details on the production and uses of RNCs for membrane protein co-translational study can be found in **Chapters 3** and **4**.

### **1.5.1. RNCs as a tool to study protein folding**

For soluble protein co-translational folding studies, many groups have utilised ribosome translational stalling to obtain partially translated, stalled ribosome nascent chain complexes which can be purified from *in vivo* or *in vitro* (cell-free) systems (Cassaignau et al., 2016, Schibich et al., 2016, Rutkowska et al., 2009, Hoffmann et al., 2006, Schaffitzel and Ban, 2007). RNCs are particularly useful for biochemical characterisation of the nascent chain at different lengths, but also for structural techniques, allowing the structure, interactions, and dynamics of the translational machinery to be characterised, and have been a key advance in the protein folding field. Typically, the *E. coli* arrest peptide sequence, SecM, is employed to halt translation (Schaffitzel and Ban, 2007, Rutkowska et al., 2009, Cassaignau et al., 2016), and resulting soluble protein RNCs, often produced using an IVTT system, are amenable to many biochemical experiments. These range from pulse-chase studies and protease protection assays of stably folded intermediates (Jensen et al., 2020, Samelson et al., 2018), to force-

sensor experiments which can determine structurally stable segments where release of stalling indicates folding (see **section 1.3.2**), additionally, time-resolved FRET analysis can be used to obtain folding kinetics (Mercier et al., 2020).

Beyond biochemical studies, *in vivo* production protocols have been developed to produce high yields of sample amenable to structural interrogation, using NMR, to observe the ensemble dynamics of the folding nascent chain as it emerges from the ribosome (Waudby et al., 2013, Cassaignau et al., 2016, Cabrita et al., 2009).

The significant drawback with much of the existing research into membrane protein folding mechanisms is that full length, isolated proteins reconstituted into a synthetic system have been used. These systems are majorly unrepresentative of the vectorial co-translational cellular biosynthesis where the protein folds and inserts as translation occurs and lacks information on key structural events like the formation of helices. It is also unclear how these results can be related to co-translational folding and membrane insertion (Harris et al., 2017b), with the closest real-time folding and insertion experimentation arising from surface-enhanced IR spectroscopy, which has recently followed the cell-free expression and insertion of bR, GlpG and DsbB into MSP-based DMPC nanodiscs (Ataka et al., 2013, Harris et al., 2017b) as discussed in **section 1.3.3**. This lack of knowledge of co-translational folding mechanisms in membrane proteins can be bridged using RNCs, coupled with sophisticated biophysical analysis.

RNCs are a recent and immensely powerful tool which further the possibility to study folding order, and assay folding intermediates at a molecular level for soluble proteins. It has also been possible to generate membrane protein RNCs to explore protein biogenesis at the membrane, and each of these systems are explored in the following sections.

## **1.5.2. An overview of stalling sequences used for RNC production**

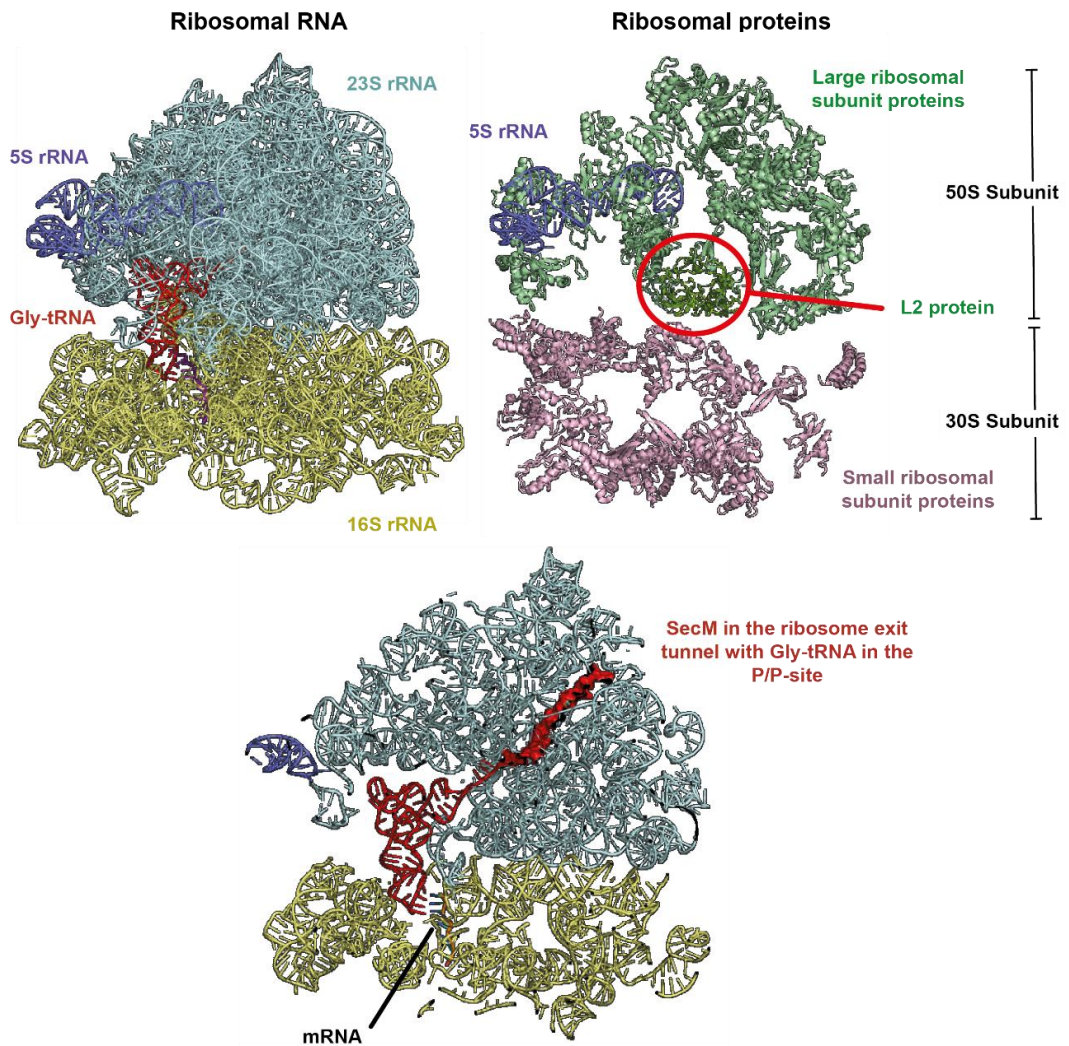
### **1.5.2.1. A brief note on the 70S ribosome**

The ribosome is an incredibly complex and dynamic machine which translates mRNA into protein, through the recognition of codon sequences by aminoacyl-tRNAs, leading to the stitching of certain amino acids together. The 70S bacteria ribosome **Figure 1-7** comprises of RNA components, and proteinaceous components. The 70S ribosome can dissociate into two subunits, the 50S (large) and the 30S (small) subunits (Zhang et al., 2015). The large subunit is made up from 23S and 5S ribosomal RNA, and 31 proteins, of which L2, at the centre of the complex is one of the identifying proteins for RNC work shown throughout this thesis (Hill et al., 1969). The smaller 30S subunit contains one strand of 16S rRNA, and 21 protein subunits,

of which the 30S identifying protein is S1 (see **chapters 3 and 4**). However S1 is often not included in ribosomal crystal structures due its flexible structure (Kutlubaeva et al., 2017).

The ribosome has three RNA binding sites for tRNA: the aminoacyl (A), peptidyl-transferase (P) and exit (E) sites. The A site binds aminoacyl-tRNA, the P site contains the peptidyl-transferase centre, where two amino acids are stitched together to form the protein peptide bond, forming a peptidyl-tRNA (tRNA with two amino acids bound), and the E site (exit site), binds used, free tRNA which is subsequently released from the ribosome. Each tRNA binds to the mRNA strand, held by the 30S subunit, through complementary codon/anticodon pair. This is the mechanism by which nucleic acid is read and translated to protein of correct primary structure.

**Figure 1-7** shows a crystal structure of a glycine-tRNA bound to the ribosomal p-site with the pre-translated SecM sequence in the ribosome exit tunnel at a 3.3 Å resolution (Zhang et al., 2015).



**Figure 1-7: Structure of the 70S ribosome**

The 70S ribosome is made up of rRNA and proteins. The bacterial 70S ribosome can dissociate into the 50S (large) subunit, with 5S (dark blue) and 23S rRNA (cyan) components, and the 30S (small) subunit consisting of 16S rRNA (yellow). A Glycine tRNA (red) is shown here in the P/P-site. Ribosomal proteins hold the structure together. The large subunit proteins are shown in green, and the small subunit in pink. The L2 large subunit protein is highlighted and is referred to throughout this chapter as an RNC quality control protein. The small S1 protein is also referred to; however, this appears rarely in ribosome crystal structures. The SecM stalling sequence has been crystallised as part of the 70S complex. This is highlighted in the slice though the mRNA structure with the peptide in the ribosome exit tunnel, and the glycine tRNA in the P/P site bound to mRNA in red. PDB: 3JBU (Zhang et al., 2015) was used to create this figure.

### 1.5.2.2. SecM peptide stalling

Stalling sequences like SecM, VemP and MifM are used in nature to regulate the translation rate of higher order proteins during co-translational folding (Evans et al., 2005, Nakatogawa and Ito, 2002), and have been incredibly useful when utilised to produce RNC complexes for the study of co-translational folding pathways. The secretion monitor (SecM) is a 170 amino acid protein which monitors protein secretion, and regulates levels of SecA, a cytoplasmic



ATPase which works with the SecYEG translocon for membrane protein insertion, as well as protein export from cells (Nakatogawa and Ito, 2001).

The SecM sequence is commonly utilised as a method to stall translation to produce RNCs, the minimal WT SecM sequence stalling motif was identified through scanning alanine mutagenesis to reveal the sequence: <sup>150</sup>-FxxxxWIxxxxGIRAGP-<sup>166</sup>. It was shown that residues Arg163 and Pro166 are essential for stalling, with the remaining important residues being partially required (Nakatogawa and Ito, 2002). The residues in this primary sequence form specific interactions with the ribosome exit tunnel, to position Gly156 in a geometry that prevents the incoming prolyl-tRNA from creating a peptide bond with glycine – in effect altering the PTC geometry and halting translation resulting in the subsequent stalling of the ribosome along the mRNA (Nakatogawa and Ito, 2002, Rutkowska et al., 2009, Bhushan et al., 2011). Pro166 remains in the A-site, but is not incorporated into the peptide, and the final incorporated amino acid (Gly165) remains in the P-site. Pro166, however still remains essential for stalling (Muto et al., 2006). The key residue comprising this translational stall is Arg163, which forms interactions with the 23S rRNA nucleotide A2062, whilst the other amino acid residues function to stabilise it (Yap and Bernstein, 2009). The signal relayed to the PTC by the interaction between Arg163 and A2062 shifts the ester linkage with the ribosome P-site by 2 Å (Bhushan et al., 2011), preventing nucleophilic attack of the incoming  $\alpha$ -amino group to the carbonyl carbon of the P-tRNA, thus halting translation. This was confirmed using molecular dynamics simulations (Gumbart et al., 2012) and high resolution cryoEM structures which show the conformation of the peptide in the exit tunnel (Bhushan et al., 2011). Stalling in this manner is also aided by the Pro166 in the A-site which naturally undergoes slow peptidyl transfer in comparison to other amino acids (Pavlov et al., 2009).

High resolution CryoEM structures also show interactions between the SecM peptide at other positions throughout the ribosome exit tunnel; namely nucleotides U2585, and U2609 of the 23S rRNA. U2585 interacts with SecM at the Ala164, and U2609 interacts with <sup>160</sup>-QAQ-<sup>158</sup>. The mid-tunnel region rRNA, nucleotide A751, interacts with bulky residues Trp155 and Ile156, further strengthening the stall (Bhushan et al., 2011).

The Woolhead group have shown that movement of SecM in the exit tunnel is key to its stalling. FRET analysis with UAA technology has shown that the C-terminus of SecM is in an extended conformation until the synthesis of the translation arrest peptide which causes the C-terminal of the emerging SecM to become compact, and position the critical Arg163 residue in the correct place for stalling (Woolhead et al., 2006).

The stall of this WT SecM sequence can be overcome when a force is large enough to move the electron deficient carbon into the proximity of proline for the nucleophilic attack to occur (Niesen et al., 2018, Ismail et al., 2012, Ismail et al., 2015, Cymer et al., 2015b, Cymer and von Heijne, 2013, Cymer et al., 2014). Folding downstream peptides outside the ribosome exit tunnel is often necessary to release this stall, in effect re-starting translation, and forms the basis of arrest-peptide folding studies as a means to observe co-translational folding of membrane proteins (Schibich et al., 2016, Ismail et al., 2012, Ismail et al., 2015, Hermansson and von Heijne, 2003, Cymer et al., 2015b, Cymer and von Heijne, 2013, Cymer et al., 2014, Cymer et al., 2015a) (**section 1.3.2**).

Alternatively, to avoid accidental release of the stall motif for structural studies, an arrest enhanced SecM (AE1) often termed SecM Str (strong) in literature; <sup>150</sup>-FxxxxWIWWPRIRAPP-<sup>166</sup> can be introduced. This provides strengthened interactions to the ribosome wall, thus avoiding accidental release of NCs (Cymer et al., 2015a, Kempf et al., 2017), and was also discovered using scanning mutagenesis. The stretch of tryptophan's in the sequence significantly increased the force required to release the nascent chain (Cymer et al., 2015a). This arrest-enhanced SecM sequences was later confirmed to strongly stall ribosomal translation in a cell-free system when measuring ribosomal activity (Kempf et al., 2017).

### **1.5.2.3. TnaC peptide stalling**

TnaC is the leader peptidase of tryptophanase (*tna*) operon, and functions to induce ribosomal stalling to control downstream expression. Similarly to SecM, TnaC can act as an arrest peptide, however, with only 24 amino acids, it is much shorter than the SecM 170 amino acid sequence, and is only inducible at high concentrations of the inducer molecule tryptophan (Gong et al., 2001). Unlike SecM, where translational elongation is prevented, TnaC prevents translation termination (Gong and Yanofsky, 2002).

The sequence: <sup>12</sup>-WxxxDxxIxxxxP-<sup>24</sup> was identified by mutational analysis (Cruz-Vera and Yanofsky, 2008), the Pro<sup>24</sup> residue is situated in the ribosome-P site, where Trp<sup>12</sup> is at the L4/L22 ribosome exit tunnel contraction site, as determined by crosslinking studies between the residue and A<sup>750</sup> of the 23S rRNA. The tryptophan at position 12 is the key residue which alters the PTC allowing the free tryptophan to bind, however it is not clear where it binds, but is suggested to interact near the ribosome A-site (Cruz-Vera and Yanofsky, 2008). Tryptophan binding also blocks methylation of the A<sup>2572</sup> of the 23S rRNA, suggesting it binds in this region (Cruz-Vera and Yanofsky, 2008).

CryoEM studies of the TnaC peptide show the interactions between the TnaC peptide and the ribosome. Unlike SecM, the TnaC peptide remains in an extended confirmation but similarly, the 23S rRNA nucleotides A2602 and U2585, which are located to the PTC adopt distinct conformations in the 70S.TnaC complex, and the U2585 shifts to interact with the essential Pro24 residue (Seidelt et al., 2009). The positioning of these nucleotides in the 70S.TnaC complex prevent the entry and binding of any release factors, and so the positioning of these residues in the PTC act to prevent the efficient hydrolysis and release of the nascent chain.

The TnaC peptide has been used in previous studies for structural interrogation of a secretory protein RNC, in complex with the SecYEG translocon in an MSP-DMPC nanodisc using CryoEM (Kater et al., 2019, Bischoff et al., 2014). This work is described in detail in **sections 1.5.4** and **4.1.4**.

#### **1.5.2.4. Other stalling methods**

There are several other methods of translational stalling which have been used in the literature. These include a depletion of certain amino acids or tRNA from an IVTT reaction, a knockout of a tRNA-synthetase from a cell line used from expression, with the latter being used by the Bustamante group (Desai et al., 2019) to study translational regulation using force and fluorescence techniques. This has been further used in the Booth group to pause and restart translation with the removal and addition of tryptophan part way through the translation cycle. Several lengths of inserted nascent chain are observed when LacY is produced *in vitro* using PURExpress, it was concluded that the length of the emerging nascent chain dictated whether the protein could continue translation when the depleted amino acid is reintroduced to the system (Harris et al., 2020).

Similarly, the truncation of mRNAs can also be used to produce RNC systems for study a technique commonly used for IVTT studies, particularly to study membrane protein biogenesis (see **section 1.5.4**) (Urbanus et al., 2001, Draycheva et al., 2018, von Loeffelholz et al., 2015).

#### **1.5.3. Current studies using soluble protein RNCs**

Co-translational protein folding is an essential process which provides timely and efficient production of new proteins in their functional state following synthesis by the ribosome, with a concerted effort by cellular chaperones to prevent off-pathway aggregation, misfolding and subsequent degradation. The ribosome exit tunnel gradually releases a nascent polypeptide, which folds in a vectorial fashion before the C-terminus has fully emerged. The basic energetic and thermodynamic principles underlying this complex process have been explored using a

wide range of techniques for soluble protein folding. Studies have focussed on secondary structure formation in the exit tunnel which are modulated by space restrictions and physiochemical properties; through to the emergence including the varied rate of peptide synthesis and ribosome modulation of folding (Cassaignau et al., 2020), as well as chaperone binding and function on the nascent peptide fold, extensively reviewed in (Waudby et al., 2019, Kaiser and Liu, 2018, Liutkute et al., 2020).

#### **1.5.3.1. Biochemical studies**

Several avenues of biochemical research have focussed on mapping the correct folding pathways of nascent chains. Experimentally measured rates of spontaneous folding of globular proteins can be measured on the microsecond to hours timescale, which can be followed biochemically, due to the relatively slow peptide elongation of up to 20 amino acids per second in bacteria and even slower in eukaryotes (Young and Bremer, 1976). The ribosome can speed up or slow down translation to modulate protein folding. This is often dependent on the identity of the codons recognised. Slowing down translation allows exposed nascent chains to fold more efficiently, whereas when sped up; most of the protein is translated and emerges from the ribosome before folding. Overall, the electrostatic environment, restricted space and speed of translation all contribute to the correct funnelling of the protein down it is correct energy landscape to prevent misfolded or aggregated protein (Liutkute et al., 2020).

Cysteine labelling at strategically placed residues was able to yield structure dependent disulphide bond formation to reveal distance dependent destabilisation of the nascent chain by the ribosome, which was similar for a set of soluble proteins (Hoffmann et al., 2012). This destabilisation effect was amplified by trigger factor binding, suggesting a mechanism by which the chaperone cooperates with the ribosome to modulate nascent chain binding.

More recently, a refined version of limited proteolysis which has been used with FLuc folding carried out by the Hartl group (Frydman et al., 1999), has been used by the Marqusee lab to quantify nascent chain destabilisation caused by the ribosome (Samelson et al., 2016). This technique used a short pulse, lasting much less than the mean lifetime of a folded protein state, of a high concentration of proteinase which digests the population occupying the unfolded state only, leaving the folded protein protected. This technique has been carried out on DHFR (Samelson et al., 2016), RNaseH (Samelson et al., 2016, Jensen et al., 2020) and Halotag protein (Samelson et al., 2018), and each protein concluded that the folded units were in close proximity to the ribosome, while lengthening the C-terminal extension from the ribosome restored nascent chain stability - confirming the ribosome destabilises emerging peptides.

Energetics of folding can consequently be extracted from these systems. Using the populations of folded and unfolded protein, calculations of  $\Delta G_U$ , as well as rates of unfolding can therefore be recorded for each protein.

Limited proteolysis has recently been combined with SecM AP force sensor experiments (**section 1.3.2**) on the RNaseH (Jensen et al., 2020), allowing a combined approach to study the thermodynamics of folding on the ribosome. RNaseH folding pathways have been extensively characterised off the ribosome using urea denaturant experiments monitored by CD spectroscopy. Unfortunately, these experiments are not amenable to RNC ‘on-ribosome’ studies due to the vast protein and rRNA content of the ribosomes which absorb strongly in the UV region. RNCs of RNaseH were produced using the SecM AP and generated using PURExpress. The rate of unfolding on-ribosome ( $0.1 \text{ s}^{-1}$ ) was faster than off-ribosome ( $0.74 \text{ s}^{-1}$ ) when treated with urea, due to the destabilising ability of the ribosome. A comparison of urea dependence (m-value) for on- and off-ribosome unfolding experiments are similar, suggesting that the same unfolding pathway is followed. It is also worth noting that ‘on-ribosome’ experiments can only tolerate 4 M urea before the ribosome unfolds – these experiments would not be suitable for membrane protein folding experiments which required much higher denaturant concentrations (Harris et al., 2017a, Paslawski et al., 2015). AP sensor experiments were also used to determine the position, and stability of a known RNaseH folding intermediate. Variants of RNaseH which are known to populate this early intermediate do release SecM suggesting this intermediate produces a significant force for release which is surprising due to its low stability of only 1-2 kcal/mol. More surprisingly, a stable variant of RNaseH which does not transition through the kinetic intermediate is incapable of release, with the authors suggesting that kinetics of folding is the reason for this, where residues involved in intermediate formation are also involved in the rate-limiting step of RNaseH folding (Jensen et al., 2020).

### **1.5.3.2. Fluorescence spectroscopy**

The incorporation of exogenous fluorophores into an emerging nascent chain is a useful tool to observe nascent chain dynamics directly. Conformational flexibility in the nascent chain can be observed through the FRET between two fluorophores, or time-resolved anisotropy, or lifetimes of fluorophore decay can give insight into the local environment. The incorporation of a N-terminal conjugated cyan fluorescent protein (CFP) donor, and a small organic fluorophore (NBD-Lysine tRNA UAA) acceptor conjugated to varying positions along the emerging nascent chain, allowed FRET to map the folding of the NBD1 region of the CFTR

protein, showing that the domain is largely folded before the entire domain was formed. Furthermore, the binding of ATP stabilised this N-terminal stable intermediate to facilitate further folding as the peptide is translated – ligand can therefore promote formation of folding intermediates during co-translational folding (Khushoo et al., 2011).

More recently, the direct time-resolved FRET studies of the *E. coli* small soluble 5-helix N5-glutamine methyltransferase (HemK) N-terminal domain and can be achieved using an IVTT system with efficient incorporation of UAA spectroscopic probes to the mRNA truncated nascent chains (Holtkamp et al., 2015). This study concluded that the HemK protein folded rapidly in a two-state fashion in solution which proceeded through a compact intermediate which forms in the ribosome exit tunnel, which then rearranges into the native like structure when the entire protein is translated and released from the ribosome. The FRET pairs incorporated into the nascent chain yield efficiencies based on proximity, allowing the mapping of extended or compact structures at each incorporation position, with compaction occurring at a peptide length of 56-70 amino acids. Additionally, thermolysis based pulse proteolysis on the RNCs gave an indication into the position of the NTD BOP-Met fluorophore and the ribosome. Successful cleavage of the nascent chain suggests the intermediate is not in vicinity of the ribosome and is not protected from proteolysis, whereas the compact intermediate was not cleaved, and thus resides in the ribosome exit tunnel. Finally, the BOP-Met fluorophore can undergo photoinduced electron transfer (PET) with tryptophan residues incorporated into the nascent chain, which is ideally suited to study conformational rearrangements on short length timescales and allows the timing of intermediate to native rearrangement to be assessed, this appeared to occur whilst still in the exit tunnel, before final emergence and release of the protein.

Although the compact nature of the intermediate was determined, a clear pathway of folding remained unclear. More recently FRET and PET have been used on the HemK helical domain to extract rapid kinetics, which showed passage through at least 4 kinetic intermediates with helices docking one after the other to form the folding core – a ‘helix-docking’ mechanism whose rate is limited only by translation, these conclusions too were drawn from FRET and PET experiments (Mercier and Rodnina, 2018).

### **1.5.3.3. Structural studies**

In addition to the sensitive proximity changes achieved using fluorescence, high-resolution NMR spectroscopy can resolve heterogeneous ensembles of intermediate states that are likely populated during co-translational folding. Selective isotopic labelling of nascent chains

attached to a ribosome (Cabrita et al., 2009, Waudby et al., 2013, Cassaignau et al., 2016, Cabrita et al., 2016) can yield high resolution structural dynamics of the emerging nascent chain.

The filamin domain 5 (FLN5) of *Dictyostelium discoideum* gelation factor has been characterised *in vivo* using NMR and molecular dynamics simulations (Cabrita et al., 2016). Structure characterisation of nascent chains at several positions using SecM mediated RNCs revealed that FLN5 attains its native structure after 47 residues of the following FLN6 domain, well after it has emerged from the ribosome exit tunnel. Isolated FLN5 ('off' ribosome), however folds even when lacking four C-terminal residues suggesting that the ribosome nascent chain interactions destabilize the native FLN5 structure until a significant portion of FLN6 has emerged. Disordered FLN6 shows significant interactions with rRNA and protein, particularly the L23 and L24 subunits (Cabrita et al., 2016). These selective labelling high resolution experiments are beginning to unfold residue specific atomistic detail of interactions that modulate co-translational folding. The major drawback of NMR experiments is the inability to access real-time actively elongating RNCs due to its long acquisition time. Nevertheless, these experiments while incredibly technically challenging, are still being developed to their full potential leading to potentially very exciting atomic resolution determination of co-translational folding interactions.

#### **1.5.3.4. Force spectroscopy**

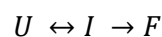
Co-translational folding is a non-equilibrium process and therefore defining kinetics of the process is crucial. Single molecule approaches to study these folding pathways are very powerful for dissecting energy landscapes and can provide information on populated folding intermediates, but also the kinetic rates associated with them (Kaiser and Liu, 2018). The mechanical force provided by optical or magnetic tweezers acts as the denaturant to promote states which are more extended and contain less structure than the native state. In contrast, chemical denaturants act on the entire protein and therefore specific intermediates which exist in the nascent chain whilst attached to the ribosome cannot be captured using bulk unfolding studies, leading to the directly applied force of tweezers being a very exciting technique for characterising co-translationally formed states.

The co-translational folding of T4 lysozyme was the first to be studied using SMFS with optical tweezer spectroscopy (Kaiser et al., 2011). The IVTT produced RNC complex was tethered to a polystyrene bead using a tagged ribosome, and a biotinylated nascent chain to streptavidin conjugated DNA handle which was in turn bound to another polystyrene bead which can be

manipulated using an optical trap. This setup showed that unfolded T4 lysozyme on the ribosome exhibited similar unfolding kinetics to isolated protein with an unfolding force of 17.0 pN to 17.2 pN between ribosome-bound (to produced full length folded protein with additional 41 amino acid unstructured region which remains in the ribosome exit tunnel), and free, when pulled at a velocity of 100 nm/s. Distances to transition states between the two experiments were also comparable, with  $\Delta x_{\text{off}} = 2.3$  nm,  $\Delta x_{\text{on}} = 2.0$  nm, as were native state lifetimes (Kaiser et al., 2011).

The refolding of T4 lysozyme on and off the ribosome however showed marked differences. Cycles of pulling and relaxation to determine the refolding probability were carried out. The free protein usually always refolded, but ‘on’ ribosome experiments refolded only 28 % of the time and lacked a well-defined folding transition observed during force ramp experiments. Time-resolved force-clamp experiments were carried out at a force of 3.6 pN to assess the apparent folding rate from measured folding times for the protein. Surprisingly, ‘on’ ribosome samples folded more than two orders of magnitude slower than free protein;  $0.012$  s<sup>-1</sup> compared with  $5.4$  s<sup>-1</sup> (Kaiser et al., 2011). Extending the C-terminal linker from 41 amino acids to 60, the apparent folding rate increased to  $0.24$  s<sup>-1</sup>, significantly faster than before, but still slower than folding off the ribosome, suggesting that even though the 41 amino acid is long enough for the free protein to emerge and fold, the restricted space and proximity to the ribosome is restrictive for folding. On careful analysis of the force-clamp data, intermediate ‘hopping’ was observed between the unfolded (U) and intermediate (I) state before collapsing into the folded state (F). This latter transition is irreversible at the refolding force and therefore the folding pathway can be written as:

**Equation 1-1: Folding pathway for T4 lysozyme determined using optical tweezer SMFS**



Using the force-extension traces, it was estimated that 96-108 residues participate in this folding intermediate, and this is consistent with the formation of the C-terminal T4 Lysozyme subdomain which may have previously been described as a hidden intermediate (Kaiser et al., 2011, Cellitti et al., 2007).

The optical set-up was further used to restart translation using mechanical force on a SecM stalled polypeptide to ‘unclog’ a blocked ribosome (Goldman et al., 2015). The Top7 protein conjugated to GFP is only translated on successful release of SecM. The force of the Top7 protein folding in proximity to the exit tunnel produces a sustained folding force and releases



SecM stalling, causing translation of the GFP reporter *in vivo*. Forces generated by different Top7 linker regions were estimated by optical trapping *in vitro* produced RNCs. This study concluded that at least 10 pN of force is sufficient to release the SecM induced stall (Goldman et al., 2015), which shows the importance of slowing down translation for domain folding, which is then upregulated when the folded protein provides a sufficient force to restart/speed up translation.

Force-spectroscopy has also been used to study the influence of complex formation during co-translational folding. The multidomain protein elongation factor G (EF-G) is a highly conserved protein of 5 domains. The first two N-domains form a stably structured unit co-translationally, which has been characterised both on and off the ribosome, with folding of the G-domain being slower on ribosome (Liu et al., 2017). This suggests that co-translationally modulating translational rates, presumably by sequestration of the nascent chain to interact with the ribosome exterior is a common feature for protein biogenesis. Furthermore, in addition to the two N-domains of EF-G, domain III appears to orchestrate folding by forming interactions with the yet-to-emerge C-terminal domains before it can fold itself. The C-domains exclusively stabilise domain III whereas the N-domains do not. This suggests a post-translational folding mechanism on the C-domain of EF-G for correct folding (Liu et al., 2019). Consequently, an accumulation of misfolded EF-G, arising from an unordered domain III, prevents productive folding, and suggests a chaperoning mechanism, similar to that that in eukaryotes, where EF-G is folded by HghI (with TRiC chaperonin, and Hsp90) which likely binds co-translationally (Monkemeyer et al., 2019). Currently, however, no such similar HghI chaperone has been identified in *E. coli*.

#### **1.5.4. Membrane protein RNCs**

A broad range of studies have been carried out on soluble protein RNCs, from biochemical assays, to structure and dynamics of an emerging nascent chain, however – very little has been achieved for membrane proteins presumably due to the already difficult nature of expressing and purifying full length proteins, the additional complexity of ribosomal stalling, and large complex purification being a daunting task. To my knowledge, nobody has yet purified large polytopic membrane protein RNC to assay the structure and dynamics of the emerging nascent chain from cells. The limited published work (in addition to biochemical analysis of membrane protein folding and insertion using SecM AP as a force sensor – **Section 1.3.2**) focusses on the use of RNCs to determine the function or structural dynamics of the translocon, and mechanisms of nascent chain delivery to the membrane and translocon machinery for

secretion across the bilayer, and most of these RNCs were generated using mRNA truncation or peptide stalling sequences *in vitro*, with nanodiscs of reconstituted insertion machinery to dictate function.

A detailed analysis of emerging nascent chain transfer from the ribosome to translocon is still unknown but significant progress using membrane protein RNC complexes are beginning to uncover mechanisms of protein biogenesis at the membrane. *In vivo*, when the nascent chain signal anchor sequence (SAS) is exposed, the signal recognition particle (SRP) binds, generating a conformational change to form a high affinity complex which prevents other chaperoning proteins from binding the exposed nascent chain. This interaction also activates SRP to reveal its Ffh binding site for the SRP receptor, FtsY, a GTPase. The recognition of SAS by SRP promotes the binding of an RNC-SRP to FtsY, and subsequently SecYEG, to form a quaternary complex which promotes nascent chain transfer to the translocon. The SRP receptor (FtsY), although not an integral membrane protein, has previously been suggested to target to the membrane co-translationally (Bibi, 2011) and requires helices 2-4 for membrane attachment as shown by production of various length FtsY SecM based RNCs *in vivo* with purification by flotation and sucrose density gradients (Bercovich-Kinori and Bibi, 2015). However, complete details of the mechanism of SRP delivery to the membrane are currently unclear with conflicting theories suggesting that SRP and SecYEG interact with the ribosome in a mutually exclusive manner, or that FtsY and the ribosome compete for the binding to a loop of the SecY channel prior to nascent chain transfer (Draycheva et al., 2018). The Wintermeyer group have used FRET analysis between an emerging Lep-RNC with signal anchor sequence (SAS), and the SecYEG embedded in a MSP-based nanodisc with *E. coli* total lipids, to determine the effect of SRP conformational change and complex formation at the membrane (Draycheva et al., 2018) to deconvolute the order of NC transfer events.

Varying length RNCs of Lep-SAS were produced using mRNA truncation using *E. coli* ribosomes in an IVTT system. Equilibrium titrations of the labelled nascent chain and FtsY-SecYEG complexes reconstituted into MSP-based nanodiscs were assayed for FRET efficiency to show the degree of interaction between the nascent chain and the translocon. RNCs were generated using truncated mRNAs in a cell-free translation system using ribosomes purified from *E. coli*. Two lengths of the exposed RNC peptide were produced of 50, or 70 amino acids (Lep50 and Lep70 respectively), leaving 15 or 40 exposed from the ribosome exit tunnel. Both constructs bound concurrently to SRP, and SecYEG as assayed by FRET. This signal was lost when analogous experiments were performed in presence of the FtsY receptor, which was due

to a rearrangement in the complex forming a pre-transfer state quaternary complex, with SRP remaining bound to Lep-RNC - causing a conformational change and a decrease in SRP-Lep affinity. This lower dissociation constant ( $k_D$ ) was more pronounced for the shorter Lep50 complex. In addition, the structure of the Lep50 RNC-SRP-FtsY complex was obtained using cryoEM previously (von Loeffelholz et al., 2015). The lower affinity is suggested to be connected to the transfer of SAS to SecYEG. This was further confirmed using Proteinase K digestion where the Lep50 nascent chain is too short to enter the translocon and was therefore degraded, whereas the Lep75 nascent chain enters SecYEG and is protected independently of SRP and FtsY binding. These experiments uncovered a mechanism of a quaternary complex of RNC bound to SRP and SecYEG with FtsY, which conformationally changes once docked at the membrane allowing transfer of the nascent chain to SecYEG for insertion (Draycheva et al., 2018).

Also residing at the bacterial inner membrane is YidC, a 6 TM helix protein insertase which aids in the transition of hydrophobic stretches of helix to cross the membrane and partition into the bilayer. Photo-crosslinking studies using mRNA truncated RNCs on the single TM protein FtsQ, highlighted the need for the translocon and to a lesser extent, YidC for insertion. This was shown using *in vivo* protein depletion studies with Proteinase K, and photo-crosslinking assays on RNCs prepared using IVTT for insertion into IMVs. This study showed that the emerging helix makes distinct interactions with protein-lipid environments very early in biogenesis, long before the entire helix has emerged. The ribosomes also modulate this early-stage process; when the ribosome is released prematurely from the nascent chain, an altered confirmation is detected in the translocase (Urbanus et al., 2001). Together these experiments allow determination of the sequential order of helix emergence and docking to the translocon and YidC for correct insertion into the membrane.

The architecture of the YidC-ribosome complex has also been studied using RNCs of the first TM helix and part of the cytoplasmic loop of F<sub>o</sub>c of ATP synthase, a common YidC substrate. RNCs purified from *E. coli* using strep-tag affinity chromatography and were titrated against MSP-nanodiscs of *E. coli* polar lipid extract, or 30:30:40 DOPG:DOPE:DOPC lipids, with reconstituted YidC with any interactions being quantified using fluorescence correlation spectroscopy (FCS). The lipid nanodisc environment strongly promoted YidC:ribosome complex formation to a higher degree than when prepared in detergent micelles, and a single copy of YidC suffices to form this complex (Kedrov et al., 2013).

Membrane protein RNCs have also been used in conjunction with high-resolution CryoEM techniques. The structure of an active SecY channel was obtained using RNCs of a DsbA N-terminal signal-peptide with SecM stalling sequence, produced in *E. coli* or *M. jannaschii* and purified with a ribosomal Strep-tag (Park et al., 2014). These RNCs were produced *in vivo* and crosslinked to Cys68 in the plug region of the SecY channel and purified in DDM detergent. All RNCs were monomeric due to the co-expression of MazF endonuclease, which cleaves the exposed mRNA between ribosomes. Minor differences in SecY channel structure were observed on RNC binding showing the open state, with the signal sequence intercalated with the open lateral gate, and the nascent chain forming a loop on the cytoplasmic surface of SecY, rather than entering the channel.

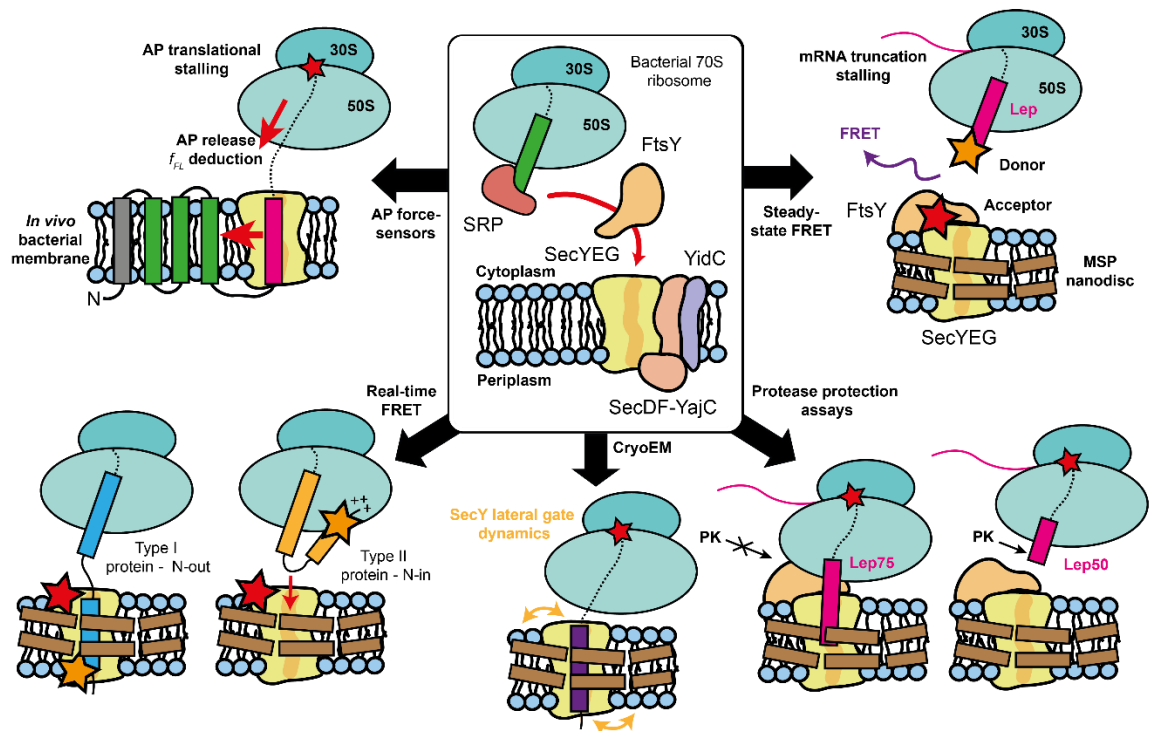
Later, CryoEM was used to visualise static interactions between 7 TM proteorhodopsin (PR) (Bischoff et al., 2014), and the SecY translocon machinery in an RNC stalled state. A tryptophan dependent TnaC stalling peptide was used to generate SecY-bound RNCs with two complete TM helices and a hydrophilic region, necessary for stalling and producing tight interactions with SecY. These RNCs were co-purified in detergent with SecYEG and an RNC-SecY structure was obtained to 7.3 Å. The structure supports the idea that the TM helices may exit the SecY channel via the lateral gate for partition into the membrane as TMs 1 and 2 are directly in front of the lateral gate. It was proposed that YidC also positions itself outside of the lateral gate (although not detected in this study), suggesting further chaperone regulation for TM partitioning, along with SecA, which cannot bind to the ribosome at the same time as SecY. Interactions between the positively charged cytoplasmic loop of the first two helices with a ribosomal rRNA helix were also observed. This suggests a role for the ribosome in retaining charged loops on the cytoplasmic side of the bilayer during TM integration into the membranes in this system.

Further studies of a hydrophobic FtsQ anchor domain helix interacting with the SecY translocon reconstituted into a MSP-nanodisc POPC:POPG lipid bilayer have also been structurally characterised using CryoEM (Seidelt et al., 2009, Kater et al., 2019), this showed that short nascent chains cause displacements in the membrane core, where SecY helices 2b, 7 and 8 tilt in the membrane to ‘unzip’ the lateral gate at the cytoplasmic side allowing the helix to transverse into the bilayer. MD simulations of this process suggest that accessory helices of SecE modulate the conformation of the lateral gate (Kater et al., 2019). However, in these studies, SecYEG was purified and reconstituted into synthetic lipids but as yet no membrane protein RNCs have been isolated directly from cells with their native lipids

surrounding the SecYEG translocon, and so these observed features may not accurately depict what happens in a native lipid environment, *in vivo*.

Most recently, the real-time insertion and topology of a membrane spanning helices explored using FRET and proteinase K accessibility of mRNA-truncation RNCs in an IVTT system with SecYEG reconstituted into nanodiscs (Mercier et al., 2020). Type I orientation of membrane protein topology describes the N-terminus of the first TM domain being on the outside (periplasmic) where type II, the inside (cytoplasmic) side of the inner membrane. Real time FRET and protease digestion were used with type I protein LepB, and type II EmrD to show that for type I topology, the TM is immediately delivered to the translocon, whereas for EmrD, a longer nascent chain is required which is then looped inside the ribosome exit tunnel, mediated by positive charges allowing the helix to flip and insert the other way around. Kinetic analysis shows that this TM 1 delivery is rapid, and the rate of insertion is only limited by translation showing the ribosome has a very important role in membrane protein biogenesis and topogenesis.

A summary of the biochemical methods to study membrane protein targeting, co-translational folding and SecYEG dynamics using RNC complexes is shown in **Figure 1-8**.



**Figure 1-8: Summary of RNC use to study co-translational membrane protein folding, targeting and translocon dynamics**

RNC complexes have been used to study co-translational folding properties using a host of different methods. The most renowned work focuses on the cloning of an arrest-peptide sequence into the gene of a membrane protein to produce a weak translational stall. As the protein folds, a force of magnitude higher than that of the stalling motif is generated and the stall is released. This can give an indication into the intrinsic stability of proteins and their interactions with the translocon machinery (section 1.3.2). Additionally, mRNA truncations can be used to produce RNCs using IVTT systems. The versatility of the IVTT technique allows nascent chain labelling with fluorophores which can undergo FRET transfer with a tagged SecY-FtsY complex in an MSP-based nanodisc (Draycheva et al., 2018). This enables spatial information on complex formation for protein biogenesis at the membrane. Additionally, protease protection assays on the emerging nascent chain have been used to determine the exposure of the peptide to the surrounding environment. In this example, a Lep75 nascent chain was protected from Proteinase K as determined by gel shift assay, whereas Lep50, a shorter nascent peptide which could not insert into the SecY channel and was therefore degraded (Draycheva et al., 2018). Similar assays using crosslinkers can be studied using SDS-PAGE to measure translocon and nascent chain interactions in vivo. Also using IVTT, it is possible to observe real-time FRET information for co-translational ‘flipping’ of helices between type I (N-out) proteins, and type II (N-in) proteins, and the interaction with the ribosome which govern this process (Mercier et al., 2020). Finally, a significant chunk of research uses AP sequence to stability stall a membrane protein in vivo, for purification in DDM and reconstitution into MSP nanodiscs with reconstituted translocon. These samples are purified to homogeneity can give high resolution structural information on the dynamics of the nascent chains effect on the SecY helix dynamics and opening of the lateral gate (Kater et al., 2019, Seidelt et al., 2009, Bischoff et al., 2014).

## 1.6. Thesis aims and layout

This thesis presents the development of a suitable method for preparing an  $\alpha$ -helical polytopic membrane protein RNC, in a native lipid environment using novel co-polymer technology. These methods provide an experimental platform for the study of the co-translational folding pathways of a characteristic membrane protein, in a native environment which is crucial for

the correct folding of membrane proteins in a host organism. The high yielding and clean preparation of these ribosome-nascent chain complexes, at varying positions along the ribosomal translation cycle will provide a basis for structural insight into the properties of the partially folded nascent chain, not only intrinsically, but also for the interactions between the nascent chain and the ribosome, as well as the interaction with lipids and chaperoning machinery resident in the native membrane environment. The samples are amenable to high resolution structural acquisition, or *in vitro* biochemical or biophysical methods which when combined, will provide a deeper understanding of the co-translational process underpinning membrane protein folding *in vivo*.

This thesis consists of 4 results chapters, **3-6**, with all materials and methods found in chapter **2**. The relatively simple method for the generation and capture of a partially stalled translating ribosome, and well-studied rhomboid protease GlpG nascent chain in DDM detergent can be found in **chapter 3**, and the development of purification of the same RNCs in a 'native' lipid nanodisc, utilising novel co-polymer technology is explored in **chapter 4**.

Additionally, I explore whether a mechanical method of protein unfolding is a feasible alternative to *in vitro* bulk unfolding methods for a large integral membrane transporter, the 12 TM helix xylose/H<sup>+</sup> symporter Xyle, which has been biochemically characterised along with necessary mutants in **chapter 5**. Using AFM, I have sought to gauge the magnitudes of forces required to unfold these proteins from a bilayer environment, and whether the protein candidate and system can be utilised for magnetic tweezer studies which will allow deconvolution of physiological forces involved in protein folding. Mechanical unfolding of Xyle is discussed in **chapter 6**.

The near-physiological representation of an emerging, folding nascent chain will transform the membrane protein folding field as people will be able to generate high quality data using their favourite biophysical technique to learn more about the mechanistic details of the co-translational membrane protein folding process. There is also the opportunity further down the line for coupling mechanical folding studies with the RNC generation to study intrinsic forces and folding pathways from a co-translational perspective.

## **2. Materials and methods**

---



## **2.1. Reagents**

### **2.1.1. Common biochemical reagents**

All reagents were purchased from Sigma Aldrich or Fisher Scientific UK and were of the highest grade unless otherwise stated in the text.

### **2.1.2. DIBMA stock preparation**

Diisobutylene/maleic acid copolymer (DIBMA) was precipitated out of Sokalan CP9 solution (BASF) with 0.6 V of 4 M HCl. The solid was centrifuged at 17000 x g for 15 minutes and washed with ddH<sub>2</sub>O. The water wash was repeated 4 times. 1.2 V of 4 M NaOH was used to solubilise the solid, before repeating the above precipitation and washing steps. The wet pellet was then lyophilised for 72 hr to produce pure DIBMA polymer (Barniol-Xicota and Verhelst, 2018, Gulamhussein et al., 2020).

The lyophilised polymer was used to produce 20 % (w/v) stocks in 2 M NaOH. Base was then added dropwise to carefully dissolve the solid and adjust the pH to 8.0. The concentration of the polymer solution was checked using a Reichert AR200 digital refractometer with  $dn/dc$  1.346 M<sup>-1</sup> (Oluwole et al., 2017a). The 20 % stock was used to produce 2.5 % stocks in the necessary buffer for the solubilisation of *E. coli* membranes.

Lyophilised DIBMA stocks were quality checked (Gulamhussein et al., 2020) using a Shimadzu IR Affinity-1s, or a PerkinElmer Spectrum Two FTIR, scanned across a complete wavenumber range 4000-400 cm<sup>-1</sup>, with 16 scans to measure transmittance. In particular, the carboxylate (1705 cm<sup>-1</sup>) and anhydride (1775 cm<sup>-1</sup>) bands were checked to ensure that my treatment of the polymer with strong acids did not result in a condensation of the anhydride ring, which would result in significantly reduced solubilisation efficiency.

## **2.2. Molecular biology for expression constructs**

### **2.2.1. Materials and plasmids**

All expression constructs presented in this thesis were cloned and donated, or synthesised and sub-cloned into the commercially available pET28a Vector (Novagen) with kanamycin resistance gene. WT XylE with C-terminal His<sub>10</sub> tag in the pET28a vector was produced by Dr Heather Findlay (Booth group, KCL), WT GlpG with N-terminal His<sub>6</sub> also in pET28a was produced by Dr Eamonn Reading (Reading group, KCL). XylE N- (M1-V275, including ICH) and C-terminal (G276-L491) domains in pET19b and pET28a respectively were produced by Dr Nicola Harris (Booth group, KCL). The HaloTag gene in the pFN18a vector (ampicillin

resistance) was donated by Dr Palma Rico Lastres (Garcia-Manyes group, KCL). The SecYEG gene in a pBAD vector (ampicillin resistance) was donated by Prof. Ian Collinson (University of Bristol), as was the empty RSF-Duet-1 vector (kanamycin resistance). The SecYEG cassette was subcloned from pBAD into pET28a for T7 promotion by Laura Blackholly (Booth group, KCL). The pBirAcm biotin ligase plasmid with chloramphenicol resistance was purchased from Avidity.

All additional cloning stages were carried out using Q5 mutagenesis or HiFi builder kits (New England Biolabs), or In-fusion (Takara). Most other molecular biology reagents (restriction enzymes, DNA T4 ligase, and any additional DNA polymerases, 1 kb DNA ladders, etc.) were purchased from New England Biolabs, as were mini-prep plasmid purification kits. The midi and maxi plasmid purification kits were purchased from Qiagen.

Agarose was purchased at the highest grade from Sigma and gels were run in TAE buffer purchased as a 50x stock from ThermoFisher. Gels were stained using Gel Red nucleic acid stain from BioRad.

Sequencing reactions were carried out by Genewiz (Stansted, UK) or MWG Eurofins (Munich, Germany), where the latter also synthesised all DNA primers used in this thesis.

### **2.2.2. Bacterial strains and buffers**

All cloning was carried out using DH5 $\alpha$ -derived one shot Top10 chemically competent *E. coli*, competent 5-alpha *E. coli* supplied with the NEB mutagenesis kits, or Turbo competent *E. coli* cells also from NEB. For protein expression, XylE and associated mutants were grown in One Shot BL21-AI chemically competent cells (Invitrogen). Isolated GlpG and associated mutants as well as GlpG RNC expression constructs were grown in BL21 (DE3) chemically competent cells (Invitrogen) without the genomic arabinose promoter system.

The most used bacterial media was Luria-Bertani (LB), which was used for both cloning and expression cell growth. LB was purchased as a premixed powder from Sigma or produced from component parts: 0.5 % (w/v) bacto yeast extract, 1 % (w/v) bactotryptone and 1 % (w/v) NaCl, pH 7.4. SOC media supplied with NEB competent cells was occasionally used for cloning. An aspartic acid/phosphate/glucose (MDG) based minimal media was also used for high density RNC expression cell growth, the details of this media composition are provided in **section 2.4.1.1.**

### **2.2.3. Construct design and PCR**

All primers, PCR conditions, gene sequences and translations for molecular biology methods are shown in **Appendices 1 to 6**

#### **2.2.3.1. Cloning of GlpG truncations and (enhanced) SecM**

His<sub>6</sub>-GlpG-SecM plasmids were constructed for GlpG RNC overexpression. pET28a was linearized with NdeI and XhoI restriction enzymes (New England Biolabs). The glpG gene (subcloned from (Harris et al., 2017b)) and a 44 amino acid sequence motif encoding both the SecM arrest motif and additional amino acids (subcloned from (Rutkowska et al., 2009)) were amplified and cloned simultaneously into the linearized pET28a vector using In-Fusion® HD cloning (Takara Bio), with its mechanism described in **section 2.2.3.3**.

The resulting pET28a-His<sub>6</sub>-GlpG-SecM vector was then used in subsequent deletion cloning reactions to generate constructs of different GlpG ‘lengths’ using the Q5® site-directed mutagenesis kit (New England Biolabs), where non-mutagenic primers are designed to flank the region to be deleted and amplify the entire plasmid, resulting in a blunt-ended linear plasmid with region deleted. The linear DNA was then treated with an enzyme mix (KLD) of kinase, to add a 5’ phosphate (K), ligase (L), which stitches the backbone together, and DpnI (D) to digest methylated parent DNA before transformation into DH5α competent *E. coli*. A pET28a-His<sub>6</sub>-GlpG plasmid was also generated for overexpression of His<sub>6</sub>-GlpG without the linker-SecM sequence using the same procedures.

To substitute the WT SecM sequence for the enhanced SecM sequence, primers were designed for the enhanced sequence (Cymer et al., 2015a, Kempf et al., 2017), where the substitution is contained in the middle of the forward primer, with at least 10 complementary base pairs on the 3’ end of the primer. The reverse primer is designed so the 5’ ends anneal back-to-back. PCR was carried out using the Q5® site directed mutagenesis kit (New England Biolabs) to amplify the entirety of each RNC vector to produce the linear DNA. This was also treated with KLD and transformed as above. These primers could also be used for arrest-enhanced SecM substitutions directly into the RSF-Duet1 plasmid.

#### **2.2.3.2. Addition of Avi-Tag to RNC constructs**

Avi-tag was added to the N-terminal end of each construct between the His<sub>10</sub> tag and the start of GlpG. Q5 addition was used, where half of the addition is included at the 5’ ends of the forward and reverse primers. The entire plasmid was amplified and blunt end KLD treated and transformation was carried out as above.

### 2.2.3.3. Cloning of SecYEG and RNC into RSF-Duet 1

Duet plasmid cloning was carried out using In-Fusion HD (Takara). Infusion is a form of Gibson assembly, and uses primers designed at <https://takara.teselagen.com/#/DesignPage> to amplify the gene with flanking homologous regions to the vector. Infusion exonuclease then digests 5' ends on the insert and a linearised vector. Linearisation of the vector can be achieved using PCR, or restriction digest, and the position of digest is not important. The digested complementary regions then combine, joining the vector and insert to form circular DNA. The plasmid is then transformed into *E coli* Stellar competent cells (supplied with Infusion) to ligate the DNA backbone together.

Firstly, the Q5 mutagenesis kit was used to replace the SecE His<sub>6</sub> tag with an Avi-tag in the pET28a vector. The avi-SecYEG was amplified from pET28a using infusion primers with CloneAmp HiFi PCR Premix polymerase before assembly into MCS2 of RSF-Duet1. RSF-Duet was linearised using NdeI and XhoI in MCS2.

The GlpG RNCs were then cloned into MCS1 also using infusion. RNCs were amplified from pET28, also using CloneAmp premix. The RSF-Duet1-SecYEGavi was used as the vector which was linearised by restriction digest overnight using NcoI and NotI in MCS1. An example agarose gel for the amplification of the insert and digest of the vector is shown in **Figure 2-1**. Sequences were confirmed using Sangar sequencing with ACYCDuetUP1 and DuetDOWN1 primers for MCS1, and DuetUP2 and T7 terminator for MCS2.

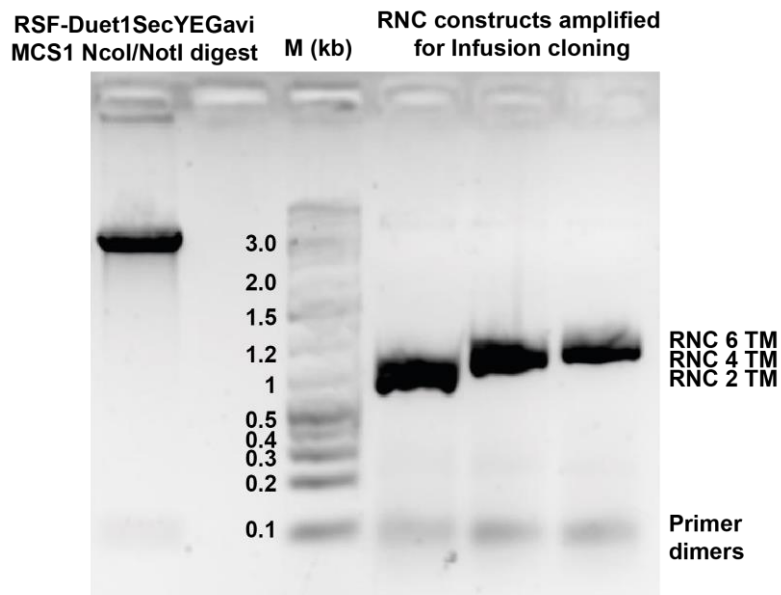


Figure 2-1: Example agarose gel for Infusion cloning of RNCs into MCS1 of RSF-Duet1-SecYEGavi

*Example 1 % agarose gel to analyse DNA fragments. Here, PCR was used to amplify GlpG-SecM constructs to include homologue regions to vector for Infusion cloning. The RNC amplicons are shown at approx. 1.2 kb with increasing size. NcoI and NotI were used to linear the RSF-Duet1-SecYEGavi vector at MSC1.*

#### **2.2.3.4. Poly-glycine linker addition to Xyle**

The addition of 28 amino acids to the C-terminal of WT Xyle was carried out using the Q5 mutagenesis kit with an annealing temperature of 60 °C and an extension time of 3.5 min. A small amount of the PCR product was run on an agarose gel to check the DNA quality and size. 1 µL PCR product was added to the NEB premixed KLD as above before 5 µL of the mixture was transformed into supplied DH5α cells, grown for mini-prep plasmid purification, and confirmed by Sanger sequencing using the T7-term primer at Eurofins Genomics or Genewiz.

#### **2.2.3.5. Addition of HaloTag and TMH to Xyle**

Firstly, Xyle-His<sub>10</sub> was amplified out of pET28a using primers with restriction sites for KpnI and BamHI, using HF Phusion polymerase. The PCR mix contained 4 µL 5x HF buffer, 0.4 µL 10 mM dNTPs, 1 µL of each forward and reverse primers, 1 µL of template DNA and 0.2 µL of HF Phusion polymerase, made up to 20 µL with nuclease free H<sub>2</sub>O, and run with PCR protocol in **appendix 5**. This allowed for the sub-cloning of the gene into the pFN18a vector upstream of the HaloTag and its TEV site, which allows the post-expression cleavage of the HaloTag from Xyle if required. Once the Vector was digested, it was subject to Shrimp alkaline phosphatase (rSAP) to remove 5' phosphates preventing re-circularisation, and T4 ligase was used after DNA clean-up. Plasmids were transformed into Top10 cells, grown, and purified, and sequenced using a custom primer (5'-TGTCGACGCTCGAGATTTCC-3'), to capture the end of Halo, and beginning of Xyle.

A *de novo* transmembrane helix was inserted between Halo and Xyle, using Q5 substitution in two stages. A poly-alanine/leucine helix (MANMGPGGPGAAAALALALLLLALAAAAGSS) was designed using two sets of primers to incorporate the long stretch of base pairs in two stages. Q5 PCR reactions were carried out as described above.

### **2.3. SDS-Page and Western blotting**

#### **2.3.1. Isolated protein SDS-PAGE**

All Xyle, SecYEG and isolated GlpG proteins and associated mutants were analysed by SDS-PAGE. Samples were run in 2x loading dye (62.5 mM Tris pH 6.5, 20 % (v/v) glycerol, 2 % (w/v) SDS 0.67 mM βME, bromophenol blue) in a Tris-Glycine based SDS buffer (10x stock: 30.8 g/L Tris, 144.2 g/L Glycine and 10 g/L SDS) on 10 % or 12 % (w/v) Tris-Glycine prime gels (SERVA). Bands were visualised with Quick Coomassie stain (Generon) or Western blot.

### 2.3.2. RNC sample SDS-Page

2-10 pmol of RNC sample was run on 10 % or 12 % (w/v) NuPAGE Bis-Tris gels at neutral pH in NuPAGE MOPS SDS buffer (20x). Low pH 5x SDS sample dye, pH 5.7 (30 % glycerol, 0.25 M Bis-Tris pH 5.7, 0.8 % (w/v) DTT, 8 % (w/v) SDS and bromophenol blue) was used to maintain the ester bond between the tRNA and the nascent chain. DIBMA samples were boiled at 95 °C for 5 min prior to gel loading to promote disruption of the nanodisc, however DDM samples were not boiled, avoiding aggregation of the membrane protein which can prevent the protein entering the gel. SDS treated membrane proteins tend to run near their correct molecular weight without additional thermal denaturation. To obtain the released forms of the nascent chain the RNC samples were treated with 10 µg of RNase A to digest the rRNA at room temperature for 5 min. Bands were visualised by Quick Coomassie stain (Generon) or Western blot.

### 2.3.3. Western blot

SDS page gels were transferred onto nitrocellulose or PVDF membranes (pore size 0.2 µm) using a semidry transfer in transfer buffer (1x stock: 3 g/L Tris, 14.4 g/L Glycine, 0.77 g/L SDS, 10 % methanol) at 45 A per gel for 1 hr 15 min on an Amersham TE 77P WR transfer block. Gels were blocked with 5 % (w/v) milk powder in PBS-Tween (0.1 % Tween-20) for 1 hour at RT and incubated with the necessary antibody (see **Table 2-1**) for another 1 hour at RT. Membranes were washed 3 x 1 min in PBST and then 3 x 15 min before developing with ECL select detection agent (Amersham), or applying a secondary antibody and repeating the PBST washing step prior to developing and imaging the blots. Developed membranes were imaged using blue fluorescence on an Amersham Imager 600.

*Table 2-1: Antibodies and dilutions for Western Blotting*

Antibody	Dilution	Used with
Poly-Histidine-HRP (Sigma)	1:10,000	XylE and associated mutants GlpG and associated mutants
Streptavidin-HRP (ThermoScientific)	1:5000	Biotinylated RNC Biotinylated Halo-TMH-XylE
Monoclonal mouse anti-SecY (Prof. Ian Collinson, UoB)	1:10,000	SecYEG + SecYEG/RNC
Polyclonal Rabbit Anti-Mouse Ig/HRP (ThermoScientific)	1:4000	Secondary to Anti-SecY

## 2.4. Ribosome nascent chain expression and purification

### 2.4.1. Growth and expression of RNCs

#### 2.4.1.1. Minimal media components

The components of the MDG and EM9 media stocks are shown in **Table 2-2** and **Table 2-3** respectively. The trace metal stocks are also shown in **Table 2-4**. The 10x stock component concentrations of MDG and EM9 salts are shown below:

**10x MDG stock:** 250 mM Na<sub>2</sub>HPO<sub>4</sub>, 250 mM KH<sub>2</sub>PO<sub>4</sub>, 500 mM NH<sub>4</sub>Cl, 50 mM Na<sub>2</sub>SO<sub>4</sub>

**10x EM9 stock:** 71 g/L Na<sub>2</sub>HPO<sub>4</sub>, 34 g/L KH<sub>2</sub>PO<sub>4</sub>, 5.84 g/L NaCl, pH 8.0-8.2

*Table 2-2: MDG media components*

MDG Component	Final stock	Volume required for 1 L
10x MDG salts (autoclave)	1x	100 mL
5 % Aspartic acid, pH 7.0 (autoclave)	0.2 % (w/v)	40 mL
20 % D-glucose (autoclave)	0.4 % (w/v)	20 mL
1 M MgSO <sub>4</sub> (autoclave)	2 mM	2 mL
1000x Trace Metals (sterile filter)	0.2x	2 mL
dH <sub>2</sub> O (autoclave in flask)		836 mL

*Table 2-3: EM9 media components*

EM9 Component	Final stock	Volume required for 1 L
10x EM9 salts, pH 8.0 (autoclave)	1x	100 mL
20 % D-glucose (autoclave)	0.4 % (w/v)	20 mL
1 M MgSO <sub>4</sub> (autoclave)	5 mM	5 mL
1 M CaCl <sub>2</sub> (sterile filter, add first)	200 μM	200 μL
1000x Trace Metals (sterile filter)	0.25x	2.5 mL
20 % NH <sub>4</sub> Cl (sterile filter)	0.1 % (w/v)	5 mL
dH <sub>2</sub> O (autoclave in flask)		876.3 mL

*Table 2-4: 1000x Trace metal stocks, pH 7.0*

<b>Component</b>	<b>Final Concentration</b>
EDTA	5 g/L
FeCl <sub>3</sub> .6H <sub>2</sub> O	0.83 g/L
ZnCl <sub>2</sub>	0.05 g/L
CuCl <sub>2</sub>	0.01 g/L
CoCl <sub>2</sub> .6H <sub>2</sub> O	0.01 g/L
H <sub>3</sub> BO <sub>3</sub>	0.01 g/L
MnCl <sub>2</sub> .6H <sub>2</sub> O	1.6 g/L

#### **2.4.1.2. Growth and expression of non-biotinylated RNCs**

The following growth conditions are based on NMR protocols yielding large quantities of highly occupied RNCs (Cassaignau et al., 2016).

Overnight LB cultures grown at 37 °C were used to seed 6 L MDG which was subsequently grown to saturation at 30 °C with shaking at 220 rpm. Cells were harvested by centrifugation before suspension of an equal volume of enhanced-M9 media for induction. All media was supplemented with 30 µg/mL kanamycin. Cells were induced for protein production using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1.5-2 hr at 30 °C before washing with ice-cold PBS\*, harvesting and resuspension in lysis buffer (**see section 2.4.2**), supplemented with 1 mg/mL lysozyme and RNase-free DNase I, and stored at -20 °C. For cryo-milling, small ‘nuggets’ of cell resuspension with lysozyme omitted were made in liquid N<sub>2</sub> using a 20 mL syringe and stored at -70 °C.

\*10x PBS stock: NaCl 160 g/L, KCl 4 g/L, Na<sub>2</sub>HPO<sub>4</sub> 28.8 g/L, KH<sub>2</sub>PO<sub>4</sub> 4.8 g/L, pH 7.4

#### **2.4.1.3. Growth and expression of Biotinylated RNCs**

For biotinylated RNC-GlpG, constructs with an N-terminal Avi-tag were co-transformed with pBirAcm, maintained with 10 µg/mL chloramphenicol, to express BirA. Cells were also grown to saturation in MDG media. However, to the enhanced M9 media, 50 µM D-biotin was supplemented. The two plasmids were both expressed using 1 mM IPTG and grown as in **section 2.4.1.2**, to produce biotinylated RNC-GlpG constructs. 30 µg/mL Kanamycin and 10 µg/mL chloramphenicol was supplemented into all media. Cells were again washed with ice-cold PBS, before harvesting and resuspension in lysis buffer (**see section 2.4.2**).



### 2.4.2. Buffers

**Lysis buffer:** 50 mM HEPES-KOH, pH 7.5, 1 M KOAc, 12 mM Mg(OAc)<sub>2</sub>, 5 % (v/v) glycerol, 5 mM EDTA, 2 mM βME, 1 mM PMSF, 250 µg/mL chloramphenicol and 1 cOmplete protease inhibitor tablet.

**Solubilisation buffer:** 50 mM HEPES-KOH, pH 7.5, 500 mM KOAc, 12 mM Mg(OAc)<sub>2</sub>, 2 mM βME, 1 mM PMSF, 20 mM imidazole, 1 % (w/v) DDM **OR** 2.5 % DIBMA (pH with KOH on addition of DIBMA).

**Washing buffer:** 50 mM HEPES-KOH pH 7.5, 500 mM KOAc, 12 mM Mg(OAc)<sub>2</sub>, 5 % (v/v) glycerol, 2 mM βME, 1 mM PMSF, 20 mM imidazole, and 0.1 % (w/v) DDM **OR** nothing for DIBMA prep.

**Elution buffer:** 50 mM HEPES-KOH pH 7.5, 500 mM KOAc, 12 mM Mg(OAc)<sub>2</sub>, 5 % (v/v) glycerol, 2 mM βME, 1 mM PMSF, 500 mM imidazole, and 0.1 % (w/v) DDM **OR** nothing for DIBMA prep.

**Tico size exclusion buffer:** 10 mM HEPES-KOH, pH 7.5, 30 mM NH<sub>4</sub>Cl, 12 mM Mg(OAc)<sub>2</sub>, 5 % (v/v) glycerol, 1 mM EDTA, 2 mM βME, 0.1 mM PMSF, and 0.1 % (w/v) DDM **OR** 0.2 M Arginine (pH with KOH for arginine with DIBMA).

All buffers were prepared using RNase-free water to prevent the degradation of the ribosomal RNA content in my RNCs. This was prepared by incubating ddH<sub>2</sub>O with 0.1 % (w/v) diethyl pyrocarbonate (DEPC) to inactivate any RNase, prior to autoclaving facilitating the removal of DEPC by decomposition into CO<sub>2</sub> and ethanol.

### 2.4.3. Purification of RNCs in DDM

Cells were defrosted slowly and lysed using sonication, or cryo-milled in a Spex 6875 Freezer/Mill® High-Capacity Cryogenic Grinder using 15 cycles of 15 cps and allowed to thaw on ice. The cell debris was removed by centrifugation at 20,000 x g for 45 min at 4 °C. This was repeated in fresh tubes if the resulting supernatant did not clear. The supernatant was then spun at 125,000 x g for 30 min at 4 °C to harvest cell membranes with associated ribosomes. The crude membrane pellet was resuspended at 25-50 mg/mL in DDM containing washing buffer with the aid of homogenization. Membranes were solubilised for 2 hr at 4°C on a blood rotator.

The solubilised membranes were spun at 125,000 x g, and the supernatant filtered before purification using an AKTA™ Pure system. Firstly, samples were separated by affinity

chromatography using a 1 mL HiTrap Nickel column (GE Healthcare). The column was equilibrated with 5 CV washing buffer. RNCs were passed over the column to allow binding. The column was washed with a further 30 CV of washing buffer to remove non-specifically bound RNCs and protein. RNCs were then eluted in elution buffer and directly loaded onto a 16/60 HiPrep Sephacryl S-400 HR size exclusion column pre-equilibrated with Tico size-exclusion buffer. According to the manufacturer's specifications the separation range of the Sephacryl S-400 is 20 kDa to 8 MDa.

Absorbance at 260 nm and 280 nm was monitored to evaluate the 70S content and its homogeneity using the  $OD_{260}/OD_{280}$  nm ratio (expected range: 1.9–2.0 (Hill et al., 1969)) and peak fractions were analysed by Bis-Tris NuPAGE and Western blotting to determine RNC purity and quality. The samples were pooled, the RNCs concentration was assessed using the  $A_{260}$  reading from a nanodrop and converted to a pmol/mL based on  $1 A_{260} = 24 \text{ pmol/mL}$  of 70S ribosomes (Hill et al., 1969).

#### **2.4.4. Purification of RNCs in DIBMA**

Steps were followed as per **section 2.4.3**, up until the solubilisation stage, where DIBMA containing solubilisation buffer was added to the harvested membranes. These however were solubilised for 1 hr at 25 °C to promote solubilisation with DIBMA but to not subject the RNC for longer at room temperature.

DIBMA solubilised RNCs were first bound overnight at 4 °C to super affinity Ni-NTA beads (Generon) pre-equilibrated with washing buffer, before further clean-up using size-exclusion chromatography. The supernatant containing unbound species was discarded and beads were washed with 20 CV washing buffer, before eluting in 5 CV of elution buffer. DDM was not present in any of these buffers. The eluate was subsequently concentrated to 2 mL using a 100 or 300 kDa MWCO concentrator and loaded directly onto the 16/60 HiPrep Sephacryl S-400 HR size exclusion column pre-equilibrated with Tico size-exclusion buffer containing 0.2 M L-arginine. The left-hand side of the elution peak, previously shown to contain the highest proportion of 70S ribosomes with nascent chain (Becker et al., 2012), was taken, concentrated, and buffer exchanged in a 100 kDa MW cut-off spin-concentrator, and passed over a PD10 G25-superdex column facilitating removal of arginine. RNC concentration was also ascertained using  $A_{260}$  and flash frozen and stored at -80 °C before use.

## **2.5. RNC sucrose gradient purifications**

Samples to be purified by sucrose density gradient were solubilised and bound/eluted to super-affinity Ni-NTA resin as above. Eluates were buffer exchanged out of imidazole using 100 kDa cut-off concentrators, PD10 columns equilibrated with washing buffer, or pelleted at 117,000 x g for 4 hr at in a Beckman Ti70 rotor and suspended in a suitable volume of washing buffer. Sucrose gradients were prepared with 5 steps of 5 %, 10 %, 20 %, 30 % and 40 % sucrose in washing buffer containing no glycerol. Sample was layered on top and were centrifuged at 202,000 x g in a Beckmann SW40 rotor for 16 hrs at 4 °C before fractions of 1 mL or 500 µL were taken for analysis.

## **2.6. Biotinylated RNC-avidin binding**

450 ng of RNC was incubated with 20 µL bead slurry (unless states otherwise) in pH 7.5 in Tico buffer with 0.1 % DDM for avidin binding tests.

Monomeric Avidin beads were added to an empty gravity flow column and washed with 10 CV wash buffer. To block non-reversible biotin binding sites, 10 CV blocking buffer containing 2 mM D-biotin was used to wash the beads. The reversibly bound biotin was removed by washing with 30 CV acidic regeneration buffer (0.1 M glycine, pH 2.8). The column was equilibrated by washing with at least 20 CV wash buffer containing no imidazole. The protein sample was added to the column after taking a sample for Western blot analysis. To maximize binding of the biotinylated protein sample to monomeric avidin, the column was incubated on a nutating mixer at 4 °C for 30 min. A sample of the flow through was taken for Western blot analysis. The column was washed with 30 CV wash buffer without imidazole and a sample for Western blot analysis was taken. The biotinylated protein was eluted from the column by adding 10 CV elution buffer, substituting imidazole for 2 mM biotin, and a sample was taken for Western blot analysis. The monomeric avidin beads were regenerated by washing with 20 CV acidic regeneration buffer to the column and a sample for Western blot analysis was taken (which was neutralized with 1/10th of 1 M Tris-HCl pH 9.5). The column was washed with another 30 CV wash buffer without imidazole and stored at 4 °C in the same buffer (optimally supplemented with 0.01 % (w/v) sodium azide).

## **2.7. Thin-layer chromatography lipid analysis for DIBMA samples**

To extract lipids of the native membrane RNCs a modified Folch method was used (Folch et al., 1957). 500 µL of DIBMA solubilised sample was added to 1.1 mL 10:23:1 (v/v/v) CHCl<sub>3</sub>:MeOH:1 M Tris pH 8.0 and mixed well for 1 hr at 25 °C. Phase separation was then

achieved by adding 1 mL 1:1 CHCl<sub>3</sub>: 1 M Tris pH 8.0 and the mixture was vortexed, and each phase was allowed to separate at room temperature. The upper aqueous phase was removed, and the organic layer washed with ion switch buffer (50 mM Tris, 100 mM NaCl and 100 mM EDTA pH 8.0) before drying under a stream of nitrogen and suspended in CHCl<sub>3</sub> to a suitable TLC concentration.

HPTLC silica gel TLC plates were run as in (Churchward et al., 2008). Plates were first washed in 1:1 CHCl<sub>3</sub>:MeOH before air-drying. The plate was immersed in 2.3 % Boric acid in EtOH, air dried and baked at 100 °C for 15 min. 1 µg of DOPE, DOPG and *E. coli* cardiolipin (CL) (all from Avanti) were spotted along with the extract samples and plates were run in a solvent mix of 65:25:4 (v/v/v) CHCl<sub>3</sub>:MeOH:CH<sub>3</sub>COOH to achieve good head-group separation of lipids. The plate was immersed in a copper char solution (10 % CuSO<sub>4</sub> and 8 % H<sub>3</sub>PO<sub>4</sub>) before air drying and further charring at 145 °C for 15 min and plates were imaged using blue fluorescence once cool. Alternatively, the dried plate was stained with iodine vapour, or sprayed with a molybdenum blue stain (commercial Dittmer-Lester stain (Dynska-Kukulska et al., 2013)) to specifically detect phospholipid. ImageJ (Schneider et al., 2012) was used for densitometry analysis, where raw spot intensity was converted to mole fraction, and percentage of total lipid composition was plotted with SD for at least 3 biological repeats (new growth and preparation of RNCs, lipid extraction and running of TLC plates).

## **2.8. Negative staining of RNCs in DIBMA and DDM**

*Methods and work on negative staining were written and carried out by Dr Sara Alvira-de-Celis of Professor Ian Collinson's group at the University of Bristol.*

### **2.8.1. Sample preparation for negative stain visualisation**

For negative stain, RNC samples were applied to glow-discharged (15 s) carbon grids with Cu 300 mesh, washed and stained with 3 % (w/v) uranyl acetate (1 min). Digital images were acquired at a Tecnai 12 with a Ceta 16M camera (ThermoFisher Scientific) at a digital magnification of 49,000 x and a sampling resolution of 2.04 Å per pixel.

### **2.8.2. Image processing**

Preliminary image processing was performed using the EM software framework Scipion v1.2 (de la Rosa-Trevin et al., 2016). For RNC-FL- DIMBA, 1305 particles were manually selected as input for automatic particle picking through the XMIPP3 package (de la Rosa-Trevin et al., 2013, Abrishami et al., 2013). 3237 particles were then extracted with the Relion v2.1 package (Kimanius et al., 2016, Scheres, 2012) and classified with a free-pattern maximum-likelihood

method (CL2D 2D-classification) (Sorzano et al., 2010). After manually removing low-quality 2D classes, a second round of 2D-classification was performed. For RNC-DDM, 2010 particles were selected manually and analysed following the same procedures described above.

## **2.9. Isolated protein expression and purification**

### **2.9.1. Expression and purification of GlpG and associated mutants**

#### **2.9.1.1. Growth and expression of GlpG and mutants**

N<sub>His</sub>-GlpG and associated mutants in the pET28a vector were transformed into BL21 (DE3) or C43 cells (ThermoFisher). An LB overnight was made, and 10 mL was added to each 1 L of LB culture. These cultures were grown at 37 °C with shaking at 220 rpm until an OD of 0.6-0.8 was achieved. The cultures were then induced with 1 mM IPTG and cooled to 16 °C and left overnight before harvesting the cells. The cells were harvested by centrifugation and washed with PBS and either stored at -80 °C or used immediately.

#### **2.9.1.2. Buffers**

**Cracking buffer:** 300 mM NaCl, and 20 mM 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), pH 7.4 at room temperature, 1 cOmplete protease inhibitor tablet (Roche)/50 mL, 0.5 mM PMSF, 1 µL Benzonase nuclease and 5 mM βME.

**Solubilisation buffer:** 500 mM NaCl, 10 % (v/v) glycerol and 50 mM Tris, pH 7.4, 1 % (w/v) DDM, 1 EDTA free cOmplete protease inhibitor tablet (Roche) and 0.5 mM PMSF.

**Buffer A:** 500 mM NaCl, 20 mM imidazole, 10 % (v/v) glycerol and 50 mM Tris, pH 7.4 and 0.025 % (w/v) DDM.

**Buffer B:** 500 mM NaCl, 500 mM imidazole, 10 % (v/v) glycerol and 50 mM Tris, pH 7.4 and 0.025 % (w/v) DDM.

**Gel-filtration buffer:** 50 mM Tris, pH 7.4, 150 mM NaCl, 10 % (v/v) glycerol and 0.025 % (w/v) DDM.

#### **2.9.1.3. Purification of GlpG and mutants**

Cell pellets were thawed and re-suspended in cracking buffer; the cell suspension was passed twice through a cell disruptor (Constant Systems) at 25 kPsi. Insoluble material was pelleted by centrifugation twice at 20,000 x g for 25 min at 4 °C. Membranes were pelleted by centrifugation at 150,000 x g for 1.5 hr at 4 °C.

Membranes were re-suspended to 40 mg/mL in ice-cold solubilisation buffer and homogenized using a Potter-Elvehjem Teflon pestle and glass tube. The membrane suspension with DDM was incubated overnight with gentle agitation at 4 °C, followed by 1 hr of centrifugation at 100,000 × g at 4 °C.

The supernatant was then filtered before loading onto a 1 ml HiTrap equilibrated in buffer A. The column was washed with 10 CV of buffer A, then 20 CV of 5 % buffer B.

GlpG was eluted with 100 % buffer B and the 1 mL elution peak was injected directly onto a Superdex 75 10/600 GL size exclusion chromatography (SEC) column (GE Healthcare) equilibrated in gel-filtration buffer. Peak fractions eluted from the SEC column containing GlpG were pooled and spin filtered before being flash frozen and stored at -80 °C or -20 °C). SDS-PAGE electrophoresis and Western blotting, with an anti-his antibody, assessed GlpG purification.

## **2.9.2. Expression and purification of XylE and associated mutants**

### **2.9.2.1. Growth and expression of XylE and mutants**

Genes encoding WT XylE, XylE-pGly28 and HaloTag-XylE from *E. coli* were cloned into expression vector pET28a producing a C-terminal His<sub>10</sub> tag, and subsequently transformed into BL21-AI cells (ThermoFisher). Cells were cultured in 6 x 1 L media in 2.5 L baffled flasks with 30 µg/mL kanamycin until cell density reached an OD<sub>600</sub> of 0.6-0.8. Cells were then induced with 1 mM IPTG and 0.1 % (w/v) L-Arabinose until growth arrest, harvested by centrifugation and washed in 1x PBS before resuspension in 50 mL PBS with 10 mM βME, and cOmplete protease inhibitor tablet and were frozen at -20 °C before purification.

### **2.9.2.2. Buffers**

**Solubilisation buffer:** 50 mM sodium phosphate pH 7.4, 200 mM NaCl, 20 mM Imidazole, 10 mM βME, 10 % (v/v) glycerol, 2 % (w/v) DDM, one cOmplete EDTA free protease inhibitor tablet and 0.1 mM PMSF.

**Gel-filtration buffer (A):** 50 mM sodium phosphate, pH 7.4, 2 mM βME, 10 % (v/v) glycerol, 0.05 % (w/v) DDM, 0.1 mM PMSF.

**Elution buffer (B):** 50 mM sodium phosphate, pH 7.4, 10 mM βME, 10 % (v/v) glycerol, 0.05 % (w/v) DDM, 0.1 mM PMSF, 500 mM Imidazole.

### **2.9.2.3. Purification of XylE and mutants**

Cells were defrosted and incubated at room temperature with 1  $\mu$ L Benzonase nuclease (ThermoFisher) for 10 min before being rested on ice. Homogenously thawed cells were then passed through a cell disruptor (Constant Systems) at 25 kPsi, 4 °C and fed into ice-chilled tubes. Cell membranes were harvested by 30 min ultracentrifugation at 100,000 x g, 4 °C. The membrane pellets were solubilised for 2 hr with mixing at 4 °C in solubilisation buffer. DDM insoluble material was removed by 30 min ultracentrifugation at 170,000 x g. Purification was carried out on an AKTA Pure system using a 1 mL Ni-NTA column, equilibrated in 96 % buffer A and 4 % buffer B (20 mM imidazole), bound protein was washed with 50 mL 85 % buffer A, 15 % buffer B (75 mM imidazole) then protein was eluted in 2 mL 100 % buffer B (500 mM imidazole). The 2 mL eluted was fed directly into Superdex 16/600 gel filtration column for protein oligomer removal. The column was equilibrated with 1.2 CV of buffer A, and 2 mL fractions of eluate were collected. Protein concentration was determined using a UV-Vis spectrophotometer (Cary) to obtain the protein  $A_{280}$ . The protein was spin-concentrated in 100 kDa cut off spin concentrators if necessary. Samples were flash frozen in liquid  $N_2$  and stored at -80 °C until use.

### **2.9.3. Expression and purification of separate XylE half domains**

The vectors containing the XylE halves were transformed into BL21-AI cells and grown as above. All purification protocols were also identical with minor variations at the chromatography stage. Samples were still loaded onto at 1 mL Ni-NTA trap, however C-terminal domain was washed with 25 mM imidazole, and N-domain was washed with 75 mM imidazole due to the differing length of histidine tags. The sample was eluted in 1 mL buffer B (500 mM imidazole) with up-flow and was directly injected onto a Superdex 10/30 increase gel filtration column. The column was equilibrated with 1.2 column volumes of buffer A and 1 ml fractions were collected. Eluted protein was spin-concentrated in 50 kDa cut off spin concentrators if necessary, flash frozen in liquid  $N_2$  and stored at -80 °C until use.

## **2.10. Reconstitution methods**

### **2.10.1. Preparation of lipid films**

5 or 10 mg DOPE, DOPC and DOPG lipids (all from Avanti Polar Lipids) were solubilised in cyclohexane aided by incubation on a 49 °C heat-block if necessary. Desired ratios were mixed, and flash frozen in liquid  $N_2$  before freeze drying overnight. On removal from the freeze dryer, a stream of  $N_2$  was used to preserve the lipid films before storage at -20 °C.

### **2.10.2. Reconstitution of FL protein into synthetic lipids**

Films were re-suspended with mixing in 50 mM sodium phosphate, pH 7.4. 10 mM MgCl<sub>2</sub> was added to AFM sodium phosphate sample buffers. Suspensions were extruded through 100 nm or 400 nm Millipore filters, with at least 21 pushes. 1 % (w/v)  $\beta$ -D-octyl glucoside (OG) from a 20 % (w/v) stock was added to aliquots of extruded liposomes and left to incubate with detergent for 30 min at RT. Protein was then added to a final concentration of 0.8 mg/mL. Excess OG was removed with buffer equilibrated spin-detergent removal columns (ThermoFisher) and then washed twice by ultracentrifugation at 440,000 x g in buffer containing 200 mM NaCl (for AFM) to remove incorrectly reconstituted protein. Salt and urea washes were not used for transport assays or synchrotron radiation circular dichroism (SRCD) data. Proteoliposomes were checked for size homogeneity using dynamic light scattering (DLS),

### **2.10.3. Solubilisation of proteoliposomes with DIBMA**

Proteoliposomes produced in **section 2.10.2** were solubilised for 3 hr at room temperature with 2.5 % (w/v) DIBMA on a rotating wheel. Insoluble material was removed using a 100,000 x g centrifugation step and the supernatant was bound to Ni-NTA beads, pre-equilibrated in reconstitution buffer with 20 mM imidazole, overnight at 4 °C. Non-specific material was removed with 20 CV of equilibration buffer, and sample was eluted in 5 x 1 CV samples of reconstitution buffer containing 500 mM imidazole. Imidazole was removed using a PD10 column pre-equilibrated with reconstitution buffer. The eluted sample concentrated in a 100 kDa spin-concentrator and stored at 4 °C.

### **2.10.4. Markwell-Lowry assay for reconstitution efficiency**

Reconstitution efficiency was checked using a Markwell-Lowry Assay (Markwell et al., 1978) with bovine serum albumin (BSA) as a standard and corrected for Xyle. Reconstitution efficiency for Xyle was 60-80 %.

BSA standards 0-25  $\mu$ g in 5  $\mu$ g intervals, and equal masses of WT Xyle or Xyle-pGly28 were treated with 10 mg/mL Na-deoxycholate and vortexed to solubilise lipids in samples. 10 % (w/v) trichloroacetic acid (TCA) was used to precipitate the protein and was pelleted. The TCA was drained, and the protein precipitate was suspended in 1 mL of alkaline copper reagent (200 mM Na<sub>2</sub>CO<sub>3</sub>, 100 mM NaOH, 7 mM KNa Tartrate, 1 % (w/v) SDS and 0.4 % (w/v) CuSO<sub>4</sub>·5H<sub>2</sub>O). The reaction was started on the addition of 100  $\mu$ L 50 % Folin reagent (v/v) and incubated for 1 hr before A<sub>750</sub> was recorded and the standard curves plotted. As 0  $\mu$ g of protein



should yield no colour change, these were used as a zero reading. The difference in gradients for BSA and XylE were used as a conversion factor for reconstitution efficiencies between proteins.

#### **2.10.5. Carbohydrate assay for OG and DDM in proteoliposomes**

A range of OG standards were made in dH<sub>2</sub>O volume of 50  $\mu$ L from 0 % to 0.04 % (w/v). 600  $\mu$ L of concentrated H<sub>2</sub>SO<sub>4</sub> was added to each sample including 50  $\mu$ L of each sample to be tested, mixed, and left to cool. 250  $\mu$ L of 5 % (v/v) phenol was added with extreme caution and a colourless to orange colour change occurred. The A<sub>490</sub> of each sample was measured and used to construct a linear calibration curve which was used to ascertain the OG concentration in each unknown sample (Taylor, 1995).

### **2.11. General biophysical methods for protein analysis**

#### **2.11.1. Isothermal titration calorimetry (ITC)**

ITC recordings were collected using a MicroCal 200 ITC (GE healthcare) with 2 x 20 injections of 2  $\mu$ L of 2.5 mM D-glucose or D-xylose, into 35.3  $\mu$ M WT XylE, both suspended in purification buffer A to avoid the large dilution signal caused by a buffer mismatch. Injections were spaced with 150 s. Data collection was carried out at 25 °C. Buffer to ligand, and ligand to buffer backgrounds were also carried out. The linear fit of the ligand to buffer was subtracted from the dataset as a baseline for the ligand-protein injections. Data was analysed and fit to a one-site binding parameter using Origin 2015. The first titration was removed from each run due to the loss of sample from the tip during equilibration.

#### **2.11.2. Circular dichroism (CD) and synchrotron-radiation CD (SRCD)**

For secondary structure characterisation, oriented CD (OCD) and synchrotron-radiation CD (SRCD) was measured using the CD12 beamline at ANKA (Karlsruhe Institute of Technology, Germany), or on an Aviv benchtop instrument, model 410.

Proteins were scanned at 0.4 mg/mL for WT XylE or XylE pGly28 in 0.05 % (w/v) DDM buffer (50 mM sodium phosphate pH 7.4 for XylE) from 2-4 times in 0.1 mm CaF<sub>2</sub> cells (SRCD) or varying path length quartz cells (0.05-0.2 mm) (Aviv instrument), from 270-180 nm at 1 nm step intervals with an averaging time of 0.5 s at 25  $\pm$  0.5 °C. Next, multiple scans of the background buffer were taken in the same pathlength cell and averaged, zeroed between 253-260 nm and smoothed using CDtool software (Lees et al., 2004).

Mean residue ellipticity (MRE) was calculated using **Equation 2-1** with an average MRW of 110 g/mol, and data analysis was carried out using CDtool (Lees et al., 2004) and GraphPad Prism 7.03. Spectra were deconvoluted using the Dichroweb web server (<http://dichroweb.cryst.bbk.ac.uk>), with reference dataset SMP180 using the CDSSTR method (Abdul-Gader et al., 2011).

*Equation 2-1: Mean residue ellipticity (MRE) calculation*

$$MRE = \frac{mdeg \times MRW}{c \left( \frac{mg}{mL} \right) \times l (mm)}$$

### **2.11.3. Dynamic light scattering (DLS) for particle size**

Approximately 10 µL of liposome or nanodisc samples were diluted into an appropriate buffer (50 mM sodium phosphate buffer pH 7.4 with or without 0.05 % (w/v) DDM) and run in particle size mode on a Malvern ZetaSizer DLS. The attenuation of the laser should typically be 9.0, and the dilution was adjusted to contain more sample if the attenuation was too high. Data was displayed as either number, intensity, or volume of particle hits. The z-average (d.nm) and the average particle size samples were obtained from the displayed data.

## **2.12. Biochemical assay methods for protein analysis**

### **2.12.1. EnzCheck protease assay for GlpG activity**

Detailed GlpG activity assay methods can be found in Harris, et al, 2017 (Harris et al., 2017b) and Reading, et al, 2017 (Reading et al., 2017). Briefly, 5 µg/mL EnzCheck BODIPY-Casein substrate (ThermoScientific) was cleaved by 1 mg/mL GlpG, overnight and in the dark at 4 °C in GlpG size-exclusion buffer (**section 2.9.1.2**). The resulting fluorescence emitted at 510 nm, when excited with 480 nm light was recorded on a Horiba Fluoromax 4, with 1 nm slit widths, and a 1 s integration time. This was a measure of the reduction of self-quenching of BODIPY as casein is cleaved, indicative of GlpG protease activity. GlpG activity was inhibited with the addition of 100 µM diisopropyl fluorophosphate (DFP).

### **2.12.2. Xylose dehydrogenase assay**

Xylose dehydrogenase (XDH)/xylose mutarotase (XMR) (Megazyme) enzyme mix was separated into component proteins using a Superdex 10/300 increase 200 PG pre-equilibrated in PBS pH 7.4. Fractions under the major elution peak were collected and concentrated in a 10

kDa MW cut-off concentrator before freezing at -80 °C. The purity of each peak was checked using SDS-PAGE analysis.

The activity of the stock solutions was calculated using scanning kinetics, from 400-240 nm every 30 s for 15 min on a UV spectrometer (Cary). 50 µL pure XDH was incubated with 2 mM NAD<sup>+</sup>, and 60 mM D-xylose in PBS buffer pH 7.4 and scanned in a 0.1 mM quartz cell. The increase in absorbance at 340 nm was recorded over time, and the Beer-Lambert equation was used with extinction coefficient (ε) of 6220 M<sup>-1</sup>.cm<sup>-1</sup> used to calculate the activity of the enzyme with respect to xylose turned over.

*Equation 2-2: Beer-Lambert equation*

$$c = \frac{A}{\epsilon L}$$

For reconstituted samples, the methods were followed as **section 2.10.2**, but with 2 mM NAD<sup>+</sup>, and 0.25 mg/mL XDH enzyme in the reconstitution buffer, and the lipids were subject to 3 freeze-thaw cycles using liquid N<sub>2</sub> and extruded to 400 nm. Excess NAD<sup>+</sup> and XDH was removed by dilution into PBS with an additional 100 mM NaCl and harvested by centrifugation at 95,000 rpm in a Beckmann TLA100.3 rotor.

Harvested liposomes containing XDH and NAD<sup>+</sup> were reconstituted with 0.3 mg/mL WT Xyle pre-saturated with 1 % (w/v) OG for 30 min as described in **section 2.10.2**. A mock reconstitution was run in parallel to assay the amount of Xyle incorporation using the Markwell-Lowry assay. Excess OG was removed using pre-equilibrated detergent removal spin columns.

To assay for activity, fluorescence kinetics were measured in a Horiba Fluoromax 4. The reaction was excited at 340 nm using 2 nm slits, and emission was recorded at 460 nm using 5 nm slits, all with a 1 s integration time with anti-photobleaching mode switched on. The 4 mm<sup>2</sup> quartz cell held a 300 µL reaction volume. 250 µL of pH 6.0 PBS (to provide a pH gradient across the bilayer) was mixed with 50 µL of the reconstituted sample and mixed. The concentration of NADH produced was quantified from the fluorescence curve, using a standard curve of fluorescence against NADH, recorded using the same experimental set up, with the same concentration of empty liposomes to account for light scattering. The concentration of NADH was used to calculate the moles of Xyle turned over, and hence the activity of the transporter as xylose is transported into the liposome as a function of time. A mock

reconstitution, without protein, but with the same OG pre-saturation treatment was used as a baseline for these experiments.

### **2.12.3. HaloLink Resin binding to Halo-TMH-Xyle**

400  $\mu\text{L}$  of HaloLink resin (Promega) bed volume was equilibrated with 3 CV of 50 mM Na-phosphate pH 7.4. 50  $\mu\text{L}$  of 0.3 mg/mL sample of DIBMA or DDM solubilised Halo-TMH-Xyle was incubated with the beads in purification buffer for 30 min at room temperature. Beads were harvested with 1 min 1500 x g centrifugation steps. The supernatant 'unbound' was taken for analysis. The beads were washed three times with 1 CV binding buffer (50 mM sodium-phosphate pH 7.4, 100 mM NaCl and 0.5 mM EDTA). The washes were also collected for analysis by SDS-PAGE.

### **2.13. Atomic force microscopy (AFM) methods**

All SMFS and imaging experiments were carried out on a Bruker BioScope Resolve BioAFM.

#### **2.13.1. Sample deposition**

50-90x dilutions of liposome or proteoliposomes (as per **section 2.10.2**) were made and subject to bath sonication separate any liposomes that had fused over time in the  $\text{MgCl}_2$  buffer. 50  $\mu\text{L}$  of each dilution was incubated for 30 min onto freshly cleaved mica substrate attached to a Teflon covered steel disc for use with a magnetic AFM stage, at 25  $^\circ\text{C}$  before washing 20 times with a total volume of 1 mL of liposome buffer. The dilution was necessary for 100 % coverage of mica and limiting the chance of depositing multilamellar lipids.

HaloTag containing constructs were tethered to functionalised glass fluid cells rather than mica. *Fluid cells were prepared by Sam Gulaidi-Breen and Dr Amy Beedle of the Booth and Garcia-Manyes groups.* Briefly, Glass coverslips were cleaned with sonication in 1 % (v/v) Hellmanex II, acetone and ethanol and thoroughly washed with ddH<sub>2</sub>O between each solution before rapid drying under nitrogen and heating to 70-80  $^\circ\text{C}$  overnight. Silanization was achieved by incubating plasma cleaned coverslips with 1 % (w/v) 3-aminopropylsilane in MeOH for 25 min and thorough washing with MeOH prior to drying and heating for 40 min at 80  $^\circ\text{C}$ . The cover was then incubated with 1 % (v/v) glutaraldehyde solution in PBS and left for 1 hour. Excess glutaraldehyde was rinsed with PBS and dried. 100  $\mu\text{g}/\text{mL}$  O<sub>4</sub> Halo ligand (Promega) was passed over the cover to a volume of 300  $\mu\text{L}$  and left to incubate at RT overnight. Covers are then passivated with 1 % (w/v) BSA in Tris and stored in the fridge overnight. On the day of use, covers are dried by pipette and Halo construct is incubated and washed as was carried out with mica.

### 2.13.2. Calibration of cantilever in fluid

An AFM requires two essential components to work, a cantilever, which acts as a flexible detector, and a piezoelectric attenuator, which provides accurate nanometre control of sample. A sample deposited on a piezo stage is approached by a sharp cantilever mounted onto a holder by a piezo electric device. A focused laser beam is reflected off the back of the cantilever onto a photodetector, and as a sample is scanned, deflections of the cantilever – hence differences in the laser detected by the photodiodes, is converted into a voltage, and the output is displayed as a volage difference as the sample surface is scanned, and the laser spot moves. See **Figure 6-1**.

The angular deviation of the laser is measured at the detector, and as the cantilever behaves and a Hookean spring, where the force ( $F$ ) is the product of cantilever deflection in the z-axis ( $\Delta z$ ), and its spring constant ( $k_s$ ),  $F$  can be calculated using **Equation 2-3**. To accurately calculate this force for each force peak in each experiment, the  $k_s$  must be determined for the cantilever for each new experiment – see **section 2.13.2**.

*Equation 2-3: Hooke's law*

$$F = -k_s \cdot \Delta z$$

If the AFM is programmed to maintain a constant force, or change the force, the feedback electronics can adjust the z-height of the piezo to correct for this based on the output voltage generated by the photodetector.

The spring constant varies between different cantilevers and must therefore be calibrated for each experiment, or each time the cantilever is moved on its holder. Calibration is a two-step process, which allows for an accurate calculation of force using Hooke's law.

The first stage of calibration describes the relationship between the photodiode output voltage ( $\Delta V$ ) and the displacement of the piezo in the z-direction ( $z_p$ ). This is achieved using a series of approach/retract cycles of relatively high force (1-1.5 nN) and velocity (2  $\mu\text{m/s}$ ), conditions such that the displacement of the cantilever is equal to the change in piezo distance ( $\Delta z_c$ ). The slope of these parameters can be calculated ( $\Delta V/\Delta z_p$ ) as the deflection sensitivity of the cantilever, which is dependent on the position of the laser spot, and the refractive index of the medium used.

The spring constant is estimated from its thermal noise using the equipartition theorem (Butt and Jaschke, 1995) which states that the total energy of a system is shared amongst all degrees of freedom in our experiments, cantilever calibration was calculated using the thermal fluctuation method, which assumes that at equilibrium, the cantilever acts as a simple harmonic oscillator, fluctuating only in response to thermal noise (Hutter and Bechhoefer, 1993, Lévy and Maaloum, 2002). For calibration, this means, kinetic energy can be related to thermal energy through **Equation 2-4**, where  $k_B$  is the Boltzmann constant,  $T$  is the temperature of the system,  $k_s$  is the spring constant, and  $z$  is the displacement of the cantilever (Hughes and Dougan, 2016).

*Equation 2-4: Simple harmonic motion with one degree of freedom for AFM cantilever*

$$\frac{1}{2}k_s\langle z^2 \rangle = \frac{1}{2}k_B T$$

To obtain the  $k_s$  using the thermal tune method, the cantilever was placed above the sample surface in media, and the power is ramped to produce a spectrum which is fast Fourier transformed and fit to a least-squares fit for a simple-harmonic oscillator model for use in fluid. Rearranging **Equation 2-4** demonstrates that the spring constant can be calculated using mean square of the displacement which is equal to area under the power spectrum **Equation 2-5**.

*Equation 2-5:  $k_s$  calculation from power spectrum*

$$k_s = \frac{k_B T}{\langle z^2 \rangle}$$

Once the calibration is complete, imaging, FX and FC experiments can all be carried out. Cantilevers can have varied properties and they should be chosen for the type of experiment and substrate accordingly, for example. Cantilevers with low spring constants, are ‘soft’ and very sensitive to the low forces experienced when unfolding proteins, but alternatively, can be significantly affected by noise, especially when close to the substrate surface, and are commonly prone to instrumental drift throughout the lifetime of an experiment. Stiffer cantilevers with high spring constants are commonly used, with a higher resonance frequency for AFM imaging to obtain crisp images of the surface but are not deflected by the surface which would result in low resolution images. Most cantilevers for high resolution imaging are ultra-sharp, which allows more detailed structures to be resolved, due to the smaller width of the tip.

### 2.13.3. AFM imaging

AFM imaging was carried out on bilayers to check for coverage and depth. High resolution imaging of proteins in bilayers and DIBMA discs was also carried out. SNL-B ( $k_s = 0.12$  N/m,  $f_o = 23$  kHz) or Ultrasharp Peakforce HIRS-B ( $k_s = 0.12$  N/m,  $f_o = 60$  kHz, tip width) Si<sub>3</sub>N<sub>4</sub> cantilevers (Bruker) were used to scan 500 nm<sup>2</sup> - 3 μm<sup>2</sup> areas of bilayer in constant tapping AFM mode with a peak force of 70 pN, and standard scan frequency of 1 Hz to obtain clear images without penetrating the bilayer.

Imaging was primarily used to check quality of bilayers and deposition prior to SMFS to protect the cantilever from damage. A good deposition was defined a 3 μm<sup>2</sup> area with no lipid islands or 'holes' in at least 3 separate scan areas – at least 200 nm<sup>2</sup> clean area is required for FX to avoid tip damage if mica is probed.

Images were analysed using Nanoscope Image Analysis 1.4 where flattening of images and colour adjustment was carried out. If required, the section tool was used to determine the depth of image features. This was used for bilayer depth and protrusion of the proteins from the membrane. When looking for the horizontal size of the images, the width of the AFM tip was considered, as half of the tip width is accounted for either side of the observed feature.

### 2.13.4. Force/extension (FX) mode

#### 2.13.4.1. Data collection

Force extension was carried out with MLCT-E ( $k_s = 0.1$  N/m,  $f_o = 38$  kHz), DNP-C ( $k_s = 0.24$  N/m,  $f_o = 56$  kHz) or its sharper SNL-C ( $k_s = 0.24$  N/m,  $f_o = 56$  kHz) cantilevers, and was carried out in Peakforce QNM (quantitative nanomechanical property mapping). Each cantilever was calibrated using a thermal tune to calculate the deflection sensitivity (nm/V) and hence spring constant for each experimental set-up, or new cantilever used (see **section 2.13.2** for details). The cantilever approached the bilayer with a velocity of 400 nm/s, held for 1 s at the surface with a contact force of 1 nN to promote non-specific attachment to the protein backbone. The tip was then retracted at 400 nm/s for 500 nm ensuring the complete protein can unfold in the trajectory.

#### 2.13.4.2. Data analysis

Data analysis of FX data was carried out using a combination of Nanoscope analysis 1.4, Igor Pro 6.37 and Origin 2015. Each trajectory was baselined from the 400-500 nm extension region and exported to Igor Pro for peak fitting. Trajectories with pick-up, and longer than 150 nm were selected for analysis, where the worm-like chain model of polymer elasticity

(WLC) **Equation 6-1** was applied to each peak for further detail on the suitability of WLC to my data, see **section 6.1.2.2**. For my experiments, a persistence length ( $p$ ) of 0.6 nm was selected as this appeared to fit my curves better than the literature value of 0.36 nm. The resulting contour length was used to back calculate the number of amino acids unfolded in each unfolding segment using a single amino acid length of 0.38 nm (Mora et al., 2020).

Each trajectory was grouped into like-characteristics and the contour lengths ( $L_c$ ) and rupture forces ( $F$ ) binned and plotted in a bivariate histogram using Origin 2015 or Igor Pro 6.37.

### **2.13.5. Force/clamp (FC) mode**

#### **2.13.5.1. Data collection**

Force-clamp data acquisition was also run in Peakforce QNM mode with MLCT-E ( $k_s$ : 0.1 N/m,  $f_o$ : 38 kHz) cantilevers. Cantilevers were also calibrated using a thermal tune method. A macro was written for the time resolved unfolding of individual proteins at a constant force of 65 pN. The cantilever was held at the surface of the bilayer for 2 s at 0 pN to attach protein. The tip was then retracted at 65 pN and held for 10 seconds, and finally 65 pN was maintained whilst retracting the tip to 800 nm displacement producing a stepwise unfolding trajectory for WT XylE.

#### **2.13.5.2. Data analysis**

Using the unfolding trajectory for XylE in the FX mode, the step size occurring at each event can be estimated, which is dependent on the force applied to the protein. Using Nanoscope 1.4 or Igor Pro 6.37, each trajectory step height was accurately measured and compared with the theorized step sizes from the WLC fits.



### 3. A strategy for the generation of membrane protein RNCs

---

**This chapter, along with Chapter 4, has been modified from publication with additional details and discussions. The full article can be found in appendix 10:**

Capturing membrane protein ribosome nascent-chain complexes in a native-like environment for co-translational studies

**Pellowe GA**, Findlay HE, Lee K, Gemeinhardt TM, Blackholly LR, Reading E, Booth PJ, *Biochemistry*, 2020

doi: [10.1021/acs.biochem.0c00423](https://doi.org/10.1021/acs.biochem.0c00423)

### 3.1. Introduction

#### 3.1.1. Rhomboid protease – GlpG

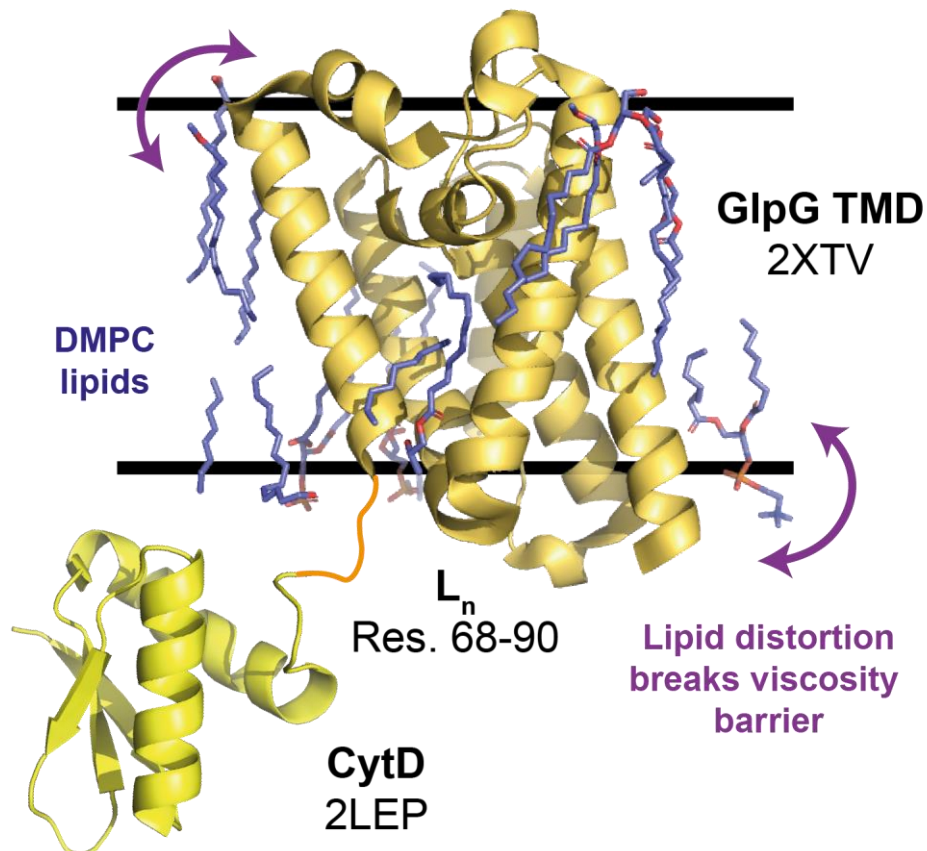
This chapter develops a method for the design and purification of a membrane protein RNC using SecM-based translational stalling in DDM detergent. The polytopic  $\alpha$ -helical protein GlpG is an intramembrane rhomboid protease of *E. coli* **Figure 3-1**, which is very well studied and forms the basis of the RNC work presented in this thesis. Rhomboid proteases degrade other membrane proteins in the bilayer, this protease family is ubiquitously expressed across all kingdoms, and has major roles in a diverse range of cellular functions, from cell signalling, apoptosis and bacterial signal transduction to name a few, and are often implicated in disease (Urban, 2016).

##### 3.1.1.1. Structure and dynamics

GlpG is formed of 6 TM helices which span the membrane, bound to a mixed  $\alpha/\beta$ -structure N-terminal cytoplasmic domain (CytD). These two domains are connected by a flexible, functionally relevant linker,  $L_n$ . Crystals structures have been solved for both the core TM domain and CytD separately, but not as a single polypeptide chain. The first crystal structure of the GlpG core TM domain was solved using x-ray crystallography to 2.1 Å resolution (Wang et al., 2006) showing the serine-histidine catalytic dyad approximately 10-12 Å deep in the bilayer, with a possible gating mechanism at the loop between helices 1 and 2. This structure was later improved to 1.7 Å of an inactive S201T GlpG mutant in the presence of DMPC/CHAPSO lipid bicelles, which showed differences in loop region structure compared to the detergent only structure. The TM helices largely remained the same, other than TMs 1 and 6 which changed in response to lipid crystal packing (Vinothkumar, 2011). It was shown that in the lipid bilayer environment, the protein was embedded asymmetrically in the bilayer, with the protein in an up/down orientation, which has now be shown as a means to distort lipids due to the irregular rhomboid shape of the protein to overcome the viscosity barrier allowing fast diffusion, of  $1.2 \pm 0.17 \mu\text{m}^2/\text{s}$  through a 70:30 POPE:POPG lipid environment (Kreutzberger et al., 2019).

The role of the cytoplasmic domain of GlpG is not clear, however the structure has also been solved and provided some clue into its function in the rhomboid family, where the cytoplasmic domain can exist on the N or C terminal (Sherratt et al., 2012). For GlpG, CytD has been solved using NMR, showing a compact 70 amino acid structure, with the remaining 20 residues being unstructured (PDB: 2LEP (Sherratt et al., 2009)). CytD removal from the TM domain does not abolish function, however the structure suggests binding interactions with the membrane and

TM domain, suggesting a possible gating or substrate recognition function for the CytD domain (Sherratt et al., 2012). There is currently no structural information for the flexible  $L_n$  linker region (Sherratt et al., 2012). Other structures of the TMD from *Haemophilus influenzae* (Lemieux et al., 2007), and the cytoplasmic domain from *Pseudomonas aeruginosa* (Del Rio et al., 2007) have also been solved using x-ray diffraction and solution NMR, respectively.



**Figure 3-1: Structure of the GlpG Rhomboid protease**

*GlpG TMD* consists of 6 helices (PDB: 2XTV (Vinothkumar, 2011)) and is shown here in gold. The associated DMPC lipids in the crystal structure are also shown. Membrane boundaries are calculated and displayed by OPM (Lomize et al., 2012). The angle and unique structure of the rhomboid distorts the lipids either side of the protein to break the viscosity barrier of the membrane to increase diffusion speed. Residues 68-90 of the full length *GlpG* sequence correspond to the flexible linker  $L_n$ , shown here in orange. Currently there is no structure for this sequence. *CytD* shown in yellow (PDB: 2LEP (Sherratt et al., 2009)) had its structure solved separately, and whilst its specific function is unknown, it is thought to have a role in substrate specificity or gating.

The crystal structures of *GlpG CytD* reveal a domain swapped dimer (Ghasriani et al., 2014). This structure may be physiologically relevant as it appears to be dimeric, not only in solution but also when in contact with the membrane domain of the rhomboid, suggesting that the dimerization is a key step for regulation of intramembrane proteolysis (Cournia et al., 2015). Models have been suggested for dimerization at the TM interface, as well as in the cytoplasmic

domain. The latter stabilizing the TMD domain, providing conformational changes required for proteolysis in the active site.

More recently however, GlpG purified into DDM micelles and SMA-based native nanodiscs (SMALPS) which contain the protein with the native lipid environment, was shown to be monomeric only, using laser Induced Liquid Bead Ion Desorption-MS (LILBID-MS), as well as native SDS-PAGE gels. The protein ran slightly higher than its expected 34.5 kDa mass on the gel, however the LILBID spectra, under soft laser conditions, showed a broad GlpG peak with associated lipids. Lipids were then stripped away by increasing laser power to reveal the expected GlpG molecular weight (Hellwig et al., 2018). This, coupled with previous assays where it was shown that GlpG contains around 100 associated lipids in a SMALP, suggest that the extra mass shown in the gel was the effect of extra associated lipids, rather than dimeric GlpG in these two membrane environments (Reading et al., 2017).

HDX-MS has also been used to study the dynamics of GlpG when captured in 3 different native lipid compositions where changes in accessibility and protein dynamics were identified. Specific regions of the protein that were sensitive to changes or were strongly influenced by changes to the native lipid environment were also observed. In particular, this was shown with the CytD,  $L_n$  and TM 1 regions as measured by a higher deuterium uptake, with TM 1 having previously been shown to be important for the structural stability of the protein (Reading et al., 2017, Reading, 2018, Ha et al., 2013, Baker and Urban, 2012). This newly established HDX-MS technique is a growing area of interest and future directions would hopefully take an approach more focused on membrane protein folding and dynamics of folding and insertion.

### **3.1.1.2. Folding studies**

GlpG has become a model membrane protein for folding and activity studies (Choi et al., 2019, Min et al., 2015, Paslawski et al., 2015, Harris et al., 2017b, Guo et al., 2016, Hong et al., 2010), due to its requirement for membrane immersion for correct substrate specificity and function (Moin and Urban, 2012) and therefore makes a good candidate for the generation of RNCs. However, the quality of the past folding work for GlpG is limited. Extensive extrapolations of very small amounts of secondary structure loss as characterised by CD, and a slight reduction in intrinsic tryptophan fluorescence on WT GlpG from a DDM to SDS denatured state has been used to extensively characterise unfolding thermodynamics and phi-value analysis (Paslawski et al., 2015). It has also been shown in an independent study that GlpG in DDM is much less stable in detergent than in a bilayer environment, also characterised using CD (Moin and Urban, 2012).

Energy landscapes have been extracted for GlpG using data from lateral membrane pulling force-ramp experiments using MT spectroscopy (Min et al., 2015, Choi et al., 2019) (see **chapter 6**). These experiments suggest that the protein inserts via rapid sequential helical hairpins with the rate limiting step being the final two helices inserting. This was also studied experimentally from a co-translational perspective using the SEIRAS technique, to uncover a time-resolved folding pathway of the protein (Harris et al., 2017b). This study suggested that GlpG forms a folding core of the first two helices, with the remaining protein collapsing into the membrane and the final helices 5 and 6 forming the inter-helix interactions to complete the fold. The Wolynes' lab sought to confirm experimental methods of GlpG folding landscapes by simulating the energy landscapes under high lateral force pulling *in silico* (Lu et al., 2018) which also suggested a folding core and hydrophobic collapse of the remaining protein.

Wolynes' methodology can predict higher resolution folding events than SEIRAS as the amide bands cannot discriminate between distinct regions of secondary structure folding. The energy landscape was constructed for a high force extraction (based on Bowie et al (Min et al., 2015)) from the membrane and showed two intermediate folding states,  $I_1$  and  $I_2$ , and two unfolded states,  $U_1$  and  $U_2$ , as determined by the free energy minima when plotted against distance between GlpG C- and N-terminus. Firstly, the large energy barrier between folded protein and  $I_1$  is overcome, this corresponds to helices 5 and 6 unfolding from helices 1-4, thus suggesting the rate limiting step for GlpG is pulling the final helical hairpin through the membrane and forming interactions with helices 1-4. Next, helices 3-6 unfold and interact with the bilayer ( $I_2$ ), followed by the intact helices 1 and 2 separating ( $U_1$ ) before  $U_2$  describes the complete unfolded state. The two unfolded states, with free energies lower than  $5 k_B T$ , agree with Harris et al, that hydropathy plots suggest that TMs 1-2 of GlpG are much more hydrophobic than the rest of the protein and thus insert rapidly (Harris et al., 2017b) forming the hydrophobic core, and solidifying most past work conclusions with both co-translational and *in vitro* folding studies of GlpG.

### 3.1.1.3. Purity

The purity of membrane protein preparations is important to check for folding studies, they are prone to degradation and aggregation when they are removed from the cell membrane. Unfortunately, very few papers show the chromatogram, or SDS-PAGE gels of the purified protein they are working on, and to my knowledge, only one (Lemberg et al., 2005) has been published outside the Booth group. Therefore, samples used in past studies may contain a degree of aggregation, or degradation which I do not detect in the protein preparations in this

thesis, with methods based on work already happening in the Booth and Reading groups at KCL (Reading et al., 2017, Harris et al., 2017b). Additionally, it is disputed that GlpG can exist as a homodimer *in vivo*, and a few mechanisms of this interaction are discussed, reviewed in (Cournia et al., 2015). However, recent studies in systems similar to experimental work presented in this thesis have shown that the *E. coli* GlpG does not exist as a dimer when purified into DDM, or when extracted from the native membrane using SMA (Hellwig et al., 2018, Reading et al., 2017). Lots of the early folding studies on GlpG were carried out in detergent micelles. One study showed that the GlpG helices in a detergent environment have flexible helices, which is not the case in a bilayer environment as determined using CD (Moin and Urban, 2012), suggesting GlpG can be fussy about the membrane environment it is used in, although it is seemingly a very stable when purified.

#### **3.1.1.4. Activity**

For folding work, it is important to check the activity of proteins being worked on as it is known that correct activity of the protein is dependent on the correct folding of the protein, and an inactive protein being studied for folding pathways may lead to following incorrect pathways. Most activity studies for GlpG have been carried out on the TMD domain only, which is active, however the full-length protein, purified into detergent from *P. aeruginosa* is more active when the CytD domain is present (Sherratt et al., 2009). It is not clear what the native substrate for GlpG is in the membrane and providing a densely packed membrane environment with rhomboid substrate is difficult to achieve, and activity assays must be altered. It has previously been shown that GlpG TMD is able to actively cleave BLA-spitz-MBP, or GFP-spitz-FLAG (Moin and Urban, 2012) construct, which use a LacY derived transmembrane region, which is cleaved resulting in a fluorescence SDS-PAGE gel shift when GlpG is active and cleaves the TM helix (Sherratt et al., 2012, Maegawa et al., 2005).

Full length, isolated GlpG (including CytD), has been shown in the Booth and Reading groups to be active through the cleavage of a soluble BODIPY-casein substrate. While not transmembrane spanning, this assay gives a large fluorescent readout as GlpG degrades Casein promoting a reduction in self-quenching of BODIPY and a larger fluorescence increase to show that protease activity is present in the purified protein (Reading et al., 2017, Harris et al., 2017b). It was shown that the native membrane surrounding GlpG is also paramount for correct full-length protease activity and can be used as an end-point assay, or kinetics assay. The activity to degrade the BODIPY-casein substrate is much higher in DDM solubilised membrane compared to that of a SMA native membranes (Reading et al., 2017), suggesting

that the native membrane is required to prevent over-digestion of substrate. It was also shown that the water-soluble substrate was cleaved to a lesser extent in the TMD only GlpG, showing that CytD is required for efficient cleavage of soluble substrates (Sherratt et al., 2012).

For correct protease activity *in vivo*, the rhomboid shape of GlpG distorts the surrounding lipids which can break the diffusion barrier (Kreutzberger et al., 2019), and further showing that the membrane environment surrounding GlpG is incredibly important.

### **3.1.2. Considerations for membrane protein RNC production**

The difficulty of working with membrane protein RNCs for folding studies is the requirement to provide and maintain a suitable membrane mimic thorough the entire preparation protocol, as well as any downstream experiment. Producing soluble protein RNCs *in vitro* using IVTT is relatively straightforward and has become routine in many labs, with the complexity increasing for production *in vivo*. Membrane proteins pose their own challenges for each of these methods, and many of the strategies and experimental methods explored in **chapter 1** are very technically challenging, even more so when membrane protein RNCs are the subject of the experiment. For this reason, very few people have carried out any folding work using membrane protein RNCs with the little that has been achieved already highlighted in **section 1.5.4**. This chapter concerns itself with the development of a new method to produce GlpG membrane protein RNCs *in vivo* and purify for *in vitro* folding studies. The methods presented here serve as a prerequisite to produce similar complexes in a native nanodisc environment, which until now has not been achieved for a polytopic protein, or by using DIBMA co-polymer as the vehicle for a native membrane nanodisc, which provides the most physiologically relevant environment for folding. This work is explored in **chapter 4**. However, initially as described in this chapter, I benchmark this method using DDM detergent.

### **3.1.3. Adapting GlpG for RNC study**

The *E. coli* rhomboid protease GlpG is used here to benchmark my strategy for RNC generation due to its well characterised structure, folding pathways, and function. An advantage of GlpG having a defined crystal structure, particularly for the TMD allows for continuity between the MPEx helix predictions, along with an accurate structure to ensuring correct protein engineering when designing stable stalling positions.

To ready GlpG for RNC study, an N-terminal affinity tag for purification away from other *in vivo* cell proteins and components is required. Current constructs (Harris et al., 2017b, Reading et al., 2017) contain GlpG with C-terminal Histidine and Myc tags, which for RNC

study will not have been translated and therefore cannot be used in this project. N-terminal tags on membrane proteins can be tricky to deal with, many membrane proteins contain sorting sequences which are recognised by insertase or translocation machinery for cellular trafficking to the correct location. If the required tag interferes with this process, low yields or cell toxicity can occur and RNCs would not be amenable to purification from *in vivo*. The N-terminal tagging may also interfere with protein activity. If the purified protein is not active, it may suggest incorrect folding and therefore studying its folding pathway is irrelevant.

Strep tags are commonly used for the purification of soluble RNCs (Schaffitzel and Ban, 2007, Kater et al., 2019, Bischoff et al., 2014) however a poly-histidine tag was chosen for this work due to ease of purification with Ni-NTA and the need for a secondary tag for purification and high affinity tethering for downstream applications, such as mechanical based tethering (Kaiser et al., 2011, Goldman et al., 2015, Min et al., 2018, Min et al., 2015, Choi et al., 2019, Chang and Bowie, 2014), or ability to withstand low pH forces for hydrogen-deuterium exchange applications (Reading et al., 2017, Jensen et al., 2013). To achieve this, I decided to incorporate an avi-tag, a 15-residue amino acid sequence (GLNDIFEAQKIEWHE), which is targeted by the biotin ligase enzyme to covalently link a D-biotin molecule into the lysine residue at position 10. Cloning this 15-residue peptide tag onto the N-terminus of the RNC construct, in addition to the poly-histidine tag, allows efficient targeted *in vivo* biotinylation when the RNC is co-expressed with biotin ligase. This allows for protein detection, immobilisation, and co-immunoprecipitation studies of the product with any associated factors (Ashraf et al., 2004).

#### **3.1.4. Aims of chapter**

This chapter explores the necessary steps for the successful design and purification of GlpG RNCs in detergent, at differing stages of translation to assess their validity for further preparation and study in native lipid membranes (**Chapter 4**). The RNCs follow a relatively simple purification workflow which utilises affinity purification chromatography, gel-filtration chromatography, or sucrose gradient purification to produce high-yielding homogenous sample amenable to structural and biochemical investigation. I can determine homogeneity of the purified RNCs using SDS-PAGE analysis for ribosome Coomassie staining and negative staining, as well as Western blotting the nascent chain directly to determine the degree of translational stalling. I also develop an *in vivo* biotinylation system for additional tagging of RNCs for further purifications, or downstream investigations using force microscopy where stable tethering is required, or HDX mass spectrometry, where the



biotin/streptavidin bond can withstand high forces, and low pH buffers which are necessary to prevent back exchange from deuterium to hydrogen. The overarching aim of this chapter is to develop the workflow to produce high yields of clean RNCs for the downstream structural and biochemical study of co-translational folding from an *in vivo* aspect at varying positions throughout the co-translational folding cycle.

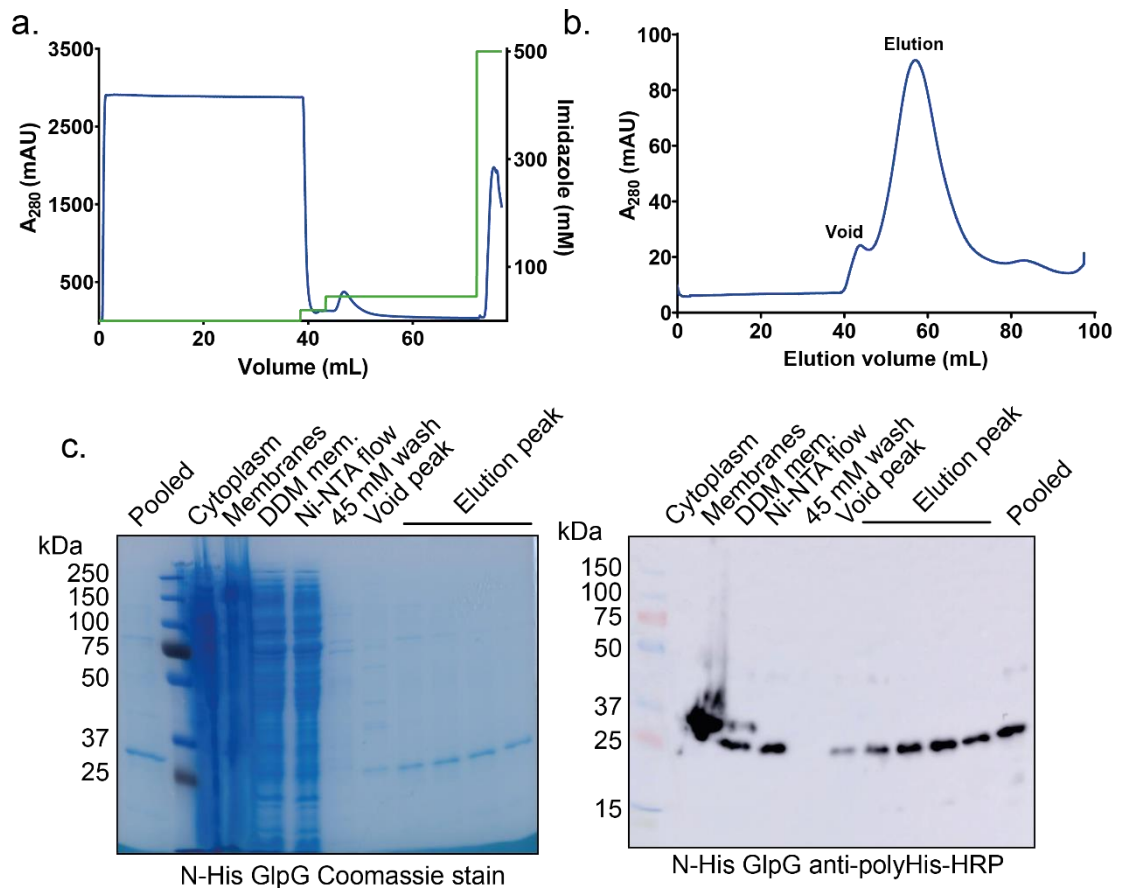
## **3.2. Results**

### **3.2.1. N-His GlpG expression and purification in DDM**

N-terminally tagged GlpG is necessary for the purification of RNCs as this is the only terminal exposed from the ribosome exit tunnel during translation. GlpG is a toxic protein and it may be assumed that including the additional N-His affinity tag was expected to interfere with the expression, folding and function of the protease, however I have shown that GlpG did in fact express and purify to the similar high yields and quality of GlpG C-His.

The isolated GlpG purification method was developed for HDX studies of GlpG in SMA-based native lipid nanodiscs (Reading et al., 2017). The protein was purified from DDM solubilised membranes using Ni-NTA. The Ni-trap was washed with 45 mM imidazole to remove non-specifically bound protein. The N-His-tagged GlpG was eluted from the trap using 500 mM imidazole (**Figure 3-2a**) and was then directly injected onto a Superdex 16/600 column for further purification which results in the characteristic chromatogram shown in **Figure 3-2b**, with a small void peak at 42 mL and a large but broad peak at 58 mL.

When collected and concentrated, the N-His GlpG peak was analysed for purity using 12 % Bis-Tris SDS-PAGE with low-pH 5.7 SDS-loading dye (for continuity with RNC requirements) and Western blotting, showing that the poly-histidine tagged protein was monomeric, successfully trafficked to the membrane, and was able to be solubilised in DDM. The fractions under the elution, and pooled and concentrated eluates were clean, and the void peak contained aggregated GlpG which was broken up by SDS in the PAGE gel (**Figure 3-2c**).



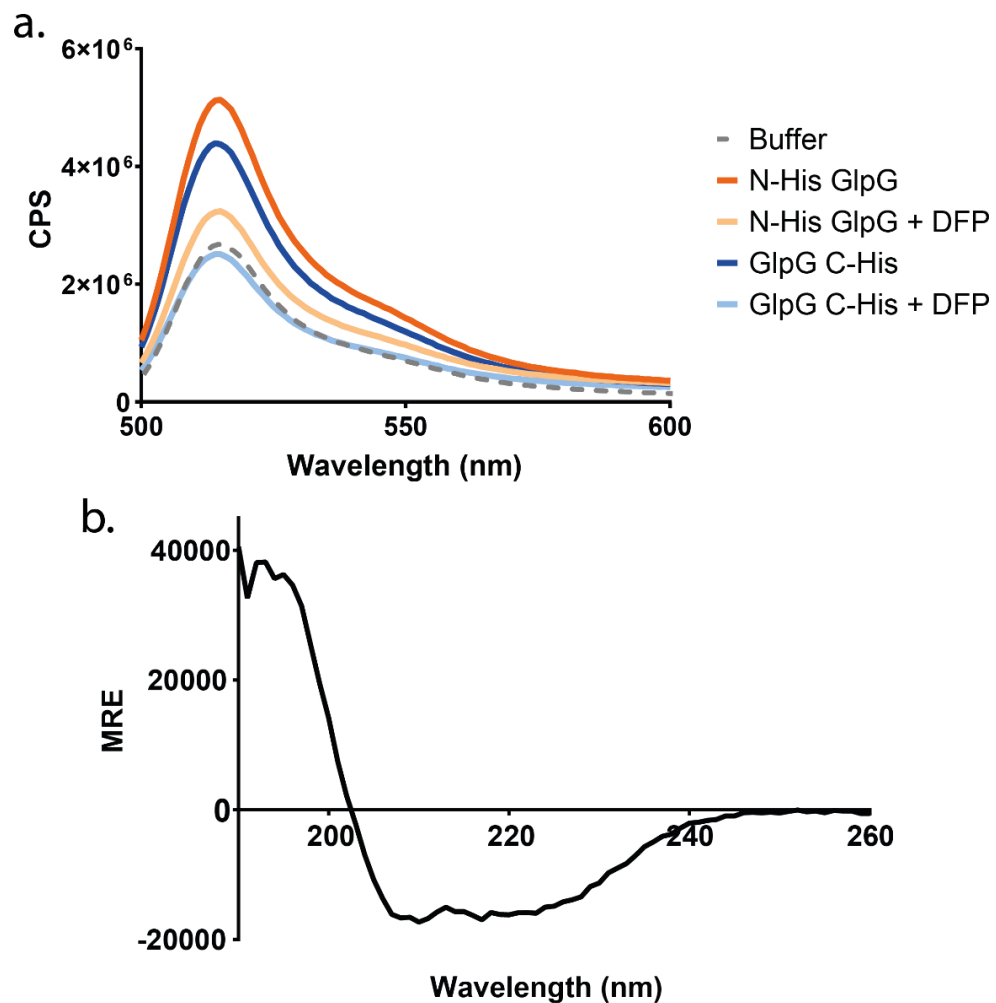
**Figure 3-2: Gel-filtration chromatogram and SDS-PAGE analysis of N-his GlpG**

a) Characteristic chromatogram for the Ni-NTA His-trap with 45 mM imidazole wash before elution with 500 mM imidazole. The 2 mL elution peak was directly injected onto the Superdex 16/60 200 PG column. b) Characteristic Gel-filtration chromatogram of N-his GlpG showing a small void peak at 42 mL and a large, broad elution peak at 58 mL for the Superdex 200 column. c) SDS-PAGE analysis of N-his GlpG, showing the protein running at 26 kDa when stained with Coomassie, and when probed with anti-histidine antibody for Western blot.

### 3.2.2. N-His GlpG is active and folded in DDM

In addition to successful purification, I confirmed that the introduction of the N-terminal tag did not affect protease activity when compared to the published C-terminal construct. An endpoint assay was used to confirm activity of N-His GlpG relative to GlpG C-His and buffer using BODIPY-Casein as a substrate. **Figure 3-3a** shows the fluorescent scans of the system after incubation at 4 °C overnight when excited with light at 480 nm. The peak at 510 nm resulting from N-his (orange) and C-his (blue) GlpG is consistent, showing activity when compared to BODIPY-casein suspended in buffer without GlpG (grey dashed). It was also possible to inhibit N-his GlpG activity with the addition of 100 μM diisopropyl fluorophosphate (DFP) for both N-his (light orange) and C-his (light blue).

CD scans of isolated N-His GlpG were also taken to assess secondary structure indicative of folding **Figure 3-3b**. This showed that most of the secondary structure is  $\alpha$ -helical, with characteristic dips at 208 and 222 nm, and a strong peak, roughly twice the height of the peak at 190 nm. Unfortunately, I was unable to scan lower than 190 nm in the benchtop machine due to the scattering of light by the detergent micelle. It was however possible to deconvolute the spectra to obtain the relative compositions of the full-length protein. Using Dichroweb, with CDSSTR and dataset SMP180 (Abdul-Gader et al., 2011), this revealed that 62 % of the structure comprises of  $\alpha$ -helix, 5 %  $\beta$ -sheet, 9 % turns, and 23 % of the structure is unordered.



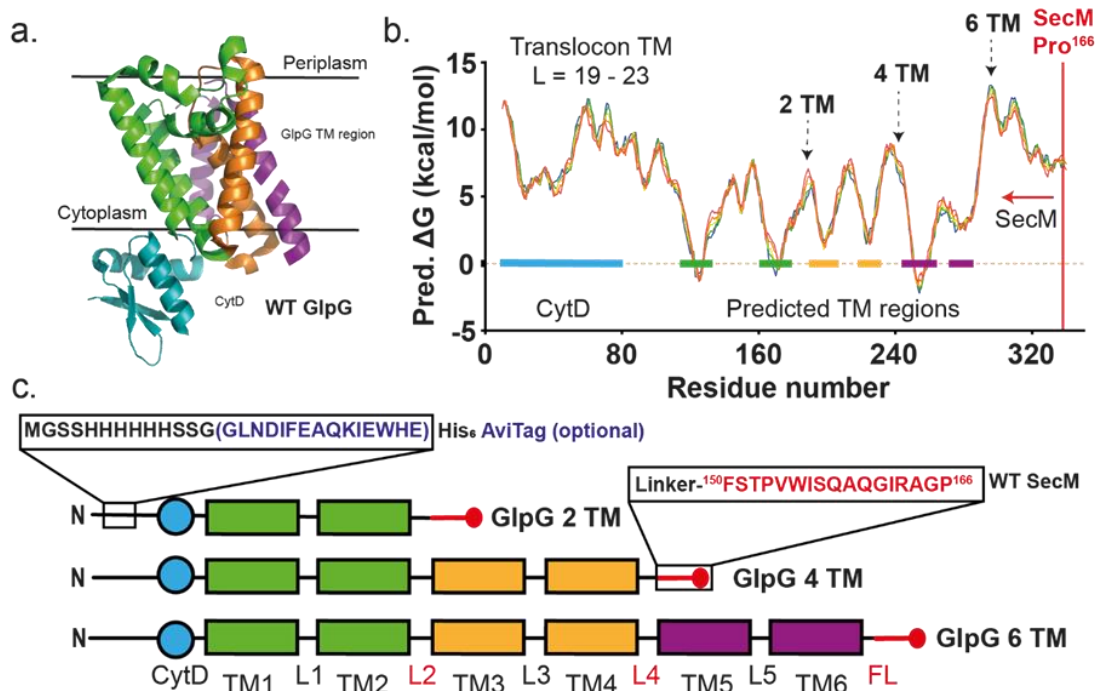
**Figure 3-3: N-His GlpG is active and folded in DDM**

a) Assessment of N- and C-terminally His-tagging on GlpG protease function. 5  $\mu$ g/mL Casein-BODIPY reagent was used as an artificial substrate for a protease end-point assay; where an increase in fluorescence emission at 510 nm indicates a destruction of BOPIPY self-quenching arising from the degradation of casein by 1 mg/mL GlpG, (Reading et al., 2017, Harris et al., 2017b). Both C-terminal GlpG (blue trace), N-His GlpG (orange trace) are active compared to a buffer control (grey dashed). Protease activity can be inhibited with the addition of 100  $\mu$ M Diisopropyl fluorophosphates (DFP) (light orange, light blue traces) (Xue and Ha, 2012), which acted as a control for assay background. b) CD trace of N-his GlpG in 0.025 % (w/v) DDM showing majority of  $\alpha$ -helical structure. Average of two scans in an AVIV benchtop machine.

### 3.2.3. RNC stalling sequence design rational

I designed the specific placement of SecM in the construct to avoid SecM release due to protein folding, which can often generate a force of sufficient magnitude to release the stalled protein. The positioning of the sequence can be directed using MPEx which is a tool designed for predicting the hydropathy of each amino acid in a protein and therefore can be used to determine the transmembrane domains of the protein (Snider et al., 2009). The isolated GlpG (**Figure 3-4a**) sequence was run through MPEx using the translocon-assisted hydropathy plot. The resulting data is displayed as predicted  $\Delta G$  insertion against residue number with a sliding scale of 19-23 amino acids (denoted  $L$ ), an MPEx default value indicative of the predicted amino acid length of a TM helix using the Sec61 translocon (**Figure 3-4b**). The regions with low  $\Delta G$  are predicted TM regions, where the residues readily insert into the hydrophobic lipid bilayer and match positioning of the  $\alpha$ -helices as described by the crystal structure. The high  $\Delta G$  peaks on the hydropathy plot correspond to residues which are least likely to insert into the bilayer and are therefore the regions which the SecM stalling sequence was incorporated. It is known that the ribosome exit tunnel holds around 40 amino acids when the emerging peptide is in an unfolded state and as such, incorporating the final stalling Pro166 downstream of a stably inserted helix (trough) by 45 amino acids, this should ensure that a complete structurally stable domain can insert into the bilayer whilst maintain attached to the ribosome, and not releasing the SecM stall when a significant folding force is generated by the protein.

The positions chosen to incorporate the SecM stalling sequence were decided after 2, 4 and 6 transmembrane helices as shown in schematic **Figure 3-4c**. In addition to the N-terminal His<sub>10</sub> tag, an optional avi-tag was also included for *in vivo* biotinylation (**see section 3.2.6**).



**Figure 3-4: Construct design of GlpG RNCs at different co-translational ‘intermediates’**

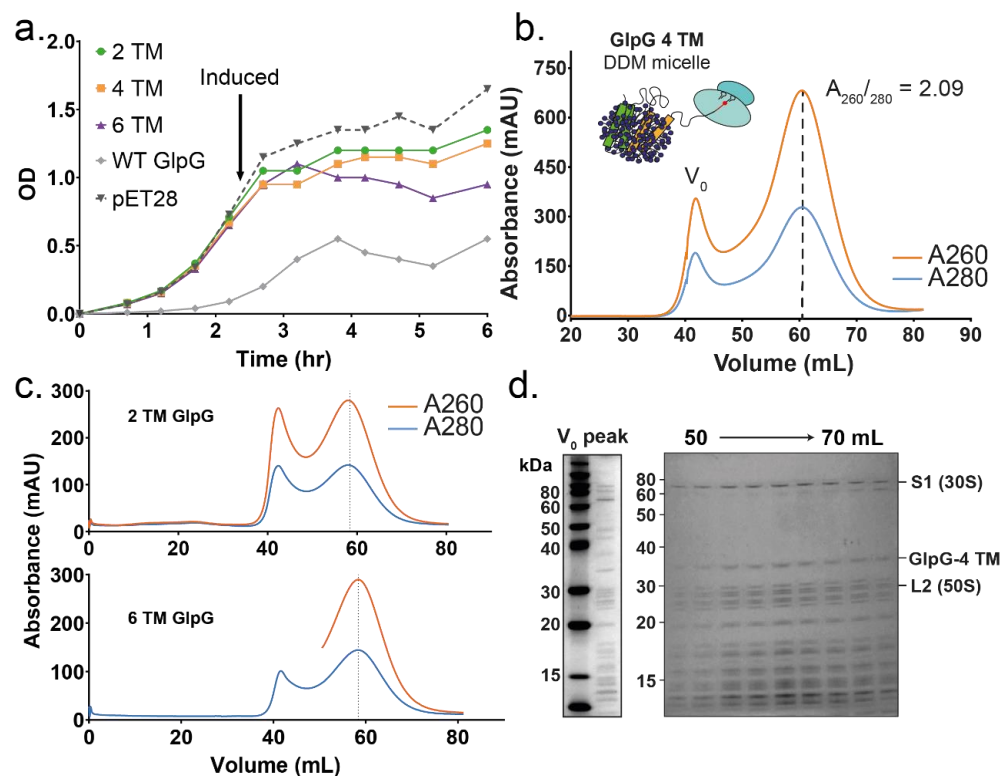
a) PDB structure of WT GlpG: TM region, with helices 1 and 2 coloured green, 3 and 4 orange, and 5 and 6 purple, PDB: 2XTV (Vinothkumar, 2011) and CytD, coloured blue: 2LEP (Sherratt et al., 2009). b) MPEX (Snider et al., 2009) translocon TM predicted hydropathy plot for isolated GlpG with linker-WT SecM. A sliding scale length (L) of 19-23 amino acids was used for MPEX hydropathy prediction. Arrows depict the position chosen for the 3' cloning of linker-WT SecM. This provides 44 amino acid residues between a stable TM helix, and the SecM Pro<sup>166</sup> stalling residue. c) Construct design for RNC truncations of GlpG. WT SecM was cloned to the C-terminal of each construct (red). A hexa-histidine tag was cloned to the N-terminal of the construct. An additional Avi-tag (purple) could also be cloned and successfully utilized for in vivo biotinylation if desired (see section 3.2.6). The colour scheme for GlpG constructs is consistent with panel a.

### 3.2.4. Preparation of $\alpha$ -helical membrane protein RNCs within DDM

Each RNC construct was transformed into BL21-AI cells and growth in either Luria-Bertani media, or MDG/Enhanced M9 media. The latter were grown to near saturation before induction and further growth to saturation after media exchange and induction. **Figure 3-5a** shows characteristic growth curves for each RNC length compared with empty pET28 vector, and isolated GlpG in LB media. Each flask was induced at an OD<sub>600</sub> of 0.8 and grown until saturation before harvest. RNCs, do not hinder the growth of the *E. coli* relative to empty pET28, but interestingly isolated N-His GlpG growth density is much lower, likely due to its toxic nature in cells and hence growth and purification in C43 cells which are tolerant to the expression of toxic membrane proteins (**Section 3.3.1**).

RNCs were prepared in a similar way to isolated protein preps using a two-stage AKTA workflow (**section 3.2.1**). From cryo-mill cracked cells, the membrane fraction was isolated

and solubilised in DDM at 4 °C before passing over a Ni-NTA column and directly injecting the eluate onto a Sephacryl S400 column. This column has a modest resolution of 20 kDa – 8 MDa and is sufficient for the purification of the large RNC from the void aggregate peak as shown in **Figure 3-5b** for the 4 TM RNC construct. Dual wavelength readings at 260 nm and 280 nm check for the presence of rRNA and protein, respectively. All three RNC constructs (2, 4 and 6 TM GlpG) eluted as two resolved peaks with SEC (**Figure 3-5bc**, **Figure 3-6a**), with an  $A_{260}/A_{280}$  ratio of  $\sim 2.1$  signifying the presence of homogenous 70S ribosomes (Hill et al., 1969). The void peak ( $V_0$ ) was found to contain aggregates of released nascent chain, observed by Western blot, and aggregated 70S ribosomes, observed by Coomassie stain showing characteristic ribosomal proteins; S1 at 71 kDa and L2 at 29 kDa corresponding to 30S and 50S subunits respectively (**Figure 3-5d**, **Figure 3-6b**).

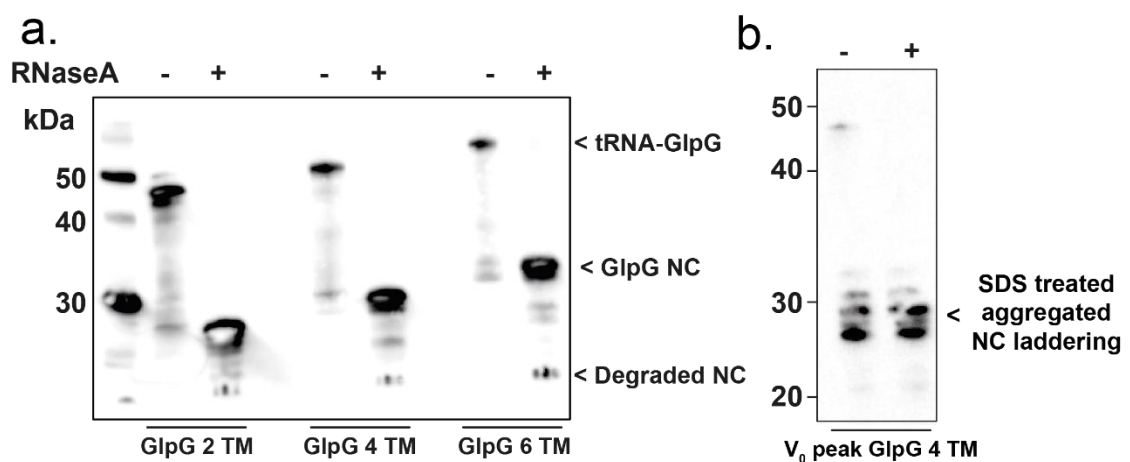


**Figure 3-5: Quality control for DDM purified RNCs**

a) Representative growth curves for RNC and WT GlpG expression against empty pET28 vector in BL21-AI cells grown at 37 °C with shaking at 220 rpm in LB media. RNC production was induced with 1 mM IPTG and 0.1 % (w/v) L-arabinose at an  $OD_{600}$  of  $\sim 0.8$ , after 130 min (2.2 hr) and the temperature cooled to 30 °C and left to induce until saturation. b) Representative SEC trace showing void and elution peak at  $\sim 60$  mL for 4 TM GlpG RNC. b) DDM solubilised GlpG RNCs were purified using IMAC followed by SEC using a 16/60 HiPrep Sephacryl S-400 column. c) SEC profiles of 2 TM and 6 TM GlpG RNCs prepared in DDM detergent show the positions of the  $V_0$  and elution peaks. The ratios for  $A_{260}/A_{280}$  at the  $\sim 60$  mL elution peak 2.01 and 1.97 for 2 TM GlpG and 6 TM GlpG respectively, indicative of homogeneous 70S ribosomes. c) SDS-PAGE of  $V_0$  peak, and fractions under elution peak in (b), show characteristic 70S ribosome protein bands; the 30S S1 protein at  $\sim 65$  kDa, the 50S L2 protein at  $\sim 30$  kDa, and a selection of other ribosomal proteins  $<30$  kDa after Coomassie staining (Dzionara et al., 1970).

The major elution peaks for each stalling length contained homogenous GlpG RNCs as confirmed by low-pH SDS-PAGE and Western blotting, with an  $A_{260}/A_{280}$  ratio of  $\sim 2.1$  signifying the presence of homogenous 70S ribosomes (Hill et al., 1969). It has been suggested that the left side of the 70S elution peak only should be collected for intact 70S RNC which contains the highest portion of 70S ribosomes (as disassembled 50S and 30S components leech into the elution peak) (Cassaignau et al., 2016, Becker et al., 2012), which sucrose gradients inherently separate.

Stalled nascent chains can be characterized by low-pH SDS-PAGE analysis which preserves the nascent chain-tRNA ester bond supplying an increased shift in mass and reduced migration (Cassaignau et al., 2016, Nakatogawa and Ito, 2001). RNase A treatment digests the bound tRNA, leaving the stalled polypeptide. The major elution peak at  $\sim 60$  mL contained homogenous GlpG RNCs as confirmed by low-pH SDS-PAGE and Western blotting (**Figure 3-6a**). For each stall length, a higher order band at 47-55 kDa (from 2 TM to 6 TM) indicates GlpG nascent chain bound to tRNA, when treated with RNase A the band decreases to 25-35 kDa (from 2 TM to 6 TM) indicating the tRNA has been degraded and confirming ribosomal stalling. For each RNC length there is a degree of low intensity bands underneath the major GlpG construct bands these were attributed to low populations of truncated GlpG.

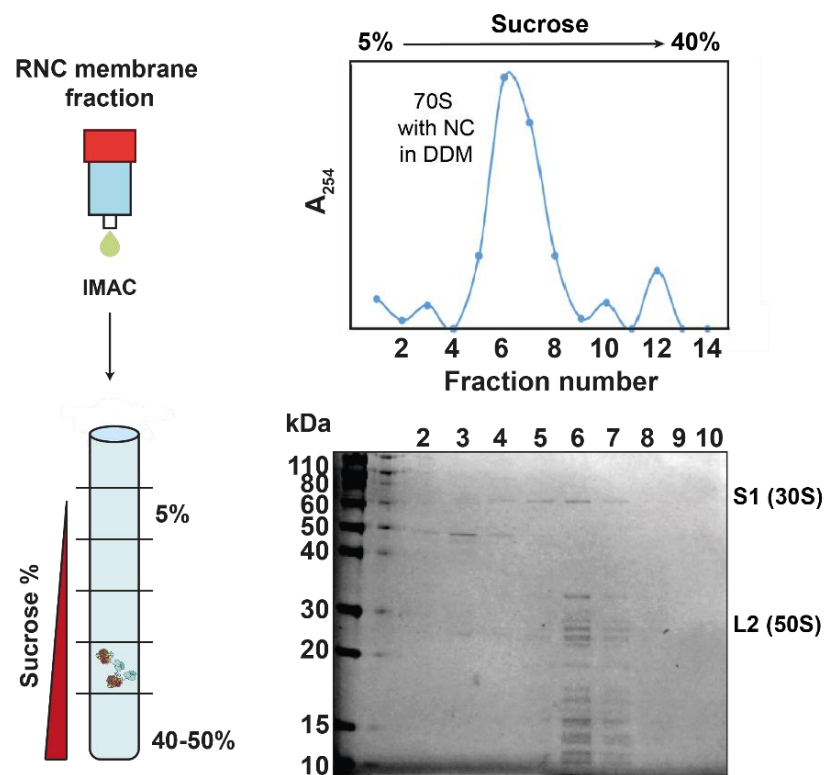


**Figure 3-6: Confirmation of purified, stalled membrane protein RNCs**

a) Low pH SDS-PAGE and Western blotting analysis of fractions under major elution peak show stalled ribosome complex pre (-) and post (+) RNase A treatment (tRNA-GlpG and GlpG nascent chain (NC) labelled bands respectively) - the stall is confirmed by a decrease in molecular weight as tRNA is digested. These bands increase in size as the protein increases by two TM helices with each stall. b) Western blot analysis of a void peak from a 4 TM GlpG RNC preparation in DDM detergent, pre (-) and post (+) RNase A treatment. This peak appears to contain aggregated, released nascent chain. All Western blots were probed using a polyhistidine-HRP antibody.

This gel-filtration based purification strategy is much faster than sucrose gradient preparations commonly used for RNC preparations (Cassaignau et al., 2016), with the shorter purification time being preferable for maintaining the integrity of membrane protein samples in detergent micelles (Seddon et al., 2004). However, my RNC constructs were also amenable to sucrose gradient purification post-IMAC (**Figure 3-7**). The  $A_{254}$  ribosomal content can be followed from a continuous sucrose gradient from 5-40 % in Tico buffer to assay the ribosome fractions. There is a large peak in fractions 6 and 7 corresponding to 70S ribosomes and confirmed using Coomassie staining. Detached 30S ribosomes are present in the early fractions (low sucrose) 2-4 and were not taken for RNC analysis. Stalling was confirmed for 4 TM GlpG 70S sucrose peak using Western blotting with RNase A treatment as shown above.

Yields of RNC varied between nascent chain lengths when overexpressed and purified under identical conditions. Purified yields for the 2 TM, 4 TM, and 6 TM constructs were approximately 30, 100, and 25 pmol/mL respectively, based on ribosomal RNA ( $A_{260}$ ) content and assuming 100 % ribosome/NC occupancy (see **section 3.2.7**) in DDM detergent.



**Figure 3-7: Sucrose preparation of DDM RNCs as an alternative to SEC purification**

A sucrose gradient (Cassaignau et al., 2016, Schaffitzel and Ban, 2007) can also be utilised after the solubilisation and IMAC stages of purification. 70S ribosomes with bound nascent chain in DDM detergent can be separated from released 50S and 30S ribosomal subunits and their positions in the gradient determined by absorbance at 254 nm (left) and checked for characteristic 70S ribosome bands by SDS-PAGE and Coomassie staining (right).



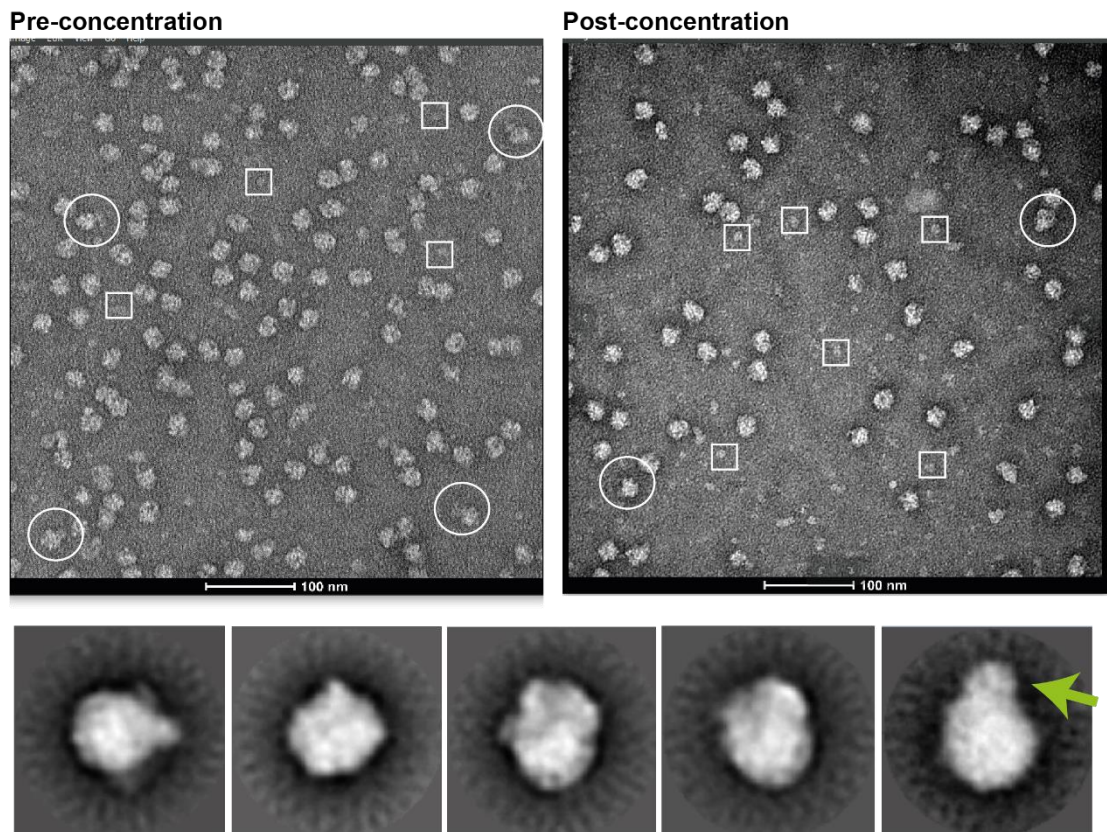
### 3.2.5. Negative stain EM of RNCs in DDM

*Negative staining detailed in this section was carried out in collaboration with Dr Sara Alvira and Professor Ian Collinson at the University of Bristol.*

Negative stain EM provides a means to check the quality and homogeneity of the RNCs, this can give an indication into whether the samples are intact, and whether any further processing steps are required for to improve sample quality after preparation. Negative staining is also the initial step towards cryoEM to study the structure of the emerging nascent chain and its interactions with the ribosome, the membrane and the translocon.

100 ng of 4<sup>TM</sup> GlpG RNC were stained with UrAc and imaged at digital magnification of 49,000 x and a sampling resolution of 2.04 Å per pixel. An even coverage of the grid was observed allowing the visualisation of ribosomes (**Figure 3-8**), disassembled and empty ribosomes (squares) and ribosomes attached to ‘something’ (circles). The 4<sup>TM</sup> GlpG RNC has a nascent chain of size 30 kDa, we would not expect to see it under this in this small number of particles, without quality data acquisition and processing, but in the few particles that were observed we saw ribosomes with something attached. We propose that this is the co-purification of the SecYEG bacterial translocon (see **section 3.2.8**). By far, most particles are empty ribosomes, we identify this by comparison with (Gabashvili et al., 2000).

I would expect a concentration step through a 100 kDa cut-off concentration to remove a high portion of released nascent chain. While this may be good for biochemical experiments where the result of an experiment relies on probing the bound nascent chain directly, for negative staining and cryoEM, as well as other structural techniques, the disassembled ribosome would also remain in the concentrator leading to a heterogeneous population of RNC sample as shown in **Figure 3-8**, post concentration.



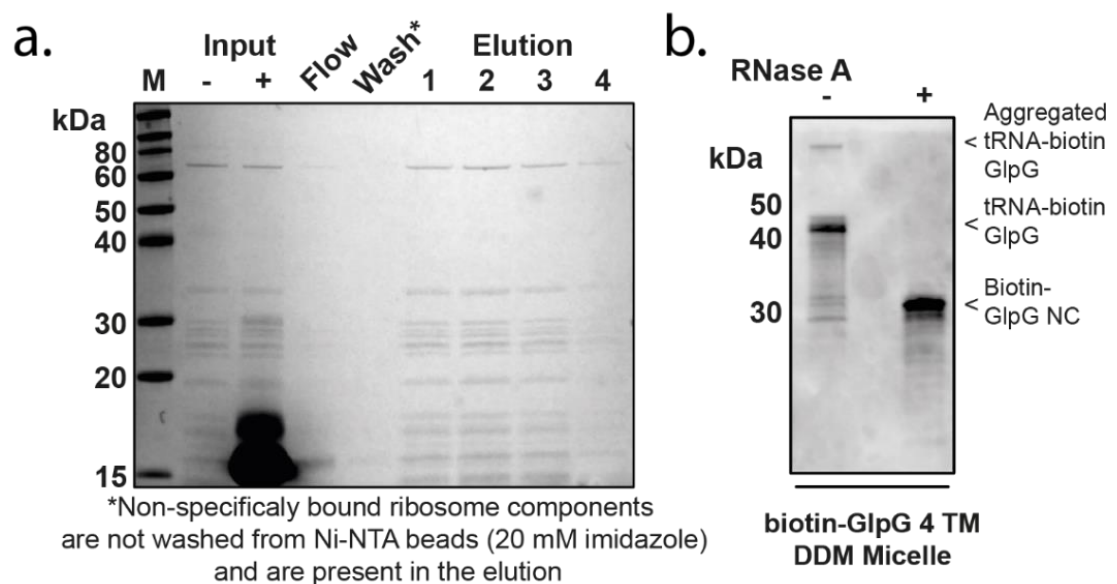
**Figure 3-8: Negative staining of 4 TM GlpG in DDM**

*Negative stains of the 4 TM GlpG RNC in DDM pre- and post-concentration. Circles show ribosomes with ‘something’ attached; squares show disassembled particles. The green arrow in 2D classification shows ribosome bound to something larger than the NC, this is likely co-purified SecYEG.*

### 3.2.6. Co-expression with BirA for efficient *In vivo* biotinylation of RNCs

Biotin conjugation provides rapid, high affinity binding to (strept)avidin molecules, which are often utilized for downstream purification strategies and biophysical characterisation (Chang and Bowie, 2014, Guo et al., 2016, Kaiser et al., 2011, Min et al., 2015, Min et al., 2016). Furthermore, nickel affinity resin has been shown to bind non-specifically to ribosomes (Schaffitzel and Ban, 2007, Bischoff et al., 2014). I show that the 4 TM GlpG RNCs treated with RNase A, were bound to Ni-NTA beads via the N-terminal GlpG His-tag. Washing the beads with Tico buffer containing 20 mM imidazole did not remove the ribosomal components which were expected to be released from the NC post-RNase A treatment. This therefore confirmed that ribosomal proteins remained non-specifically bound to the beads. These proteins were however eluted in 500 mM imidazole with the released nascent chains (**Figure 3-9a**), overall suggesting a significant non-specific binding of the ribosomal proteins to Ni-NTA affinity purification beads.

To counter this, it was pertinent to test whether incorporation of a high affinity tag within membrane protein RNC constructs was viable as an additional, non-IMAC purification method which could be advantageous to provide further flexibility in purification and downstream biophysical interrogation. To achieve this, the 4<sup>TM</sup> GlpG RNC was co-expressed with a biotin ligase (BirA) enzyme in the presence of D-biotin supplemented into the growth media. This drove the *in vivo* biotinylation of the RNC when an additional avi-tag is cloned onto the N-terminal of the RNC as shown in **section 3.2.3, Figure 3-4** and as previously described in (Ashraf et al., 2004). When the 4<sup>TM</sup> biotin-GlpG construct is prepared as above using gel filtration, I confirmed successful biotinylation by Western blotting using streptavidin-HRP, which binds biotin directly, and stalling of the biotinylated RNC was confirmed as before, with digestion of bound tRNA with RNase A, and a decrease in molecular weight (**Figure 3-9b**).

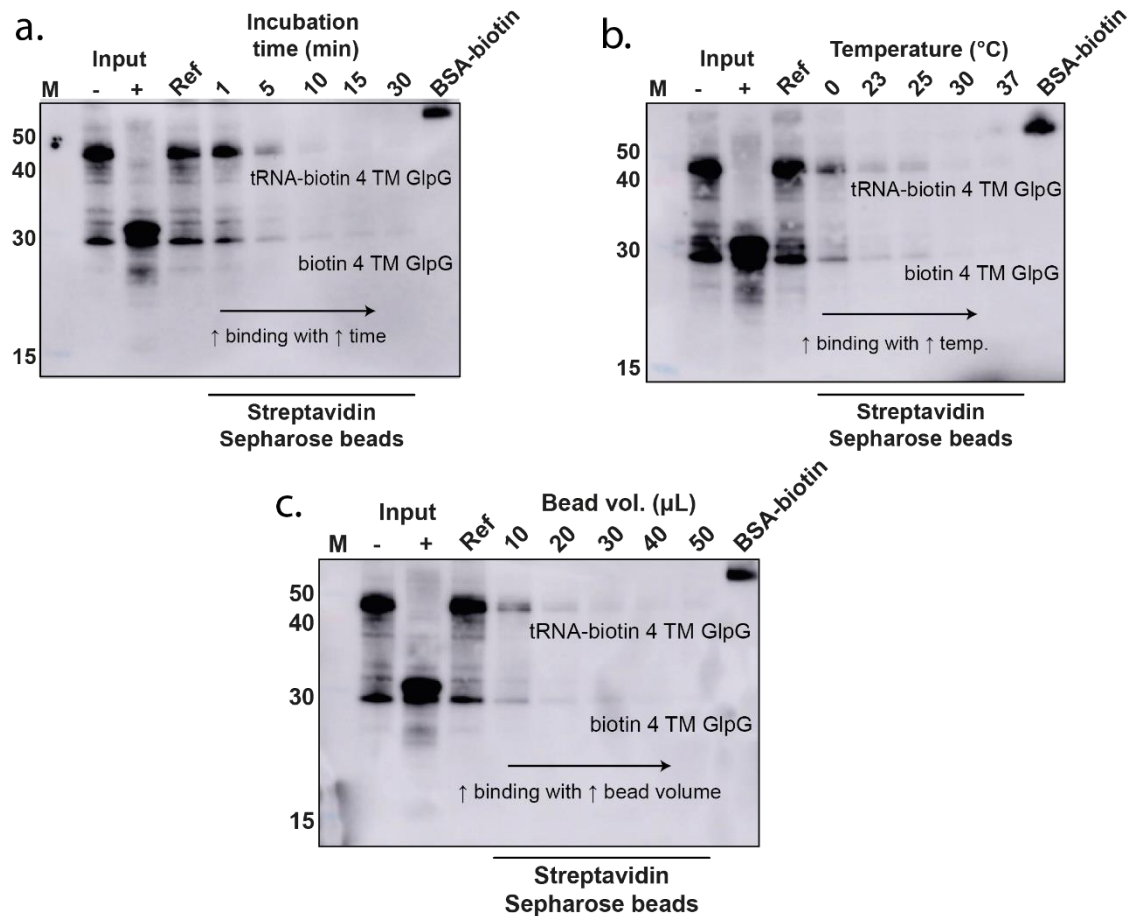


**Figure 3-9: In vivo biotinylation of RNCs to prevent non-specific Ni-NTA binding**

a) 4<sup>TM</sup> GlpG RNC binds to Ni-NTA beads pre- (-) and post (+) RNase A treatment as shown by Coomassie stain. RNase A treated samples were added to the Ni-NTA beads. Ribosomes are not detected in the flow through, or 20 mM wash, only the nascent chain should bind to the beads via the His-trap. Ribosomes are however detected in the 500 mM imidazole elution, suggesting non-specific attachment of the ribosome to Ni-NTA beads which I expected to be removed in the wash step. b) Streptavidin-HRP blotted Western confirming biotinylation of the 4<sup>TM</sup> RNC and correct stalling, with a decrease in molecular weight with RNase A treatment.

A selection of biotinylated 4<sup>TM</sup> GlpG RNC bead binding tests to streptavidin coated sepharose beads were carried out with varying incubation time, bead volume and incubation temperature. For each experiment, there was an increase in RNC binding to beads, observed by the disappearance of 4<sup>TM</sup> GlpG tRNA band at 45 kDa with increased time (1 to 30 min) **Figure 3-10a**, temperature (0 - 37 °C) **Figure 3-10b**, and bead volume (from 10 - 50 µL)

**Figure 3-10c**, when probed using a streptavidin-HRP antibody, which detects the biotinylated species. The best conditions from these experiments suggest that at neutral pH, a 10-15 min incubation time with 20-30  $\mu$ L bead slurry at 30  $^{\circ}$ C captures biotinylated 4 TM GlpG RNC with the greatest efficiency. There is still however a degree of released GlpG nascent chain (at 30 kDa) for each condition resulting from thawing of the intact RNC samples.

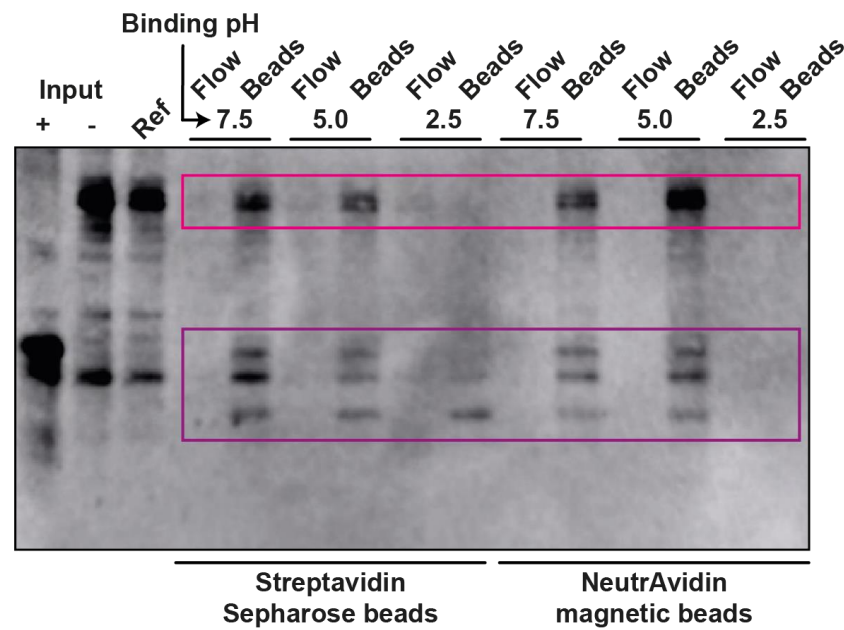


**Figure 3-10: Biotinylated 4 TM GlpG RNC binds to streptavidin Sepharose beads**

A selection of biotinylated 4 TM GlpG RNC bead binding tests to streptavidin coated sepharose beads with varying incubation time (left), bead volume (middle), and incubation temperature (right). Biotinylated BSA is run as a control, and Western blots were probed with streptavidin-HRP. For each gel, the disappearance of the tRNA-bound 4 TM GlpG at 45 kDa is indicative of bead binding. There is a degree of released biotin 4 TM GlpG nascent chain in each gel at 30 kDa.

A variety of conditions and different avidin varieties were also tested. A biotinylated 4 TM-GlpG RNC was bound to either NeutrAvidin magnetic beads, or streptavidin sepharose beads at varying low pH's 7.5, 5.0 and 2.5 (**Figure 3-11**). Both neutravidin and streptavidin beads bind biotinylated RNC, observed by the disappearance of band from the flow through (flow), and reappearance once beads are boiled at 95  $^{\circ}$ C for 10 min (beads). Elution of the biotinylated RNC was not observed on boiling when bound at pH 2.5. tRNA biotin 4 TM GlpG (intact RNC)

bands are highlighted by a pink box, the released biotin 4<sup>TM</sup> GlpG species are shown in a purple box. There is some laddering of this released GlpG, however, without high resolution molecular weight mass spectrometry analysis, I cannot identify the nature of these bands.  $\alpha$ His-HRP was used to probe the Western blot as the biotin was already interacting with the streptavidin, and so streptavidin-HRP interactions may not be possible.

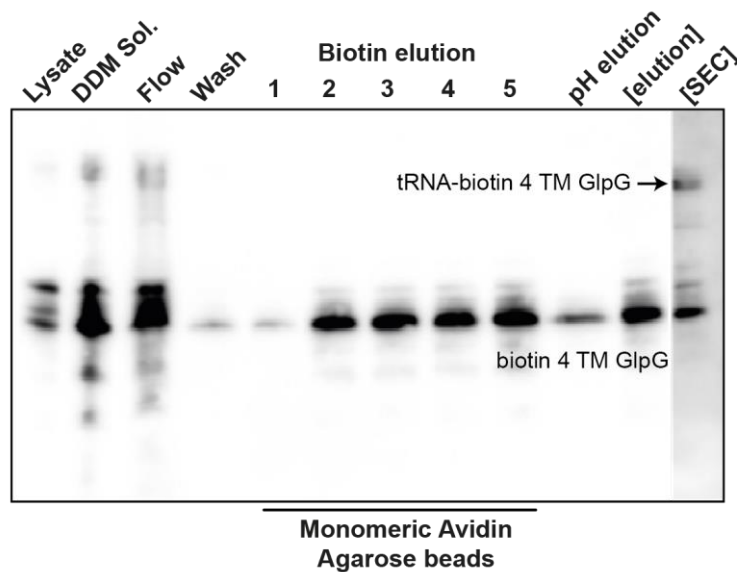


**Figure 3-11: Low pH RNC binding to streptavidin and neutravidin coated beads**

A biotinylated 4<sup>TM</sup>-GlpG RNCs were bound at varying low pH's to either NeutrAvidin magnetic beads, or streptavidin sepharose beads. Both neutravidin and streptavidin beads bind biotinylated RNC, observed by the disappearance of band from the flow through (flow) and reappearance once beads are boiled at 95 °C for 10 min (beads). tRNA biotin 4<sup>TM</sup> GlpG (intact RNC) bands are highlighted by a pink box; the released biotin 4<sup>TM</sup> GlpG species are shown in a purple box.  $\alpha$ His-HRP was used to probe the Western blot.

I can also bind and elute RNC samples to monomeric avidin. Elution with excess biotin is gentler than 95 °C boiling of the samples from streptavidin beads, which maintains the structure and integrity of the RNC for downstream experiments. **Figure 3-12** shows somewhat low efficiency binding (as judged by a considerable amount of GlpG in the flow-through) of the GlpG RNC to these beads, with 5 CV of excess biotin elution. The sample input showed lots of released GlpG RNC from the ribosome, but when the eluates are pooled passed over the Sephacryl S400 gel filtration column, I do see evidence of purified intact RNC suggesting that monomeric avidin does in fact provide an alternate tagging and purification strategy for biotinylated RNCs, and with further optimisation into binding capacity and biotinylation efficiency, it is possible to harness this technique for downstream purification or enrichment.





**Figure 3-12: Monomeric avidin binding and eluting of biotinylated 4 TM GlpG RNCs**

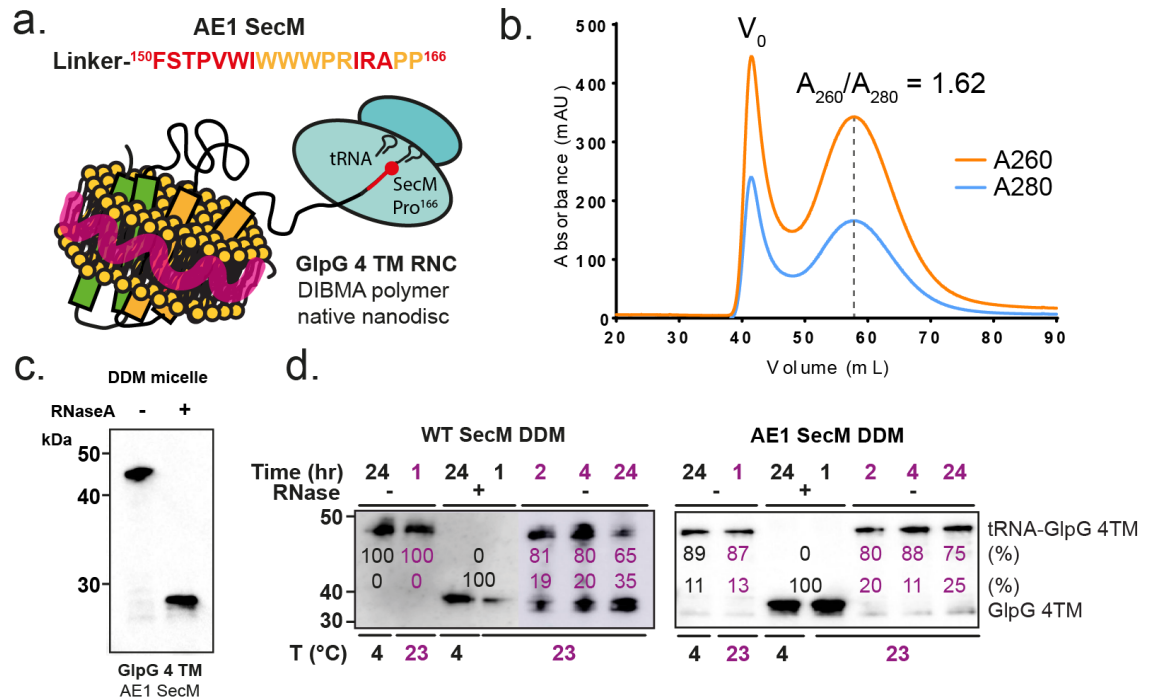
Using monomeric avidin to purify biotinylated 4 TM GlpG RNCs, shows that DDM solubilised RNCs do bind to the beads, observed by the reappearance of the released GlpG during the elution fractions. Sample is eluted with excess concentration of *D*-biotin and the resulting biotin-4 TM GlpG can be pooled and separated from the intact RNC using SEC. When blotted, the tRNA bound and released nascent chain are observed.

### 3.2.7. An enhanced SecM arrest sequence to improve RNC stability

An arrest enhanced SecM sequence (<sup>150</sup>-FSTPVWIWWPRIRAPP-<sup>166</sup>) (Cymer et al., 2015a, Kempf et al., 2017) was employed to improve the stability of RNCs and mitigate the release of the nascent chain during and after purification, thus enhancing RNC quality and longevity. Compared to WT SecM, the enhanced SecM sequence contains an additional proline at the 165 position and a substituted stretch of tryptophan amino acids (**Figure 3-13a**), which forms more interactions with the ribosome exit tunnel and further occludes peptide bond formation in the PTC (Bischoff et al., 2014).

4 TM GlpG RNCs were first prepared in DDM with the enhanced SecM stalling sequence. The gel-filtration chromatogram (**Figure 3-13b**) shows a similar yield in RNC based on the elution peak as the WT 4 TM GlpG RNC, but the  $V_o$  peak is larger. The  $A_{260}/_{280}$  ratio is lower than the expected 1.8, at 1.6, with the elution maxima at 58 mL. Enhanced SecM 4 TM GlpG RNC samples are of much higher quality than WT SecM, as evident by negligible nascent chain release and degradation pre-RNase A treatment with Western blot (**Figure 3-13c**). Improved nascent chain retention was also found for enhanced SecM RNCs; with only 12 % spontaneous release of nascent chain from enhanced SecM RNCs observed over 24 hours at 23 °C, compared to 35 % release for WT SecM, for DDM purified samples under the same conditions (IMAC and

SEC purification) (**Figure 3-13d**). These percentages were estimated using densitometry analysis on each time point and calculating the relative ratios of tRNA-bound 4 TM GlpG and naturally released 4 TM nascent chain.

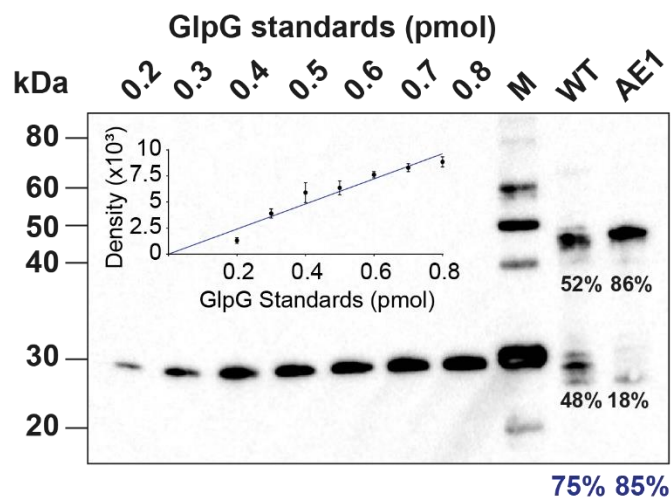


**Figure 3-13: Arrest-enhanced SecM yields RNCs with a greater stability, occupancy, and purity in DDM**

a) Arrest-enhanced SecM sequence, with residues differing from the WT SecM coloured orange. b) Gel-filtration chromatogram for the purification of 4 TM GlpG RNC with enhanced SecM. The RNA:protein ratio was 1.62 and the elution volume was 58 mL. c) Low pH SDS-PAGE and Western blotting analysis shows that the enhanced SecM RNC constructs were consistent with WT SecM for GlpG 4 TM in DDM, but the degree of spontaneously released nascent chain is significantly reduced. d) GlpG 4 TM RNCs produced with WT (left) and enhanced SecM (right) in DDM detergent were incubated at 23 °C for 24 hr, samples were taken and blotted after 1, 2, 4 and 24 hr before running SDS-PAGE gels without RNase A treatment (-). The ratio between intact RNC (~45 kDa band, tRNA-GlpG 4 TM) and spontaneously released nascent chain (~30 kDa band, GlpG 4 TM) was calculated for each lane using band density and ImageJ. Percentages of intact and released are quoted next to the corresponding band, with purple text being used for analysis. For WT SecM, from 1 hour to 24 hr, 35 % nascent chain was released (100% to 65% intact) for enhanced, only 12 % (87 % to 75 %) was released.

I also estimated the fraction of ribosomes in the samples that contain bound nascent chain to understand purified RNC occupancy to aid with negative staining and cryoEM experimentation. Occupancy was calculated by blotting a known  $A_{260}$  derived concentration of each RNC sample (where  $1 A_{260} = 24 \text{ pmol/mL}$ ) against isolated GlpG standards (**Figure 3-14**) and calculating pmol of GlpG NC for the tRNA-bound band (Cassaignau et al., 2016). There was an increase in occupancy of enhanced SecM RNCs in DDM (85 % occupancy) compared to WT SecM RNC (75 % occupancy). Moreover, 4 TM GlpG RNCs containing WT SecM released

48 % of its nascent chain upon only one freeze thaw cycle, whereas enhanced SecM released only 18 % (**Figure 3-14**).



**Figure 3-14: RNC occupancy calculation using GlpG standards**

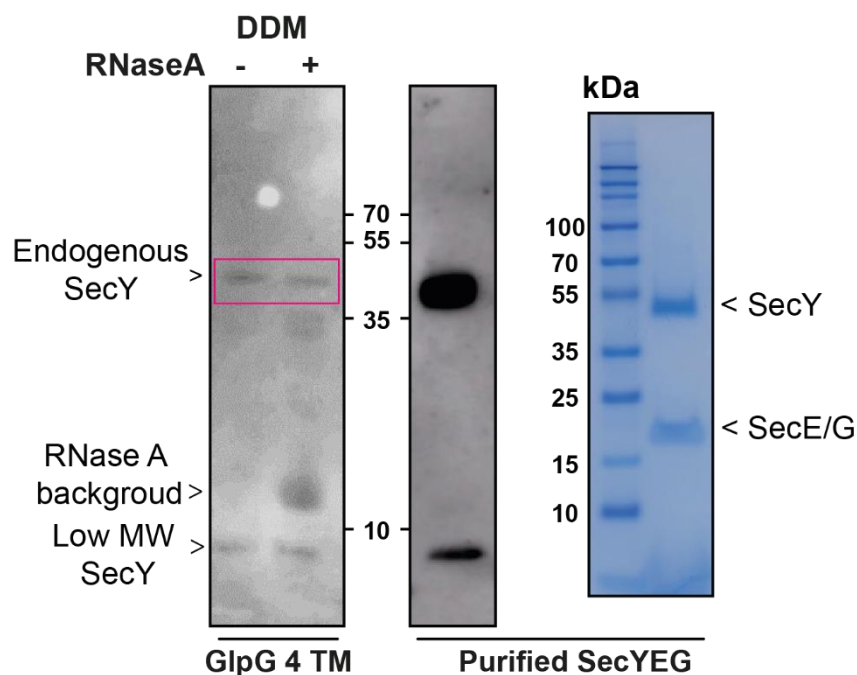
Nascent chain occupancy of ribosomes can be determined by loading an  $A_{260}$  estimate nascent chain concentration against known concentrations of purified isolated GlpG. A calibration curve of band density is plotted against concentrations and a percentage occupancy is determined and shown in the inset. Occupancy was calculated for the intact RNC band and was calculated to be 75 % and 85 % respectively for WT and enhanced SecM when prepared in DDM (where 0.78 pmol of each sample was loaded and underwent one freeze-thaw cycle).

### 3.2.8. RNC overexpression with SecYEG for structural studies

To explore whether any SecYEG remained associated with GlpG RNCS, I probed 4 TM GlpG RNC purified in DDM for the presence of any endogenous Sec translocase machinery using a monoclonal SecY antibody (Whitehouse et al., 2012). SDS-PAGE and Western blot analysis revealed the presence of endogenous SecY at 37 kDa (which has a lower blot intensity than overexpressed SecYEG), with the low molecular weight SecY degradation product at 5 kDa in both pre- and post-RNase A treatment (**Figure 3-15**). In addition to the correct running of the SecY bands, SecYEG was purified and blotted using the monoclonal SecY antibody and stained with Coomassie to show the presence of SecY, but also counterparts E and G, as comparison. SecE and SecY co-migrate in this Bis-Tris 12 % acrylamide gel system. This co-purification of SecY supports that a GlpG RNC, Sec translocase interaction *in vivo* is maintained upon isolation in DDM.

*SecYEG was purified by Laura Blackholly, Booth group, KCL.*





**Figure 3-15: Endogenous SecYEG co-purifies with RNC in DDM detergent**

Endogenous SecY was found to be present in pre- (-) and post- (+) RNase A treated samples (highlighted with a pink box, left panel). The low molecular weight SecY band is characteristic of a C-terminal cleavage product (Whitehouse et al., 2012). The SecY 37 kDa band was confirmed by comparison with Western blot (middle panel) and SDS-PAGE Coomassie stain (right panel) of purified SecYEG - SecE and SecG, which co-migrate in this SDS-PAGE environment. All Western blots were probed using a monoclonal antibody for SecY.

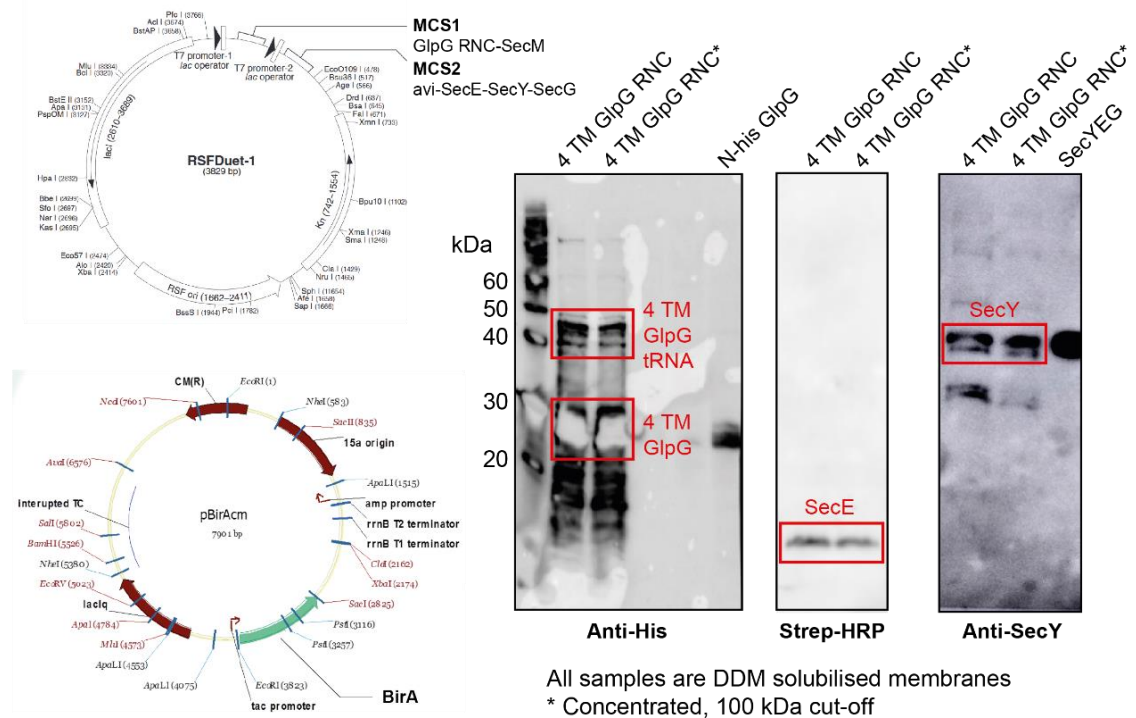
Now I have means to increase the occupancy of the RNC samples and have observed evidence that my purification strategy can co-purify RNCs with SecYEG, I wonder whether it is possible to overexpress the SecYEG translocon with the RNC and utilise *in vivo* biotinylation for the two-stage purification of these samples for clean homogenous and occupied sample for CryoEM and other biochemical analysis.

Using the RSFDuet-1 plasmid, it is possible to clone two genes into one expression vector. This allows only two antibiotics (rather than 3) for growth conditions, reducing bacterial stress, as well as having two different origins of replication to form a stable genotype with the correct expression ratios for the three proteins when co-expressing these proteins with BirA. GlpG RNCs with WT SecM were initially cloned into MCS1 of Duet, where the SecYEG ORF, with an avi-tagged SecE protein was cloned into MCS2 (**Figure 3-16**).

When the duet plasmid is co-expressed with BirA, each of the three components were expressed and targeted to the membrane as shown in **Figure 3-16**. The membranes were harvested and solubilised with DDM before analysing the expression products on Western blot. The 4 TM GlpG RNC was observed as normal using anti-poly histidine antibody, showing

strong expression of the 4 TM GlpG-tRNA band at 45 kDa, and a large portion of released nascent chain as is common at this stage of the preparation, and could not be removed with a 100 kDa spin concentrator polishing step at this point (denoted by Asterix in **Figure 3-16**). It was also possible to check for *in vivo* biotinylation by probing the Western blot with streptavidin-HRP. This shows that the SecE protein was in fact biotinylated and was correctly expressed. Finally, SecY was probed for using the monoclonal SecY antibody, this also expressed and ran at its correct molecular weight when compared with overexpressed SecYEG as before.

With the successful expression and DDM solubilisation of each component, I can now purify the solubilised membranes with Ni-NTA to isolate the RNCs, and a further polish step using monomeric avidin beads to pull out RNCs with SecYEG associated. Cloning is also in progress to swap the WT SecM stalling sequence for the enhanced, to prevent further degradation of the nascent chain from the ribosome which should drastically improve the preparative sample.



**Figure 3-16: Overexpression of biotinylated SecYEG and 4 TM GlpG RNCs**

Plasmid maps showing the RSFDuet-1, where GlpG RNC-SecM is cloned into MCS1, and avi-SecE-SecY-SecG cloned into MSC2, and pBirAcm for BirA expression. When co-expressed and membranes are harvested and solubilised in DDM, the fraction can be run on SDS-PAGE to confirm expression of RNC when probed with anti-poly histidine, *in vivo* biotinylation and expression of SecE when probed with Streptavidin-HRP, and SecY when probed with monoclonal SecY antibody.

### 3.3. Discussion

#### 3.3.1. GlpG preparation and activity

A C-terminal hexa-histidine tagged *E. coli* GlpG protein was first purified (Lemberg et al., 2005) with a TEV protease cleavage site and expressed and purified to homogeneity using nickel affinity chromatography, with further removal of the affinity tag using TEV protease, and clean-up using anion-exchange chromatography. These highly clean GlpG samples in DDM were monomeric and active, when assayed using the Spitz-TMD substrate, showing the *in vitro* assay accurately depicts the proteolytic activity in cell membranes (Lemberg et al., 2005, Urban et al., 2002).

The additional TEV cleavage of the hexa-histidine affinity tag and further clean up are cumbersome experiments, and as such were removed from preparation in my protocols. I maintain the histidine tag and remove any aggregates or degradation using size-exclusion chromatography immediately after the Ni-NTA elution, in a two-step process. This avoids the need from prolonged incubation with imidazole and removes the concentration step which can often result in sample loss and can be time consuming.

The requirement of the N-terminal histidine tag rather than C-terminal for GlpG is due to the mechanism by which proteins are expressed, whereby the nascent chain emerges from the ribosome N-terminal first and the C-terminal is not expressed or exposed, and thus, the affinity tag must be cloned onto the N-terminus. It was possible to purify GlpG homogeneously with the N-His tag using Ni-NTA affinity chromatography with no significant set-back in quality or yield, with the standard 1 mg per litre of growth being produced (Reading et al., 2017). The tag did not affect trafficking to the membrane, as shown by the majority of the his-tagged sample residing in DDM solubilised membranes by Western blot, or affect activity of the protein when assessed using the BODIPY-casein endpoint assay, in comparison with the previously published and active C-terminal histidine tagged GlpG (Reading et al., 2017, Harris et al., 2017b) – in fact, the N-terminal GlpG showed a larger fluorescent signal indicative of a higher degree of proteolysis than C-terminal. The activity of the protein can be inhibited on addition of DFP, similarly to C-His tagged protein.

I used an endpoint assay in these experiments to show GlpG is active. Previously activity has been shown using a BODIPY-casein kinetic assay in both DDM and SMALP, and the rate of protease activity is highly influenced by the surrounding lipid bilayer (Foo et al., 2015, Reading et al., 2017). GlpG has been shown to have a much higher activity in detergent micelles than

in SMALPs, this is likely due to the lipids modulating the activity of the enzyme, which slow the diffusion of the enzyme down to reduce activity (Reading et al., 2017). The BODIPY-Casein substrate used here is also a soluble protein, rather than a natural TM spanning substrate for GlpG. This may also explain the higher activity in micelles compared to native nanodiscs. Contrary to this, it has been suggested that the rhomboid shape of the protein is able to distort the flanking lipids, allowing the viscosity barrier across the membrane to be broken to allowing fast diffusion and efficient proteolytic activity in a membrane (Kreutzberger et al., 2019). This was ascertained using a single TM spanning protease substrate *in vivo*, rather than the soluble substrate shown here, in membranes with many more lipids and space to diffuse than the SMALP environment would allow.

Overall, the production of N-terminal GlpG which is required for the purification of membrane protein RNCs has been purified to a similar standard as past purified samples of C-terminal GlpG, whilst maintaining the high (>95 %) purity, and activity of the protein. The circular dichroism trace shows characteristically  $\alpha$ -helical structure ( $\alpha$ -helix: 62 %,  $\beta$ -sheet: 5 %, other: 31 %), albeit slightly less than reported in the PDB structure (67 %  $\alpha$ -helix) when deconvoluted, but CD spectra are additive and the mixed structure CytD domain takes away from the helical content. Nevertheless, this domain is known to alter lipid composition and activity of GlpG, so I, and the Booth group, have included it in our experiments which other folding studies do not include (Paslawski et al., 2015, Guo et al., 2016, Choi et al., 2019). CytD must also remain on the cytoplasmic side of the membrane when folding, and if not present it is not clear whether topology of the protein will be affected which consequently may adversely affect function and change folding pathways. Cysteine labelling studies on the isolated protein produced in a cell-free system and void of the translocon, showed that GlpG topology was consistent with cellular topology, with the CytD domain remaining on the outside of the liposome (cytoplasmic) – it did not cross the bilayer (Harris et al., 2017b)

### **3.3.2. Designing RNCs with membranes in mind**

The generation of polytopic membrane protein RNCs relies on the relatively simple cloning strategy of the WT SecM sequence into a desired position throughout the protein. This can be guided by MPEx (Snider et al., 2009), to determine the transmembrane regions and decide on the best position to include the stalling sequence so to not induce folding induced stall release. Much of the work that has already been carried out on RNCs has been on soluble proteins and as such many of the many of the published methods used to produce RNCs for folding studies much be adapted and developed for membrane protein study.

Deciding where to stall a protein for folding studies depends on both the protein of interest and the biological question that the researcher wants to answer, and with soluble protein biogenesis, the regions in which a stalling sequence can be cloned and tolerated by the bacteria are more diverse than that for membrane proteins. Membrane proteins need to be surrounded by a membrane, or a suitable membrane mimic for correct folding, to prevent aggregation and in many cases, to protect function. Placing a flexible SecM stalling sequence in the middle of a hydrophobic stretch of amino acids which preferentially form  $\alpha$ -helices will likely disrupt folding and function of the protein and will not accurately portray the folding intermediates we try to capture in the process. To aid in my decision to place SecM stalling sequences at suitable positions in the GlpG protease, I rely on both the crystal structure (PDB: 2XTV (Vinothkumar, 2011)), and a bioinformatics tool, MPEx (Snider et al., 2009), which describes the thermodynamic costs of insertion of the primary protein sequence into a hydrophobic membrane environment. As my RNCs are produced *in vivo*, I have used the translocon-assisted model in MPEx. This model is based off the Sec61 translocon interactions with TM helices. This mode not only considers the individual amino acid properties, but also stretches of amino acids which form helices, and the flanking loop regions of the protein, as residues often have position dependent hydrophobicities. This is perhaps a better hydropathy model as I would expect the RNCs to be translocon dependent in their native membranes, as shown by a comprehensive proteomics study of GlpG binding SRP (Schibich et al., 2016). Based on the MPEx output, I observe the 6 transmembrane helices of GlpG as expected per crystal structure (PDB: 2XTV), where helices are determined by low  $\Delta G$  predictions equating to favourable insertion kinetics.

Analysis of the primary sequence in this way allows me to pinpoint the cloning target for SecM insertion. Areas of high  $\Delta G$ , likely loop regions in the protein can be selected for the downstream cloning of the 44 amino acid WT SecM and linker residues. These 44 amino acids between a stably inserted transmembrane helix and the final stalling proline-166 in SecM slightly exceeds the  $\sim 40$  fully extended amino acids which can fit inside the ribosome exit tunnel. This successfully prevented the release of stalling due to force exerted by a co-translationally folding protein, which may have occurred if the SecM sequence was cloned into the middle of the TM helix, as opposed to the flanking helix loops, after helices 2, 4 and 6 to achieve a stable RNC which does not release on co-translational folding into the membrane as previously shown.

Other groups have used the SecM stalling system to study membrane protein co-translational folding *in vivo* from a biochemical perspective, rather than a preparative method by utilising its stalling release when the folding of the protein produces a force with a magnitude that physically pulls SecM out of the PTC. Inserting the SecM sequence at varying positions throughout the protein sequence can produce a force-trace for the protein of interest from the fraction of folded nascent chain to released nascent chain using SDS-PAGE analysis (Ismail et al., 2012, Cymer et al., 2015b, Cymer and von Heijne, 2013, Cymer et al., 2014). Interestingly, these ‘force-traces’ nicely match up with the predicted hydropathy plots determined using MPEx (Cymer and von Heijne, 2013, Snider et al., 2009).

### **3.3.3. Advantages of *in vivo* generated MP RNCs purified in DDM**

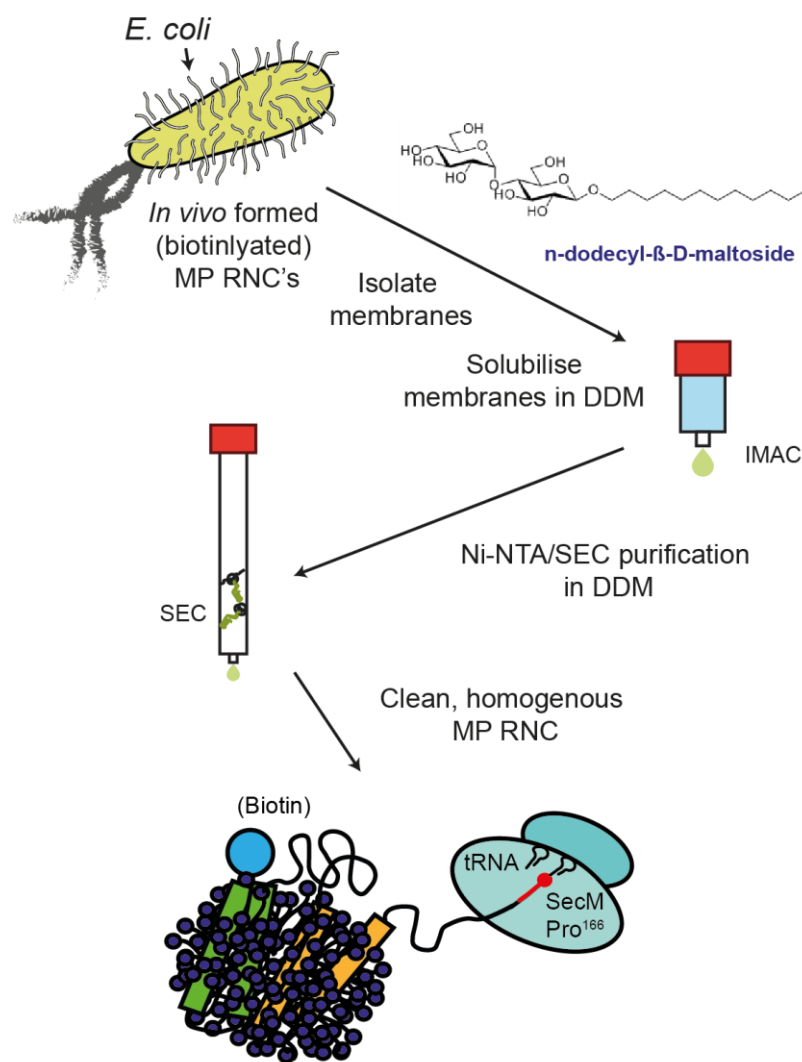
Soluble protein RNCs are commonly generated both *in vivo* and *in vitro* for structural and biochemical study of protein interactions and folding. The first examples of the production of RNCs come from *in vitro* methods, using cell-free extract systems with truncated mRNA to halt translation as first shown with co-translational Globin folding (Komar et al., 1997). Cell-free systems are excellent to study fundamental principles of folding, due to the ease of addition or removal (if recombinant systems like PURExpress are chosen over host cell extracts) of additional chaperones or labels/dyes to study how the proteins interact and fold. These cell-free approaches can produce a relatively clean sample, the yield is often much lower and for biophysical techniques such as NMR, which requires high concentrations of good quality sample (Waudby et al., 2013, Rutkowska et al., 2009, Cassaignau et al., 2016, Cabrita et al., 2009), this is very difficult to achieve. More importantly, for membrane protein work a membrane mimic is required for correct helix insertion. These are often produced from synthetic lipids or directly into detergent which have vast differences to the native membrane.

To increase the yield of RNCs, and to obtain them from a more native environment than a cell-free kit or lysate, I have sought to grow and prepare these designed RNC samples *in vivo* and purify them using DDM detergent. Detergents are very easy to work with and are commonly used for membrane protein study, often for the purification and some structural analysis to assess the quality of the preparation before reconstitution in a synthetic lipid mimic of choice for further study. Unfortunately, the preparation of membrane proteins in this way can often result in a diminishing activity or structure over time (Seddon et al., 2004), and are a poor mimic of the native membrane, where they lack a bilayer environment and compartmentalisation, as well as a lack specific chemical properties intrinsic to the vast array of lipid types which the protein may need to function. Nevertheless, the capture of GlpG RNCs

in a DDM detergent micelle is a major advance for the membrane protein field and is a significant milestone in the reconstitution of RNCs into a lipid bilayer, or purification of the RNCs directly from a native lipid membrane - all methods shown here can be adapted for lipid study, as shown in **Chapter 4**.

### 3.3.4. Polytopic $\alpha$ -helical membrane protein RNC purification in DDM

The generation and purification of GlpG RNCs requires some modifications to the usual production of a membrane protein. A basic schematic of the workflow for DDM purified RNCs is shown in **Figure 3-17**, with a detailed summary of trialled methods to improve preparation quality shown in **Table 3-1**.



**Figure 3-17: Basic summary of DDM RNC preparation steps**

Methods of RNC purification are similar to WT membrane protein purifications, with cell cracking and harvesting of membranes, followed by DDM solubilisation, purification by Ni-NTA affinity chromatography and further clean-up using Sephacryl S400 gel-filtration to produce homogenous RNC samples which can be biotinylated in vivo if desired.

*Table 3-1: Methods summary to obtain homogenous RNC preparation in DDM micelles*

	Growth/expression		Cell cracking			Affinity chromatography/purification		
	MDG/EM9	LB media	Probe sonication	Cryo-milling	Cell disruptor	Online Ni-NTA	Sepacryl-S400	Sucrose
Outcome	Good growth, overnight saturation in MDG, and media exchange into EM9 for expression.	Good growth, induction at 1-1.2 OD <sub>600</sub> . Occasionally more degradation of RNCs.	Inefficient solubilisation, viscous lysate with addition of lysozyme, hence sample loss in debris removal spin. Leads to high portion of released NC due to aggressive cracking strategy.	Very efficient and gentle cracking under liquid N <sub>2</sub> . Prevents repeated freeze thawing which releases NC.	Loss of sample due to foaming under cracking pressure – likely due to required buffer components.	Ni-NTA for DIBMA RNCs worked well, some non-specific ribosome binding but these are removed in purification stage. no alternatives trialed.	Good and efficient separation of void and elution peak, 70S ribosomes intact with bound nascent chain.	Good purification of DDM RNCs in a 5-40 % gradient, however these are time consuming and samples contain sucrose after purification.
Experimental observation	High cell mass on cell harvest which yielded clean RNC as reported by SDS-PAGE Western blot.	Lower cell mass, and overall yield of RNC. No significant difference in RNC quality determined using Western blot and SDS-PAGE for 70S ribosomes.	Large cell-debris pellet, low Ni-NTA elution peak, degraded/released nascent chain on Western blot.	Smaller debris pellet and more material entering solubilisation stage. Higher yield and clean RNCs with lower degree of released nascent chain.	No sample left after cracking to purify from.	Good chromatogram, good SDS-PAGE and Western blots of RNC.	Good quality SEC chromatograms, each peak taken for analysis by SDS-PAGE and Western blot to determine 70S content and quality of intact RNC.	Quality of sucrose purified samples are on par with SEC purification as determined by SDS-PAGE and Western blot and position of RNA in the gradient.



The first stage of RNC generation is the growth and expression in *E. coli*. The RNCs were grown in LB or minimal media and are much less toxic to the cell than isolated N his-GlpG when grown in the same cell line in the same media (**see section 3.2.1**). The cells containing overexpressed RNC were then cracked using a cryo-mill, which uses an electromagnet with bar to pulverise small ‘nuggets’ of cells frozen in liquid nitrogen. This gentle process of cracking avoids repeated freeze-thaw cycles of the cells which can release nascent chain from the ribosome. The cracking is much more efficient than probe sonication, preventing sample loss during debris centrifugation, as well as preventing foam formation and thus sample loss when using a cell-disruptor or French press and overall increasing the yield of the RNC purification.

To streamline the membrane protein RNC purification process, the use of a Sephacryl-S400 resin gel-filtration column was validated for the purification of membrane protein RNCs. Sephacryl S400 was first used to purify ribosomes in (Becker et al., 2012), which replaces the standard sucrose preparation, with the shorter purification time being preferable for maintaining the integrity of membrane protein samples in detergent micelles (Seddon et al., 2004). The Sephacryl resin has a reasonable separation resolution of 20 kDa – 8 MDa as suggested by the manufacturer which is ideal for separating the empty and aggregated ribosomes (which elute at ~43 mL) from the intact RNC (~60 mL) and the released nascent chain (~86 mL) in the late stage of the column. This purification strategy produces a similar, if not better quality RNC sample as previously published RNCs produced *in vivo* (Cassaignau et al., 2016, Rutkowska et al., 2009, Schaffitzel and Ban, 2007).

The quality of the purified GlpG RNCs was checked using SDS-PAGE analysis with Coomassie and Western blotting. RNase A treatment digests the bound tRNA, leaving the stalled polypeptide and a shift in molecular weight is observed when blotted. The presence of the 70S ribosome is also checked on SDS-PAGE by checking for presence of the characteristic ribosome banding (Dzionara et al., 1970). Low intensity bands underneath the major GlpG construct bands were observed by Western blotting and attributed to low populations of truncated GlpG. This is commonly observed within preparation of RNCs by SecM based stalling (Bercovich-Kinori and Bibi, 2015, Cabrita et al., 2009), caused by digestion with remaining proteases during purification. Notably, when the GlpG constructs were released from the ribosome they became more prone to degradation, suggesting that the constructs are stabilised and/or protected by the ribosome to protease digestion, this is commonly observed in soluble protein emergence from the ribosomes (Samelson et al., 2018, Jensen et al., 2020) and in soluble RNC

production (Schaffitzel and Ban, 2007, Rutkowska et al., 2009, Cassaignau et al., 2016), and in the small amount of work on short membrane protein RNCs, although its often hidden in supplementary material (Kater et al., 2019, Bischoff et al., 2014).

Yields of RNC varied between nascent chain lengths when overexpressed and purified under identical conditions. Purified yields for the 2 TM, 4 TM, and 6 TM constructs were ~ 30, 100, and 25 pmol/mL respectively (around 0.1 mg per litre of culture). Yields are not often discussed in the limited literature, however these individual yields may be linked to the extent of expected membrane force-pulling of the different nascent chain lengths (Cymer et al., 2015b, Ismail et al., 2015, Ismail et al., 2012), with an increased “insertion force” of some sequences reducing the stalling effect of SecM and in turn quantity of stable RNCs obtained. MPEX analysis indicates that the 4 TM construct is preceded by TM regions TM 3 and TM 4 that have the least favourable insertion energetics. As a result, SecM stalling is effective and this is the highest yielding construct. Whereas the 2 and 6 TM constructs are preceded by TM’s with much more favourable insertion energetics (lower energy minima), which may lead to decreased RNC stalling stability and, therefore, the reduced yields observed. Of course, for folding studies, forcing a naturally transient folding intermediate into a lowest energy barrier is perhaps not supported by the cell, and may skew representation of folding pathways. Currently, however, this method of RNC design and production is the closest the field has come to recapitulating the *in vivo* folding pathway.

The yields achieved for GlpG RNCs are sufficient for further structural and biochemical analyses but are around 100-fold less than those previously quoted for small globular protein RNCs (Cassaignau et al., 2016). This is likely due to the toxic nature of overexpressing membrane proteins, caused in part by saturation of available SecYEG translocon (Schlegel et al., 2013). Indeed, GlpG has been shown to insert into the *E. coli* inner membrane using SecYEG and YidC (Schibich et al., 2016), however details of these interactions are currently unknown (Komar et al., 2016).

There is a significant release of nascent chain from the ribosome in purified WT SecM samples as judged by both Western blot and negative stain EM. The released nascent chain, or indeed ribosome-bound nascent chains are far too small to see without further processing using negative stain. To improve this study, a much higher proportion of occupied ribosomes is required. The 2D class with the green arrow (**Figure 3-8**) does show density, this is very large and is likely to be something much bigger than nascent chain and may be presence of the bacterial SecYEG translocon which co-purifies with the RNC. For successful cryoEM

experiments observing the structure and possible interactions of the nascent chain during folding, a high occupancy of ribosomes with homogenous sample is required. It does appear that although my purification strategy produces clean sample, there is a large degree of released nascent chain which occurs post-purification on the thawing of stored samples. This will need to be improved to streamline 2D classification during the image processing stage to get better quality images, particularly of the small 30 kDa nascent chain which is unable to be seen by eye, which should be possible to achieve with incorporation of enhanced SecM, although this has not yet been trialled.

### **3.3.5. *In vivo* biotinylation for purification and structural experimentation**

Nickel affinity resin has been shown to bind non-specifically to ribosomes (Schaffitzel and Ban, 2007, Bischoff et al., 2014) and so an additional, non-Ni-NTA purification method could be advantageous for an alternative purification strategy. Others in the field use Strep tags (Schaffitzel and Ban, 2007, Bischoff et al., 2014, Rutkowska et al., 2009) which can be altered to improve yield of RNC from affinity purification 20-fold when a sequence of three strep tags are interspersed with glycine and serine spacers of at least 5 amino acids (Schaffitzel and Ban, 2007). It would, therefore, be pertinent to test whether incorporation of a high affinity tag within membrane protein RNC constructs was viable to incorporate further flexibility in purification and biophysical interrogation in a single co-expression at the growth stage of RNC generation.

*In vivo* biotinylation of the RNCs is a very useful method for a further (or alternative) purification strategy. The co-expression of BirA with the RNC expressing an avi-tag is rapidly biotinylated when D-biotin is added to the growth media. I can sufficiently tag the RNC with biotin *in vivo* and confirm this by Western blotting with streptavidin-HRP. It would also be necessary to determine the efficiency of tagging using mass-spectrometry to gauge how many RNC molecules are tagged correctly for downstream analysis (Rutkowska et al., 2009). To do this, the RNC would be re-bound to Ni-NTA beads and treated with RNase A before washing away the disassembled ribosomal components and leaving behind the nascent chain which can be eluted and precipitated for MS analysis. This is also a good way of validating the correct nascent chain length based on the mass of the peptide released, as you may expect to see the presence of polysomes resulting from the queueing of ribosomes after the first translationally stalled one (Schaffitzel and Ban, 2007) which cannot be identified on the low resolution SDS-PAGE gels, or separated by gel-filtration or sucrose gradient. Initial experiments to precipitate the nascent chain away from the ribosome proved difficult and yields of nascent chain were

too low for MS analysis after enrichment and precipitation steps. Polysomes have been removed previously using the co-expression of MazF endoribonuclease when preparing transmembrane spanning RNCs (Park et al., 2014). This is another plausible method for improving RNC production to confirm polysomes are not present, however over-expressing many proteins in a cell is often incredibly toxic and can result in very low yield.

Secondary tagging (in addition to His<sub>10</sub> tag) is also important for co-purification strategies where multiple proteins require multiple but different tags. I have also successfully biotinylated the SecE protein of the SecYEG translocon complex using this *in vivo* strategy. Biotinylation again was confirmed using Western blot with a streptavidin conjugated HRP.

In addition to further purification, biotinylated samples can be useful for structural and functional assays (Chang and Bowie, 2014, Guo et al., 2016, Kaiser et al., 2011, Min et al., 2015, Min et al., 2016). Biotinylated proteins are particularly useful for hydrogen/deuterium exchange mass spectrometry (HDX/MS) which may be of some use to study the dynamics of the emerging nascent chain. The biotinylation of the N-terminal Avi-tag allows the selective enrichment of these nascent chains with deuterium, which can be quenched with a low pH dye, of which the biotin-streptavidin bond is tolerant (Jensen et al., 2013). GlpG dynamics have been previously studied in native nanodiscs using SMA (Reading et al., 2017), but how are helices arranged throughout the co-translational folding cycle, and whether there any particular regions of interest through the protein which may be prone to uptake or protection which aren't observed when studying isolated protein, may be highlighted with this powerful technique.

The biotin-streptavidin bond is also useful for mechanical force experiment tethering due to the high affinity nature of the interaction (Kaiser et al., 2011, Goldman et al., 2015). The biotin-tagged biomolecule conjugates to a streptavidin coated magnetic bead and for the study of low force intrinsic folding intermediates in a magnetic tweezer (or optical trap) set-up. For GlpG in particular, this mechanical folding method was pioneered by the Bowie Lab, where isolated GlpG, reconstituted into bicelles and MSP-based nanodiscs was used to study the intrinsic folding forces at each intermediate of lateral protein unfolding (Min et al., 2016, Min et al., 2015, Choi et al., 2019). Soluble protein RNCs have also been studied using mechanical force, where the ribosome is tethered at one end, followed by the nascent chain at the other (Kaiser et al., 2011, Goldman et al., 2015, Liu et al., 2019). Therefore, my biotinylation study could provide a means to study membrane protein RNC folding using mechanical force.

### 3.4. Chapter summary

Throughout this chapter I have explored the development of a novel method to design, express and purify stable polytopic membrane protein ribosome-bound nascent chain complexes which will revolutionise the study of co-translational membrane proteins. The major themes successfully detailed in this chapter are as follows:

- Development of a design strategy to produce stable GlpG RNCs using (enhanced) SecM at varying positions through the translation cycle which could be applied to other membrane proteins using MPEX.
- Development of the purification protocols using affinity chromatography and size-exclusion chromatography to produce homogenous GlpG RNCs in DDM detergent.
- Development of the *in vivo* biotinylation labelling strategy for downstream RNC purification or structural experimentation.
- Improving the stability of GlpG RNCs using an arrest enhanced SecM sequence.
- Capture of endogenous SecY with overexpressed RNCs, and the co-expression and solubilisation of SecYEG with the GlpG 4 TM RNC construct.

## 4. Capturing RNCs in a native-like environment for co-translational studies

---

**This chapter, along with Chapter 3, has been modified from publication with additional details and discussions. The full article can be found in appendix 10:**

Capturing membrane protein ribosome nascent-chain complexes in a native-like environment for co-translational studies

**Pellowe GA**, Findlay HE, Lee K, Gemeinhardt TM, Blackholly LR, Reading E, Booth PJ,  
*Biochemistry*, 2020

doi: 10.1021/acs.biochem.0c00423

## 4.1. Introduction

Following on from the proof of principle study in **chapter 3** with the preparation of membrane protein RNCs in detergent, the logical next step is to adapt the methods to produce GlpG RNCs in a lipid environment. Lipids come in many forms, with many different chemistries and currently mixes of synthetic lipids with specific chemistries can achieve this. However, the most ideal environment for the protein is its native lipid environment. Native lipid environments can be obtained by solvent extraction of cell membranes and prepared for protein reconstitution, but recently a much more accessible system in the form of polymer nanodiscs have been employed to purify proteins directly from the membrane without first extracting them using detergent. This forms a polymer ring around a disc of native lipids with the membrane protein embedded inside, producing a sample in a native membrane environment that is highly amenable to biochemical and structural interrogation.

### 4.1.1. Native nanodiscs for *in vivo*-like *in vitro* study

The most common system for this membrane capture utilises styrene-maleic acid copolymer lipid particle (SMALP) system. The use of SMALPs (Pollock et al., 2018) has led to the study of membrane proteins in a much more native membrane environment - as well as capturing the protein, the polymer captures the surrounding lipids, and the discs are often small enough for use with powerful biophysical and structural techniques. A great advantage of membrane protein extraction by SMA, and other polymers, is the applicability to proteins from a wide range of proteins from different organisms, with proteins from bacteria, archaea, yeast, insect, and human cells which are all purified relatively easily using affinity chromatography before downstream biophysical studies (Bada Juarez et al., 2019). An overview of the types of polymer used in the field is detailed in **chapter 1**. Here I explore specific studies of membrane proteins in polymer-nanodiscs which give further insight into the structure and function of proteins when immersed in a native-like environment. The SMALP field is rapidly growing, and extensive reviews can be found (Overduin and Esmaili, 2019, Lemieux and Overduin, 2021, Bada Juarez et al., 2019, Ravula et al., 2019, Pollock et al., 2018) and at [www.smalp.net](http://www.smalp.net).

The first use of SMA to isolate protein containing native nanodiscs showed the successful a purification of the  $\alpha$ -helical bR, and the  $\beta$ -barrel PagP reconstituted DMPC liposomes in the SMA polymer (Knowles et al., 2009). Solubilisation was confirmed by size-exclusion chromatography and TEM, showing discs of approximately 11 nm, containing 11 molecules of PC, as shown by NMR (Knowles et al., 2009). The first examples of solubilisation directly from the native membrane were shown using KcsA, the K<sup>+</sup> channel (Dorr et al., 2014), showing

purification using gel filtration chromatography, testing lipid content of the discs using thin layer chromatography (TLC), and testing the functionality of the protein in the native environment using an electrophysiologic assay potassium transport across a planar bilayer (Dorr et al., 2014). SMALPs are also much more stable than detergent micelles, maintaining the protein structure, and can extract a stable tetrameric form of KcsA in the native lipid nanodisc (Dorr et al., 2014). Although an advance over protein reconstitution and production of nanodiscs, the styrene-moiety is still highly absorbent in the far-UV region, and so spectroscopic techniques like CD cannot obtain a clean signal below 200 nm and deconvolution of secondary structure is not possible for folding studies (Oluwole et al., 2017a).

SMALPs can be utilised alongside a multitude of different techniques such as NMR spectroscopy, native MS, HDX-MS, and single molecule fluorescence imaging to study membrane proteins in a near native environment (Swiecicki et al., 2020, Lloris-Garcera et al., 2020, Bibow, 2019) and such experiments are explored in the following introduction.

#### **4.1.1.1. Stability and functional studies**

Using detergent to purify membrane proteins can often destabilise a protein of interest leading to irreversible conformational change or aggregation. This makes biochemical/physical studies of the protein impossible. SMA technology has removed many issues which are faced when using detergent for membrane protein purification. The presence of lipids in the sample may be expected to increase the stability of the protein over a detergent purified sample, and many sources have sought to quantify this inherent stability of the native nanodisc. Furthermore, mass spectrometry has been successfully used for lipidomics to ascertain which lipid species are captured with the native nanodisc (Reading et al., 2017, Reading, 2018, Pyle et al., 2018).

Thermostability of a protein can be measured using any technique which can quantify structure reduction. CD, which measures the relative absorption between right- and left-handed circularly polarised light gives an indication into chiral molecules, much like protein secondary structure, and is commonly used alongside intrinsic fluorescence quenching to follow secondary structure loss. Unlike large proteoliposomes, SMALP discs have negligible CD absorbance, and do not scatter light significantly (Knowles et al., 2009) between wavelengths 200 and 300 nm which encompasses characteristic regions for  $\alpha$ -helical, or  $\beta$ -strand proteins. The lack of absorption in this region makes following unfolding of proteins in a native nanodisc



easier than detergent purified and reconstituted proteins, or with MSP-nanodiscs which interfere with this signal due to the proteinaceous structure of MSP itself.

The A<sub>2A</sub>R GPCR from *Pichia pastoris* exhibits an increased thermostability between SMALP and DDM solubilised A<sub>2A</sub>R of ~ 5.5 °C, from a  $T_{50}$  of 44.4 to 49.9 °C between the DDM to SMALP when characterised by CD (Jamshad et al., 2015). Similar experiments were run on KcsA, using CD as well as intrinsic tryptophan fluorescence with thermal denaturation. Similarly, to A<sub>2A</sub>R the protein showed an increased stability when the protein is surrounded by SMA compared to DDM, which loses around 50 % of its secondary structure compared to around 20 % with SMA. (Dorr et al., 2014). This stabilisation effect for KcsA may be explained by its tetrameric state which is first disrupted by heat in the micellar environment of DDM, leading to more pronounced unfolding of the individual monomers, unlike SMA which preserves the oligomeric state with surrounding lipids. Tolerance to repeated freeze thaw cycles and longer storage times are also characteristic of SMALPs, with A<sub>2A</sub>R showing protein half-lives of >16 days for native nanodiscs, compared to 1.8 days for DDM samples, binding affinity of the GPCR was also maintained, with binding efficiency lasting in excess of 5 freeze-thaw cycles, compared to DDM samples which lost efficacy after a single freeze thaw (Jamshad et al., 2015).

As well as stability of the protein, activity must also be maintained to avoid characterising an inactive (and therefore potentially misfolded) protein structure. Functional assays are specific for the protein of interest and therefore many ways of achieving this have been tested. The PagP barrel which phospholipase activity for example can be measured using NMR by detection of a <sup>13</sup>C-lysoPC product as a breakdown of a radiolabelled phosphocholine lipid (Knowles et al., 2009). NMR can also be used with <sup>19</sup>F unnatural amino acid incorporation to observe autophosphorylation of the ETK bacterial tyrosine kinase in SMALPs (Li et al., 2015). A selection of radioligand binding assays have also be reported for GPCRs in SMALPS to test pharmacological properties of the eukaryotic proteins (Jamshad et al., 2015), as well as single channel conductivity experiments to observe transport activity of potassium through the KcsA tetramer – this confirmed activity was identical to detergent purified protein reconstituted into lipids (Dorr et al., 2014). Unfortunately, the lack of compartmentalisation when using SMALPs disrupts the possibility of assaying transport by any other means, such as fluorescence or enzyme-linked assays (see **section 5.2.6** for work on Xyle activity assays), and single-channel patch clamp experiments with well-placed probes are only suitable for certain proteins, and require advance set-ups which not all labs have access to.

More complex functional assays using larger complexes are also possible. FRET analysis of SecYEG function in SMA has been shown by mixing a fluorescently labelled proOmpA-DHFR substrate with SecYEG in SMA, in the presence of ATP. As the substrate entered the translocon, the fluorescence signal increased as expected, and this was not the case in the presence of a non-hydrolysable ATP analogue (Prabudiansyah et al., 2015). Deduction of the rate of hydrolysis of ATP however is not compatible with SMALPs, this is a typical measure for ABC transporters activity which requires presence of magnesium for function.  $Mg^{2+}$  is not compatible with SMA (Smirnova et al., 2016). These ABC transporters, often of 12 TM helices undergo large conformational changes in response to nucleotide binding and hydrolysis, and it is currently unknown whether the 10 nm diameter of SMALPs will allow for this extensive conformational change due to the inability to test for function. Perhaps a DIBMA based nanodisc which are larger in diameter (~ 25 nm, compared to SMA ~ 10 nm), and tolerant to divalent cations (Gulamhussein et al., 2020) would be useful in this respect.

#### **4.1.1.2. Structure and dynamics of proteins in SMALPs**

Structural studies of membrane proteins are notoriously difficult to achieve due to the requirements of high yields of stable and homogenous samples. The enhanced stability of SMALPs make these protein samples ideal for such studies, and samples are amenable to negative stain and cryo-electron microscopy, x-ray crystallography and solid-state NMR. However, speculation over the difficulty to crystallise proteins for structural acquisition has been highlighted due to the high number of lipids around a protein. Fortunately, Cryo-electron microscopy and NMR do not require protein crystals, and sample preparation and equipment have vastly improved to increase resolution which was previously only attainable with x-ray crystallography.

Solution NMR has been thought not to be compatible to SMALPS due to the larger size of the discs which limit tumbling speed of the proteins – already a drawback of any NMR experiment on membrane proteins in detergent (Pollock et al., 2018). Magic-angle spinning solid state NMR which magnetically aligns the nanodiscs with uniformed orientation was used with the zinc diffusion facilitator (CzcD) and was able to generate 3D NMR spectra using  $^1H$ - $^{15}N$ - $^{13}C$  correlation experiments in an attempt to achieve atomic resolution. However, due to spectral overlap it was not possible to assign the structures (Bersch et al., 2017). More recently, a selection of proteins; the coat protein Pf1 from bacteriophage (1 TM helix), a truncated bacterial mercury detoxification protein, MerFt (2 TM helix), outer membrane protein Ail from *Yersinia pestis* (8 stranded  $\beta$ -barrel) and human chemokine receptor CXCR1 (7 TM helices),

were all SMALPed from reconstituted DMPC and DMPG lipids and subjected to solid-state NMR to yield high resolution 2D  $^1\text{H}/^{15}\text{N}$  spectra. These experiments show it is possible to achieve spectra which directly study structure, membrane orientation and dynamics of the proteins in their native environment (Park et al., 2020).

Cryo-electron microscopy has been extensively used to obtain structures of proteins to high resolution in native nanodiscs. Thus far, in addition to MSP-derived nanodiscs which have recently yielded a 1.9 Å resolution of Connexins-46/50 solubilised in two nanodiscs (Flores et al., 2020), the SMA polymer is the only polymer which has been used for high-resolution structural acquisition. A summary of these experiments and corresponding PDB entries are highlighted in **Table 4-1**.

Table 4-1: Structures of membrane proteins in SMA polymer nanodiscs

Protein	Function	TM-spanning helices	Host Organism	Expression organism	Method	Resolution (Å)	PDB	References
P-Glycoprotein/ABCB1	Multidrug resistance protein	12	<i>Homo sapiens</i>	<i>Trichoplusia ni</i>	Negative stain CryoEM	35	N/A	(Gulati et al., 2014)
HwBR	Proton pump bacteriorhodopsin	7	<i>Haloquadratum walsbyi</i>	<i>E. coli</i>	X-ray	2.0	5ITC	(Broecker et al., 2017)
TRPV1	Capsaicin receptor; cation channel	6 per monomer; homo-4-mer	<i>Rattus norvegicus</i>	HEK293S	CryoEM	3.4	3J5P	(Liao et al., 2013)
KimA	K <sup>+</sup> /H <sup>+</sup> symporter	12	<i>Bacillus subtilis</i>	<i>E. coli</i>	CryoEM	3.7	6S3K	(Tascon et al., 2020)
ZipA	Bacterial cell division protein	1	<i>E. coli</i>		Negative stain CryoEM	16	N/A	(Lee et al., 2019)
Alternative Complex III	Key component of bacterial electron transport chain	23 total, hetero-6-mer	<i>Flavobacterium johnsoniae</i>		CryoEM	3.4	6BTM	(Sun et al., 2018)
AcrB	Multidrug efflux pump	12 per monomer; homo-3-mer	<i>E. coli</i>		Negative stain and CryoEM	23 and 3.2	N/A and 6BAJ	(Postis et al., 2015) (Qiu et al., 2018)

#### **4.1.2. GlpG in native nanodiscs**

The membrane environment is key for the structure and function of GlpG, unfolding experiments have shown that GlpG has low thermodynamic packing stability, and the membrane is therefore thought to have a stabilising effect on the protein (Baker and Urban, 2012). Part of the reason for selecting GlpG for the RNC native nanodisc work was due to its previous investigation and suitability for SMALP, and DIBMALP solubilisation. SMALP technology was used for the investigation of GlpG using HDX-MS technology (Reading et al., 2017), and DIBMA solubilised GlpG is functional (Barniol-Xicota and Verhelst, 2018).

The HDX-MS study on GlpG in SMA sought to determine changes in GlpG dynamics in response to a varied lipid environment (Reading et al., 2017). Lipid conditions from three cell lines, BL21(DE3), and C43(DE3), grown at 16 or 37 °C to determine any observable changes in GlpG dynamics with lipid environment. Interestingly, where peptides corresponding to TM 2-6 were relatively well protected from HDX, part of TM1, and linker region connecting CytD, was unprotected and showed dynamic behaviour (Reading et al., 2017, Reading, 2018). Lipid head group did not directly affect HDX, however chain length and degree of lipid tail saturation did. TM 1, linker and CytD all showed a degree of lipid sensitivity when subject to 16 °C cold shock, resulting in an increased chain length, and tail un-saturation, hence a more fluid membrane (Morein et al., 1996). These three regions have previously been suggested to play a role in the gating of the substrate and function of the protease (Foo et al., 2015).

In addition to solubilisation using SMA, GlpG has also been solubilised into DIBMA and is functional and contains the correct lipid composition expected at the *E. coli* inner membrane (Barniol-Xicota and Verhelst, 2018).

#### **4.1.3. DIBMA and its suitability for RNC capture**

Whilst SMA is the front runner in polymer nanodisc encapsulation, it does have significant drawbacks based mainly around its low tolerance for divalent cations, its high absorption in the UV wavelength region, and its small disc size which may only allow for smaller proteins to be solubilised (Gulamhussein et al., 2020). The DIBMA co-polymer was discovered as a tool for the solubilisation of membrane proteins and improves some of the issues faced when using SMA.

DIBMA is an alternating co-polymer of maleic acid and diisobutylene with 1:1 stoichiometry (compared to 2:1 or 3:1 in SMA) and is commercially available under the tradename Sokalan CP9 from BASF. When considering preparation and analysis of ribosome containing samples,

the SMA co-polymer has its disadvantages: SMA contains a large styrene aromatic ring which contributes to high absorption spectra in the low UV region, which makes concentration and purity analysis of RNCs using  $A_{280}/A_{260}$  difficult (Oluwole et al., 2017b, Oluwole et al., 2017a). SMA is also less tolerant to divalent cations, with the precipitation of the polymer occurring at concentrations  $>5$  mM  $Mg^{2+}$  (Simon et al., 2018). An *E. coli* 70S ribosome contains more than 170  $Mg^{2+}$  atoms, with loss of bound- $Mg^{2+}$  leading to 70S dissociation and degradation (Nierhaus, 2014). Therefore, to maintain 70S integrity *in vitro*  $Mg^{2+}$  concentrations  $>5$  mM is typically required, making SMA unsuitable for RNC study.

To circumvent these spectroscopic and divalent cation issues, the recently utilized DIBMA co-polymer was selected (Oluwole et al., 2017a). Its polymer structure contains an aliphatic diisobutylene moiety instead of the aromatic styrene component found within SMA. This allows for spectroscopic study as the styrene component which absorbed strongly within 260-280 nm has been removed. Importantly, DIBMA also has a proven tolerance to high concentrations of divalent cations (up to 20 mM for  $Mg^{2+}$  and 35 mM for  $Ca^{2+}$ ) (Oluwole et al., 2017b, Danielczak et al., 2019), making it a suitable candidate for ribosome study, it is unclear why this is the case however, as DIBMA contains a higher percentage of maleic acid groups – and it's been suggested that the larger discs formed with DIBMA are less strained when binding cations (Gulamhussein et al., 2020). Alternatively, it has been suggested that the higher degree of maleic acid makes the polymer more hydrophilic and can therefore tolerate a higher degree of neutralization by  $Mg^{2+}$  or  $Ca^{2+}$ , it was actually shown that increasing the concentration of cations led to a greater solubilisation efficiency of DIBMA (Danielczak et al., 2019). DIBMA is unfortunately much less efficient at solubilising membranes than SMA. The pH, cation concentration, and method of polymer purification can all influence the efficiency of solubilisation and subsequent yield of protein. Comparing solubilisations of two proteins; single TM spanning ZipA, and 12 TM BmrA in SMA200 and DIBMA from *E. coli* showed that for DIBMA, solubilisation efficiency was 1.8x less than SMA, yielding 2.5x less protein. However, when extracting GPCR  $A_{2A}R$  from *P. pastoris* or COS7 cells, there appeared to be no significant difference between SMA and DIBMA solubilisation as deduced by specific binding of a radiolabelled agonist to the GPCR (Gulamhussein et al., 2020).

It has also been suggested that the solubilisation efficiency of DMPC bilayers by DIBMA is lower than SMA likely due to the neopentyl moieties being less able to intercalate among the lipid acyl chains (Grethen et al., 2017). Solubilisation tests of *E. coli* membranes were also carried out using SMA(2:1) and SMA(3:1) compared to DIBMA at varying pH which is also more

efficient from a thermodynamic perspective at pH 8.3 than at physiological pH. DIBMA appears to preferentially extract much larger proteins, which is likely due to its formation of larger nanodiscs than the two SMA variants (Oluwole et al., 2017a) and is milder in its solubilisation as determined by the effects on lipid thermotropic behaviour as tested using differential scanning calorimetry (DSC) (Grethen et al., 2017).

Additionally, lipid exchange is a known feature of polymer-based nanodiscs. This has recently been compared for SMA and DIBMA using stopped-flow FRET which resulted in observed a high degree of lipid exchange between discs, revealing that rather than a static snapshot of proteins in a native environment, are in a highly dynamic equilibrium assembly with exchange after initial solubilisation (Danielczak and Keller, 2020). Lipid dynamics in DIBMA discs have also been studied using EPR spectroscopy. Nitroxide spin labels placed on the fatty-acid tails of synthetic phospholipids allowing the motility of the lipid chains to be studied at various bilayer depths in synthetic mixes of DMPC and POPC. These results showed that DIBMALPs indicate a weaker interaction between the polymer and higher lipid mobility compared to SMALPs however this could be a byproduct of the DIBMALPs larger size (10-35 nm for DIBMALPs compared to 8-10 nm for SMALPs) (Colbasevici et al., 2020). Both DIBMA and SMA copolymers produce thermostable MP nanodiscs, but DIBMA has a lower impact on acyl chain order and lipid phase in solubilised membranes than SMA, maintaining a nearer physiological membrane environment (Oluwole et al., 2017a).

This chapter describes the first time that polymer nanodiscs have been isolated in the presence of the ribosome, and DIBMA lends itself perfectly, barring its lower solubilisation efficiency compared to SMA, to this work. And the larger discs allow the possible capture of not only the RNC helices, but also the SecYEG translocon which is a much larger complex with functional lipids. SecYEG with associated proteins of the holo-translocon have previously been isolated using SMA (Komar et al., 2016), however the small disc diameter may contain too few functional lipids. Cardiolipin, for example, is required for successful translocon function. Additionally, the smaller disc size of SMA may conformationally restrict dynamics, or space in the bilayer for RNC intermediates to form, and therefore the larger disc size of DIBMA may solve these issues.

#### **4.1.4. Non-native nanodiscs and membrane protein RNCs**

Membrane protein RNCs are considerably more challenging to isolate than those for water-soluble proteins, as they are complicated by the requirement of a lipid membrane and thus

far, nobody has purified an RNC into a native nanodisc. Nevertheless, notable achievements have been made in the CryoEM structural characterisation of membrane protein RNCs using MSP-based nanodiscs with synthetic lipids. For example, the Beckmann group uncovered, at high resolution, that interaction of a TM nascent chain with SecY can cause opening of its channel upon the insertion of the FtsQ anchor domain TnaC-mediated RNC into translocon (Kater et al., 2019), likely to facilitate its insertion into the membrane. To achieve this interaction, they used detergent solubilised RNCs which were then subsequently reconstituted with SecY-containing membrane scaffold protein (MSP)-based nanodiscs consisting of non-native lipid DMPC (Seidelt et al., 2009, Kater et al., 2019, Bischoff et al., 2014). Membrane mimetics such as these provide a lipid bilayer but lack the native lipid composition and, therefore, may not truly represent its functional environment.

These experiments provide valuable insight of the SecY lateral gate opening in response to its RNC substrate, and therefore I believe capturing an emerging polytopic protein in a native nanodisc with the SecYEG translocation machinery present, rather than a synthetic system, will yield a highly amenable sample to high resolution structural techniques has been demonstrated above.

#### **4.1.5. Aims of chapter**

In view of the importance of lipid composition and membrane properties in defining the behaviour of membrane proteins, I have developed an approach to extract RNCs directly from their cellular membrane. I present a strategy, building on from those introduced in **Chapter 3**, to produce high yields of homogenous polytopic membrane protein RNCs within their native lipid environments. Membrane protein RNCs are isolated from native membranes using DIBMA co-polymer native nanodisc technology (Denisov and Sligar, 2017, Simon et al., 2018, Ravula et al., 2019). Overall, my approach enables the structural and biochemical interrogation of cellular co-translational membrane protein translation and folding within a more physiological environment than is currently possible with existing *in vitro* or reconstituted systems. I have split this chapter into two sections. The first, Part A, describes the necessary steps required to prepare the DIBMA polymer for solubilisation of membranes, and to check its ability to do so, and Part B, which describes the use of this polymer to capture GlpG RNCs from the native membrane.



## 4.2. Part A: DIBMA - Moving towards an *in vivo* environment

The diisobutylene-maleic acid co-polymer is advantageous for membrane protein study as it captures the native lipids around a protein, and lacks the aromatic ring characteristic of SMA, which allows spectroscopic analysis of the protein using CD or fluorescence in the far-UV region. I have optimised the production of DIBMA for ample lipid solubilisation of a synthetic lipid bilayer, comprising of a 40:40:20 mix of DOPE:DOPC:DOPG to confirm that the polymer can correctly solubilise a membrane of expected composition. This was analysed by HPTLC analysis. MFS transporter Xyle (see **chapter 5** for introduction and reconstitution results) was reconstituted into the same synthetic bilayer, was solubilised, and was also correctly folded as assayed by CD. This confirmed that DIBMA is suitable for capturing large proteins with the correct surrounding lipid environment, providing a sound basis for the preparation of GlpG RNCs from their native lipid environment using DIBMA in **Part B, section 4.3**.

### 4.2.1. Preparation of DIBMA

DIBMA is a co-polymer of maleic acid and diisobutylene. The maleic acid must be hydrolysed to produce the membrane active form of the polymer. Unlike SMA, DIBMA is shipped in its hydrolysed state, as the 'Sokalan CP9 alkaline solution' from BASF. It is unknown what the composition of the alkaline solution is, and so preparation of the polymer into a known buffer of choice is advantageous for protein study. Many methods have been published to do just this (Oluwole et al., 2017a, Gulamhussein et al., 2020), mostly relying on dialysis of the polymer into a desired buffer. This however causes an increase in volume of the polymer, altering its concentration which can only be determined using refractometry (**Equation 4-1**). Additionally, small variations in volumes between the outer and inner buffer can lead to large variations in the degree of effectiveness of solubilisation (Gulamhussein et al., 2020).

*Equation 4-1: Molar refractive index increment for DIBMA*

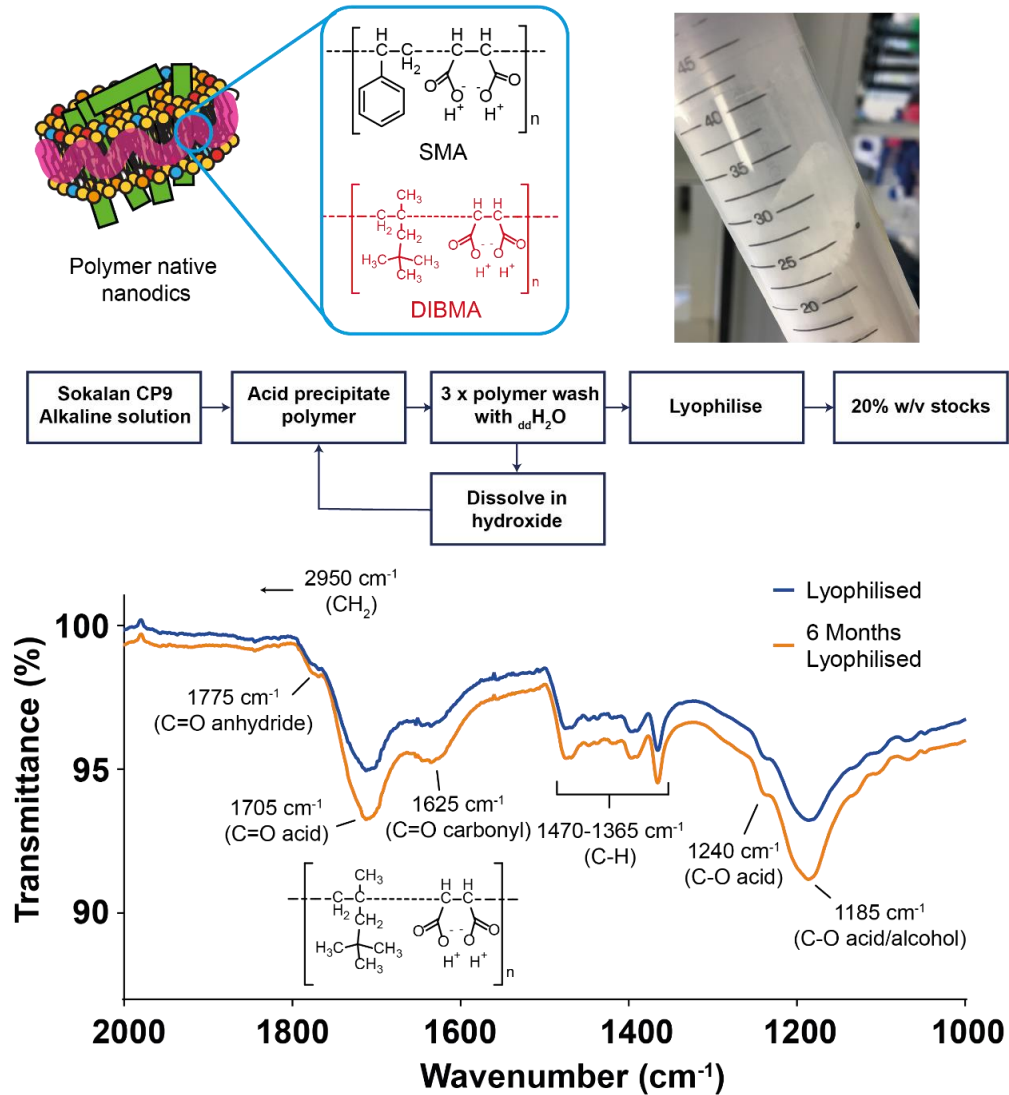
$$\frac{d_n}{d_c} = 1.346. M^{-1}$$

To remove any impurities and inconsistencies in the protocol and polymer, I have trialled an acid-based precipitation, wash, and lyophilisation of the polymer before resuspending in the buffer of choice. This prevents the dilution of the buffer mix which occurs with excess volumes of DIBMA stock when producing the required final polymer concentration. To do this, 4 M HCl was used to precipitate the polymer, and a series of double-distilled water washes with centrifugation harvest were carried out. The polymer was then re-dissolved into 4 M NaOH

and the acid precipitation and wash steps were repeated. Finally, the wet powder obtained from the water wash was lyophilised to a fine white dust, containing pure DIBMA with residual salts removed. A flowchart of workflow and a photograph of the lyophilised DIBMA are shown in **Figure 4-1**.

To ensure my product is consistent, I ran FTIR in the fingerprint region on the powder where I confirm the structure matches the spectra, and also matches published work (Barniol-Xicota and Verhelst, 2018). Each wavenumber displayed in the FTIR spectra was assigned and corresponded well with published data (Gulamhussein et al., 2020). In particular, I checked that polymer processing does not interfere with the structure of DIBMA - complete hydrolysis (the membrane active form) of the maleic acid anhydride ring was confirmed with a carboxylic acid C=O stretch  $1705\text{ cm}^{-1}$ , rather than an anhydride band at  $1775\text{ cm}^{-1}$ . C-H stretching vibrations of diisobutylene are also present at  $2950\text{ cm}^{-1}$  (not shown) and in the  $1470\text{-}1365\text{ cm}^{-1}$  region. After 6 months, there was no change in the molecular structure of lyophilised DIBMA **Figure 4-1**.

To achieve solubilisation, the lyophilised powder was dissolved into 250 mM NaOH and adjusted to pH 8.0 to form 20 % (w/v) stocks. This was subsequently be used to produce solubilisation buffers of the desired concentration and composition for purification.



**Figure 4-1: DIBMA preparation and IR quality control**

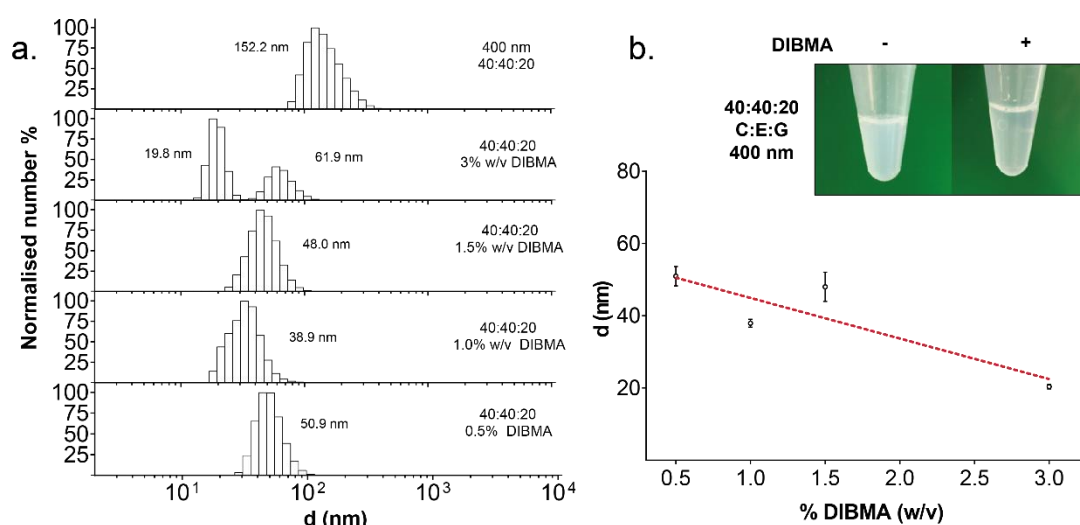
Schematic of polymer native nanodiscs with structures of SMA and DIBMA. Photograph of lyophilised DIBMA after processing as per flow-chart. FTIR analysis of lyophilised DIBMA recorded at room temperature before and after 6 months showing no degradation and maintaining hydrolysed structure 1775 cm<sup>-1</sup> and 1705 cm<sup>-1</sup>.

#### 4.2.2. DIBMA solubilises synthetic lipids

Although the potential of DIBMA lies with solubilising native *E. coli* membranes directly, I first trialled solubilising mixes of synthetic liposomes containing 40:40:20 mixtures of DOPE:DOPC:DOPG to confirm that DIBMA does indeed solubilise the expected lipids in the correct composition.

**Figure 4-2** shows DLS traces to assess the size of 400 nm liposomes of the 40:40:20 synthetic lipid mixes with varying concentrations of DIBMA. The 400 nm liposomes, when run in a standard sodium-phosphate buffer, had an average diameter of 152.2 nm, less than the expected 400 nm but this is relatively common when using mechanical extrusion. When

solubilised in 0.5 %, 1 %, 1.5 % and 3 % (w/v) DIBMA, there was a decrease in average diameter of the lipid nanodisc from 55 to 20 nm when following the average diameter of the highest populated sample in the cuvette. I also observed a clear translucency of the liposome sample after a 2-hr solubilisation with 3 % DIBMA from a cloudy white, to a much more translucent white, which is a good readout of successful lipid solubilisation. Discs were also imaged using AFM (see **section 6.4.3**) and the diameters agree with DLS data shown here. Altering polymer concentration is particularly useful for altering the disc size, allowing the adjustment of the number of lipids surrounding a protein.



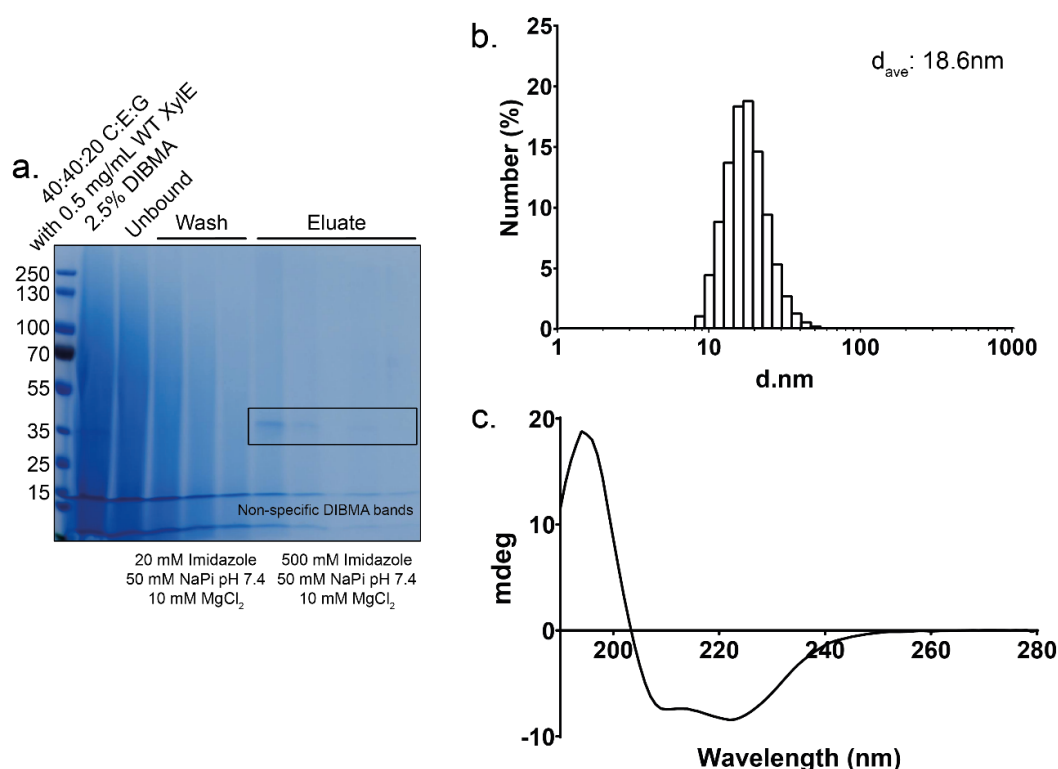
**Figure 4-2: DIBMA solubilisation trials of synthetic liposomes**

a) Dynamic light scattering of 40:40:20 C:E:G lipid samples when solubilised in varying concentrations of DIBMA, 0, 3, 1.5, 1 and 0.5 % (w/v) from top to bottom, yielding diameters of 152.2, 19.8, 48.0, 8.9 and 50.9 nm, respectively. The normalised number of populations detected is plotted against diameter, and these are a mean of 3 biological repeats. b) Photograph of before and after of lipids solubilised with 3 % (w/v) DIBMA after 2 hr. The mean diameters  $\pm$  SD of discs were plotted against % (w/v) DIBMA and a linear trendline was fit.

### 4.2.3. DIBMA captures reconstituted protein with the correct lipid composition

The ability of DIBMA to solubilise a large polytopic protein from a known lipid composition was also checked. Proteoliposomes of 40:40:20 DOPC:DOPE:DOPG and WT XylE (see **chapter 5, section 5.2.5**) were solubilised with 2.5 % (w/v) DIBMA and the insoluble material was removed. The sample was bound to pre-equilibrated Ni-NTA beads for 2 hr, and the flow through was collected. The beads were washed with 20 mM imidazole in standard sodium-phosphate buffer, before eluting the reconstituted beads with buffer containing 500 mM imidazole. **Figure 4-3a** shows a typical Coomassie gel for XylE polymer discs in synthetic lipid mixes, with a characteristic XylE band at the expected molecular weight of 37 kDa. The

smearing in the solubilisation, flow and wash lanes is due to excess DIBMA in the sample. Elution samples were pooled and concentrated before removal of imidazole using a PD10-desalting column. DLS of the sample was run with the average diameter calculated to be 18.6 nm from an average of three biological repeats **Figure 4-3b**. A SRCD scan of Xyle in DIBMA showed the characteristic (see **chapter 5, section 5.2.3**) shape of Xyle in DDM **Figure 4-3c**. It was not possible to deconvolute the spectra for assay of protein secondary structure due, to the lack of concentration information due to light scattering from DIBMA. It was also much more difficult to get a CD scan of DIBMA discs on a benchtop machine as scattering is majorly increased, resulting in low signal:noise.

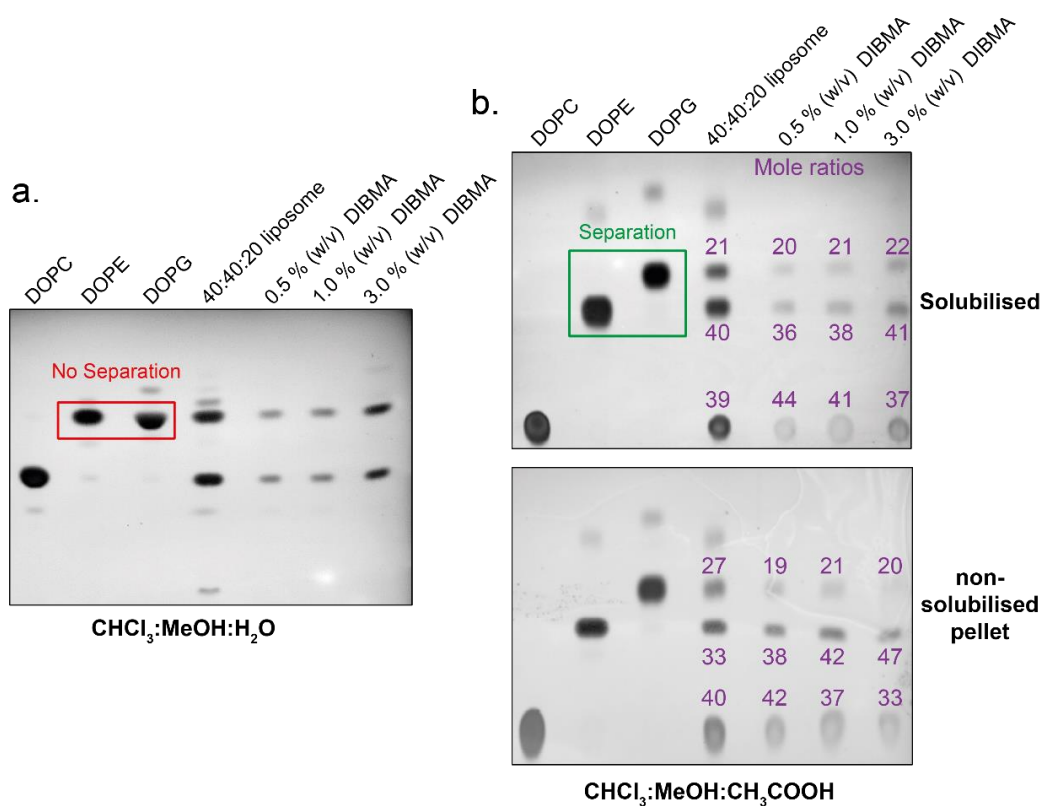


**Figure 4-3: Preparation of Xyle DIBMA nanodiscs in a synthetic lipid mix**

a) Purification of solubilised 40:40:20 PC:PE:PG proteoliposomes with reconstituted Xyle. Solubilised liposomes are passed over super-affinity Ni-NTA beads and then washed with 20 mM imidazole containing buffer and then eluted with 500 mM imidazole elution. The DIBMA polymer band is observed at 12 and 5 kDa, like SMA identified in (Lee et al., 2016). b) The elution was concentrated and run in DLS showing a diameter of 18.6 nm. c) The discs are also amenable to SRCD showing the correct shape for Xyle in DDM. It was not possible to deconvolute the structure due to difficulty of protein concentration acquisition in the polymer disc.

To ensure I have efficiently solubilised the expected lipids surrounding the protein with DIBMA, HPTLC analysis was carried out on the discs. Optimisation of the mobile phase solvent was necessary for good separation between DOPE and DOPG headgroups. The Avanti Polar Lipids website suggested a 52:25:4 (v/v/v) ratio of chloroform:methanol:water for good

separation of these lipids, (RF values: 0.34, 0.79, 0.60 for PC, PE and PG respectively). However, I found that DOPE and DOPG headgroups co-migrated in this system (**Figure 4-4a**). Separation was improved by the replacement of water with glacial acetic acid in the same ratio (**Figure 4-4b top**). Plates were stained with  $\text{CuSO}_4$  and the visualisation of each spot which was subject to ImageJ analysis, where the band intensity was converted to the mole ratio of lipid present (calculated using the lipid molecular weight), which was used to calculate the percent lipid composition. Solubilised Xyle proteoliposomes show the correct lipid composition was captured in the solubilised fraction as well as the non-solubilised pellet fractions (**Figure 4-4b bottom**), albeit a significant amount of lipid which was not solubilised.



**Figure 4-4: DIBMA solubilisation of 40:40:20 synthetic lipids with Xyle**

a) HPTLC plate for synthetic lipid standards and liposomes of 40:40:20 PC:PE:PG against varying concentrations of DIBMA solubilising these liposomes. A solvent system of 65:25:4 chloroform:methanol:water was used, however this led to no separation of the PE and PG headgroups (red box). b) Top: Replacing water with glacial acetic acid resulted in separation of PE and PG headgroups. For each sample, ImageJ analysis was carried out and mole-fractions of each lipid worked out (purple text next to each spot). These closely match the 40:40:20 expected lipid ratios showing DIBMA does in fact solubilise the correct lipids in correct composition for the solubilised fraction. Bottom: The pellet (insolubilized matter) also contains the correct composition of lipid suggesting longer solubilisation times would also solubilise these lipids.

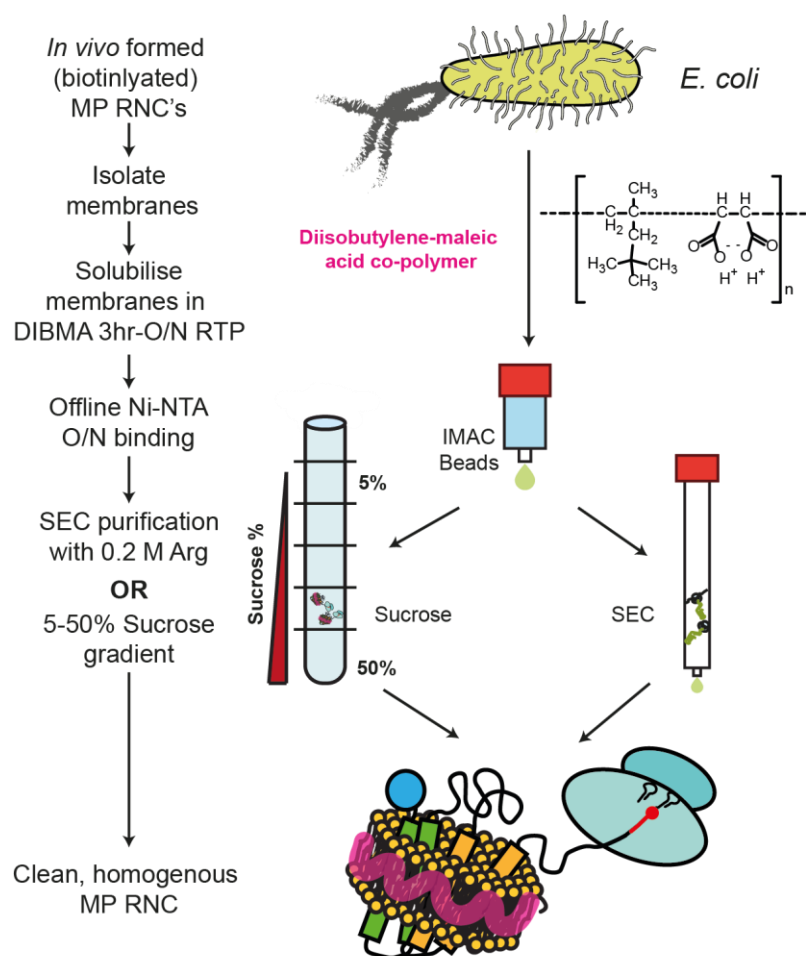
Reconstituted Xyle into 40:40:20 liposomes were indeed solubilised with the correct lipid compositions as determined by HPTLC analysis, and these Xyle nanodiscs were used for AFM

analysis in **Chapter 6**. Safe in the knowledge that DIBMA can isolate the correct 40:40:20 PC:PE:PG composition of synthetic lipids with a larger 12 TM transporter protein, DIBMA was then used to isolate GlpG RNCs with their native membranes directly from cells.

### 4.3. Part B: GlpG RNCs in DIBMALPs

#### 4.3.1. Capturing membrane protein RNCs within DIBMA nanodiscs

To capture RNCs in a native lipid environment, GlpG RNCs of 2, 4 and 6 TM length were isolated with the DIBMA co-polymer using a similar protocol to DDM in **chapter 3**, but with some important modifications (**Figure 4-5**).



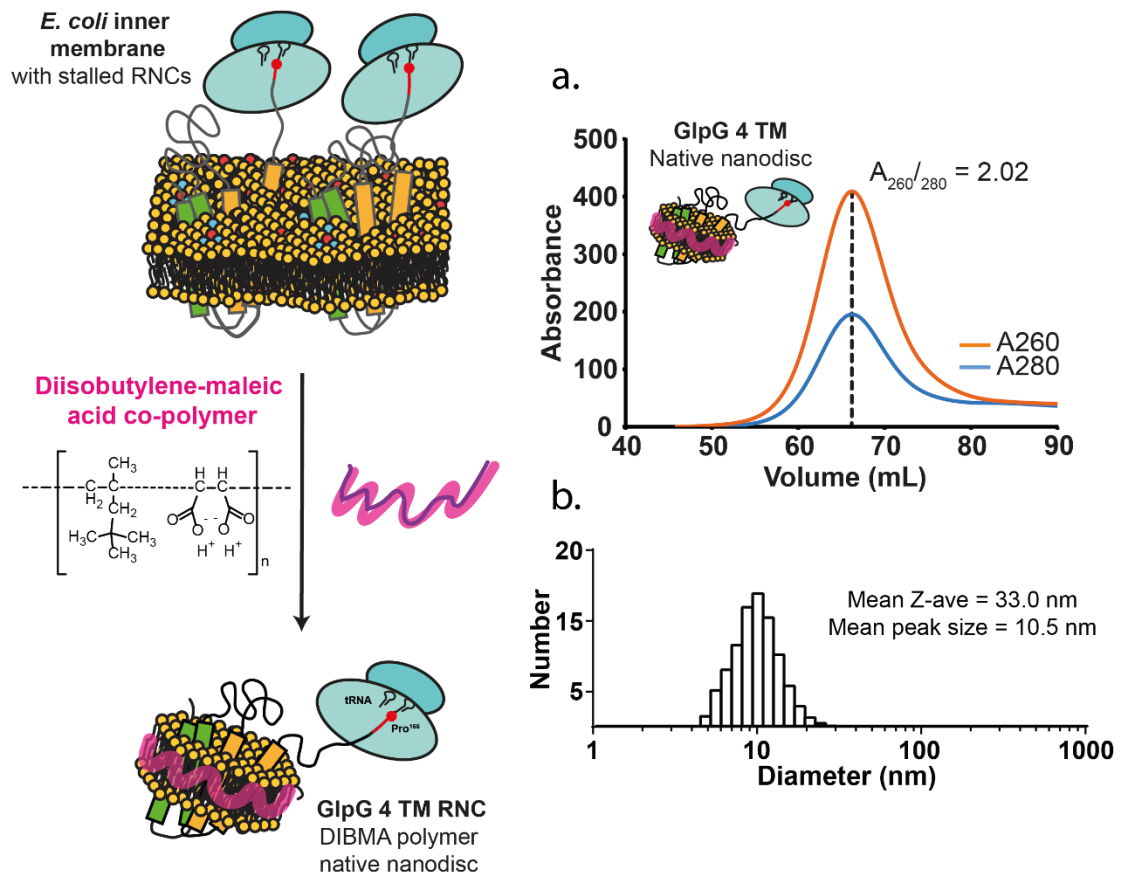
**Figure 4-5: Schematic of DIBMA RNC preparation steps**

Preparation of DIBMA RNCs is largely similar to detergent samples, with a few important differences: DIBMAs lower solubilisation efficiency required a longer, room-temperature incubation of 3 hr to 16 hr for efficient solubilisation. Solubilised membranes were batch bound to Ni-NTA beads overnight to increase incubation time to promote His-tag binding rather than non-specific disc or ribosome binding. The affinity tag elution was purified using the Sepachryl S400 column similarly to DDM, however 0.2 M arginine was included to promote entry of the large hydrophobic disc to the resin. Alternatively, a sucrose gradient can be used to isolate intact RNCs from released nascent chain and dissociated ribosomes.

The harvested and cryo-mill cracked *E. coli* is solubilised with DIBMA and subject to ‘off-line’ Ni-NTA affinity chromatography to harvest discs containing the RNC. DIBMA RNCs were incubated with the Ni-NTA overnight to ensure specific binding, unlike DDM which bound RNC under a 1 mL/min sample flow rate ‘on-line’. Samples were eluted in 500 mM imidazole were then concentrated and injected directly onto a Sephacryl S400 column. L-Arginine was included in the gel-filtration buffer to promote disc entry into the resin. This was required as initially the sample was not able to pass through the column, and no purified sample could be obtained. I determined this using a 2 M NaOH up-flow column wash which removed a significant amount of aggregated RNC from the top of the column. The inclusion of arginine is commonly used as an additive to promote entry of hydrophobic biomolecules into the gel-filtration resin (Ejima et al., 2005) and this proved successful for my GlpG RNCs which are embedded within a large hydrophobic lipid disc and surrounded by polymer. However, it is not clear whether the lipid disc, or unknown interactions between DIBMA and the resin are the cause of this issue. Nevertheless, the arginine-containing size exclusion trace (**Figure 4-6a**) showed a major elution peak at 66 mL with an  $A_{260}/A_{280}$  ratio of 2.02. No void peak was present in DIBMA solubilised samples in arginine containing buffer as was shown with DDM solubilised samples (**chapter 3**). The elution peak was taken and concentrated before buffer exchange into a non-arginine containing buffer using a 100 kDa concentrator or PD10 desalting columns.

Dynamic light scattering (DLS) was used to assess size and homogeneity of solubilised, IMAC-purified 4<sup>TM</sup> GlpG RNCs and gave a mean peak size of 10.5 nm, which is similar to the 12.7 nm previously reported for isolated GlpG in DIBMA (Barniol-Xicotá and Verhelst, 2018). The average z-average particle size (which is the intensity weighted mean hydrodynamic size of the ensemble) was determined to be 33.0 nm suggestive of populations of ribosome bound and unbound discs causing sample polydispersity in the sample (**Figure 4-6b**).





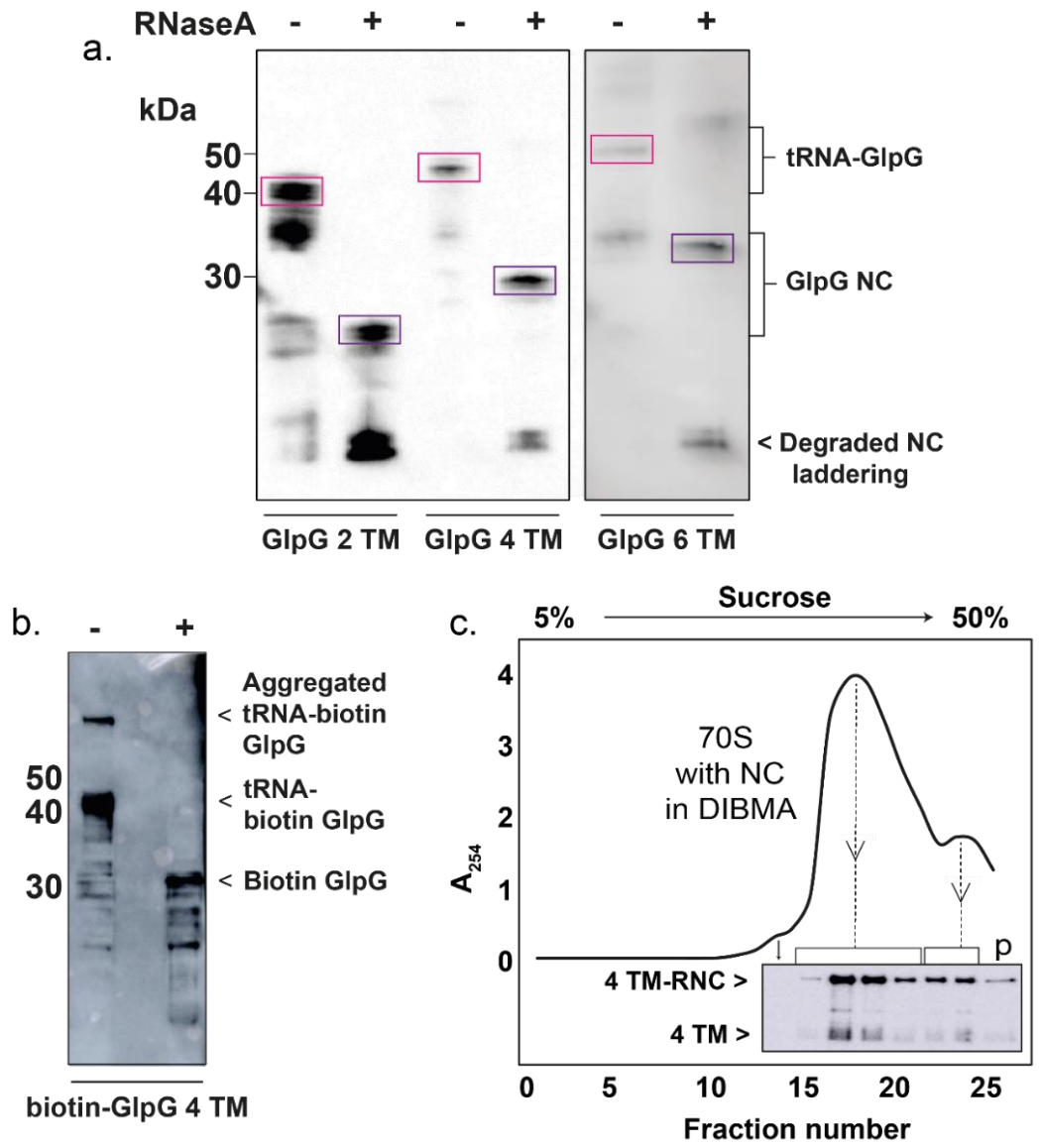
**Figure 4-6: Solubilisation and gel-filtration purification of GlpG RNCs in DIBMA**

Schematic of RNC DIBMA nanodisc preparation. *E. coli* inner membranes packed with membrane protein RNCs are subject to DIBMA solubilisation to produce homogenous RNC nanodiscs containing a native lipid environment. a) SEC purification with 0.2 M arginine for the 4 TM GlpG RNC construct purified in DIBMA nanodiscs. The characteristic void peak has leeched into the major peak, suggesting a reason for a higher proportion of released, degraded, and aggregated nascent chain in the DIBMA Western blots below. The dashed vertical line shows the peak maxima at 66.2 mL where the  $A_{260}/A_{280}$  ratio is 2.02, suggesting homogenous 70S RNCs. b) Dynamic light scattering of IMAC purified 4 TM GlpG RNC.

Low-pH SDS-PAGE and Western blotting analysis showed that 2, 4 and 6 TM RNCs within DIBMA nanodiscs (pink box) were successfully obtained at their correct molecular weights, 40, 48 and 55 kDa for 2 TM, 4 TM and 6 TM respectively (**Figure 4-7a**). However, the Western blot shows an increased presence of nascent chain release (purple box), particularly in 2 TM GlpG RNC, even before treatment with RNase A, and degraded nascent chain post-RNaseA treatment (**Figure 4-7a**). These blots are similar to RNCs prepared in DDM (**Chapter 3**), however here there is more degradation and lower order bands for each RNC length. Like with RNCs prepared in DDM, I can also biotinylate the 4 TM GlpG DIBMA RNC *in vivo*, as shown in **Figure 4-7b**. The samples are probed with streptavidin-HRP to confirm biotinylation and decrease in molecular weight with RNase A treatment to confirm stalling.

In addition to gel-filtration purification. It was also possible to use sucrose gradients to clean IMAC-purified DIBMA RNCs, this is a common method when preparing soluble proteins (Schaffitzel and Ban, 2007, Cassaignau et al., 2016, Rutkowska et al., 2009) to remove the intact RNC from the release nascent chain and separated 30S and 50S ribosomal subunits. However, this strategy can be time consuming, which can have negative effects on sample stability, and can also be difficult to reproduce without the correct equipment to generate gradients. While I have shown that sucrose gradients can be used to purify GlpG RNCs in DDM (**chapter 3**), the more convenient gel filtration protocol can produce samples of the same or better quality as determined by Western blot.

For DIBMA, purification using gel-filtration can be difficult due to the requirement of L-arginine, and therefore further buffer exchange steps are required to remove this additive which can may interfere with nascent chain release due to increased handling of the sample. A continuous 5-50 % (usually 5-40 % for DDM) gradient was used to eradicate released nascent chain artefacts caused by SEC-arginine preparation of RNC DIBMA nanodiscs following IMAC purification. The increase in sucrose concentration was necessary due to the increased density of the lipid discs, which did not separate when 40 % was used and the RNCs all pelleted to the bottom of the tube. The increase to 50 % resulted in DIBMA nanodisc samples residing in the bottom fifth of sucrose (**Figure 4-7c**), as has been shown previously for membrane associated RNCs (Bercovich-Kinori and Bibi, 2015). Together this method provided pure and homogenous samples with better quality than is shown for gel-filtration based purification (**Figure 4-7a**). Ideally a higher concentration gradient would be used to increase resolution of separation, however beyond 50 % sucrose is very viscous and difficult to work with, and removing the sucrose is more difficult. This supports that sucrose gradient purification is also a viable purification strategy for RNC DIBMA nanodisc samples (DIBMA samples being more stable than those in DDM).

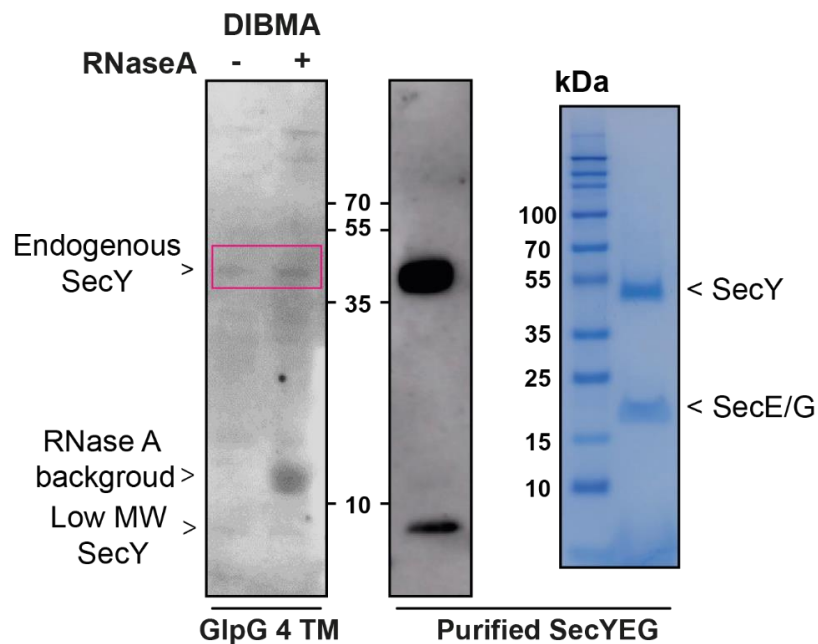


**Figure 4-7: Preparation of DIBMA RNCs using SEC and sucrose gradient**

a) Low pH SDS-PAGE and Western blotting analysis was used to assess the three RNC lengths from DIBMA preparations. Intact DIBMA solubilised RNCs (pre-RNase A) for each stall length (2 TM, 4 TM and 6 TM) are highlighted by pink boxes, released GlpG nascent chains due to RNase A treatment (+) are boxed in purple. Any unlabelled bands are identified as released, truncated, or aggregated GlpG caused by arginine-facilitated migration of the void peak into the SEC elution peak b) biotin-GlpG 4 TM purified using native nanodiscs and SEC and blotted and with Streptavidin-HRP which also shows stalled, biotinylated nascent chains pre (-) and post (+) RNase A treatment. c) GlpG RNCs in DIBMA nanodiscs purified by sucrose gradient. A<sub>254</sub> was plotted for each fraction to determine their position within the 5-50 % gradient. The inset shows that 4 TM GlpG RNCs are present underneath both peaks and in the sucrose gradient pellet. There is also a small degree of release nascent chain (labelled 4 TM) which is subsequently removed upon RNC pelleting and resuspension.

### 4.3.2. Evidence of endogenous SecYEG translocon association with GlpG RNCs

With GlpG RNCs prepared in DDM, I explored whether endogenous translocon machinery was able to be co-purified with the RNC by blotting with a monoclonal anti-SecY antibody (Whitehouse et al., 2012). The native lipid environment captures using DIBMA should also be able to capture SecYEG. The same experiment was carried out on 4 TM GlpG RNCs purified into DIBMA (**Figure 4-8**). When probed with anti-SecY, the Western blot did show presence of endogenous SecY at 37 kDa, with the expected 5 kDa degradation products both pre- and post-RNase A treatment. This was confirmed by comparison with Western blot and Coomassie stain of purified SecYEG, which show the two characteristic SecY bands when blotted with anti-SecY, and SecY with SecE and SecG counterparts when stained with Coomassie. SecE and SecG co-migrate in this Bis-Tris 12 % acrylamide gel system, however this gel type is necessary for RNC work. Preliminary evidence of co-purification of SecY supports that a GlpG RNC-Sec translocase interaction *in vivo* is maintained upon isolation in both DDM (**chapter 3**) and DIBMA.



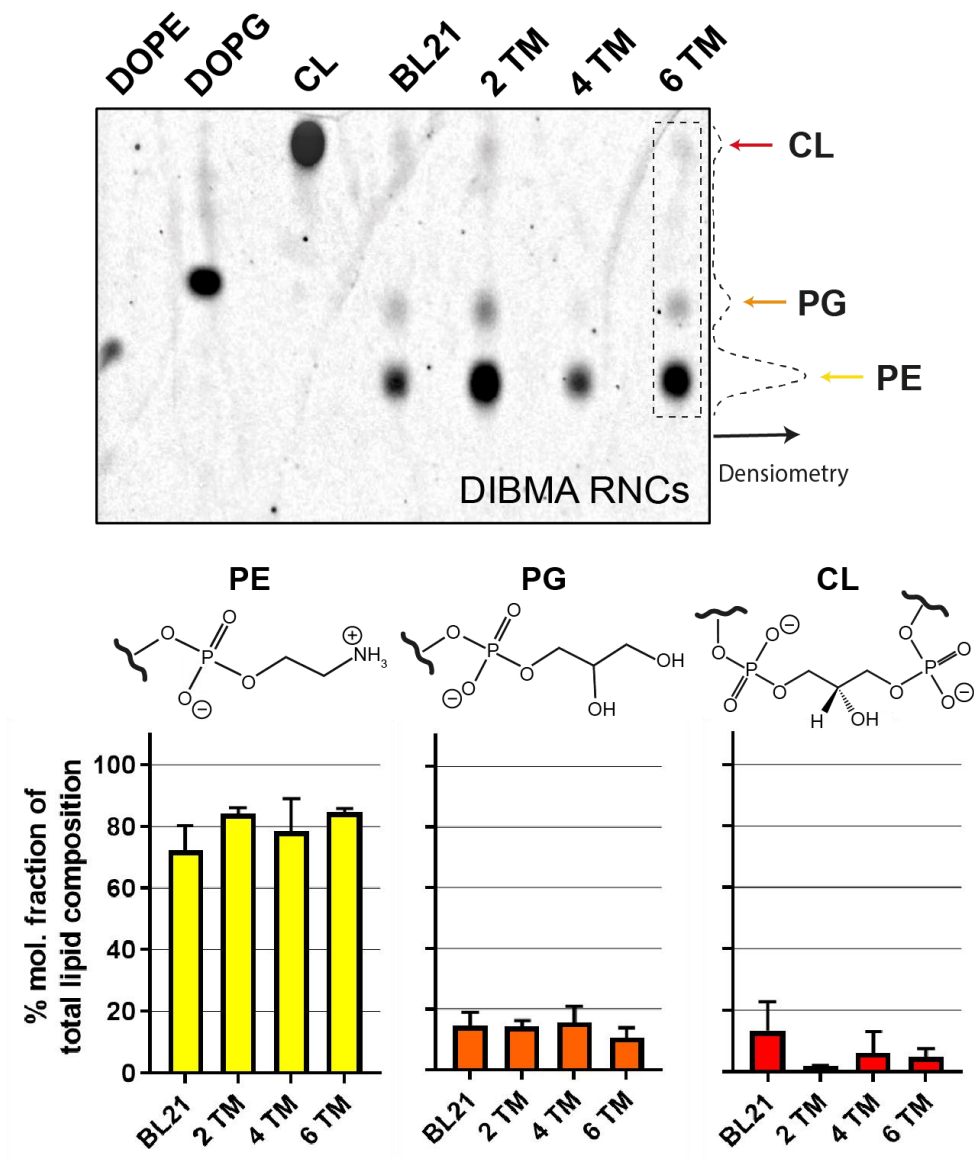
**Figure 4-8: Endogenous SecYEG is captured by DIBMA**

Evidence of endogenous SecYEG was detected within GlpG RNC samples prepared in DIBMA nanodiscs, as well as DDM shown in **Chapter 3**. Endogenous SecY was found to be present in pre- (-) and post- (+) RNase A treated samples (highlighted with a pink box, left panel). The low molecular weight SecY band is characteristic of a C-terminal cleavage product (Whitehouse et al., 2012). Its 37 kDa band was confirmed by comparison with a Western blot (middle panel) and SDS-PAGE Coomassie stain (right panel) of purified SecYEG - SecE and SecG co-migrating in this SDS-PAGE environment. All Western blots were probed using a monoclonal antibody for SecY.

### 4.3.3. DIBMA captures membrane protein RNCs in a native lipid environment

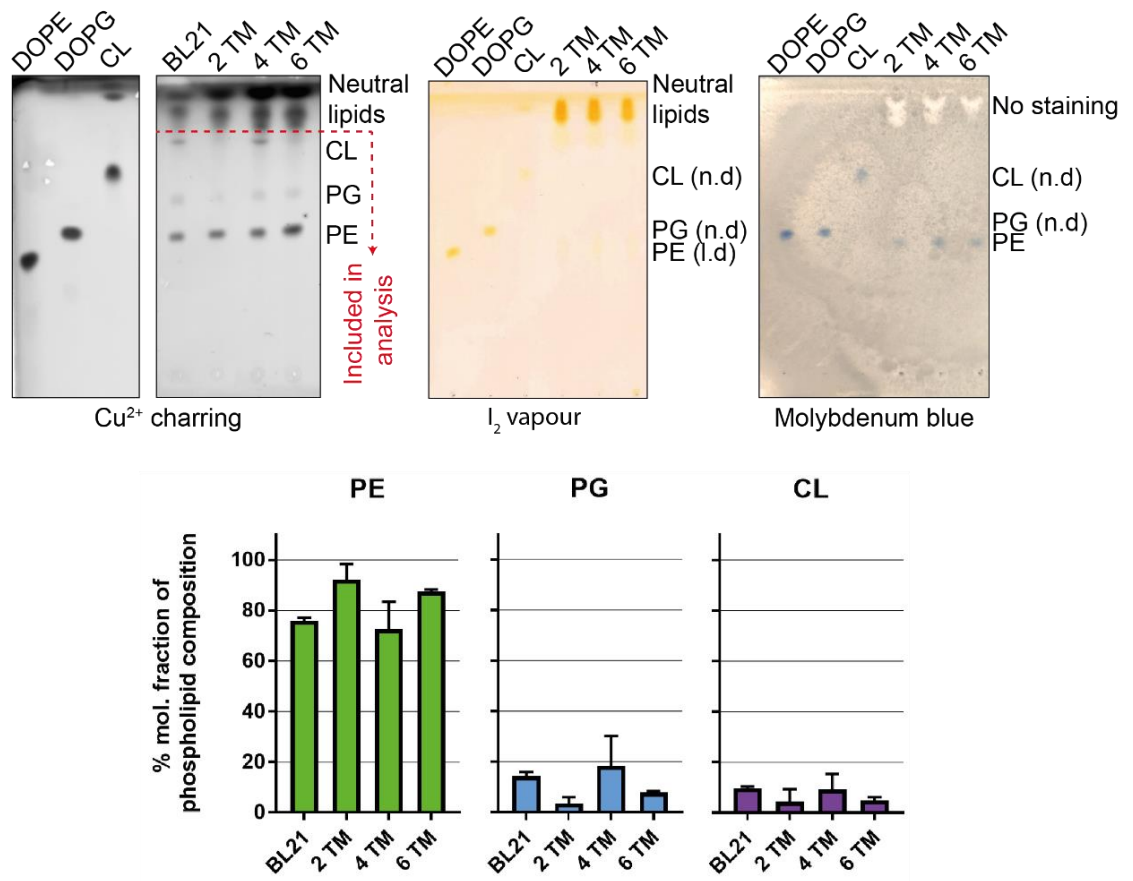
To ensure that the DIBMA polymer had solubilised GlpG RNCs within a native lipid environment I assessed the lipid compositions using high performance thin layer chromatography (HPTLC) with the same protocol as described in **section 4.2.3**. Purified DIBMA RNC samples and cell membranes from non-induced cells grown at the same temperature were subject to Folch lipid extraction. Lipid identity and relative mole ratios of each observed lipid were then assessed using HPTLC analysis (**Figure 4-9**). The three major phospholipid types expected within the *E. coli* inner membrane are: PE, PG and CL. I found that each length of DIBMA GlpG RNC nanodiscs contained similar relative abundances of PE, PG and CL to each other, and to the non-induced membrane extract when grown in MDG/EM9 media (**Figure 4-9**). A slight increase in PE and decrease in CL abundance was observed between BL21 and the 2 TM GlpG RNC construct.

It was also possible to produce RNCs grown in LB media. When HPTLC analysis was run on these samples, a high proportion of total lipid content was determined to be neutral lipids in addition to the expected phospholipid composition (**Figure 4-10**). To confirm this, a selection of different TLC plate stains were required. Copper charring binds to all lipid tails. Iodine vapour, which also stains all lipid species was used to confirm copper charring specificity to all lipid species, including the unknown lipids spots near the solvent front. To verify that these additional spots did not correspond to phospholipid, a molybdenum blue stain (Dyńska-Kukulska et al., 2013) which specifically stains phospholipid headgroups, was unable to stain these spots thus confirming that they were not phospholipid, and therefore were not included in the mole lipid composition analysis. These higher order spots have previously been observed in DIBMA nanodiscs of isolated GlpG (Barniol-Xicota and Verhelst, 2018).



**Figure 4-9: GlpG RNCs prepared in DIBMA nanodiscs contain native lipid compositions**

DIBMA GlpG RNC samples and cell membranes, prepared from BL21 *E. coli* cells grown in minimal media, were subject to Folch lipid extraction. Extractions were run alongside standards of DOPE, DOPG and CL. A representative TLC plate from samples grown in MDG/EM9 media is shown. The three bar charts show a similar relative abundance of each lipid for each stall length (2 TM label denotes 2 TM GlpG RNC extraction) as determined using densitometry analysis. Mean % lipid composition and SD are plotted for three or more biological repeats on each stall length (see methods). Phospholipid headgroups for phosphatidylethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (CL) are shown.



**Figure 4-10: HPTLC plate staining and quantification of DIBMA nanodisc RNCs prepared from membranes extracted from cells grown in Luria-Bertani media**

Copper charring, iodine vapour and molybdenum blue stains were used to determine DIBMA extracted lipids from *E. coli* BL21 cells grown in LB media. Copper charring and Iodine vapour detect C=C bonds in phospholipids, sterols, and neutral lipids, where the modified Dittmer-Lester Molybdenum stain specifically stains phospholipids (Dyńska-Kukulska et al., 2013). RNC sample extracts grown in LB media contain a significant portion of extra lipid bands near the solvent front on the HPTLC plates. These bands were not stained by molybdenum blue but stained by  $\text{Cu}^{2+}$  charring and therefore were determined to be non-phospholipids or neutral lipids (n.d = no detection, l.d = low detection). Densitometry analysis of the phospholipid bands, alongside standards of synthetic PE, PG and CL which were run similarly to samples generated in minimal media, and bar-charts with SD plotted for three or more biological repeats. These results show similar relative abundances of the phospholipid species when compared to minimal media grown samples and confirms DIBMA still extracts a similar lipid composition to native *E. coli* membranes.

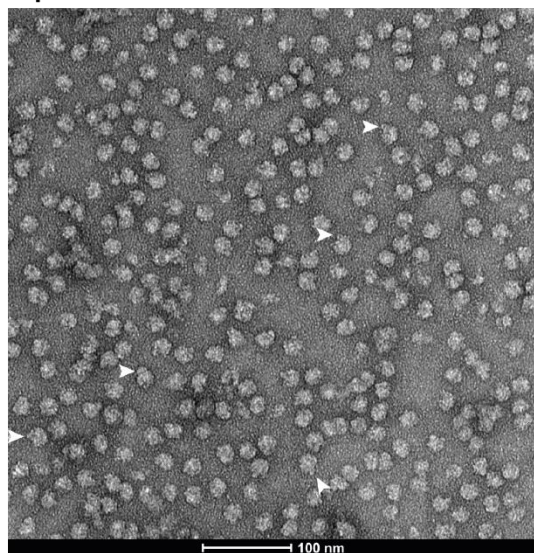


#### 4.3.4. Negative stain EM of RNCs produced in DIBMA

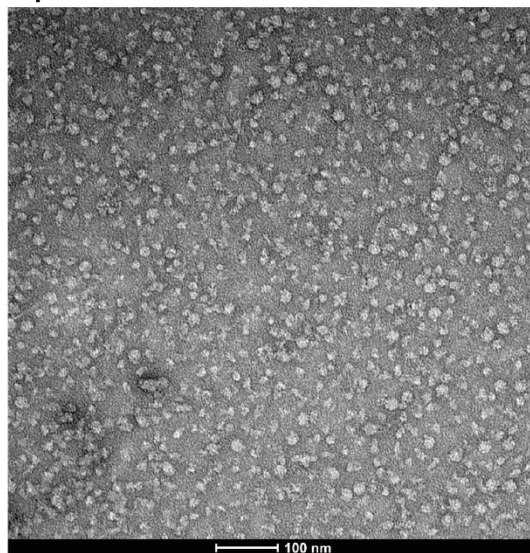
DIBMA RNC samples appear to be more stable than those prepared in DDM when imaged by negative stain EM. Grids were prepared as before with 150 ng of sample, stained with 3 % UrAc. The 2<sup>TM</sup> GlpG RNC in DIBMA was pure and clean, with few disassembled complexes. In contrast, the 4<sup>TM</sup> GlpG RNC sample was very dirty with many disassembled ribosome particles and no distinguishable ribosomes, or disc complexes (**Figure 4-11**). This sample needs further purification, and investigations into why the sample may be disassembling post purification.

6<sup>TM</sup> GlpG RNCs were clean and pure (**Figure 4-12**). There were few disassembled complexes compared to the 4<sup>TM</sup> GlpG sample, and there were clear discs in complex with the ribosome as shown by green triangles. A small selection of 1400 particles were 2D classified, with around 15 % showing occupied ribosomes (green boxes) but unfortunately, 85 % of these ribosomes were empty (red boxes). The green arrows identify clear density attached to the ribosome and the identity of this may be the DIBMA disc with nascent chain, or SecYEG which was co-purified with the RNC. For a more detailed analysis of these systems, both here and in DDM, ribosomal occupancy must increase, and nascent chains must be prevented from detaching from the ribosome.

**GlpG 2<sup>TM</sup> RNC DIBMA**



**GlpG 4<sup>TM</sup> RNC DIBMA**

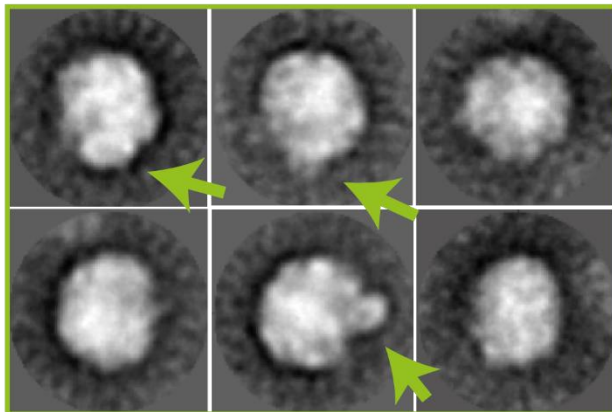
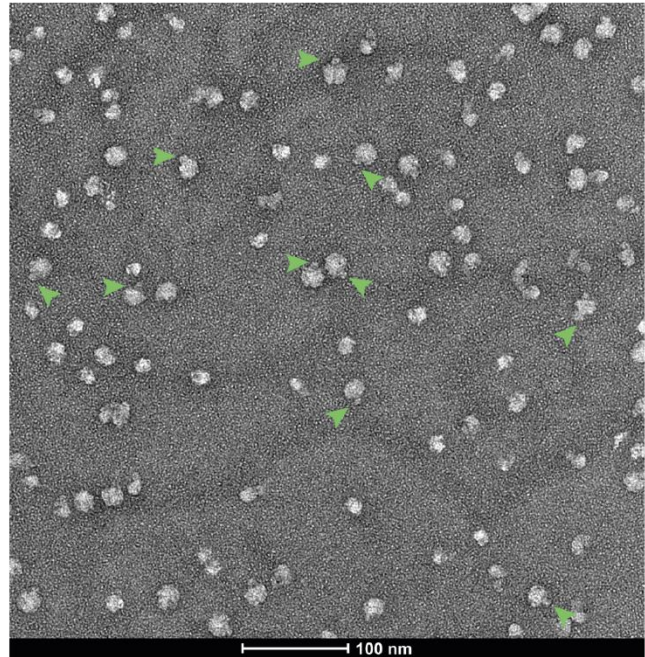
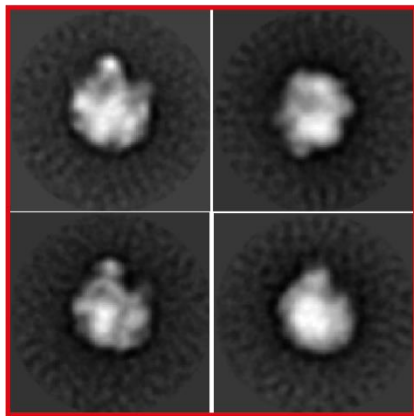


**Figure 4-11: Negative stains of 2<sup>TM</sup> and 4<sup>TM</sup> GlpG in DIBMA native nanodiscs**

*Negative stains of 2<sup>TM</sup> GlpG and 4<sup>TM</sup> GlpG are deposited onto grids. 2<sup>TM</sup> show a near homogenous selection of discs, whereas 4<sup>TM</sup> show many released, dissociated ribosomes, and require further purification.*



**GlpG 6 TM RNC  
DIBMA native nanodisc**



**1400 particles**

**85% empty ribosomes**  
**15% occupied (SecYEG?)**

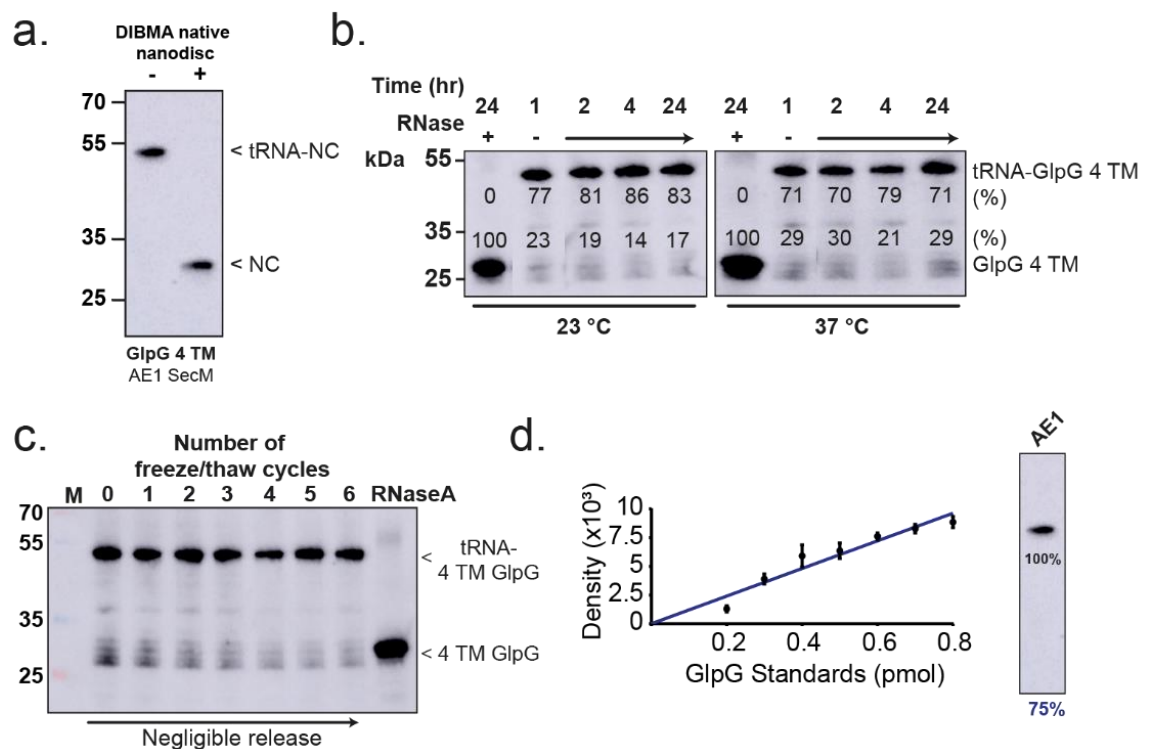
*Figure 4-12: Negative stain of 6 TM GlpG RNC in DIBMA native nanodiscs*

*The 6 TM GlpG RNC, when deposited on grids, green arrows show ribosomes with something attached. Of A 2D classification of 1400 particles, 85 % showed empty ribosomes (red) and 15 % showed occupied ribosomes (green) with the disc or SecYEG attached.*

**4.3.5. Arrest-enhanced SecM improves DIBMA purified RNC quality**

Purifying the GlpG RNCs with enhanced SecM drastically improved the quality of the samples. The 4 TM GlpG enhanced SecM RNC prepared in DIBMA is incredibly clean with no released nascent chain (**Figure 4-13a**), and were incredibly stable, with no significant release of NC found after 24 hr at both 23 °C and 37 °C (**Figure 4-13b**), compared to the 12 % decrease in intact RNC for enhanced 4 TM GlpG RNC in DDM at 23 °C (**Chapter 3**). Furthermore, these samples were able to withstand at least 6 freeze-thaw cycles with no significant release of the nascent chain (**Figure 4-13c**). The occupancy of the RNC was calculated in the same way as

DDM samples, GlpG standards were blotted and ImageJ densitometry was used to construct a calibration curve. The occupancy for the DIBMA purified 4<sup>TM</sup> GlpG RNC with enhanced SecM was determined to be 75 %, however there is no significant release (100 % intact) of this DIBMA nascent chain compared to DDM detergent samples (86 % intact with 85 % occupancy (chapter 3, Figure 3-14). Taken together, this signifies that incorporation of the enhanced sequence leads to a significant increase in RNC stability which will aid future structural, biochemical, and biophysical analysis of not only DDM samples, but also samples prepared in their native lipid environment using DIBMA.



**Figure 4-13: GlpG RNCs are stable in DIBMA nanodiscs and tolerant to freeze thaw cycles**

a) Western blot for enhanced SecM 4<sup>TM</sup> GlpG RNC is consistent with WT and enhanced samples in DDM and DIBMA respectively, and the degree of release nascent chain is significantly reduced. b) Enhanced SecM GlpG 4<sup>TM</sup> RNCs purified in DIBMA nanodiscs were heated at 23 °C and 37 °C for 24 hours, samples were taken and blotted after 1, 2, 4 and 24 hr before low pH SDS-PAGE and Western blotting analysis. The ratio between intact RNC (~50 kDa band, tRNA-GlpG 4<sup>TM</sup>) and spontaneously released nascent chain (~30 kDa band, GlpG 4<sup>TM</sup>) was calculated for each lane using band density and ImageJ. Percentages of intact and released are quoted next to its corresponding band. There was no significant release for each temperature from 1-24 hr. c) Enhanced SecM GlpG 4<sup>TM</sup> RNCs purified in DIBMA are tolerant of freeze-thaw cycles. 6 freeze-thaw cycles generated by flash freezing and thawing at room temperature show no significant decrease in release nascent chain. d) For enhanced SecM prepared in DIBMA, occupancy was calculated to be 75 % (where 0.65 pmol of sample was loaded). Estimated occupancy is presented as blue text. The GlpG standard raw blots can be found in Chapter 3, Figure 3-13. The ratio of intact to release 4<sup>TM</sup> GlpG RNC was estimated using ImageJ for each lane, for enhanced 4<sup>TM</sup> GlpG prepared in DIBMA, there was no release of nascent chain. All Western blots were probed using a polyhistidine-HRP antibody.

## **4.4. Discussion**

### **4.4.1. Preparation and functionality of DIBMA**

The method of preparation of DIBMA is very important for the correct function of the polymer, with small variations in protocol drastically affecting its solubilisation efficiency. The precipitation and washing of the raw polymer away from the unknown 'alkaline solution' is the safest way to achieve a reproducible method, and FTIR quality control checks of stocks were comparable to previous studies, despite the altered method of production (Gulamhussein et al., 2020). Many groups use dialysis methods which can change the volume and salt conditions of the polymer and can be difficult to reproduce. The production of a lyophilised powder allows concentrated stocks of the correct pH and buffer constituents to be produced, which can be used directly, or freeze-dried in the correct salt conditions for long-term storage. This avoids introducing variations in dialysis protocols and correct buffer components are crucial in RNC work to maintain integrity of the ribosome, therefore I believe this to be the most reproducible way of achieving a functional DIBMA stock.

DIBMA was used to isolate a larger 12 TM transporter, which was reconstituted into a known lipid composition of 40:40:20 DOPE:DOPC:DOPG. The large protein was able to be solubilised and purified using Ni-NTA affinity chromatography and I was able to acquire a SRCD spectra of the Xyle transporter. When these samples were subject to HPTLC, the correct 40:40:20 ratios were observed from mole-ratio calculations from the TLC plate. Together, this suggests that DIBMA can create discs, of a minimum of 18 nm with a large helical protein embedded in a correct lipid environment, and was amenable to spectroscopic study which has largely been ruled out for SMA based polymers. Overall, this data and past literature (Barniol-Xicotá and Verhelst, 2018) suggests that DIBMA may be capable of solubilising *E. coli* native membranes, with a GlpG based RNC complex embedded in the correct surrounding lipid environment.

### **4.4.2. Purification of RNCs into DIBMA native nanodiscs**

Building on from protocols developed to produce DDM purified nascent chains, I have successfully purified RNCs in lipid discs matching their native lipid environment. The correct lipid properties and environment are crucial for the correct folding and function of the membrane protein and so native lipids will provide the most realistic snapshot of co-translational folding *in vitro*. The DIBMA co-polymer was used for the solubilisation of the inner *E. coli* membranes with attached RNC. DIBMA, unfortunately has a lower solubilisation efficiency when compared to DDM detergent and SMA co-polymer. However, higher divalent

cation concentrations (presence of 10 mM Mg<sup>2+</sup> or 7.5 mM Ca<sup>2+</sup>) have been shown to improve DIBMA-mediated solubilisation of *E. coli* membranes by up to two-fold, through association and neutralisation of the polymer's carboxylate groups (Danielczak et al., 2019). To this end, the magnesium acetate concentration was increased in my membrane solubilisation buffers from 6 to 12 mM. DIBMA is also much more tolerant to divalent cations than SMA (Oluwole et al., 2017b, Oluwole et al., 2017a, Gulamhussein et al., 2020). When using SMA, higher magnesium concentrations, which are imperative to the correct function and stability of the 70S ribosome, cause the polymer to precipitate out of solution.

Solubilized membranes containing His-tagged GlpG RNCs were then batch bound to super-affinity Ni-NTA beads overnight at 4 °C to allow nickel-histidine coordination. This step was necessary as I found that DIBMA could interfere with nickel-histidine coordination (Schaffitzel and Ban, 2007, Rutkowska et al., 2009) reducing His-tag binding to the Ni<sup>2+</sup>-resin, resulting in very low yields of remaining RNC. In contrast, DDM detergent preparations did not require long incubation times for adequate binding to occur. Alternative affinity tags could be used for purification. For example, strep tags are commonly used for both membrane and soluble protein RNC production (Kater et al., 2019, Bischoff et al., 2014, Schaffitzel and Ban, 2007, Rutkowska et al., 2009), which are not thought to compete with ribosome binding, however the ease of using the poly-histidine tag and the subsequent gel-filtration stage which will remove the non-specifically bound ribosomes from the sample, as well as the secondary biotinylation tag which is also shown to work with the DIBMA purified samples suggests this is unnecessary in work shown here.

The major drawback of the DIBMA polymer compared to SMA based polymers is its low solubilisation efficiency. Although I can mitigate this slightly by increasing the magnesium concentration in my buffer (Danielczak et al., 2019) and increasing the solubilisation time at room temperature – particularly when using arrest-enhanced SecM to avoid premature nascent chain release, I still achieve only a third yield of each of the RNC lengths compared to purification using DDM detergent (**chapter 3**). New alternative polymers are regularly synthesised by the SMALP community. A polymer which has a higher solubilisation efficiency, and increased tolerance to divalent cations (magnesium) would be the optimum choice for this project, such as SMA-QA or SMI, highlighted in (Overduin and Esmaili, 2019), unfortunately, these are produced elsewhere and are not available commercially.

Affinity purified DIBMA solubilised GlpG RNCs were subject to SEC for further purification from protein-free DIBMA discs, released nascent chains and aggregated RNCs. However,

during SEC the DIBMA polymer was found to interact with the Sephacryl S400 resin which prevented its elution from the column – leading to no product eluting from the column. To counter this, 0.2 M arginine was added to the SEC buffer mobile phase to mitigate protein disc interactions with the resin and eliminate possible aggregation (Ejima et al., 2005). With arginine present, the DIBMA nanodisc samples could now be resolved by SEC but with a different elution behaviour than found within DDM detergent preparations: the void peak disappeared, and the major peak shifted to a larger elution volume of 66 mL instead of 60 mL for 4 TM GlpG RNC (**Figure 4-6**). 70S ribosomal controls with and without the presence of arginine in the SEC buffer showed that this shift was caused by arginine, whilst the  $A_{260}/A_{280}$  ratios were unaffected.

The quality of the DIBMA purification, like DDM, was analysed using Western blot. The gel-filtration purified samples did produce the expected length nascent chains, however there was a significant degree of degradation of released nascent chains associated with each stall length. I suggest these bands are released nascent chain, previously found as aggregates in the void peak in DDM preparations, which migrate into the major elution peak during SEC caused by the inclusion of arginine into the mobile phase. When purified using sucrose gradient from 50-5 % sucrose, the sample quality was much cleaner, confirming that the arginine causes the aggregate leeching issue, and sucrose is perhaps a better method to produce RNC in DIBMA nanodiscs. A higher sucrose concentration gradient (50 % compared to 40 % for DDM) was required for the lipid nanodiscs due to their higher density than DDM micelles, this produced a better separation and less sample was found in the ‘pellet’ at the bottom of the gradient, although this has been shown before when using sucrose to prepare membrane bound RNCs (Bercovich-Kinori and Bibi, 2015).

The use of the arrest enhanced SecM (Cymer et al., 2015a, Kempf et al., 2017) sequence dramatically improved RNC stability and longevity for both DDM and DIBMA purified RNCs. When prepared in DDM detergent, samples eluted with a similar elution profile to WT SecM, however the void peak containing aggregated released nascent chain was much larger. This results from a more significant translational stalling which cannot be released on co-translational folding, and required the possibility of additional quality control factors *in vivo* which actively remove these stalled ribosomal complexes as they arise, removing the formation of aggregate likely through an *in vivo* degradation pathway.

#### 4.4.3. Capture of native lipids and endogenous insertion machinery

Across each nascent chain length, DIBMA appears to be capable of solubilising the correct lipid composition from the native lipid environment. There was however a small increase in PE and decrease in CL abundance was seen between BL21 and the GlpG RNC DIBMA preparations (especially for the 2 TM GlpG construct). Lipid membrane solubilisation by DIBMA is relatively unexplored, but these effects may arise from collision induced lipid transfer (Danielczak and Keller, 2018), preferential solubilisation by DIBMA, or reflect the 2 TM local lipid environment as previously shown with SMA (Teo et al., 2019). When RNCs were grown in LB media, the presence of neutral lipids was also found in addition to the expected phospholipid composition as confirmed by iodine vapour and molybdenum blue staining (Dyńska-Kukulska et al., 2013). This has also previously been observed in DIBMA nanodiscs of isolated GlpG (Barniol-Xicota and Verhelst, 2018). Taken together, I confirm that DIBMA is capable of extracting MP RNCs within a composition akin to its native lipid bilayer.

GlpG is predicted to use SecYEG translocon to aid insertion machinery *in vivo* (Schibich et al., 2016). I probed the DDM and DIBMA solubilised samples with a monoclonal antibody for SecY and for both systems, there was presence of endogenous translocon machinery when compared with blotting of an overexpressed purified SecYEG, hence the much lower blot intensity for the protein. Overexpression of the translocon with the RNC constructs at different lengths may not only be a useful field for study of structural interactions during the co-translation cycle, but also improve yield and reduce protein aggregates if inserted into the membrane correctly by the translation machinery. It has also been suggested that when SecYEG is isolated using SMA, the SecA ATPase motor protein which is involved in protein translocation, is also co-purified (Prabudiansyah et al., 2015). It would be interesting to determine whether GlpG requires SecA with SecYEG activity to transverse the inner membrane – however, currently it is thought that SecA is involved in across the membrane secretion only. It would also be interesting to assay the lipid compositions of the RNC samples with overexpressed SecYEG. The translocon requires certain lipids to function and maintain stability of the dimeric SecYEG and SecA oligomerisation. These factors are significantly impaired when studied using a cardiolipin deficient *E. coli* strain (Ryabichko et al., 2020) and perhaps cardiolipin, at the 2 TM helix stage of field, which appears to have reduced CL when assayed for lipid composition, suggests that SecYEG is not in-play at this stage, and the RNC is separated from SecYEG:CL in the purification process.

With the possibility of producing very clean DIBMA RNC samples, particularly for the 6 TM GlpG RNC complex, the use of cryoEM is a very real and imminent possibility to study the folding pathways in a native lipid environment. This coupled with the *in vivo* biotinylation protocols presented in **chapter 3**, as well as the possibility of overexpression of SecYEG, enables the structure of the polytopic nascent chain to soon be determined in a native lipid environment highlighting any interactions with the SecYEG apparatus at advancing stages of protein biogenesis.

#### 4.5. Chapter summary

I have demonstrated that pure, stable RNCs of different nascent chain length can be prepared for a polytopic  $\alpha$ -helical integral membrane protein, in both detergent micelle (**chapter 3**) and native lipid surroundings. This work advances membrane protein RNC preparations by establishing that:

- 1) SecM stalling can successfully stall a polytopic  $\alpha$ -helical MP at various points throughout its synthesis to yield stable RNCs for structural interrogation; thus, allowing folding of the full chain to be probed as opposed to single, short TM constructs previously investigated.
- 2) RNC constructs are tolerant to different purification tags, such as a His- and Avi-tags, enabling diverse purification and experimental strategies to be devised.
- 3) The nascent chain is maintained within a lipid bilayer comprised of its cellular lipid mixtures rather than synthetic lipids.
- 4) RNCs can be extracted directly from native membranes, escaping reconstitution of isolated RNCs prepared in non-native environments, and permitting co-purification of endogenous chaperones, such as the SecYEG translocon, which are all amenable to structural investigations - particularly Cryo-electron microscopy.

Overall, I reveal that homogenous membrane protein RNCs can be captured within a native lipid environment using DIBMA native nanodisc technology. These samples are stable for timeframes and conditions required for future biochemical and structural studies (Parmar et al., 2018). *In vivo* formed membrane proteins RNCs generated here provide snapshots of co-translational folding, paused at various nascent chain lengths and are a major stepping-stone towards studying the structure and dynamics of co-translational membrane protein folding in an increasingly more *in vivo* context.

## 5. Expression, preparation, and characterisation of XyleE

---

### **XyleE half-domain preparations contributed to:**

Cell-free expression tools to study co-translational folding of alpha helical membrane transporters

Harris NJ, **Pellowe GA**, Booth PJ, *Scientific Reports*, 2020

doi: 10.1038/s41598-020-66097-4

### **XyleE protein preparation and biophysical characterisation contributed to:**

Hydrogen-deuterium exchange mass spectrometry captures distinct dynamics upon substrate and inhibitor binding to a transporter

Jia R, Martens C, Shekar M, Pant S, **Pellowe GA**, Lau AM, Findlay HE, Harris NJ, Tajkorshid E, Booth PJ, Politis A, *Nature Communications*, 2020

doi: 10.1038/s41467-020-20032-3



## 5.1. Introduction

This chapter introduces the D-xylose:H<sup>+</sup> *E. coli* symporter XylE, which follows an interesting alternate access mechanism for solute transport with many functional mutants. Here XylE is prepared and assayed for purity, correct secondary structure and binding capacity using standard biochemical/physical techniques, both WT and a mutant contained a C-terminal extension for use with sensitive single-molecule AFM folding experiments carried out in **Chapter 6**, and future studies using magnetic tweezer (MT) spectroscopy. I also explore the possibility of developing a novel enzyme-linked assay as a fluorescence readout for XylE transport activity, as an alternative to expensive radiation experiments detailed in the literature which not all research groups may have access to.

### 5.1.1. D-xylose/H<sup>+</sup> symporter – XylE

#### 5.1.1.1. Structure and stability

XylE is a D-xylose:H<sup>+</sup> symporter of the Major Facilitator Superfamily (MFS) which resides in the *E. coli* inner membrane and acts to transport sugar into the cell with the aid of a proton-motive force for uphill transport (Quistgaard et al., 2013, Madej et al., 2014) **Figure 5-1**. MFS transport is necessitated through a cycling of different conformational states between two distinct domains, each of 6  $\alpha$ -helical transmembrane bundles with 3+3 symmetry. The domains are connected by a cytoplasmic loop, with a central channel for the binding and release of substrate (Harris et al., 2017a, Martens et al., 2018).

This structure is characteristic to MFS proteins (Hediger et al., 2013, Yan, 2015) and allows for the ‘rocking’ activity within the membrane resulting in a dynamic opening and closing on either end for substrate entry, binding, and release on the opposite side of the membrane. However, the order that the protein cycles through these conformations as sugar and proton associate and disassociate very difficult to decipher, due to the difficulty in ascertaining the allosteric coupling between the proton and sugar binding states (LeVine et al., 2016).

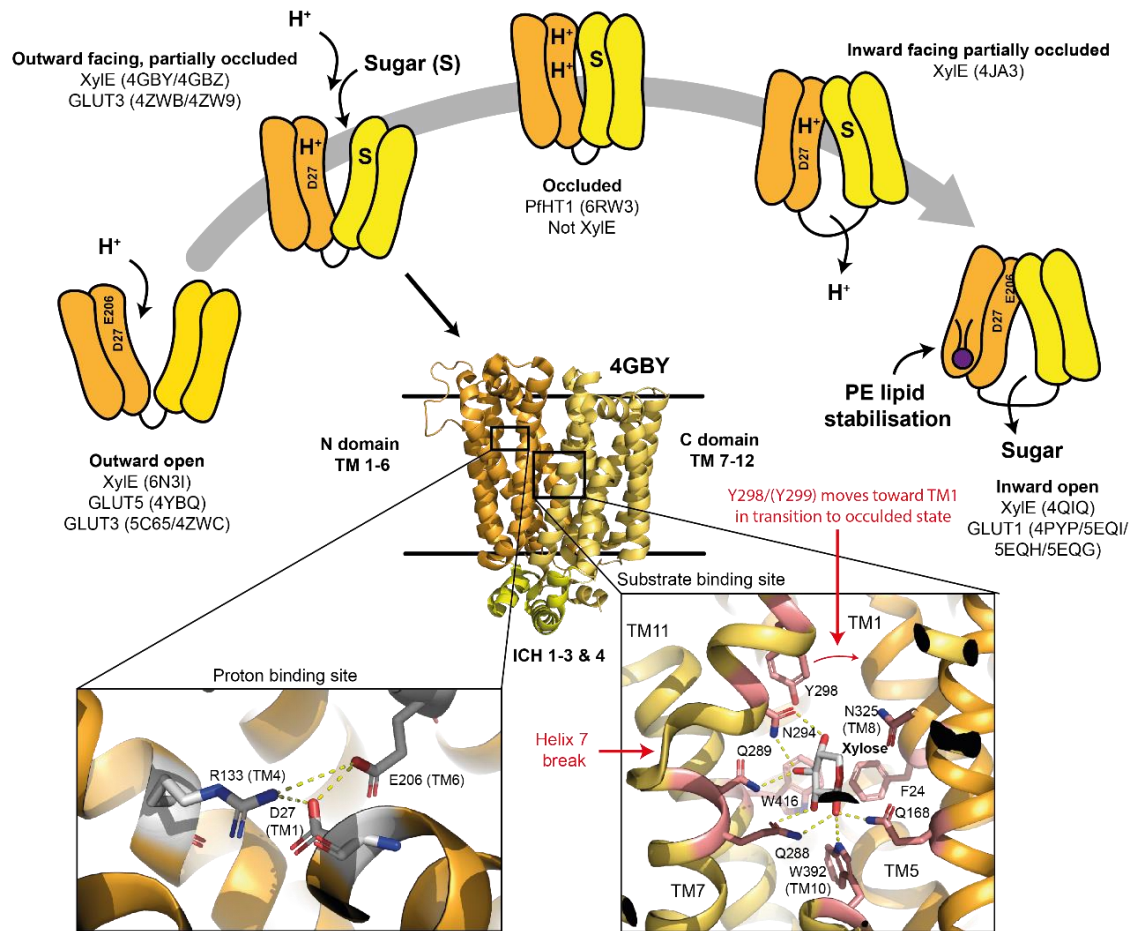
A handful of MFS proteins may have extra domains to provide stability for function (Madej et al., 2014). For example, XylE has an additional two intra-cytoplasmic helices (ICH), and two extra-cytoplasmic helices (ECH), which contribute to protein stability and transport cycle, particularly at the ICH domains which form a ‘latch’ mechanism (Deng et al., 2014). These ICH domains appear to contribute heavily to the stability of the protein (Harris et al., 2017a) as determined using SRCD with urea and GuHCl denaturation and following the loss of secondary structure. This study showed that XylE was more resistant to urea than other MFS

transporters LacY and GlpT, but with a similar thermal stability to PepT<sub>so</sub>. XylE did however unfold in the presence of GuHCl irreversibly, and the addition of xylose to the denaturation buffers improved the stability of the protein.

Some groups have sought to determine whether the two domains of MFS transporters have a degree of independent stability such that the domains can express and fold separately. So far, LacY has been expressed *in vivo* as two separate domains (Sahin-Toth et al., 1996, Bibi and Kaback, 1990, Wu et al., 1996), and were found to be susceptible to protease digestion unless co-expressed together (Bibi and Kaback, 1990). The C-domain was less stable when expressed alone (Sahin-Toth et al., 1996). XylE to my knowledge has not been separated and expressed in cells and subsequently purified outside the Booth group. The two halves of LacY and XylE could be produced and identified using an *in vitro* transcription/translation system which does showed stable half domains without the requirement of the translocon. However, when purified from cells this was not the case with large presence of aggregate and degradation, with most protein eluting in the void of chromatography columns, with very poor expression and low yields (Harris et al., 2020) and therefore the half domains could not be probed using AFM SMFS.

#### **5.1.1.2. Binding and transport**

Isolated XylE has been heavily structurally studied due to its close structural homology with the Human Glucose uniporter (GLUT1) which has disease related implications when mutated, leading to glucose storage disorders and diabetes (Sun et al., 2012). The XylE crystal structure has been solved in many conformations, all of which are crucial for understanding the MFS transport cycle. These conformations: inward open (4JA4) and inward partially occluded (4JA3 (Quistgaard et al., 2013)), outward partially occluded with substrate D-xylose (4GBY (Madej et al., 2014)) and inhibitor D-glucose (4GBZ (Sun et al., 2012)) bound, and inward open with detached intracytoplasmic helices (4QIQ (Wisedchaisri et al., 2014)), have all been very useful for structural study, however the final substrate bound fully occluded structure has only recently been solved for the homologous malarial hexose transporter PfHT1 (6RW3) (Qureshi et al., 2020)). The occluded structure acquisition, coupled with the powerful mutagenesis experiments, HDX-MS analysis and molecular dynamics simulations have recently allowed for the proposal of the mechanism of which MFS transporters cycle to transport solute across the bilayer **Figure 5-1**.



**Figure 5-1: MFS alternate access transport mechanism for XylE**

*XylE* structure: *D*-xylose bound, outward facing partially occluded state (4GBY) (Sun et al., 2012). N-domain helices 1-6 are coloured orange, ICH domains yellow, and C-domain helices 7-12 in gold. Membrane boundaries calculated using orientations of proteins in membranes (OPM) (Lomize et al., 2012). Proton binding site: The D27-E206-R133 salt bridge is the site of proton binding. Proposed transport cycle: the D27-E206-R133 salt bridge is first disrupted by binding proton at E206, this promotes the transition from outward open, to the partially occluded state. Substrate then binds, and the D27 is also protonated. This shifts the protein into an inward facing state, via the occluded state where a significant movement of the broken helix 7 is observed (see main text). The inward facing partially occluded site releases D27 proton first, followed by the substrate, with the N-terminal domain being stabilised by a PE lipid as determined by HDX-MS characterisation of *XylE* within MSP-lipid nanodiscs. The final release of the E206 proton reverts the structure back towards the outward open conformation ready to restart the cycle. For each conformation, the PDB structure of either *XylE*, GLUT1-5, or malaria hexose transporter PfHT1 is listed, which aid in the proposal of this MFS transport cycle for *XylE*. The occluded structure has only been solved for PfHT1 (Qureshi et al., 2020), which shares roughly 30% sequence homology with GLUT and *XylE* transporters. Proton binding mechanisms stem from HDX studies on the *XylE* transporter (Martens et al., 2018). Xylose binding site: *D*-xylose bound adjacent to broken helix 7, bound to several polar and aromatic residues which are coordinated by 8 hydrogen bonds. On transition to the occluded state from the outward facing partially occluded state, the extracellular half of broken TM7 shifts the conserved Y298 and Y299 residues towards TM1, trapping the substrate in the binding site (Qureshi et al., 2020).

It is suggested that protonation must occur before xylose binding for the correct catalysis of the transporter. Two suggested residues for this protonation are D27 (helix 1), and E206 (helix 6), of which D27 has been proven to be essential for transport by biochemical experimentation,

where mutation to D27N removes the charged residue. E206 is suggested to modulate the pKa of D27 which affects the binding and dissociation of the proton (Wisedchaisri et al., 2014, Bazzone et al., 2016). The D27-E206-R133 salt bridge network is shown in the partially occluded xylose bound structure (4GBY) in **Figure 5-1**.

A mechanistic role for the conserved network of conserved charged residues (D27-E206-R133) on the intracellular side, as well as the effects of lipid, using XylE reconstituted within MSP-based lipid nanodiscs, and ligand binding for XylE transport have recently been characterised in a benchmark study using HDX-MS and molecular dynamics simulations. Synchronisation of the transport cycles was achieved by locking the protein into an outward-facing conformation, promoted by the introduction of a bulky tryptophan at G58W at the outside face between the two domains. Starting with the G58W outward facing conformation with substrate bound, it was shown that H<sup>+</sup> first binds D27, which then moves to disrupt the charged network, and causes an opening of the protein on the intracellular side. Spontaneous deprotonation then occurs resulting in H<sup>+</sup> release. PE lipid then binds and stabilises the inward facing state, which facilitates release of substrate into the cytosol, prior to transitioning back to outward facing conformation to restart the transport cycle. This transport scheme has been summarised in **Figure 5-1**.

D-xylose binds adjacent to the break in helix 7 and is recognised by 8 hydrogen bonds to both polar and aromatic residues (**Figure 5-1**) including: Q168(TM5), Q288/Q289/N294(TM7), W392(TM10) and W416(TM11). Y298(TM7) and Q415(TM11) also contribute to substrate binding via water mediated hydrogen bonding. Aromatic residues commonly surround the binding pocket, these include F24(TM1), Y298(TM7), W392(TM10) and W416(TM11) and are involved in substrate gating and coordination (Sun et al., 2012). Binding assays for WT XylE with D-xylose have been carried out using ITC (Sun et al., 2012, Madej et al., 2014), and microscale thermophoresis (MST) (Jiang et al., 2019) and all yield a similar  $k_D$  of around 0.3 mM. This value is similar for D-glucose binding; however, glucose is not transported by XylE, and acts as a natural inhibitor for the protein although all the molecular contacts between XylE and the sugar are similar. However, the C<sub>6</sub>-OH group of D-glucose is hydrogen bonded to Q175 and the C<sub>1</sub>-OH group is hydrogen bonded to the carbonyl oxygen of G388. These interactions are not observed with D-xylose (Madej et al., 2014).

The GLUT transporters, unlike XylE, are uniporters which do not require protonation to transport their glucose substrate. Nevertheless, coupling the above XylE focused experiments

with structural studies of the GLUT transporters and the malarial hexose transporter PfHT1, allow a detailed insight into the substrate induced mechanism and the function of the broken helices 7 and 10 which are conserved throughout the MFS transporters of transport. The Drew lab (Qureshi et al., 2020) proposed that the extracellular region of broken TM7 helix modulates substrate binding and gating at the extracellular gate, through a substantial movement of a universally conserved asparagine residue (N294 in XyleE) during the transition to a sugar-bound occluded state. The asparagine binds the C3 and C4 hydroxyl sugar groups of substrates in the binding site (**Figure 5-1**). In GLUT transporters 1-5 and XyleE, but not PfHT1, two conserved tryptophan residues in same broken helix 7 (Y298/Y299 in XyleE) move in concert with N294 towards TM1 to occlude sugar exit - this was shown structurally with GLUT3 (Qureshi et al., 2020). Additionally, broken helix 10 (related by inverted symmetry to helix 7) gates the intracellular interface as an additional method of allosteric coupling (Qureshi et al., 2020). However, the fact that PfHT1 can be crystallised in this fully occluded state, suggests that the malarial transporter must be more stable than it is for XyleE and GLUT transporters in the occluded state, and with the dynamism of broken helix 7 possibly exaggerated compared to mechanism in XyleE and GLUT transporters. Although, it is likely still important in these for these proteins, suggesting that substrate gated allosteric coupling is possibly more important in transitioning between conformations than first thought, particularly for XyleE which also requires protonation at a separate site for transport.

Transport activity of XyleE can be determined using an active transport or counterflow assays (Madej et al., 2014, Sun et al., 2012), often using radiolabelled xylose which can be very expensive. For counterflow experiments, radiolabelled uptake assays into liposomes yield activities of 2 nmol xylose/mg WT XyleE, or 0.02 nmol glucose/mg WT XyleE. For active transport of xylose, XyleE deficient cells were resuspended in media containing radiolabelled xylose, and uptake was measured. Data was quoted as a percentage of WT transport, where the D27N mutant, which prevents protonation abolished active transport of xylose, whereas counterflow transport remained similar, if slightly lower. XyleE mutant E206Q was also tested, which disrupts the D27 and R133 salt bridge network resulting in a similar active transport compared to WT, however this time, counterflow increased by around half of activity compared to WT XyleE transport (Madej et al., 2014).

The ICH domains in XyleE not only contribute to protein stability but are also functionally important, this was determined by alanine screening and testing functionality. Experiments showed that most residues in the ICH domains are largely hydrogen bonded, and when these

residues were not present, activity was significantly reduced compared to WT XyleE (Sun et al., 2012). The mechanism of ICH function is unclear, however they appear to form a latch which tightens the intracellular gate (Bai et al., 2017), with the extensive contacts on the interface with the ICH and the cytosolic face of XyleE TM domains. The interface includes a second network of charged residues which form salt-bridges; D337-R341 (Quistgaard et al., 2013, Madej et al., 2014), which are also necessary for transport.

### **5.1.2. Aims of chapter**

This chapter aims to express and purify XyleE, and associated XyleE tagged mutants. These proteins are then characterised for structure and function, required for downstream study. In this thesis, XyleE has been prepared for further study using AFM and MT single-molecule force spectroscopies. Details of which can be found in **chapter 6**.

#### **5.1.2.1. Purification, purity, and reconstitution**

It has been possible to prepare XyleE to high yields and purity in the Booth group, which was first published on as a comparative stability between MFS transporters (Harris et al., 2017a). Preparation of XyleE in this thesis was primarily for downstream AFM projects, and due to the sensitive nature of this technique, the quality of the protein should be high, free from aggregates and degradation. However, papers which prepare similar proteins, like LacY for AFM studies, do and are perhaps not as clean as they possibly could be (Serdiuk et al., 2014).

Further modifications were made to WT XyleE for future downstream single-molecule force spectroscopy studies, with preliminary studies using AFM described in **Chapter 6**. It was necessary to clone an additional 28 amino acid poly-glycine based linker onto the C-terminus of XyleE, based on previous work studying LacY using AFM (Serdiuk et al., 2014). The SDS-PAGE gels, chromatography traces, and structure and activity data for XyleE-pGly28 are shown through this chapter. Additionally, it was necessary to manipulate the protein construct for future magnetic tweezer experimentation. Purification and protein characterisation for this construct is also shown here.

Finally, using detergent purified XyleE for reconstitution into a synthetic lipid bilayer is paramount for downstream *in vitro* studies – I investigate whether this can be achieved for XyleE, to a similar or better standard than previous studies on similar proteins.

#### **5.1.2.2. Determining purified XyleE functionality**

To determine the functionality of purified XyleE and mutants, secondary structure was first cross-checked to published PDB structures of XyleE using Dichroweb CD spectra deconvolution

(Abdul-Gader et al., 2011). Additionally, binding assays were carried out using isothermal titration calorimetry (ITC) and compared with previously published values (Madej et al., 2014).

This chapter also addresses attempts to measure protein transport of D-xylose across a membrane using a novel enzyme-linked transport assay. Fluorescence assays can be a much cheaper alternative than commonly used radiolabelled substrates; however transport assays of this nature are difficult to achieve due to the requirement of tight compartmentalisation in the bilayer to prevent leakage of the transporting substrate or the co-substrate – and the MFS family requires maintenance of a pH, or solute gradient across the membrane. Additionally, a fluorescent analogue of D-xylose does not exist, and therefore a different approach must be taken. A xylose dehydrogenase (XDH) enzyme has been used in previous research for the quantification of D-xylose in urine (Sanchez-Moreno et al., 2016). I attempt here to determine whether it is possible to harness this reaction; from D-xylose to D-xylonoactone, with subsequent turnover of NAD<sup>+</sup> to NADH in a 1:1 stoichiometry, for a XylE transport assay when the XDH is compartmentalised within a liposome system and can only catalyse the reaction when XylE imports D-xylose into the vesicle.

## 5.2. Results

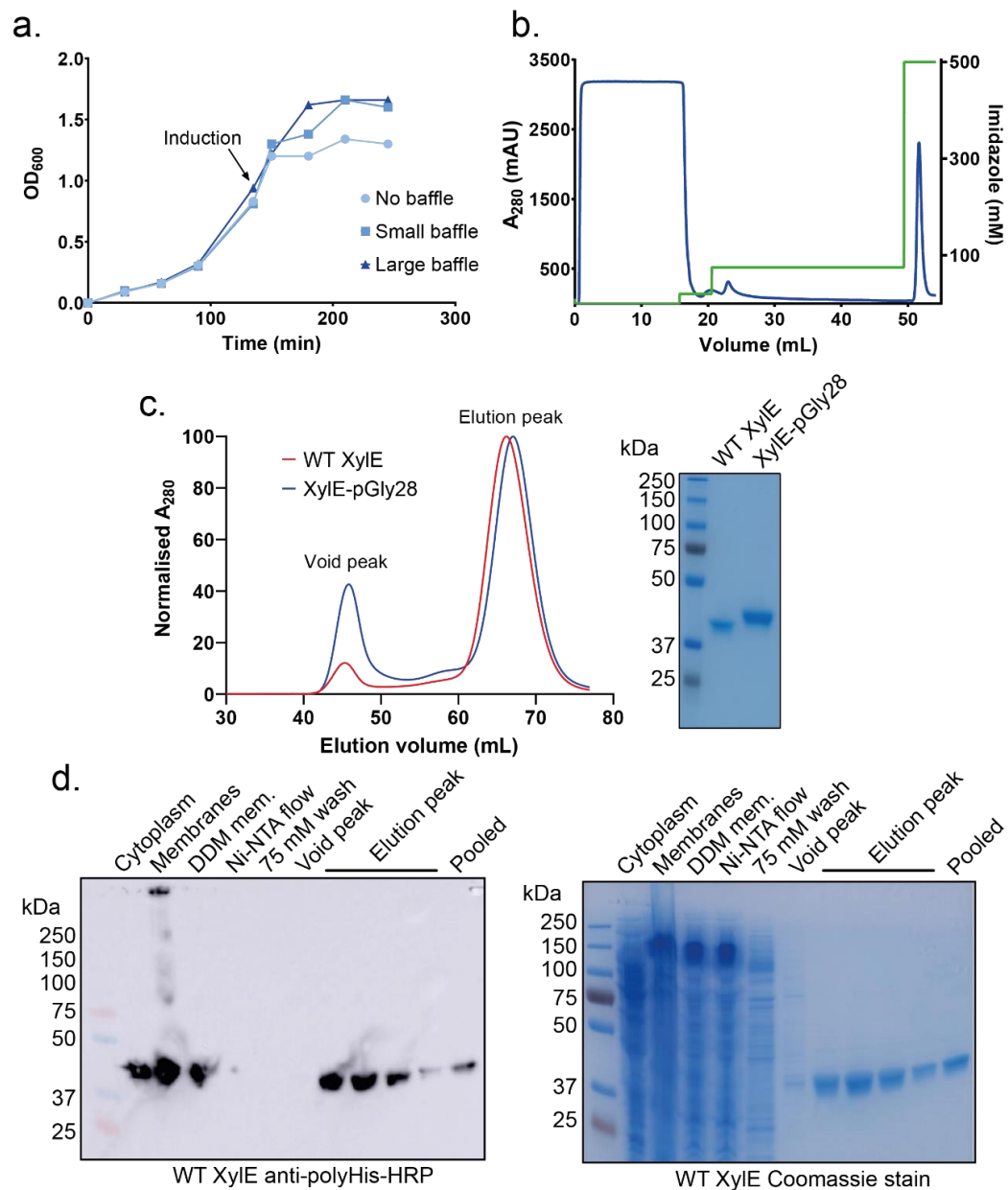
### 5.2.1. XylE expression and purification in DDM

XylE from *E. coli* with C-terminal His<sub>10</sub> was overexpressed from a pET28 vector in BL21-AI cells as described in **Chapter 2** before purification in DDM detergent. **Figure 5-2a** shows typical growth curves for WT XylE expression.

For XylE purification, DDM detergent was used to solubilise the *E. coli* inner membrane and was included in each purification buffer. **Figure 5-2b** shows a characteristic chromatogram for the loading of the Ni-NTA His-trap and elution of the solubilised membranes for affinity protein purification of WT XylE. Once the sample was loaded, the trap was washed with buffer containing 75 mM imidazole to further remove non-specifically bound protein from the trap. This is slightly higher than a usual 25 mM imidazole wash which is commonly used for hexahistidine tags, unlike the His<sub>10</sub> tag which binds Ni-NTA with higher affinity. The protein was then eluted from the trap and directly injected onto a Superdex 16/600 200 PG gel filtration column. A characteristic WT XylE purification chromatogram from the gel-filtration step is shown in **Figure 5-2c**, where the void peak containing protein aggregates eluted at ~45 mL, and the major elution peak at ~68 mL. Peaks were collected and concentrated and run on a 12

% Tris-Glycine SDS-PAGE gel to confirm the high purity and ~40 kDa molecular weight of purified XylE, and slightly larger 42 kDa XylE-pGly28 (see below) as shown in **Figure 5-2d** with Coomassie staining and Western blotting of the His<sub>10</sub> tag.

For SMFS experiments detailed in **Chapter 6**, a 28 amino acid poly-glycine sequence was cloned onto the C-terminal of XylE to provide a tag with a known length, and to remove symmetry in the structure of the protein. The XylE-pGly28 construct was also prepared using the same procedure as above and was able to purify to the same standard as the WT XylE protein, albeit with larger degree of aggregation present in the void peak (**Figure 5-2c**).



*Figure 5-2: Growth, expression, and purification of WT XylE and XylE-pGly28*

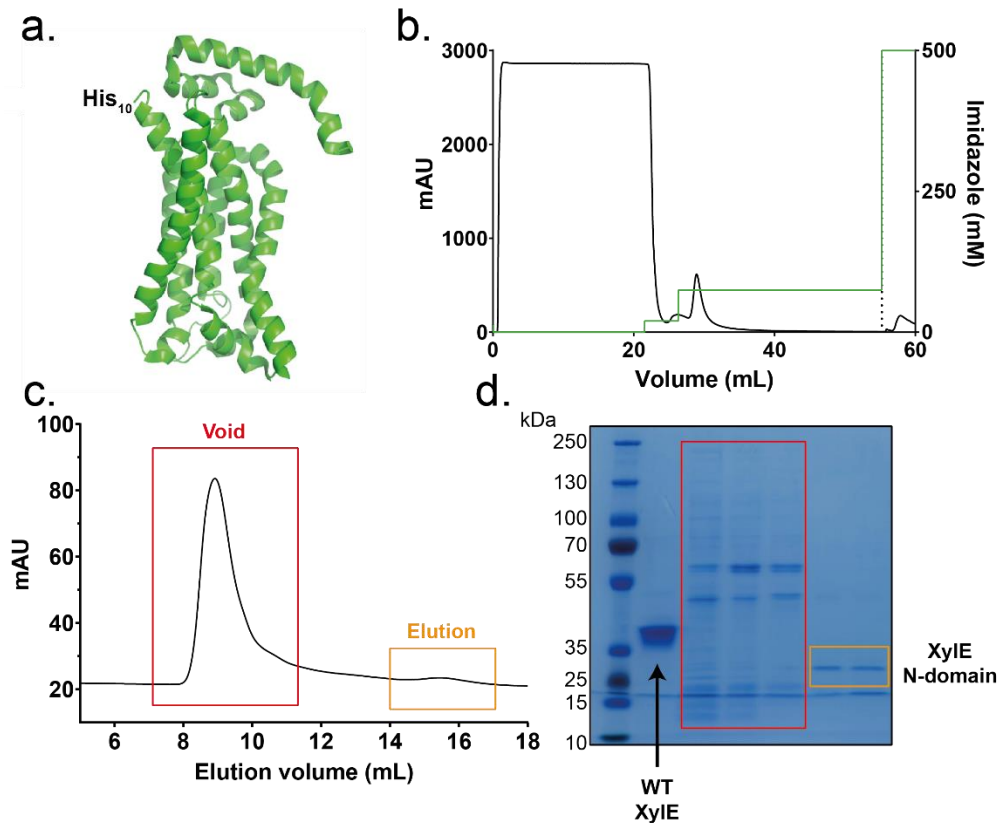


a) Characteristic growth curves for WT XylE in LB media, in flasks with different sized baffles. b) Characteristic AKTA chromatogram for the loading of solubilised E. coli membranes onto a 1 mL Ni-NTA His-trap. The 2 mL 500 mM elution is injected directly onto the Superdex 16/600 200 PG gel-filtration column. c) Characteristic elution profile for WT XylE and XylE-pGly28 gel filtration showing the aggregate void peak at 45 mL and the major elution peak at 68 mL. d) The elution peak was collected and concentrated before running on an SDS-PAGE gel to confirm high purity and size of XylE at around 40 kDa, and XylE-pGly28 at 42 kDa when stained with Coomassie.

### 5.2.2. Expression and purification of separate XylE domains

It was also possible to separate XylE into its N and C domains and attempt to purify each domain separately *in vivo*. XylE was split at V275, producing an N domain containing residues M1-V275 (**Figure 5-3a**) and the C domain containing G276-L491 (**Figure 5-4a**). The ICH domains of XylE were included in the N domain half.

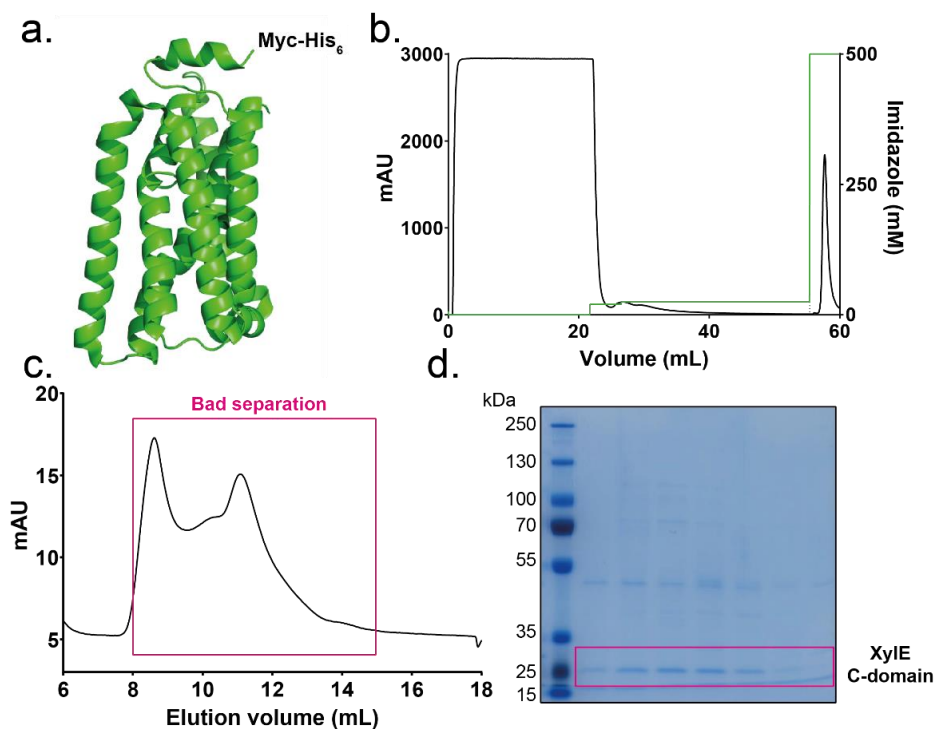
The N domain construct Ni-NTA trace showed normal loading onto the column, followed by an elution of non-specifically attached protein when washed with 75 mM imidazole (**Figure 5-3b**). A very small elution peak, indicative of low expression in the cell, was injected onto a Superdex 10/300 increase column, where the majority of sample eluted in the void volume and purifying only suitable N domain XylE in the elution peak (**Figure 5-3c**). The elution peak was run on a 12 % Tris-Glycine SDS-PAGE gel and stained with Coomassie to reveal that the void contains mainly oligomers or aggregates of the N domain, but I was able to isolate a small amount of the purified N domain (**Figure 5-3d**).



**Figure 5-3: Purification of the Xyle N-domain**

a) PyMol structure of N domain half of XyleE with N-terminal His<sub>10</sub> tag made using PDB: 4GBY. b) Ni-NTA chromatogram showing loading, 75 mM imidazole wash, and elution of N domain XyleE. c) Gel filtration chromatogram of N domain XyleE with large void peak (red box) at 9 mL and a small elution bump (orange box) at 15.5 mL. d) SDS-PAGE gel of WT XyleE and the proteins eluted under each peak. N domain XyleE runs at approximately 25 kDa.

The same protocol was carried out for C domain XyleE purification but was washed with 25 mM imidazole, due to the shorter His-tag to prevent washing away good protein. The Ni-NTA chromatogram (**Figure 5-4b**) shows a much larger elution peak for injection onto the Superdex 10/300 increase, perhaps explaining the poor separation (**Figure 5-4c**) of the void and monomer peaks shown here at 9 mL and elution at 11 mL. This is possibly indicative of mixed oligomers of the C-terminal domain which cannot be resolved from monomeric protein.



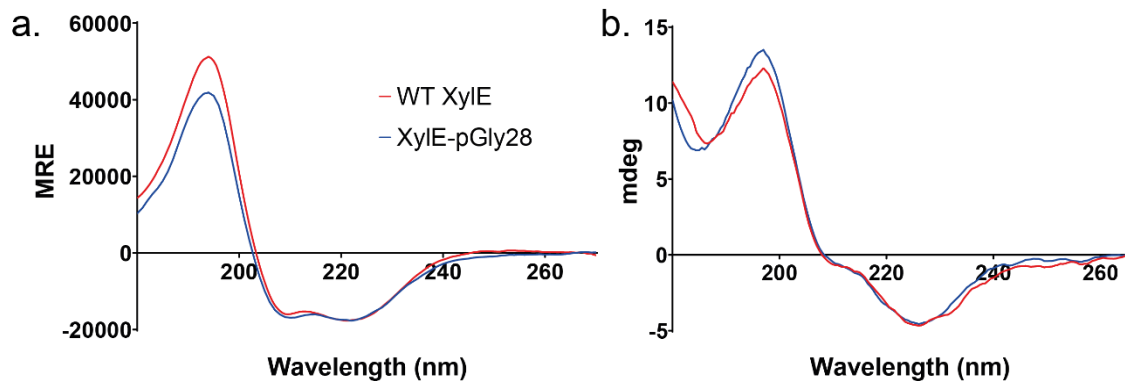
**Figure 5-4: Purification of the Xyle C domain**

a) PyMol structure of C domain half of Xyle with C-terminal His<sub>10</sub> tag made using PDB: 4GBY. b) Ni-NTA chromatogram showing loading, 25 mM imidazole wash, and elution of C domain Xyle. c) Gel filtration chromatogram of C domain Xyle with poor separation between the void peak and the elution peak (pink box). d) SDS-PAGE gel of C domain proteins eluted under each peak. C domain Xyle runs at approximately 25 kDa.

### 5.2.3. WT Xyle and Xyle-pGly28 are folded DDM

WT Xyle and Xyle-pGly28 in DDM detergent were subject SRCD to check correct folding after purification (**Figure 5-5a**). Both proteins showed characteristic  $\alpha$ -helical structure with troughs at 209 and 222 nm, and a peak at 190 nm which is twice the magnitude of the negative ellipticities. Trace deconvolution using program CDSSTR with dataset SMP180 (Abdul-Gader et al., 2011) and CDtool (Lees et al., 2004) suggested an  $\alpha$ -helical content of 73 %,  $\beta$ -sheet of 4 %, 7 % turns and 16 % unordered. For Xyle pGly28 the  $\alpha$ -helical content was 66 %,  $\beta$ -sheet of 5 %, 9 % turns and 20 % unordered. I compared these ratios with the PDB crystal structure 4GBY, which suggests 76 % of the structure is helical and confirm that my purified proteins are correctly folded. There was a drop in helicity for Xyle-pGly28, which I would expect due to the unordered linker region. Oriented CD was also carried out on WT Xyle and Xyle-pGly28 after reconstitution into a lipid bilayer consisting of a 40:40:20 mixture of synthetic DOPC:DOPE:DOPG (see **section 5.2.5.**). The traces of WT and pGly28 overlap **Figure 5-5b** and therefore suggest that they reside in the same orientation in the bilayer. Together, these CD

experiments confirm that WT Xyle is folded post-purification and suggest that presence of pGly28 does not affect Xyle folding.

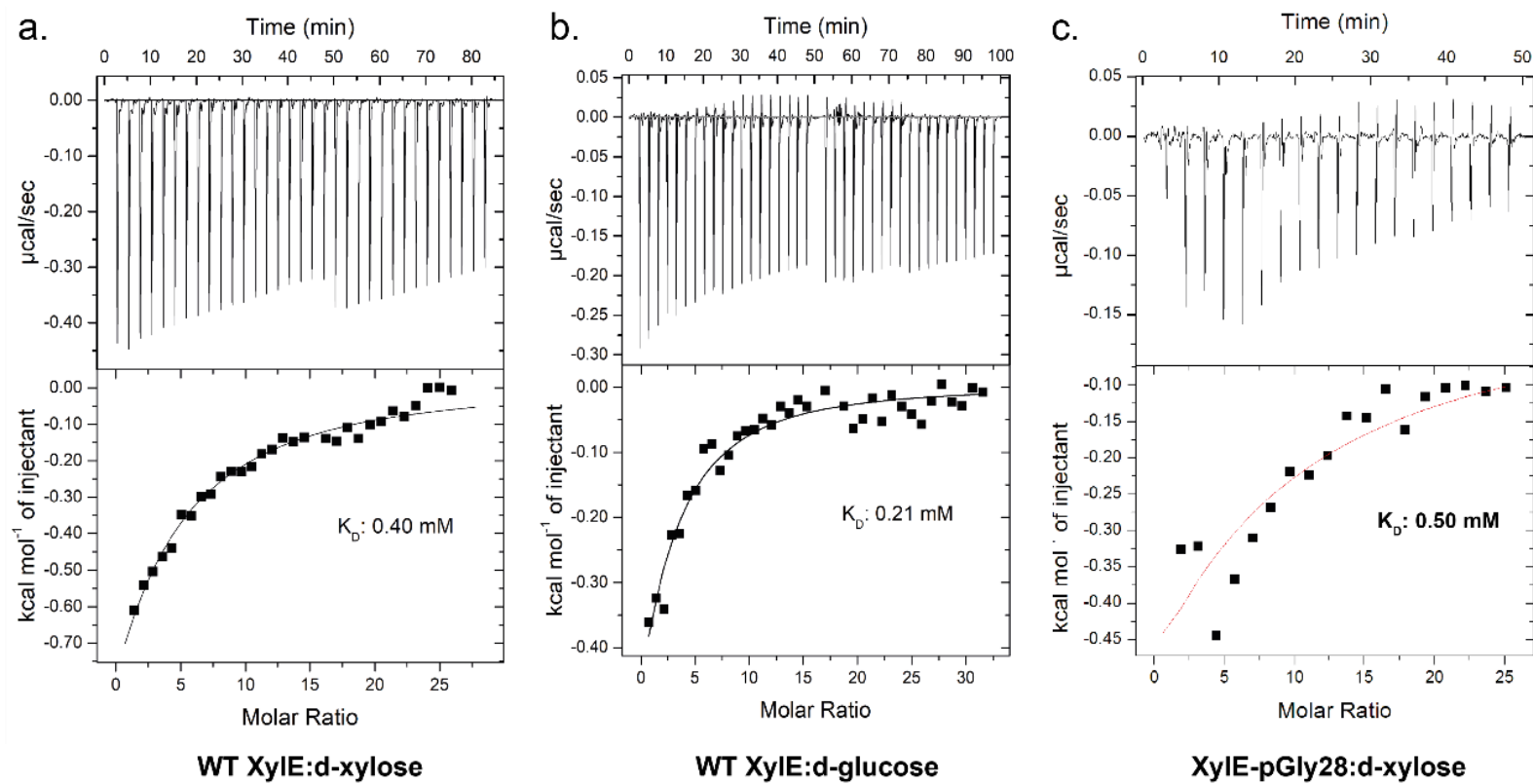


*Figure 5-5: CD and OCD of WT Xyle and Xyle-pGly28 in DDM*

*a) SRCD for WT Xyle and Xyle-pGly28 showing characteristic  $\alpha$ -helical structure with maxima at 190 nm, and two minima at 208 nm and 222 nm. b) OCD spectra of Xyle and Xyle-pGly28 in 40:40:20 PC:PE:PG bilayers. Both traces overlap suggesting both proteins are in the same orientation in the bilayer, and the additional pGly28 linker does not affect Xyle folding.*

#### 5.2.4. WT Xyle and Xyle-pGly28 bind ligand in DDM

Xyle binding capacity to its natural ligand, D-xylose and inhibitor D-glucose was checked using ITC. 2.5 mM ligand was titrated into 35  $\mu$ M Xyle or Xyle-pGly28, this causes molecular structure changes as the ligand interacts with its binding site, these steric changes can be measure by 'heat-differences' which are detected between the reaction chamber and a reference. The binding isotherms were fit to a one-site binding model and the  $k_D$  of WT Xyle binding to D-xylose was determined to be 0.4 mM (**Figure 5-6a**), and for D-glucose, 0.21 mM in 50 mM NaPi pH 7.4 with 0.05 % DDM detergent (**Figure 5-6b**). For Xyle-pGly28 binding to D-xylose, the  $k_D$  was calculated to be 0.50 mM (**Figure 5-6c**), showing a greater effect on the binding properties of Xyle than we expected, however the protein did still bind xylose, but additional experimentation is required to determine why the additional pGly28 linker affected binding capacity of Xyle, compared to LacY – with a likely explanation being an interaction with the ICH domains which are involved in conformational switching and transport.



**Figure 5-6: ITC for WT Xyle and Xyle-pGly28 in DDM**

a) Binding isotherm characterised using ITC between Xyle and *D*-xylose. The  $k_D$  was determined to be 0.4 mM after fitting to a one-site binding model. b) As panel (a) but with titration of *D*-glucose.  $k_D$  was determined to be 0.21 mM. c) Binding isotherm of *D*-xylose titrated into Xyle-pGly28. Data was again fitted to a one-site binding model and the  $k_D$  was determined to be 0.50 mM.

### 5.2.5. Protein reconstitution into synthetic lipid vesicles

Although very convenient for purification and simple denaturant folding studies, detergent micelles are not the best membrane mimic and are often destabilising to membrane proteins. The reconstitution of DDM solubilised membrane proteins into synthetic lipid bilayers is common and there are multiple methods to achieve this, mostly surrounding a pre-saturation of liposomes with detergent and mixing with protein before a fast removal of the detergent allowing the protein to enter the lipid bilayer.

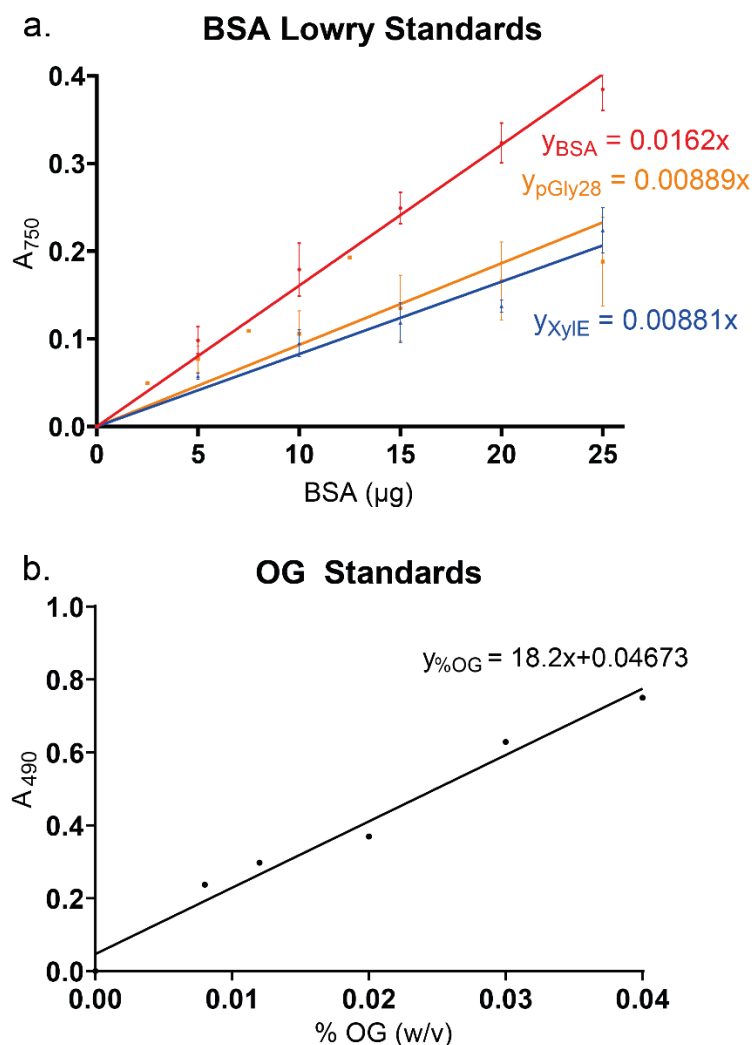
Octyl-glucoside (OG) detergent is commonly used as a pre-swelling detergent and was added to 400 nm extruded liposomes consisting of a 40:40:20 ratio of synthetic lipids DOPC:DOPE:DOPG. This lipid mix was chosen due to its higher fluidity than native *E. coli* membranes because of the inclusion of DOPC, due to its lamellar structure, rather than cone-shaped structure, allowing fluid bilayers to form. This fluidity allows for high protein reconstitution efficiencies of functional protein which was previously determined for LacY (Findlay and Booth, 2017).

To the swelled liposomes, a known amount ( $\mu\text{g}$ ) of protein was mixed and the detergent was removed quickly using either BioBeads, dilution, or detergent removal spin columns to promote protein insertion into the bilayer. The efficiency of reconstitution was determined using the Markwell-Lowry assay (Markwell et al., 1978), which precipitates protein out of the system and reacts with the Lowry reagent to generate a blue colour development. This was compared with a standard curve of BSA, or the protein of interest; XylE or XylE-pGly28. The linear trendline ( $y=mx+c$ ) passed through the origin for each protein, and conversion factor between the BSA gradient and the protein gradient was calculated for a quick conversion and meaning BSA could be used each time to save purified protein (**Figure 5-7a**). Reconstitution of XylE into 40:40:20 liposomes was determined to be between 60-80 % efficient for high lipid:protein 12.5:1 (w/w) required for AFM (**see chapter 6**), and 80-90 % when aiming for lower 33:1 (w/w) lipid:protein ratios for use with transport assays.

It is suggested in the literature that a 5:1 (w/w) ratio is possible for LacY (Serdiuk et al., 2014) equating to  $1.1 \times 10^{26}$  lipids around a single protein however, reconstitution efficiency was not measured in these published protocols. Efficiency calculations use starting lipid concentrations and does not consider a possible loss of lipid which occurs in the clean-up stage of my proteoliposomes. Urea or salt washes with proteoliposome centrifugal pelleting stages were

used to further clean reconstitutions for AFM and this can lead to liposome bursting. However, this eventuality is likely to increase protein:lipid, rather than decrease in my samples.

The degree of removal of OG (and DDM) sugar-based detergents was also checked by a colorimetric carbohydrate assay. OG standards (**Figure 5-7b**) were run to produce a standard curve, which was used to determine the remaining OG concentration in a sample. An example sample was determined to contain 0.0122 % remaining detergent, far below the 0.05 % usually found in DDM purified samples. This suggests proteoliposome samples contained a tight bilayer with protein reconstituted through it.



**Figure 5-7: Markwell-Lowry and OG detergent standard curves**

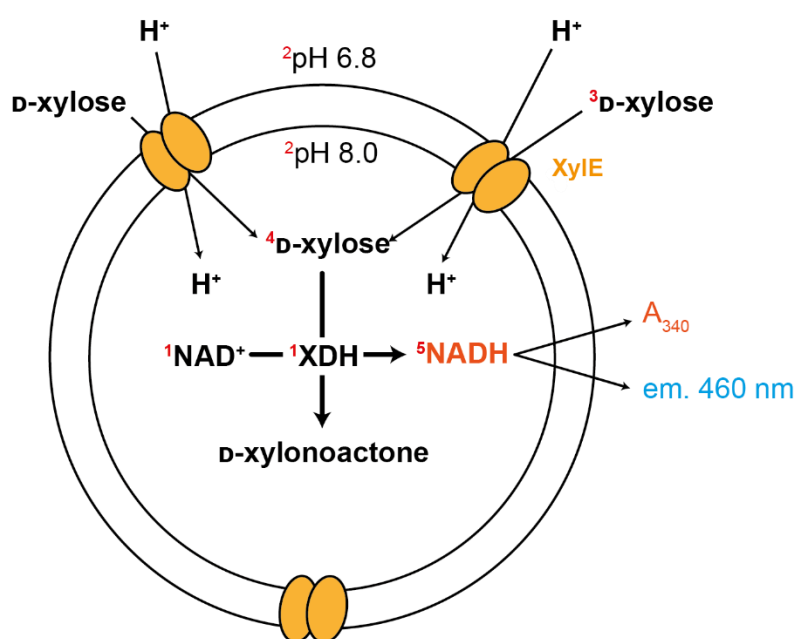
a) Markwell-Lowry assay standards of BSA ( $n = 17$ ), XylE ( $n = 3$ ) and XylE-pGly28 ( $n = 6$ ). Means are plotted with SD error bars and  $y=mx+c$  expression shown. All reconstitutions were run against BSA and a conversion factor of 1.83 applied to the gradient as each curve was forced through the origin. This allowed an estimate of  $\mu\text{g}$  XylE in a known volume of reconstituted sample. b) Carbohydrate OG assay for the presence of excess OG a reconstituted sample. One reconstitution was tested and 0.0122 % (w/v) OG was detected suggesting efficient removal of the pre-swelling detergent.

## 5.2.6. A novel NAD<sup>+</sup>-linked XylE functional assay

### 5.2.6.1. Preparation of xylose dehydrogenase (XDH)

Using the same reconstitution methods for low protein:lipid ratio shown in **section 5.2.5**. I tested the activity of WT XylE using a novel XDH assay by utilising the redox states of NAD<sup>+</sup> and NADH and observing their absorbance or fluorescence. I first trapped XDH inside a 40:40:20 PC:PE:PG liposome with NAD<sup>+</sup> at a relatively low concentration to avoid disrupting the bilayer, and reconstituted XylE into the system. When placed in an outside buffer of lower pH with D-xylose, the sugar is transported across the proteoliposome membrane by XylE and turned over to D-xylonoactone inside the liposome, as NAD<sup>+</sup> is reduced to NADH in a 1:1 stoichiometry. NADH yields a fluorescence response of transport for observation which can be observed spectroscopically.

**Figure 5-8** shows a schematic of the reaction set-up.



**Figure 5-8: Schematic for novel XDH assay for XylE transport**

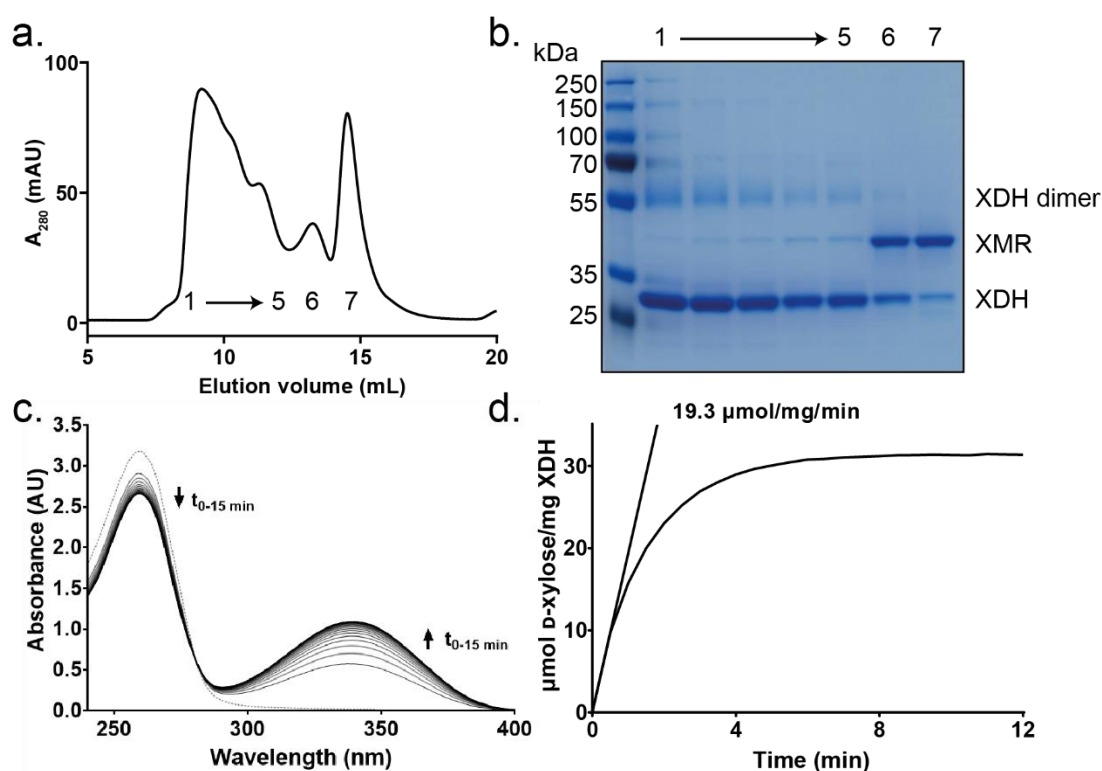
(1) NAD<sup>+</sup> and XDH are trapped inside a 400 nm liposome. These liposomes can be reconstituted with XylE and put in a lower pH buffer (2) with D-xylose (3) to initiate xylose transport across the bilayer using XDH. D-xylose gets converted into D-xylonoactone (4) and NADH is formed (5) in the process. This provides a fluorescence, or absorbance readout of xylose turnover, and hence XylE transport.

A commercial dehydrogenase/mutarotase enzyme mix (XDH/XMR) for the D-xylose substrate is separated into its two proteins using gel-filtration chromatography. When injected into a Superdex 10/300 column, the majority of the XMR was removed from the XDH monomer and dimer in solution and checked via 12 % Tris-glycine SDS-PAGE gel, where the XDH ran at 26



kDa, the XMR at 38 kDa and XDH dimer at 52 kDa (**Figure 5-9ab**). Separating the proteins allowed for more efficient trapping of XDH in the lumen of the 400 nm liposome.

Once separated, XDH was collected and concentrated before assaying for activity. UV spectroscopy was used as a direct measure of NADH produced and activity of XDH was calculated. XDH was mixed with NAD<sup>+</sup> and D-xylose to start the reaction, scanning kinetics across the absorbance of NAD<sup>+</sup> at 260 nm, and NADH at 340 nm was monitored as the reaction progressed (**Figure 5-9c**). The moles of NADH produced were calculated using Beer-Lambert and a molar extinction coefficient of NADH at 340 nm of 6220 M<sup>-1</sup>.cm<sup>-1</sup>. The reaction scheme suggests a 1:1 stoichiometry of D-xylose turnover to NADH production and thus μmol of D-xylose produced against time can be plotted, and the specific activity of the enzyme can be calculated using the initial rate of the curve, and mass of XDH used in the reaction for each batch of XDH prepared (**Figure 5-9d**). For the representative data in **Figure 5-9**, the activity was calculated to be 19.3 μmol/mg/min. See **section 2.12.2** for further details.



**Figure 5-9: Preparation and calculating the activity of XDH**

a) The XDH/XMR mixture was passed over a Superdex 10/300 gel-filtration column for separation of the two proteins. b) Fractions 1-5 containing XDH and fractions 6-7 containing XMR were boiled and run on SDS-PAGE gels. Coomassie stained gels showed XDH monomer at 26 kDa, XMR at 38 kDa and XDH dimer at 52 kDa. c) Pooled and concentrated XDH was added to a reaction mixture of NAD<sup>+</sup>, D-xylose and scanning kinetics absorption traces are run from 240-400 nm. Over 15 minutes, NAD<sup>+</sup> absorbance at 260 nm decreases, and NADH at 340 nm increases. The absorbance change of NADH was plotted over time and converted to μmol of D-xylose turned over per mg of XDH, and the activity of the batch was calculated using the initial rate as shown in panel d.

### 5.2.6.2. Encapsulation of XDH, and XylE activity

WT XylE proteoliposomes (produced using protocol described in **section 2.12.2**) were added into 300  $\mu\text{L}$  of PBS at pH 6.4 with varying concentrations of D-xylose, and the reaction was followed using fluorescence with excitation at 340 nm, and emission at 460 nm. A calibration curve linking fluorescence with NADH concentration **Figure 5-10a**; hence xylose turnover concentration, was used for conversion, and initial rates ( $v_0$ ) for each D-xylose concentration were calculated to work out activities; milligrams of XylE reconstituted was calculated using a Markwell-Lowry assay with subtraction of the mock reconstitution from the XylE reconstitution to determine encapsulated XDH concentration. A comparison of two baseline conditions with a corresponding 50 mM xylose reaction are shown in **Figure 5-10b** and **Figure 5-10c**. The mock reconstitution contains no protein but was subject to the same additional detergent swelling and removal. This showed a significantly lower activity, likely due to xylose leaking across the bilayer. The no pH gradient control shows only downhill transport – any transport which occurs in absence of the  $\text{H}^+$  co-substrate. The latter was used to construct a Michaelis-Menten curve **Figure 5-10d** between the initial rates of each condition and the corresponding concentration of D-xylose. MM relationship was applied to the data to yield a  $V_{\text{max}}$ :  $489.38 \pm 87.5 \text{ min}^{-1}$  and  $k_m$ :  $166.71 \pm 50.75 \text{ mM}$ .

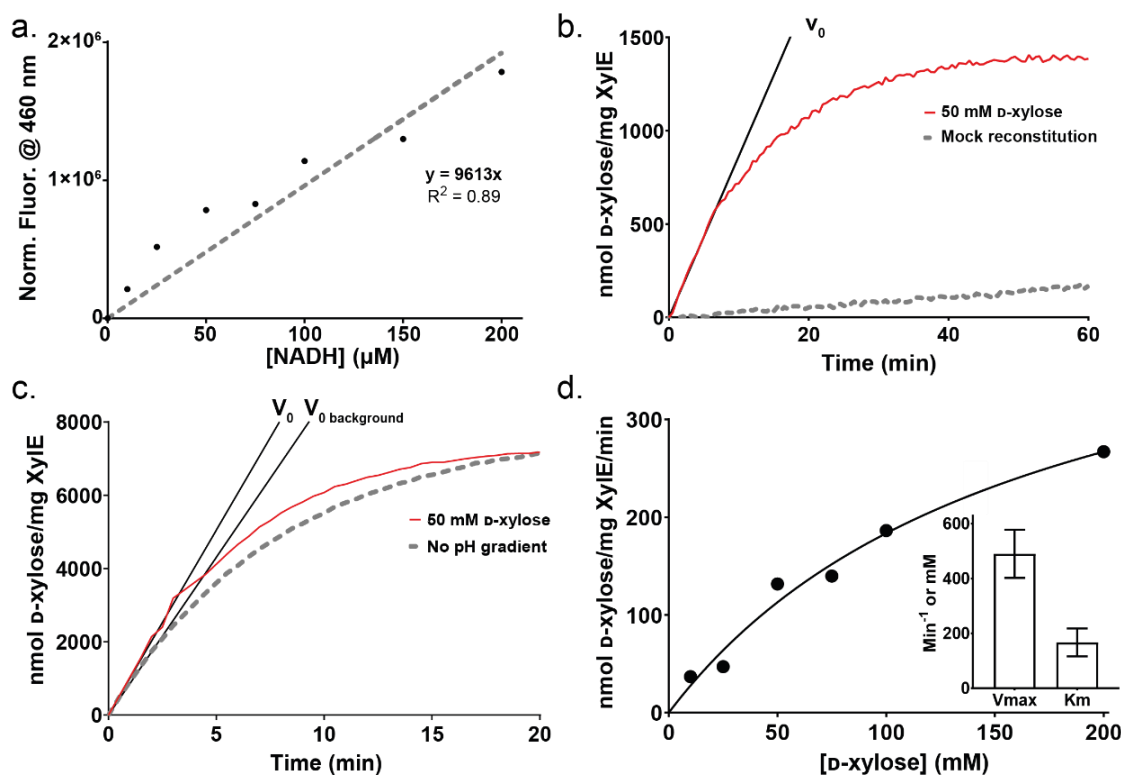


Figure 5-10: Transport of D-xylose across a bilayer assayed using XDH

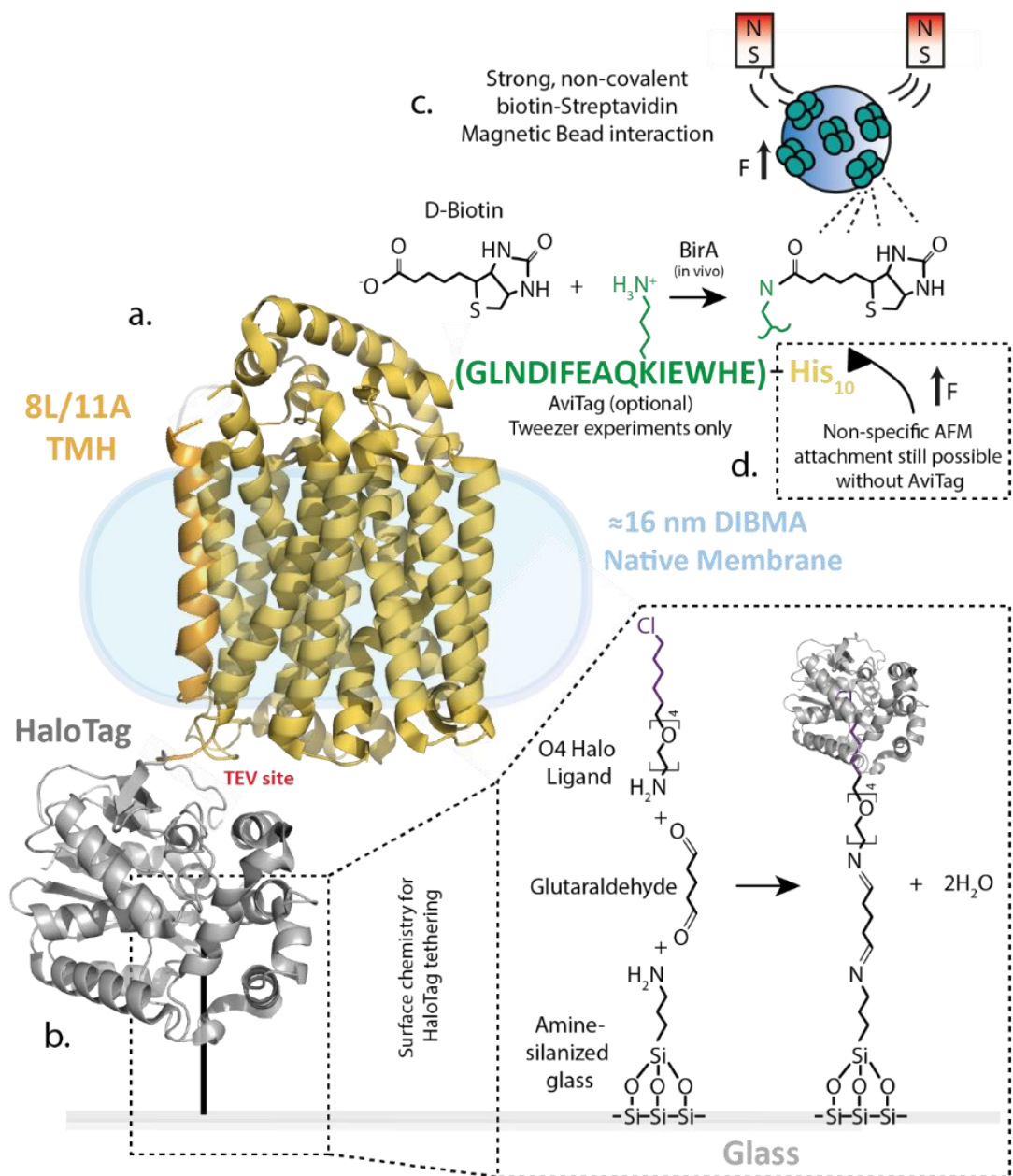
a) Calibration curve relating fluorescence output with concentration in the presence of empty liposomes to mimic light scattering. A linear fit was forced through the origin for easy conversion of fluorescence to NADH concentration. b) Transport of xylose by XyleE down a concentration gradient, with a mock reconstitution background. The initial rate of transport ( $v_0$ ) can be calculated by plotting a line through the first 2.5 min of transport. c) The same experiment was repeated with no pH gradient as a background to prove downhill transport. d) 6 xylose concentrations were tested, and the no pH gradient background was subtracted. The initial rates were plotted against xylose concentration and data was fitted to a Michaelis Menten relationship to calculate  $V_{max}$ :  $489.38 \pm 87.5 \text{ S}^{-1}$  and  $K_m$ :  $166.71 \pm 50.75 \text{ mM}$  (mean  $\pm$  SEM). This data should be taken with a pinch of salt, as only one repeat was performed, and the background was very high.

### 5.2.7. Modifying XyleE for single-molecule experiments

In addition to pGly28 tagging of XyleE to remove symmetry of the protein for AFM studies, it was necessary to produce a construct with protein terminals on different sides of the membrane to enable the use of MT spectroscopy; in WT XyleE, both protein terminals are on the same side of the membrane. This allows the protein to be tagged and tethered on either side of the membrane, allowing unfolding in the same vector as an AFM study. To achieve this, an additional helix was cloned onto the XyleE N-terminus. A *de novo* helix of leucine and alanine, an 8L/11A helix (Cymer et al., 2014) with a GSSG linker (termed TMH from here on), was chosen as this has been shown previously to maintain the native topology of similar MFS proteins when grown in *E. coli* (Cymer and von Heijne, 2013). To tether the *de novo* N-terminal TMH to a substrate, the TMH-XyleE construct was subcloned into pFN18 to incorporate an N-terminal HaloTag protein (Popa et al., 2013a), which can strongly bind via glutaraldehyde to a functionalised glass surface. This also enabled us to test the HaloTag tethering using AFM imaging (**section 6.4.3**).

Further down the line and beyond the scope of this thesis, C-terminal tethering for the tweezer construct can be achieved by incorporating an Avi-Tag which can be biotinylated *in vivo* (co-expression with BirA) as discussed in **Chapter 3**. This enables the strong covalent bonding to streptavidin coated magnetic beads and generating a construct which is tethered at both ends.

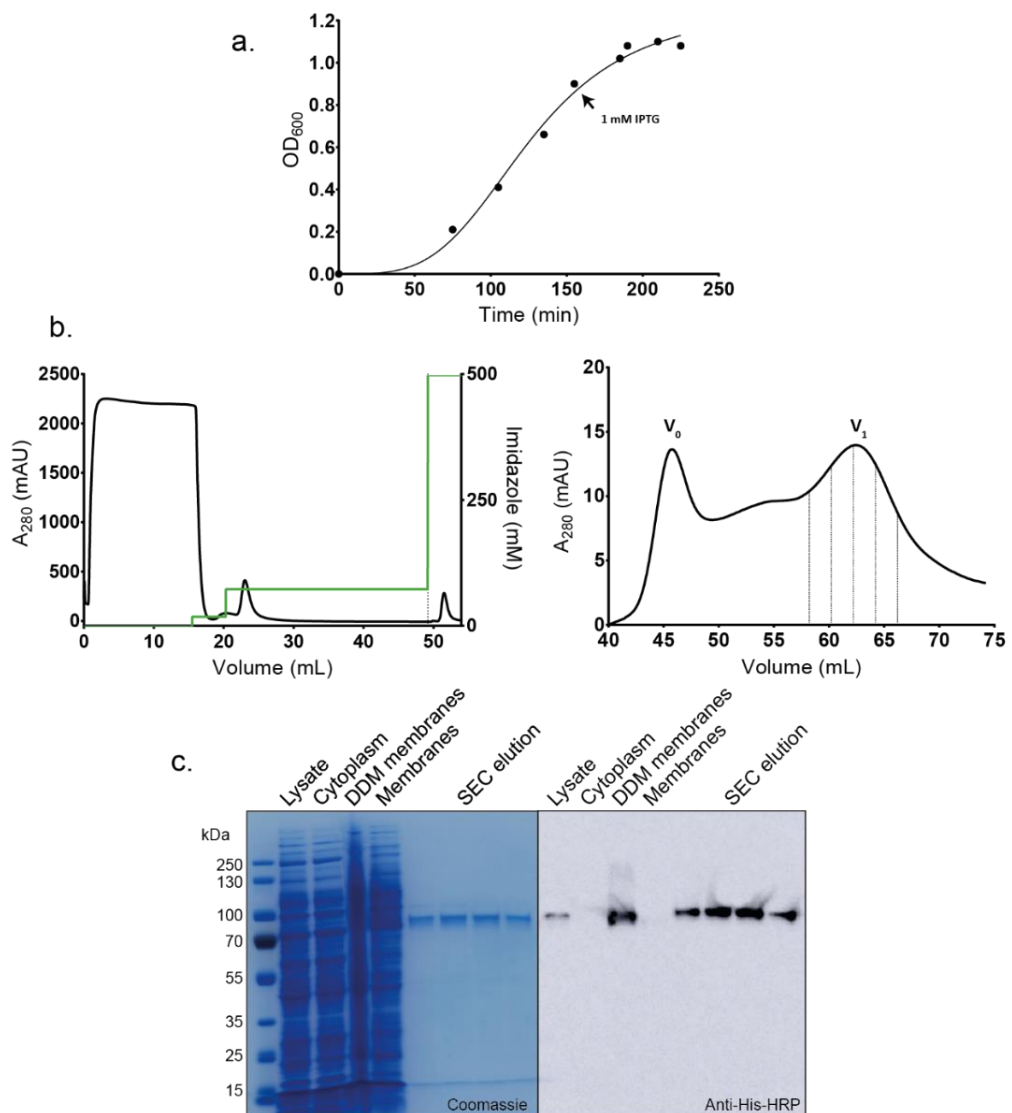
A schematic for the entire planned tweezer construct is shown in **Figure 5-11**.



**Figure 5-11: Schematic of doubly tagged Xyle for perpendicular AFM or MT pulling**

a) A schematic of the entire HaloTag-TMH-Xyle-Avi-tag construct purified in DIBMA native nanodiscs. The HaloTag tethers the protein to glass, and the C-terminal tethers either a streptavidin coated magnetic bead when the Avi-tag is present and biotinylated, or by a non-specific interaction with the AFM tip. A TEV site was also engineered between the Halo protein and TMH. The TMH is produced by an additional 8A/11L helix on the Xyle N-terminus. This bridges the membrane allowing tethering of Xyle on both sides of the disc. b) Details of the HaloTag surface chemistry are shown here. First, glass is silanized using 3-aminopropylsilane, next glutaraldehyde and the O4 amine Halo ligand are added allowing covalent bond formation with the expulsion of two water molecules. The Halo protein binds the chloroalkane ligand to complete the tethering. c) At the C-terminal end of the construct, when the Avi-tag is present the lysine (K) in the sequence can be biotinylated in vivo when co-expressed with a plasmid encoding the biotin ligase enzyme (BirA). This biotinylation forms strong non-covalent interactions with a streptavidin coated magnetic bead. d) If the Avi-tag is not present (or not biotinylated) the  $Si_3N_4$  AFM tip can also bind non-specifically for AFM SMFS studies.

The Halo-TMH-XylE construct (without Avi-Tag) was initially prepared in DDM detergent to test the validity of the construct for preparation without the additional DIBMA purification which can cause more complex issues as explored in **Chapter 4**. **Figure 5-12a** shows the characteristic growth curve of BL21-AI cells expressing the construct which were chilled to 30 °C post induction. The two step purification traces for Ni-NTA, followed by gel-filtration on a Superdex 16/600 200 PG column are shown in **Figure 5-12bc**. 1 mL fractions of the elution peak were run on SDS-PAGE to test the purity of the sample **Figure 5-12c**. Western blotting with the anti-poly-histidine antibody showed that all the protein resided in the membrane, and all protein was successfully solubilised using DDM. This shows that the construct is as expected, and while yield is low, the protein is clean and with quantity enough for single molecule experiments.



**Figure 5-12: DDM preparation of Halo-TMH-XylE construct for AFM**

a) Growth curve of BL21-AI cells containing the pFN18a vector with Halo-TMH-XylE gene. This was expressed on addition of 1 mM IPTG at an OD of 0.9 and the temperature dropped to 30 °C to slow down expression of the protein. b) Left: Chromatogram of the loading, washing and elution of the 1 mL His-trap at corresponding imidazole concentration on the right-y axis (green). Right: The 2 mL elution was loaded directly onto a 16/600 Sepharose column to remove aggregated Halo-TMH-XylE which elutes around 46 mL ( $V_0$ ). The  $V_1$  elution peak was taken in 4 x 2 mL fractions as denoted by vertical dotted lines. c) The cell fractions and four elution fractions from size-exclusion chromatography were run on SDS-PAGE gels and Coomassie stained or Western blotted for anti-Histidine tag. DDM efficiently solubilised membranes containing the protein and was subsequently purified to homogeneity with minimal degradation or aggregation.

A similar method was used for the purification of this construct in native nanodiscs directly from the *E. coli* inner membrane (see **Chapter 4, section 4.2.3**). The DDM was substituted for the DIBMA co-polymer and the sample was bound to Ni-NTA beads overnight to separate membranes solubilised containing protein from discs which do not contain protein. These samples were deposited onto halo-ligand functionalised glass and imaged using AFM to confirm functionality of the Halo protein (**section 6.4.3**).

### 5.3. Discussion

#### 5.3.1. XylE preparation and activity

I have been able to prepare XylE to a high purity, and the purified protein is active in DDM micelles, assayed by way of binding capacity to its natural ligand D-xylose, and its natural inhibitor D-glucose, using ITC. I also develop the makings of a fluorescence activity assay for XylE using an enzyme-linked xylose dehydrogenase enzyme and NADH production as the output.

**Table 5-1** shows the biochemical results obtained for XylE compared to those recorded in the literature.

Table 5-1: Summary and literature comparisons of biochemical values obtained for WT XylE and XylE-pGly28

	Purity	Secondary structure (CD) in DDM detergent	Binding capacity (ITC) In DDM detergent	Reconstitution efficiencies	Activity
<b>Thesis result</b>	>99 % pure monomer determined using SDS-PAGE for both WT and XylE-pGly28.	<b>WT XylE:</b> $\alpha$ -helix: 73 %, $\beta$ -sheet: 4 %, turns: 7 %, unordered: 16 %. <b>XylE-pGly28:</b> $\alpha$ -helix: 66 %, $\beta$ -sheet: 5 %, turns: 9 %, unordered: 20 %.	<b>WT XylE <math>k_D</math>:</b> D-xylose: 0.40 mM, D-glucose: 0.21 mM <b>XylE-pGly28 <math>k_D</math>:</b> D-xylose: 0.5 mM.	60-80 % for high (12.5:1 wt:wt) lipid:protein ratios, and 80-90 % when aiming for lower (33:1 wt:wt) lipid:protein ratios.	<b>WT XylE</b> Michaelis-Menten kinetics for downhill transport across a bilayer: $V_{max}$ : $489.38 \pm 87.5 \text{ min}^{-1}$ and $k_m$ : $166.71 \pm 50.75 \text{ mM}$ .
<b>literature results</b>	Not shown in vast majority of XylE structure or biochemical work. Although my result is comparable with Harris, et al.	PDB 4GBY: $\alpha$ -helix: 76 %.	For <b>WT XylE:</b> D-xylose: $k_D = 0.35 \pm 0.03 \text{ mM}$ , glucose: $k_D: 0.77 \pm 0.01 \text{ mM}$ .	Nothing published for XylE. However similar proteins like LacY are suggested to reconstitute to very high lipid:protein of 5:1 (wt:wt). However, efficiencies are not determined in these papers.	Counterflow of $^3\text{H}$ -D-xylose by WT XylE yielded $k_m$ of $0.47 \pm 0.05 \text{ mM}$ . Glucose is not a transport substrate for XylE carried out using proteoliposomes of <i>E. coli</i> polar lipids.
<b>Advantages and limitations of thesis work</b>	My XylE preparations are very pure when purified, chromatograms and gels of XylE purified in the literature has not been assayed – I can only presume my samples are to a similar, if not better quality than published. Although XylE has not been	My XylE preparations are on par with published work in the field in terms of secondary structure content. The pGly28 construct has less $\alpha$ -helical content and more unordered structure. This is expected due to the additional 28 amino acid	Most binding assays in the literature for XylE concentrate on its natural ligand xylose, and inhibitor glucose and functional mutants of the protein are identified. My $k_D$ values are consistent with literature results for both xylose and glucose, with glucose in my	Actively characterising the amount of protein inserted into the membrane allows for calculation of proteins per area, which is useful for AFM, as well as accurately defining mg of XylE involved in transport assays for activity. More information can be found on increasing XylE	Although incredibly sensitive, radiation studies of protein activity can be very expensive and limited due to lab resources. A fluorescence alternative would provide a more universal approach. This XDH Transport assay is a work in progress with many difficult steps to encapsulate and maintain gradients. The controls used in this

	<p>published as an AFM study, comparing this purity with preparations of LacY in the literature would suggest my protein is much cleaner, with LacY showing evidence of dimerization in the prepared samples. Interestingly, LacY-poly-glycine samples improve the dimerization artifact, but there is a small degree of degradation in the sample.</p>	<p>poly-glycine structure which I would expect to not have any defined secondary structure. This however have not been assessed in the literature.</p>	<p>case binding with a slightly stronger affinity.</p>	<p>reconstitution efficiency in <b>Chapter 6.</b></p>	<p>new assay can also vary depending on whether downflow or up-flow transport is being observed. The MM data shown in the thesis is based off one repeat with a high background and should be taken with a degree of caution. However, the makings of a novel fluorescence activity assay are shown which could easily be optimised for accurate determination of transport activity, particularly in cell-free applications, and/or DIB systems.</p>
<b>References</b>	(Harris et al., 2017a, Serdiuk et al., 2014)	(Sun et al., 2012, Harris et al., 2017a)	(Madej et al., 2014, Sun et al., 2012)	(Serdiuk et al., 2014)	(Madej et al., 2014, Sun et al., 2012)



In addition to the preparation of pure and active XylE and XylE-pGly28, which bind ligand, and with WT XylE appearing to transport xylose down its concentration/electrochemical gradient as determined using a novel enzyme linked XDH assay. I have also optimised of reconstitution methods to achieve relatively high and consistent reconstitution efficiencies is crucial to successful AFM experiments where data collection is limited by how much protein it can 'find' in the bilayer, see **Chapter 6** for more details.

Quantifying and reporting reconstitution efficiencies is particularly important to deducing how many lipids flank each protein, which may affect its unfolding trajectory and stability. For example, more lipids around a single protein may unfold under a higher force than a 2D crystallisation of a membrane protein with very few lipid molecules – an additional experiment to check phospholipid content in my reconstitutions would be useful to yield accurate lipid:protein ratios after quantifying protein in the sample using the Markwell-Lowry assay, rather than quoting an unusually high and possibly unrealistic ratio as has been reported in the past – I am unsure how much lipid is lost through the reconstitution process. Additionally, checking that the pre-swelling detergent for reconstitution has been removed is important to keep the bilayer 'tight', and to ensure the protein does not get pulled out as one molecule when a mechanical force is applied this also is not quoted in the literature.

### **5.3.2. XylE separate domain preparation and relevance**

Prior to experiments shown here on expressing XylE half domains, LacY of the MFS family has been expressed and purified in both an extract-based cell free system, where the N domain was shown to be stable when detached from the ribosome, but when attached, the ribosome destabilised the protein and was susceptible to protease degradation (Nagamori et al., 2003). When these half domains were expressed *in vivo*, the LacY halves were also shown to be unstable (Sahin-Toth et al., 1996, Bibi and Kaback, 1990, Wu et al., 1996).

Due to the increased intrinsic stability with XylE compared to LacY as shown by urea denaturation experiments (Harris et al., 2017a), I assessed the possibility to overexpress and purify the halves of XylE for downstream biophysical analysis. The N domain expression levels were very low, and most of the protein ended up eluting at the void. When run on SDS-PAGE, this domain appeared to run as oligomers, with a very small elution of monomeric N domain XylE - an amount too small for reconstitution to assess using standard biophysical techniques and AFM. These low expression levels suggest a significant degree of degradation *in vivo* when the C domain is not present as has previously been shown with LacY. The C domain did express

to much higher levels, and although SEC separation was poor, a higher concentration of C domain Xyle could be identified when the fractions under each peak were run on SDS-PAGE. The domain ran at 25 kDa, but with higher order structures, likely protein aggregates which I would expect to be removed in the gel-filtration step. It is likely that the C domain is not stable without the presence of the N domain first being expressed in the cell, and this makes *in vivo* expression and purification of half domains very difficult.

However, when these half domains were expressed in the PURExpress cell free system, void of any quality control machinery the domains were expressed well, and resistant to urea extraction and protease degradation, showing that it is possible to get these domains stability in a bilayer (Harris et al., 2020), unfortunately the yields produced here are likely too low for biophysical analysis or AFM experiments. The cell-free systems however do open avenues to study the folding cooperativity between transporter domains which cannot be accessed by expression *in vivo*.

Attempting to prepare independent halves of Xyle was initially designed for AFM experiments to determine whether the domains exhibit a degree of independent stability. However, expressing these domains *in vivo* is very low yielding, likely due to cellular degradation as their low stability makes them a target for quality control mechanisms (Marinko et al., 2019), and as such it was not possible to pursue these experiments.

### **5.3.3. Tagging Xyle for downstream single-molecule work**

It was possible to modify Xyle for downstream mechanical force unfolding experiments. An additional 28 amino acid linker was cloned to the C-terminal of the protein for AFM based SMFS (**Chapter 6**) and was able to be purified, could bind ligand to a similar affinity as WT Xyle and was similarly folded. It was also possible to clone an additional helix onto the N-terminal of Xyle to have the proteins terminals on one side of a membrane. This was further modified with an N-terminal HaloTag for surface chemistry tethering and could be altered to include an avi-tag on the C-terminal, which would be amenable to *in vivo* biotinylation (**chapter 3**) – however, this step is beyond the timeframe of my PhD. The modified Halo-TMH-Xyle construct was prepared to good quality and adequate yields for single-molecule work in both DDM detergent, and DIBMA native nanodiscs. This work will enable downstream single-molecule work for Xyle, or any other biophysical process which requires protein tethering at both ends.

#### 5.4. Chapter summary

- XylE and XylE-pGly28 can be purified to a high standard of purity and retains its secondary structure as determined by CD, which is comparable to literature values. Additionally, XylE proteins can bind xylose and glucose with a similar affinity to that shown in the literature in a detergent micelle.
- XylE can be reconstituted into 40:40:20 DOPE:DOPC:DOPC proteoliposomes with 50-80 % efficiency when aiming for high 12.5:1 wt:wt lipid:protein ratios, and 80-90 % efficiency when aiming for lower 33:1 wt:wt ratios. Detergent removal was shown to be very efficient leaving negligible amounts of OG/DDM present in the samples.
- Commercial XDH can be separated from XMR and encapsulated into proteoliposomes enabling a fluorescent readout of xylose transport. XylE downhill transport is observed, however showed high background fluorescence and further optimisation for proteoliposome clean-up, and maintenance of a 'tight' bilayer are necessary to achieve accurate transport data.
- It was possible to express and purify XylE with an additional N-terminal synthetic helix, and N-terminal HaloTag protein within DDM detergent and DIBMA directly from the native membrane for future MT experimentation.

## 6. Single-molecule mechanical unfolding of Xyle

---

*The work in this chapter was carried out in collaboration with Prof. Garcia-Manyes (Physics, KCL) with the assistance of Dr. Yongjian 'Jim' Wang whose training and assistance in AFM data collection, data analysis and fundamental physical principles has been invaluable for this project and I am extremely grateful for the assistance.*

*The project is being continued by another PhD student in the Booth and Garcia-Manyes groups*

## **6.1. Single-molecule spectroscopy techniques for membrane proteins**

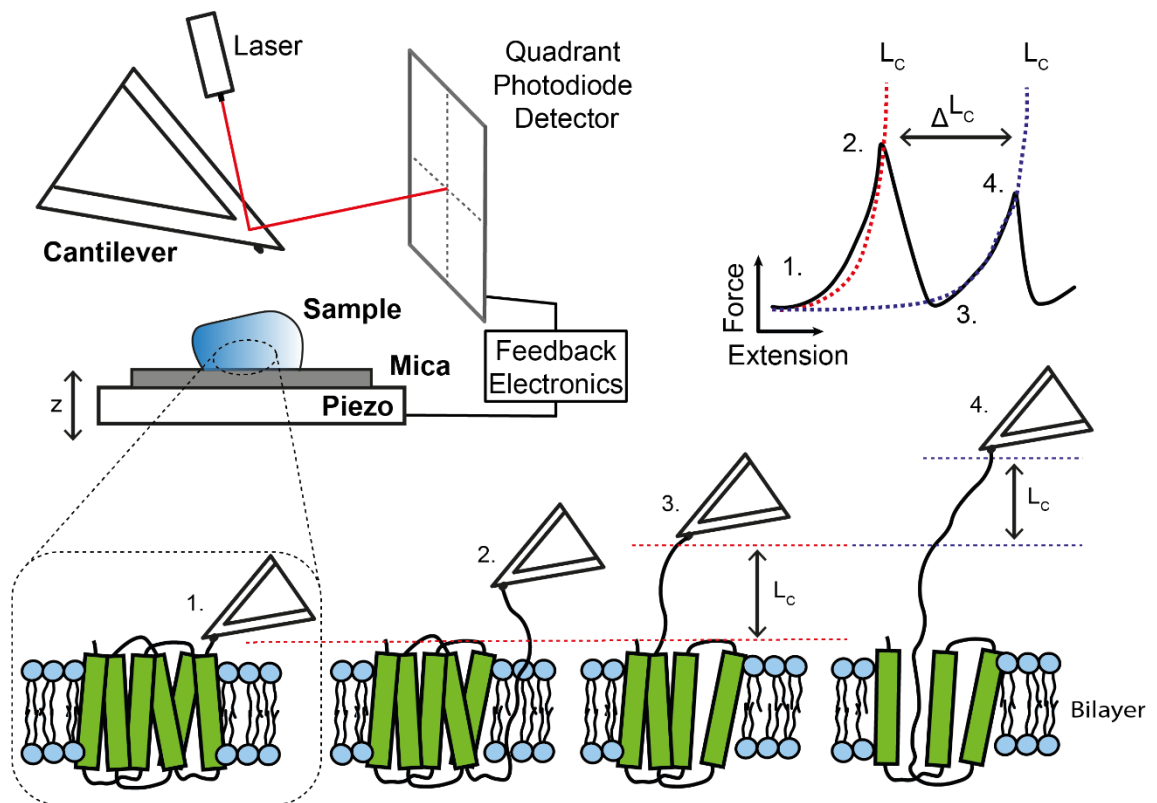
AFM is a powerful technique in biophysics. It can be used for sample imaging of anything from small molecules up to the cellular level, as well the mechanical manipulation of molecules with atomic resolution. The predominant use of AFM in biology is for imaging of surfaces – where the AFM measures the ‘roughness’ of a substrate, and nanoindentation, to work out mechanical properties of soft substrates (Hughes and Dougan, 2016). One particular use is for unfolding single molecules using force spectroscopy in an aqueous environment, or in air (Neuman and Nagy, 2008). SMFS allows us to define a reaction coordinate between an unfolded and folded state which cannot be assessed using typical biochemical thermal or chemical denaturant studies (Popa et al., 2013b, Borgia et al., 2008), thus providing an additional technique in physical biochemistry to study the unfolding of peptides.

Further details into physical properties and parameters of AFM setup, and models used to extract energetics from AFM experiments can be found in the following reviews: (Hughes and Dougan, 2016, Janovjak et al., 2008, Bippes and Muller, 2011, Mora et al., 2020).

### **6.1.1. Past AFM studies on membrane proteins**

The most common method to study membrane proteins using SMFS is force-extension (FX) AFM, where the attachment of an AFM tip to a protein, emersed within a lipid bilayer, is extracted at a constant velocity from the membrane to yield a characteristic stepwise unfolding trajectory as stable structural segments are unfolded. Unlike soluble polyproteins which unfold stochastically, membrane segments can only unfold after the previous segment as unfolded due to the additional stability provided by the bilayer environment. These unfolding events can be described as topologically shielded. A schematic of a typical AFM setup for membrane protein spectroscopy is shown in **Figure 6-1**.

Details of FX spectroscopy data collection, fitting, and methods to extract energetics can be found in **section 6.1.2**.



**Figure 6-1: AFM SMFS to (un)fold membrane proteins**

A typical AFM experiment set-up is shown. A membrane protein sample deposited on a piezo stage is approached by a sharp cantilever which scans the surface of the sample. The force generated/experienced deflects the cantilever which in turn is detected by a laser and photodiode detector. The laser voltage detected is converted into the distance deflected by the cantilever and the feedback electronics adjust the z-height of the piezo to maintain, or alter the force experienced by the sample. When the cantilever is retracted from the surface at a constant velocity, a bound membrane protein unfolds in a stepwise manner which is characteristic of the proteins' intrinsic stability. A 'saw-tooth' pattern of unfolding is generated, where each peak is fit to the worm-like chain model of polymer elasticity. The end-to-end length of the unfolded protein, or contour length ( $L_c$ ) is calculated from WLC for each structural segment which unfolds at the given pulling velocity. The  $L_c$  can be converted into the number of amino acids unfolded at each force peak for mapping of the protein stable segments.

The first experiments on mechanical unfolding of membrane proteins were carried out in the Gaub group on purple membranes of bR (Oesterhelt et al., 2000). The densely packed bR purple membranes were directly extracted from *Halobacterium salinarum* and deposited directly onto mica without the need for additional purification and reconstitution into synthetic liposomes. It was shown that bR helices G and F, and E and D always unfold pairwise, whereas the middle helices B and C can unfold as a pair, or separately using AFM SMFS (Oesterhelt et al., 2000). When extracting unfolding kinetics from each helical pair, the  $\Delta G_U$  for bR was calculated to be 290.5 kcal/mol (Muller et al., 2002, Cisneros et al., 2005, Oesterhelt et al., 2000), and for the common fold in halorhodopsin, 205.7 kcal/mol (Cisneros et al., 2005, Preiner et al., 2007). These values are much higher than the unfolding free-energies obtained using bulk-unfolding methods ( $\approx 20$  kcal/mol in detergent (Curnow and Booth, 2007) and  $\approx$

11.2 - 12.3 kcal/mol in bicelles (Chang and Bowie, 2014)), most likely due to unfolding into aqueous environment corresponding to first-stage folding, rather than tertiary structure formation (second stage folding).

Biochemical analysis on bR unfolding using denaturant revealed higher level transition intermediates (Curnow and Booth, 2007, Lu and Booth, 2000). The Perkins lab sought to understand these intermediates using ultra-short cantilevers which reduce thermal noise allowing the characterisation of the 'hidden' transition states of bR folding (Yu et al., 2017) at a 1  $\mu$ s temporal resolution (compared to a usual 50-100  $\mu$ s in other SFMS) (Neuman and Nagy, 2008). Data returned a 100-fold improvement over previous unfolding studies of bR using SMFS, showing intermediate states that refolded to, and were populated for < 10  $\mu$ s. In total, the ED helix pair showed 14 intermediates where only 2 were previously observed using standard AFM cantilevers, 7 intermediates for the CB pair instead of 2 were observed, and 3 intermediates when unfolding helix A instead of the 1 described previously (Yu et al., 2017, Bippes and Muller, 2011, Janovjak et al., 2003). The quasi-equilibrium refolding of these systems can be used to determine free-energy landscapes, which can be used as a comparison alongside other chemical methods of equilibrium refolding to gain a fuller picture of membrane protein folding.

After the seminal work on bR, the Müller group continued their studies on larger helical transporters; the first being the 12 TM spanning  $\text{Na}^+/\text{H}^+$  antiporter NhaA from *E. coli*. 2D membrane crystals of NhaA were deposited on mica and extracted from the bilayer yielding two unfolding families (Kedrov et al., 2004). It was later deduced that these two families were representative of unfolding from the N- and C-terminal of the protein. Later, dynamic force spectroscopy (DFS) was used to determine the free-energy of unfolding of NhaA, the energy landscape was reconstructed using the Hummer-Szabo model to yield a value of  $\Delta G_U = 485.0$  kcal/mol for the entire protein (Kedrov et al., 2006a). Others often quote specific unfolding stabilities for the structural segment which unfolds.

A summary of stabilities extracted for a selection of proteins using FX AFM is shown in **Table 6-1**.

**Table 6-1: Energetics extracted from AFM based SMFS on membrane proteins**

*\*In absence of ligand, displayed as single value for the entire proteins, or lowest and highest values for the average free energy of each unfolding intermediate. DFS model used affects the units presented here. Hummer-Szabo units are kcal/mol, whereas Bell-Evans-Richie derived units use multiples of  $k_B T$ .*

<b>Protein Family</b>	GPCR						MFS		APC
<b>Protein</b>	Bacteriorhodopsin	Halorhodopsin	Bovine rhodopsin	Mouse rhodopsin	Mouse opsin	Human $\beta 2$ adrenergic receptor	NhaA	LacY	SteT
<b>Number of helices</b>	7	7	7	7	7	7	12	12	12
<b>Mimetic</b>	Native membrane		Rod outer segment (ROS) disc membranes			DMPC:CHS liposomes	<i>E. coli</i> lipids		DMPC liposomes
<b>DFS model</b>	Hummer-Szabo		Bell-Evans-Richie				Hummer-Szabo	Bell-Evans-Richie	
<b>Stability by DFS: <math>\Delta G_U</math> (kcal/mol)*</b>	290.5	205.7	Unfolding segments of 22-28 $k_B T$	Unfolding segments of 22-25 $k_B T$	Unfolding segments of 20.6-24.5 $k_B T$	Unfolding segments of 20-23 $k_B T$	485	Unfolding segments of 19-23 $k_B T$	Unfolding segments of 15-18 $k_B T$
<b>References</b>	(Preiner et al., 2007, Oesterhelt et al., 2000)	(Preiner et al., 2007, Cisneros et al., 2005)	(Kawamura et al., 2010)	(Kawamura et al., 2010)	(Kawamura et al., 2013)	(Zocher et al., 2012a)	(Preiner et al., 2007, Kedrov et al., 2004)	(Serdiuk et al., 2014)	(Bippes et al., 2009)



LacY has also been extensively studied by the Müller group which is the first protein to be purified and reconstituted in a synthetic lipid bilayer for AFM experiments. The effect of  $\alpha$ NPG ligand binding to LacY drastically changed the kinetic, energetic, and mechanical properties of LacY in the N-terminal domain of the protein (Serdiuk et al., 2014). This was not observed in the C154G LacY mutant known to conformationally restrict LacY.

Additionally, LacY is known to invert its N-terminal in response to its surrounding lipid environment. This was first observed by the Dowhan lab with PE-deficient cell strains (Bogdanov and Dowhan, 2012). It was also possible to observe the inverted LacY structure using SMFS, whereby two distinct unfolding trajectories for LacY were observed in C-terminal unfolding WT protein, corresponding to the fully folded conformer, and the inverted N-terminal domain. This topological shift can be promoted by the addition of 3:1 POPE:POPG as opposed to bilayers consisting of POPG only (Serdiuk et al., 2015) – reducing the DOPE content increases the net-negative charge promoting a greater degree of domain flipping.

It was also possible to determine the effects of chaperone binding on LacY unfolding and refolding. Insertase YidC and LacY were co-reconstituted into synthetic liposomes and LacY was unfolded as previously shown. YidC reduced the energy landscape of LacY unfolding (Serdiuk et al., 2016, Serdiuk et al., 2017, Serdiuk et al., 2019), and when the fully unfolded LacY was moved towards an area of bilayer with YidC, the protein started to refold, which was not observed in the absence of YidC (Serdiuk et al., 2017). The refolded LacY protein took between 1-5 s to fully refold in 12 % of cases, which folded through 10 insertion and folding intermediates and was stable in the membrane.

A YidC-SecYEG fusion protein was also reconstituted into a lipid island to observe how they work together to insert LacY. It was shown that any one of the 10 LacY insertion steps could initiate refolding by both YidC and SecYEG (Serdiuk et al., 2019), much unlike the first helix which is observed in co-translational folding. The rest of the protein was inserted with preference for the C-terminal domain but helix insertion occurred in a random order when assisted by SecYEG only, but this was not the case with YidC (Serdiuk et al., 2016, Serdiuk et al., 2019), showing that proteins can use YidC or SecYEG for insertion purposes. SecY, however requires energy to insert TM helices post-translationally *in vivo* (as is the case in this system). This is usually provided by the ATPase function of SecA which pushes the nascent chain across the bilayer, often through, or in association with the SecY channel (Gold et al., 2007). These mechanisms are difficult to reconstitute into these single molecule systems and have not yet been included in these bottom-up single-molecule experimental approaches. Therefore, the

observations here likely come from non-specific interactions between TM helices and the exposed exterior of the insertase channel.

### **6.1.2. Force-extension spectroscopy for membrane proteins**

#### **6.1.2.1. Collecting and analysing force-extension data**

Proteins in a bilayer can be deposited on mica or glass substrates, and proteins embedded are able to adsorb to an AFM cantilever through non-specific interactions, or by using chemical or biological tags and modified cantilevers. Commonly, cantilevers with silicon nitride tips are used for these experiments and although the mechanism of interaction of the protein to Si<sub>3</sub>N<sub>4</sub> isn't known, it is assumed that the nature of interaction is electrostatic, and adsorption can be promoted by providing a larger force at the membrane surface (Hughes and Dougan, 2016). A schematic of the AFM set-up for membrane protein extraction is shown in **Figure 6-1**.

FX AFM approaches and retracts the cantilever tip to and from the sample surface at a constant velocity. Should the cantilever attach a protein, the protein is unfolded in a stepwise manner as the tip retracts. This results in saw-tooth 'fingerprint' trajectory of the protein **Figure 6-1**, as structurally unstable segments unfold. This occurs when the pulling force is greater than the stability of the interactions stabilising the protein and yields a measurement of force against the extension of the protein. The apex of each peak captures the force required to rupture each structurally stable segment, where the distance between two apexes corresponds to the extension of the amino acids (Hughes and Dougan, 2016, Jefferson et al., 2018) (where 1 AA  $\approx$  0.38 nm) (Mora et al., 2020) between two anchor points. This distance can be mapped to the number amino acids unfolded, giving a detailed analysis of the intrinsic properties of the membrane protein and its inherent stable structural segments – helices in the case of membrane proteins. The trajectory can be used to determine the anchor point of the tip to the membrane protein if the structure of the protein is known.

There are however drawbacks to FX spectroscopy technique itself; protein unfolding is fundamentally a probabilistic event with many multiple pathways that may co-exist, these pathways are often dependent on temperature (Janovjak et al., 2003), oligomerisation (Sapra et al., 2006) and the vector at which force is applied (Muller et al., 2006). Most AFM set-ups do not necessarily have full control over these features and so many (in excess of 100) trajectories of a protein unfolded in a certain condition must be collected, aligned and fit to find the most probable unfolding pathway. This is extremely difficult to achieve for membrane protein SMFS as protein pick-up rate is often very low, often due to low protein:lipid ratios

which are achieved when membrane proteins are purified and reconstituted into bilayers. To circumvent this issue, early studies on the unfolding of membrane proteins have used membrane extracts of purple bacteria, which contain densely packed rhodopsin protein with very low lipid content (Oesterhelt et al., 2000). 2D lipidic crystals reconstituted with proteins has also been trialled with success (Kedrov et al., 2004), but this is often protein dependent and can make sample preparation time consuming and very tricky.

#### 6.1.2.2. Fitting of force-extension data

Each saw-tooth force peak in an FX unfolding trajectory **Figure 6-1** can be fit to an array of different polymer physics models, enabling calculations of the contour length from the extension of the protein. The protein extension is dependent on the force applied to the protein and can be subsequently be used to map the unfolding segments onto the structure of the protein to determine structurally stable segments. The assignment of a structure is usually calculated using histograms of each contour length (more than 100 trajectories) to determine the most probable unfolding length of a particular segment (Bippes and Muller, 2011).

The most common and conventionally used polymer model is the worm-like chain model of polymer elasticity (WLC) **Equation 6-1**, where the force applied to the polymer ( $F$ ) is dependent on the product of the Boltzmann constant ( $k_B$ ) and absolute temperature of the system ( $T$ ), the extension ( $z$ ), the contour length of the polymer ( $L$ ), which defines the end-to-end length of the stretched polymer, and the persistence length ( $p$ ) which can vary between 0.25-0.6 nm in accordance with the average length of a fully stretched amino acid.

*Equation 6-1: WLC model of polymer elasticity*

$$F = \frac{K_B T}{p} \left[ \frac{1}{4 \left(1 - \frac{z}{L}\right)^2} - \frac{1}{4} + \frac{z}{L} \right]$$

The WLC model describes the behaviour of a semi-flexible polymer under an applied force (Bustamante et al., 1994, Muller et al., 2002). When there is an absence of force, the model assumes that the polymer will reside in a random collapsed state, however when force is applied there is a reduction of the number of possible configurations as the polymer is stretched, resulting in an entropic force which resists this external elongation.

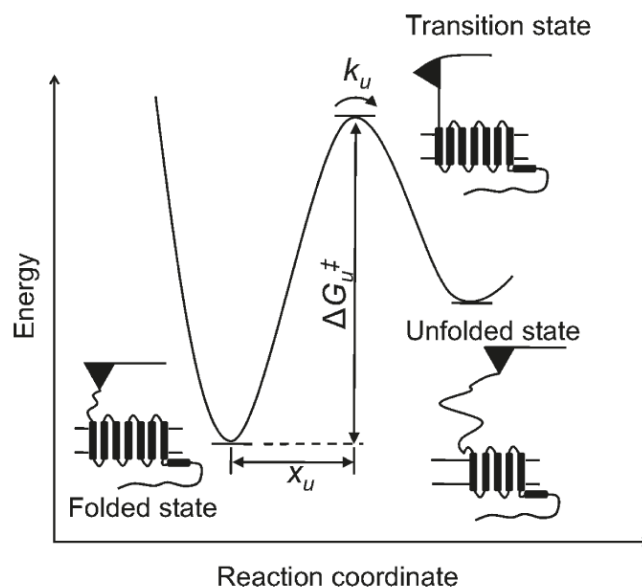
The WLC is not a perfect model of unfolded polypeptides. One limitation, for example, results in an overestimation of a monomer bond length when overstretched at forces higher than 60 pN (Bouchiat et al., 1999). Additionally, the model assumes that  $p$  is independent of extension,

this is a crude estimation as stiffness of the polymer will vary between a fully stretched helix and a fully folded helix in a bilayer (Su and Purohit, 2009). Additionally, the transition of a helix from a hydrophobic bilayer into a completely aqueous environment will also alter the mechanical properties of the peptide when extracted from the bilayer, as will the pulling of hydrophilic loop regions of the protein through the membrane. To counter this, several modifications to the WLC model have been considered and have been extensively reviewed in (Hughes and Dougan, 2016).

### 6.1.2.3. Extracting energetics from AFM experiments

A characteristic energy landscape for force-unfolding membrane proteins is shown in **Figure 6-2**. Mechanically unfolded proteins are acted on by two components, thermal and mechanical forces, which can only be decoupled when the same construct is pulled at different velocities. This is known as dynamic force spectroscopy (DFS). Several models can be used to fit DFS data (reviewed in (Hughes and Dougan, 2016)) enabling calculation of the parameters shown in **Figure 6-2**; the distance to the transition state ( $x_U$ ), and the rate of unfolding in the absence of force ( $k_U$ ). The height of the transition barrier, or Gibbs free energy of unfolding ( $\Delta G_U$ ) can then be calculated using the Arrhenius equation.

An example calculation of energy extraction using Bell-Evans-Richie is shown in **Appendix 7**.



**Figure 6-2: Unfolding energy barrier for a membrane protein**

*Idealised energy barrier profile for a stable structural segment within a membrane protein.  $x_U$  denotes distances between the folded and transition state,  $k_U$  denotes the rate of unfolding of the folded segment in the absence of applied force.  $\Delta G_U$  denotes the height of the transition state barrier. Diagram reproduced from (Kawamura et al., 2010).*

### 6.1.3. Force-Clamp spectroscopy

In addition to FX, force-clamp (FC) spectroscopy can be employed to unfold these proteins at a constant force, employing the use of a force-feedback-loop. This allows us to observe the stepwise 'stair-case' unfolding of these proteins and determine end-end distance of protein unfolding as a function of time (Mora et al., 2020, Garcia-Manyes et al., 2007). These step sizes correspond to the differences in  $L_c$  ( $\Delta L_c$ ) at a particular force across the FX saw-tooth trajectory. Using these techniques together allow us to define the unfolding of structural segments as a function of time, enabling the calculations of kinetics and thermodynamics of protein unfolding under a quasi-equilibrium, from which an energy landscape can be determined. Occasionally, entropic hopping between two transition states is observed and maybe indicative of reversible mechanical unfolding-refolding of a protein (Berkovich et al., 2010, Schönfelder et al., 2018). However, there are currently no published works on the FC unfolding of integral membrane proteins, although soluble polyproteins have been extensively studied and reviewed in (Garcia-Manyes et al., 2007) and (Mora et al., 2020).

Again, FC comes with some drawbacks. Once the piezo corrects the distance to achieve the constant user-defined force as the protein unfolds, there is a brief (ms) period where the protein experiences no force, which perhaps allows the protein to sample lower energy configurations as it begins to refold. There is also a significant amount of drift on the AFM piezo when forces are held for an extended time. This is also observable in the force-sensitivity of the cantilevers used in biological research, which must be soft enough to respond to forces relevant to protein unfolding events (Hughes and Dougan, 2016). FX spectroscopy also experiences this, but these issues can often be overcome using magnetic or optical tweezers which can operate at much lower loading rates.

Both AFM modes can also be used to refold the protein by holding the AFM tip close to the membrane allowing refolding in FX mode, or reducing the constant force to 0 pN near the membrane in FC mode. By modifying the protein environment, lipid compositions or temperature, substrate binding or functional mutants, mechanical manipulation of the protein can be used to gain insight into the protein stability.

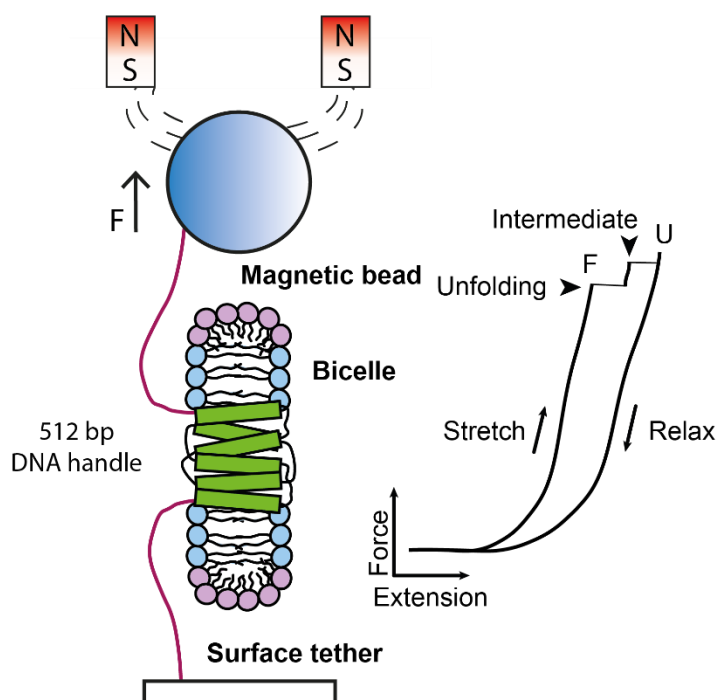
AFM SMFS has proved to be a versatile tool in characterising membrane proteins and folding **into** bilayers, it does however have limitations which need to be considered. AFM transfers the unfolded segments of the protein in to the surrounding aqueous environment, relating it more closely to the first stage of folding, as opposed to just distorting the tertiary structure of the

protein whilst it is embedded in the membrane, which would relate more closely to the second stage of folding (Jefferson et al., 2018), and therefore an alternative method of mechanical folding which maintains the protein in a membrane environment would be a preferred method to AFM high-force extraction of proteins from a bilayer – a magnetic tweezer system for example see **section 6.1.4**.

#### **6.1.4. Magnetic tweezer spectroscopy**

MT experiments have been used to investigate the mechanical properties of many globular proteins and have been paramount in probing mechanical unfolding kinetics and elucidating many mechanobiological mechanisms (Zhao et al., 2017, Popa et al., 2016, Le et al., 2016, del Rio et al., 2009). In a typical protein MT SMFS experiment, the molecule of interest is tethered between a superparamagnetic bead and the surface of a glass flow cell. Forces can be applied to the magnetic bead via a pair of permanent magnets positioned vertically above the flow cell. The excellent force sensitivity and stability of this instrument allows for long term, stable measurements, and superb low force resolution ( $<2$  pN) provided an appropriate tethering strategy for the molecule of interest is used (Zhao et al., 2017, Min et al., 2016). In MT SMFS experiments **Figure 6-3** a force can be applied along the plane of the membrane, allowing for folding measurements to be made whilst the protein is embedded in the membrane (Min et al., 2015, Min et al., 2018, Jefferson et al., 2018). These experiments follow on from methods developed in the Bustamante and Marquese labs (Cecconi et al., 2005, Cecconi et al., 2011).

### Magnetic tweezers



**Figure 6-3: MT setup to unfold membrane proteins**

*A MT set-up can unfold a membrane protein along the plane of the membrane when reconstituted into a bicelle. DNA nano-tethers bind the protein termini to a functionalised magnetic bead, and stage. When the pulling velocity is increased, the protein is stretched along the membrane and an unfolding intermediate is observed. This unfolding is reversible, and the protein refolds as the force is decreased.*

Lateral pulling is possibly more physiologically relevant for  $\alpha$ -helical membrane protein folding than AFM methods, with resolution down to much lower forces experienced within the cell. This denaturant free approach can characterise the unfolding of the protein laterally as the N- and C- terminal domains are separated by an external force. GlpG was shown to unfold in a cooperative manner with considerable hysteresis during force-ramp, followed by one-step unfolding event at 25 pN, characteristic to complete protein unfolding (Min et al., 2015). Subsequent refolding of the protein occurred at forces below 5 pN. It was also shown that the bicelle environment contributes to mechanical stability, as rupture forces of only 5 pN were measured when the bicelle environment was removed from the protein. This data was extrapolated to determine a thermodynamic stability of  $\Delta G \approx 6.54 k_B T$ , but a large unfolding barrier of  $\approx 21.3 k_B T$ , suggesting that although GlpG is not very thermodynamically stable, once folded, its half-life of unfolding ( $t_U = 3.5$  hours) is greater than the division time of *E. coli* and thus the organism would rarely see an unfolded GlpG protein.

In addition to cooperative unfolding in force-ramp conditions, four transient intermediates were noticed when the force was ramped and maintained at a pre-set force. At 21 pN, around 60 % of all trajectories showed no intermediates, another third showed one intermediate with extension of 10 nm, and in 8 % of trajectories, two intermediates were observed. Dwell times for trajectories with two intermediates were calculated and compared with the overall GlpG unfolding time, to this end, Bowie concluded that GlpG unfolds fully with only brief pausing in these positions.

The same system was used to study the unfolding pathways of two domains of the ClC Chloride transporter (Min et al., 2018), and concluded that the ClC-ec1 transporter could be separated into its two stable halves which could be unfolded independently when pulled from DMPC/CHAPSO bicelles. This agreed with the previously suggested hypothesis that the inverted topology of the two domains likely evolved from an ancient genome duplication where the domains folded separately and later fused. As with GlpG, 0.5 pN/s force-ramp experiments were carried out on ClC, and two large jumps of 55 nm extension were observed above 25 pN corresponding to complete unfolding of the two domains. Before this, a smaller pre-transition of 5 nm was observed, this was shown to be the two domains dissociating before the complete domain unfolding. This was confirmed by fusing an additional 16 amino acids added into the linker region and observing in same transition with addition 5 nm increase in extension.

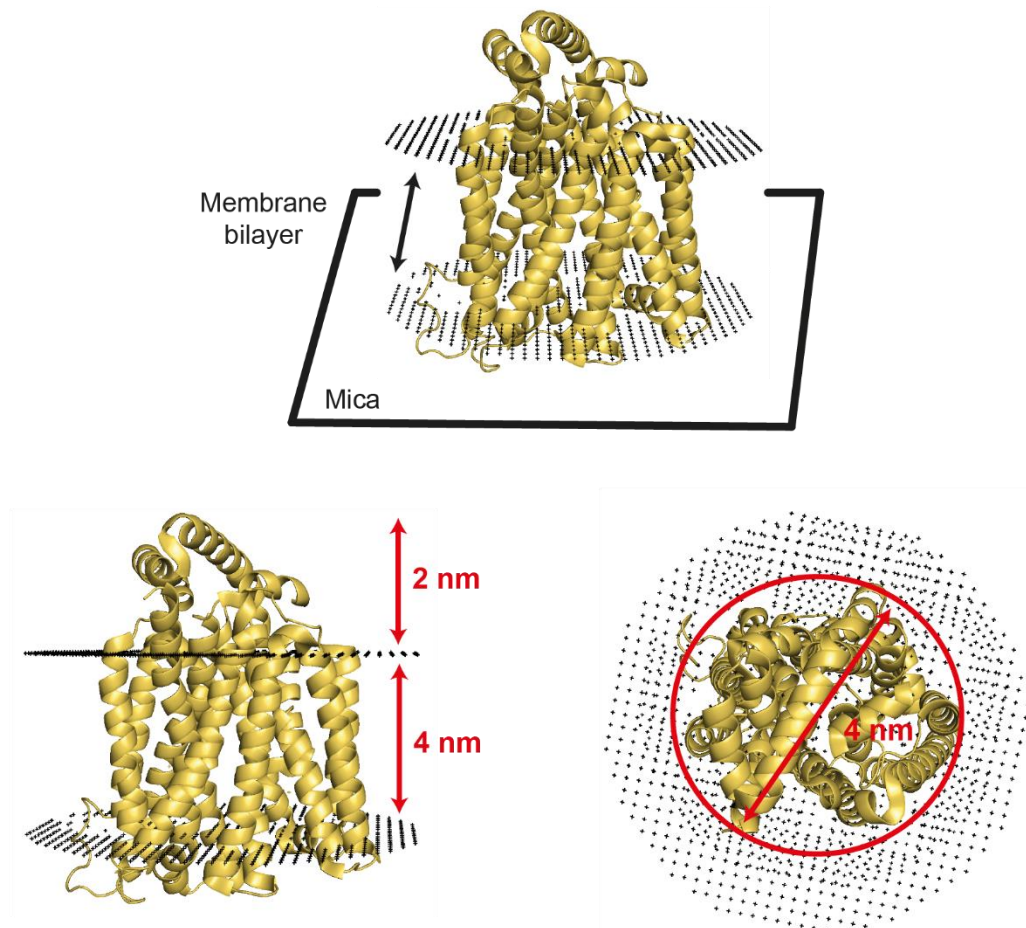
Similar experiments were carried out in vesicles of DMPC/DMPG on GlpG and the  $\beta$ -adrenergic receptor. Both proteins showed folding in a strict N- to C-terminal direction (Choi et al., 2019), and with a higher propensity of insertion using liposomes with intrinsic membrane curvature rather than bicelles. For GlpG, a free-energy difference of  $15.2 k_B T$  was calculated, which is larger than the estimates ( $13.9 k_B T$ ) for less-favourable folding conditions (Paslawski et al., 2015, Guo et al., 2016) as expected. In addition, liposome experiments showed a smaller  $\Delta G$  compared to earlier bicelle experiments ( $21.3 k_B T$ ) (Min et al., 2015), further suggesting that the planar bicelle provides a greater mechanical stability onto the protein.

### 6.1.5. Xyle as a model candidate for SMFS

Xyle was chosen as the candidate for SMFS due to its high stability and tolerance to chemical denaturants, and the relative ease of purification and biochemical characterisation (see **Chapter 5**). WT Xyle PDB structure 4GBY (Sun et al., 2012) is shown in **Figure 6-6**, with



calculated membrane boundaries (OPM (Lomize et al., 2012)) and PyMol measurements of different regions of the protein. The depth of the transmembrane region was calculated to be 4 nm, by measuring the vertical distance between the plotted membrane boundaries, the cross-sectional diameter of the protein was measured to be 4 nm. The intracytoplasmic helices (ICH) provide an additional 2 nm on top of the bilayer. These expected dimensions allow us to identify XylE in the bilayer when imaged using AFM.



**Figure 6-4: PyMol structure of XylE showing protein dimensions**

a) Schematic of XylE reconstituted in a bilayer and deposited on mica substrate, PDB: 4GBY with membrane boundaries plotted by OPM, labelled with PyMol measured distances of the transmembrane domain (4 nm), circumference from a top-down view (4 nm), and the membrane extruding ICH region (2 nm).

Unfolding XylE using AFM force-spectroscopy allows us to determine a fingerprint unfolding pathway for XylE and the corresponding rupture forces for the unfolding of each stable structural segment – does this correspond with previous observations that the MFS family and bR unfold as pairs of helices? XylE, like LacY, is a symmetrical protein which makes defining the cantilever binding site by unfolding trajectory very difficult. To counter this, an additional 28 amino acid linker was cloned onto the C-terminal of the protein, which provides an

additional  $\sim 10$  nm shift in  $L_C$  values, which are otherwise not detected when N-terminal tethered. This linker was based on previous AFM experiments using LacY and was shown not to affect activity of the protein (Serdiuk et al., 2014), as I have also confirmed for XylE in **Chapter 5**. XylE also reconstitutes into lipids to relatively high protein:lipid ratios, also shown in **Chapter 5**. High ratios are required to increase the probability of protein pick-up through the non-specific attachment to the AFM cantilever tip – more protein in the bilayer results in an increased pick-up rate. High reconstitution ratios, coupled with the additional poly-glycine linker is expected to increase protein pick up 100-fold compared to WT protein, as was previously shown with LacY (Serdiuk et al., 2014).

XylE also has a well-defined binding partner with its natural substrate D-xylose, and inhibitor D-glucose. Does D-xylose alter the XylE unfolding trajectory and rupture forces on binding? This interaction however is of much lower affinity binding  $k_D$ : 0.40 mM (**chapter 5**), compared to  $\alpha$ NPG with a  $k_D$ : 10  $\mu$ M (Harris et al., 2014), for WT XylE and LacY respectively. The latter being explored using AFM previously (Serdiuk et al., 2014).

Furthermore, I try to unfold this same protein at a constant force, using an appropriate force revealed in FX experiments. This will eventually allow us to observe the time resolved mechanical unfolding allowing the stepwise fingerprints of the protein to be determined, and ascertaining whether the two variants of spectroscopy match, so the group can pursue this technique further, leading to a calculation of unfolding rates and associated  $\Delta G$  of unfolding and refolding from lipid bilayers.

## 6.2. Aims of chapter

The aim of this chapter is to ascertain whether existing SMFS unfolding experiments on MFS proteins readily translate to XylE, and to gauge the magnitude of forces required to unfold the transporter. Is it possible to use XylE as a model protein to trial the SMFS folding of isolated protein using AFM? I also make the necessary adjustments to experimental setup and construct design as the group aims to move from studying isolated proteins in artificial systems, to unfolding membrane proteins mechanically from a more biological perspective with appropriate tethering strategies, leading to studies of more physiologically relevant forces associated with XylE folding when fully immersed in a bilayer environment, by using MT spectroscopy rather than AFM.

I have split the results section into two parts. Part A concerns itself with the preparation and optimisation of proteoliposome deposition conditions onto the AFM stage, which is necessary

for the collection of single-molecule unfolding data for XylE, which is consequently explored in Part B.

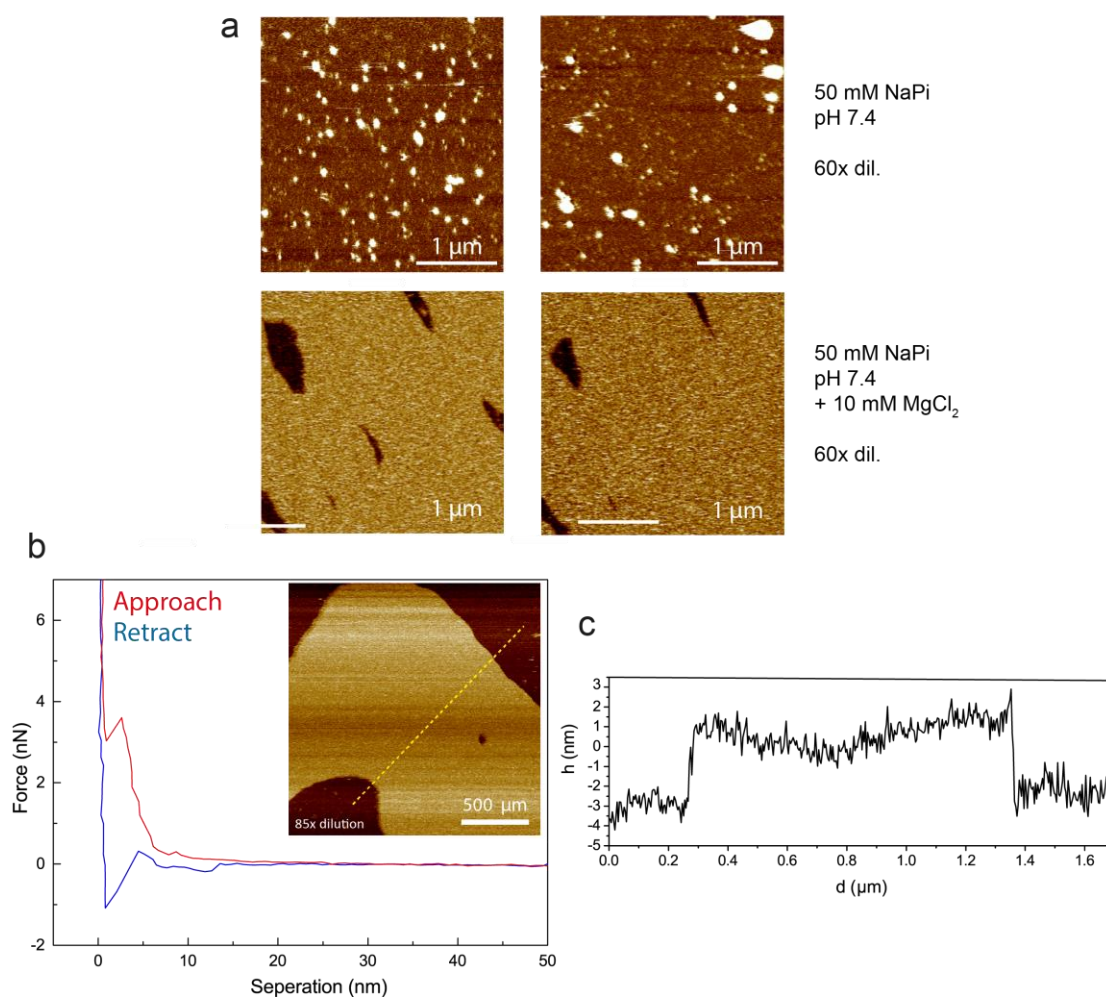
### **6.3. Part A: Bilayers, reconstitution, and discs for AFM**

#### **6.3.1. Determining mica deposition conditions using liposomes**

For AFM analysis of membrane proteins, it is necessary to deposit proteins reconstituted into synthetic liposomes onto a freshly cleaved mica substrate. Deposition of bilayers with and without protein required some optimisation to achieve a single flat bilayer with few holes and few unburst liposomes. To determine the best conditions to do this, empty bilayers of a 40:40:20 mix of DOPE, DOPC and anionic lipid DOPG were mixed to produce a net-negatively charged bilayer. These were deposited onto the mica in a magnesium rich buffer which binds the grooves of the mica and interacts with the negative lipid charge to create a homogenous flat bilayer across the surface of the substrate (Lind et al., 2015). **Figure 6-5a** shows the 40:40:20 DOPE:DOPC:DOPG mix deposited in the presence and absence of 10 mM MgCl<sub>2</sub>. In the presence of MgCl<sub>2</sub>, the bridging by Mg<sup>2+</sup> of the negatively charged membrane to the grooves of the mica substrate produced a smoother bilayer with fewer unburst liposomes and removes lipid-pooling of like species.

To ensure complete coverage of the mica surface whilst avoiding stacked bilayers and unburst liposomes. A 60x dilution of the 10 mg/mL lipid stock (0.17 mg/mL) appeared to yield the best coverage, and 30 min incubation was deemed best to allow time for liposome ‘splating’ whilst not letting bilayer stacking occur. However, mica coverage was often uneven across the complete 1 cm diameter of the substrate and dilutions ranged anywhere from 40x to 90x (0.25 – 0.11 mg/mL) across each experiment (with the 30 min incubation time remaining constant) due to variation in reconstitution quality and recovery for each sample (see **Chapter 5**).

To ensure the deposited bilayer was indeed a lipid bilayer after imaging, rather than an area of bare mica, a sharp tipped cantilever was used to pierce the bilayer at a forward velocity of 2 μm/s. This produced a force of 1.4 nN which travelled 1.27 nm in the approach curve - approximately half of the bilayer depth, and indicative of upper leaflet penetration of the bilayer. Additionally, measuring a slice through a bilayer island allowed determination of bilayer depth to be 4 nm using Nanoscope Analysis **Figure 6-5bc**.



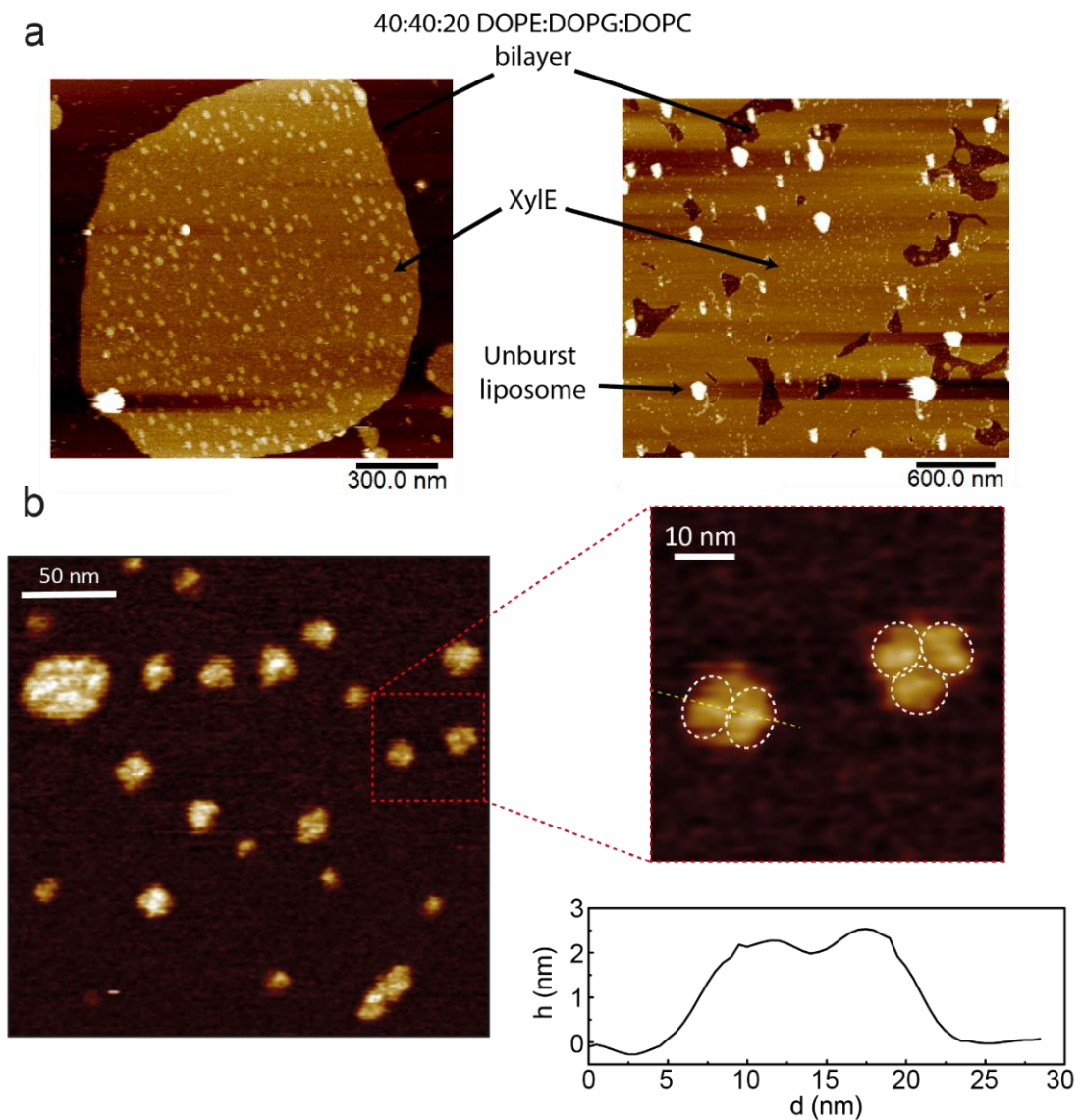
**Figure 6-5: Optimisation of liposome deposition conditions onto freshly cleaved mica**

a) Top: AFM Images of a  $3 \mu\text{m}^2$  area of a 40:40:20: ratio mix of DOPE:DOPC:DOPG liposome deposition in 50 mM sodium-phosphate buffer at pH 7.4, without (top) and with (bottom)  $\text{Mg}^{2+}$  containing buffer. Magnesium promotes formation of a more consistent bilayer with fewer unburst liposomes. Images were obtained in constant-tapping mode with SNL-B ( $k_s = 0.12 \text{ N/m}$ ) cantilever and 70 pN peak-force. b) Using AFM we are able to probe the surface of the lipid at a constant velocity of  $2 \mu\text{m/s}$ , this pierced the outer leaflet of the bilayer by 1.27 nm, with an approach force of 1.4 nN, c) The bilayer depth across a section of the bilayer island in panel b (yellow dashed line) was measured using Nanoscope analysis. This was determined to be 4 nm.

### 6.3.2. Imaging Xyle in 40:40:20 liposomes

Before starting SMFS experiments, we confirmed Xyle was present in the bilayers, and that the samples deposited as monolayers and were free from unburst liposomes and contaminants. Deposited proteoliposomes were imaged as per **section 6.3.1**, however lower dilutions were required, due to presumed lipid loss when reconstitutions underwent stringent washing protocols to remove incorrectly reconstituted proteins: 200 mM NaCl buffer dilutions, 4 M urea buffer dilutions or 60-15 % sucrose gradient floats.

Protein spots in the membrane were clearly visible as shown in **Figure 6-6a**, as were unburst liposomes. These however could be removed with further sample wash cycles at the AFM stage, or with a new deposition with a larger dilution factor. A smaller window was scanned on top of the protein rich region and individual proteins could be observed. These showed a membrane protrusion region of 2 nm as expected, **Figure 6-4**, and with a diameter of approximately 15 nm, which includes the width of the cantilever **Figure 6-6b**.



**Figure 6-6: High-resolution imaging of single Xyle proteins in bilayers**

a) High-resolution bilayer images obtained using ultrasharp Peakforce-HIRS-F-B cantilever ( $k_s = 0.05$  N/m) in constant tapping mode reveal Xyle proteins dispersed through the 40:40:20 DOPE:DOPG:DOPC bilayer. Unburst liposomes are also shown, as are bilayer islands with the bare mica behind. b) A smaller high-resolution area of a protein rich region was scanned, and individual proteins can be identified and measured. The height profile of the yellow dashed line is shown, each protein shows the 2 nm bilayer extrusion region of two single proteins as expected.

### 6.3.3. Improving AFM protein-pickup rate

Increasing protein pick-up in the following AFM experiments is crucial to increase data collection, allowing determination of the most probable unfolding pathways at different unfolding velocities to ultimately produce an extensive DFS study of protein folding. Attempts to improve protein pick-up for AFM SMFS were trialled in three ways: 1) increasing protein to lipid ratio by way of reconstitution protocols (**section 5.2.5**), 2) increasing protein-per-lipid area at the AFM stage using proteins embedded in nanodiscs, rather than bilayers (**section 6.3.4**), and 3) attaching a longer linker to the end of the protein to improve probability of tip attachment.

In addition to removing symmetry from the structure of Xyle to determine its cantilever binding point (see **section 6.4.2**), the 28 amino acid poly-glycine based linker (termed pGly28) based on the work by Serdiuk, et al (Serdiuk et al., 2014, Serdiuk et al., 2015, Serdiuk et al., 2016, Serdiuk et al., 2017, Serdiuk et al., 2019), was incorporated between the ICH<sub>4</sub> helix and the His<sub>10</sub> affinity tag on the C-terminal of Xyle in an attempt to increase the pick-up rate due to a higher portion of protein being accessible to the tip.

The total trajectory count for the duration of this project was in the region of 80,000. Of these, the tip only adhered to protein 223 times equating 0.28 %, this is more than twice the pick-up quoted in the literature for LacY, at 0.1 % (Serdiuk et al., 2014). The remaining traces were blank meaning no protein was extracted. 48,000 trajectories were attempted for WT protein, whereas 32,000 were collected for Xyle-pGly28. However, from WT to pGly28 there was only a marginal pickup increase from 0.24 % to 0.26 %, much unlike the expected 10-fold increase (0.01 % - 0.1 %) as suggested for LacY in the literature (Serdiuk et al., 2014, Serdiuk et al., 2015). In addition, my refined reconstitution strategy (see **Section 5.2.5**) and removal of poor-quality protein had already yielded 0.24 % pick up without the linker, which is considerably higher than quoted literature values. The addition of 5 mM D-xylose to the buffer showed a pickup rate of 0.34 % of a total of 6500 attempted traces. Data collection however remained very slow due to limited machine hours, and an extensive study was not possible in my remaining PhD time.

Of our 228 pickups, only 58 trajectories were able to be assigned to one of three groups based on the pGly28 addition; C-terminal pick up (n = 38), C-terminal Xyle-pGly28 pick up (n = 10) or N-terminal pick up (n = 10). And for D-xylose binding, 8 trajectories could be assigned, 5 for C-terminal binding and 3 for N-terminal binding. The remaining traces were omitted as it

was not clear whether they were N- or C-terminal without the entire unfolding trajectory due to early tip-detachment.

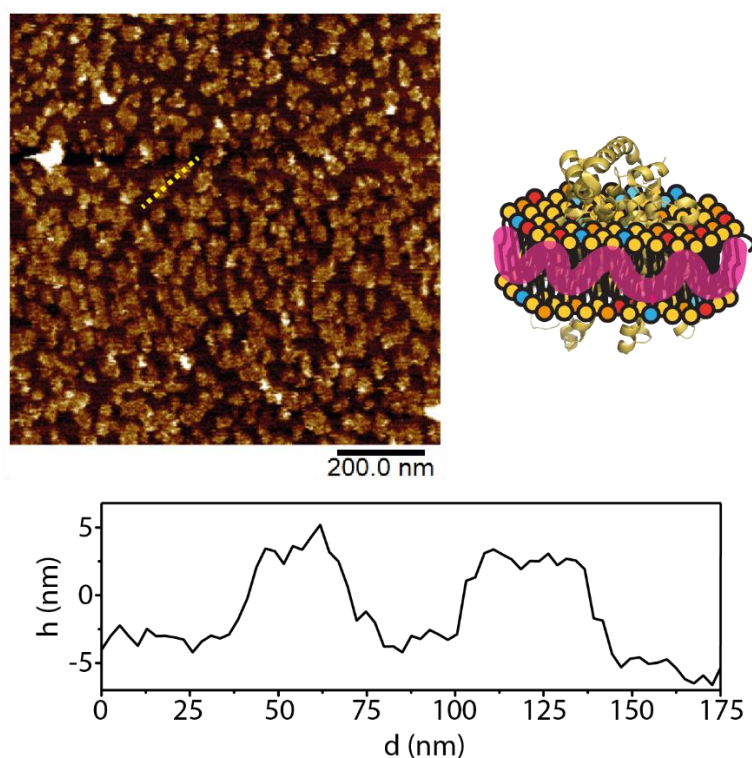
Even after attempting to increase pick-up probability, the N- and C-terminals of XylE remain on the same side of the membrane, and it is unclear which topology the protein is in the membrane after the liposomes are deposited. Depending on how the liposomes 'splat' – it is likely that I lose approximately 50 % of 'available' protein to dual-topology in the membrane, resulting in 50 % of the terminals not being exposed to the cantilever.

#### **6.3.4. Polymer nanodiscs for AFM**

DIBMA based nanodiscs explored in **Chapter 4** were used in an attempt to achieve a higher protein-per-lipid area, which removes the excess empty lipids which is more likely in a bilayer environment to leave more protein in the  $1 \mu\text{m}^2$  area we are searching in. XylE nanodiscs were purified from the same synthetic DOPC:DOPC:DOPG lipid bilayer, to maintain a similar environment as bilayer unfolding, by using affinity chromatography to remove excess lipid and empty discs before deposition.

**Figure 6-7** shows a high-resolution AFM image of the deposited XylE reconstituted DIBMA discs. The depth along the yellow dashed line shows showed the heights of two individual discs of 6.76 nm and 7.09 nm which showed an additional  $\approx 3$  nm on top of the empty bilayers suggesting that protein is present within the discs, which may promote a conformation where the ICH helices extend upwards, accounting for the additional 1 nm observed with is not present in bilayers. The diameter of these discs was determined to be 18.2 nm as judged by DLS (**section 4.2.3**) which is in conclusion with the AFM image presented here, although the width of objects under AFM cannot accurately be obtained due to the additional width of the AFM tip used, and the drift of the piezo.





**Figure 6-7:** AFM high-resolution image of WT Xyle reconstituted in DIBMA nanodiscs

$1 \mu\text{m}^2$  AFM image of Xyle in DIBMA 40:40:20 DOPE:DOPC:DOPG nanodiscs (schematic shown). Images were obtained using ultrasharp Peakforce-HIRS-F-B cantilever ( $k_s = 0.05 \text{ N/m}$ ,  $f_{o,fluid} = 60 \text{ kHz}$ ) in constant tapping mode. The height of two discs were determined to be 6.76 nm and 7.09 nm from left to right along the yellow dashed line.

## 6.4. Part B: Unfolding Xyle

### 6.4.1. Mapping the expected unfolding pathway of Xyle

Based on past experiments (Kedrov et al., 2004, Serdiuk et al., 2014, Oesterhelt et al., 2000) we can map the theoretical unfolding pathway of Xyle by assuming it follows the characteristic helix-pair unfolding pathway shown by similar proteins. The total length of the Xyle construct is 503 amino acids, equating to a total extension of 191.14 nm ( $1 \text{ AA} = 0.38 \text{ nm}$ ) – this includes the WT protein, and a His-10 tag at the C-terminal. Using a simple schematic of Xyle **Figure 6-8**, which shows the 12 major TM helices, and 4 extra-membranous helices, and assuming each helix pair is a stable structural segment, I can map the unfolding segments when extracted from the C-terminal under constant velocity to 5 stable segments, whereby the first pair of helices (11 and 12) unfold, followed by the second (9 and 10), the third (7 and 8) including the ICH domains, the next two helices of the N-terminal domain (5 and 6), and the middle two (3 and 4), leaving helices 1 and 2 in the bilayer which were unlikely to be detected in unfolding as these are likely to be freely suspended in buffer once the rest of the protein has



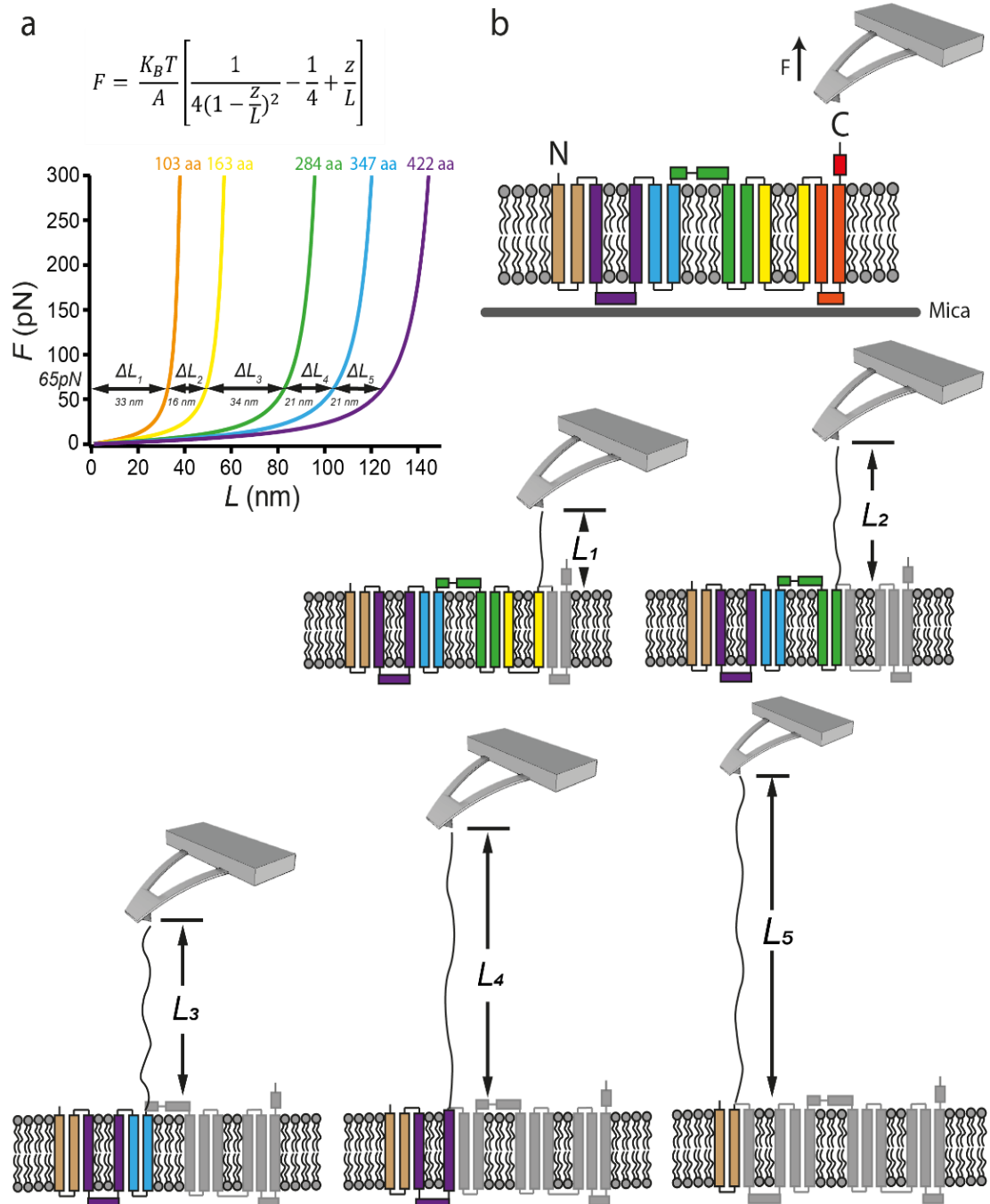
unfolded. The following predictions make assignment of the protein easier – and allow us to determine the cantilever binding terminal for Xyle.

The 4GBY structure suggests that very little of the N-terminal protein is exposed from the membrane, whereas the C-terminal His<sub>10</sub> tag and ICH4 helix are exposed and thus the non-specific covalent interaction of the tip is expected to bind to the C-terminal region of the protein. This was later shown to be the case with the addition of the poly-glycine tail being captured (see **section 6.4.2**). Assuming the tip binds to the final histidine residue of the sequence, I predict that the five major force peaks will occur after the unfolding of 103, 163, 281, 347 and 422 amino acids shown by the series of schematics **Figure 6-8ab**. A complete list of expected unfolding  $L_C$  and corresponding amino acids is displayed in **Appendix 6**.

To calculate predicted contour lengths for each of these force peaks, the number of amino acids unfolded was multiplied by the average length of a single stretched amino acids of 0.38 nm, and these were calculated to be:  $L_{C1} = 39$  nm,  $L_{C2} = 62$  nm,  $L_{C3} = 107$  nm,  $L_{C4} = 132$  nm,  $L_{C5} = 160$  nm. Using these contour lengths, we can simulate expected fits of the WLC model, which were generated using the above  $L_C$  values, a  $p$  of 0.6 nm\*, and displacements of 1 nm. These fits are shown in **Figure 6-8a**, which aid in the designation of force-clamp data in **section 6.4.7**, as the  $L_C$  is dependent on force, and this force is determined by experimental FX data.

*\*The persistence length is a measure of polymer stiffness. For flexible regions of polymer,  $p$  is low, and for more rigid polymers  $p$  will be larger as the elastic properties are maintained through a larger distance along the polymer chain. A persistence length of 0.6 nm, rather than the usual 0.4 nm (Serdiuk et al., 2014), was used in these experiments as this appeared to be a better fit for the curve of experimental unfolding trajectories shown in throughout this chapter. This difference likely arises from the limitations of the WLC, briefly discussed in **Section 6.1.2.2**.*

When a constant pulling force of 65 pN is applied, the theoretical differences in contour length ( $\Delta L_C$ ) correspond to the length of a fully extended stable segment (predicted here to be a pair of helices). Using the generated WLC fit at 65 pN, I estimate these to be  $\Delta L_1 = 33$  nm,  $\Delta L_2 = 16$  nm,  $\Delta L_3 = 34$  nm,  $\Delta L_4 = 21$  nm,  $\Delta L_5 = 21$  nm when extracted from the C-terminal end of the protein, and these lengths should be observed in experimental FC data (see **section 6.4.7**) should the protein unfold we would expect.

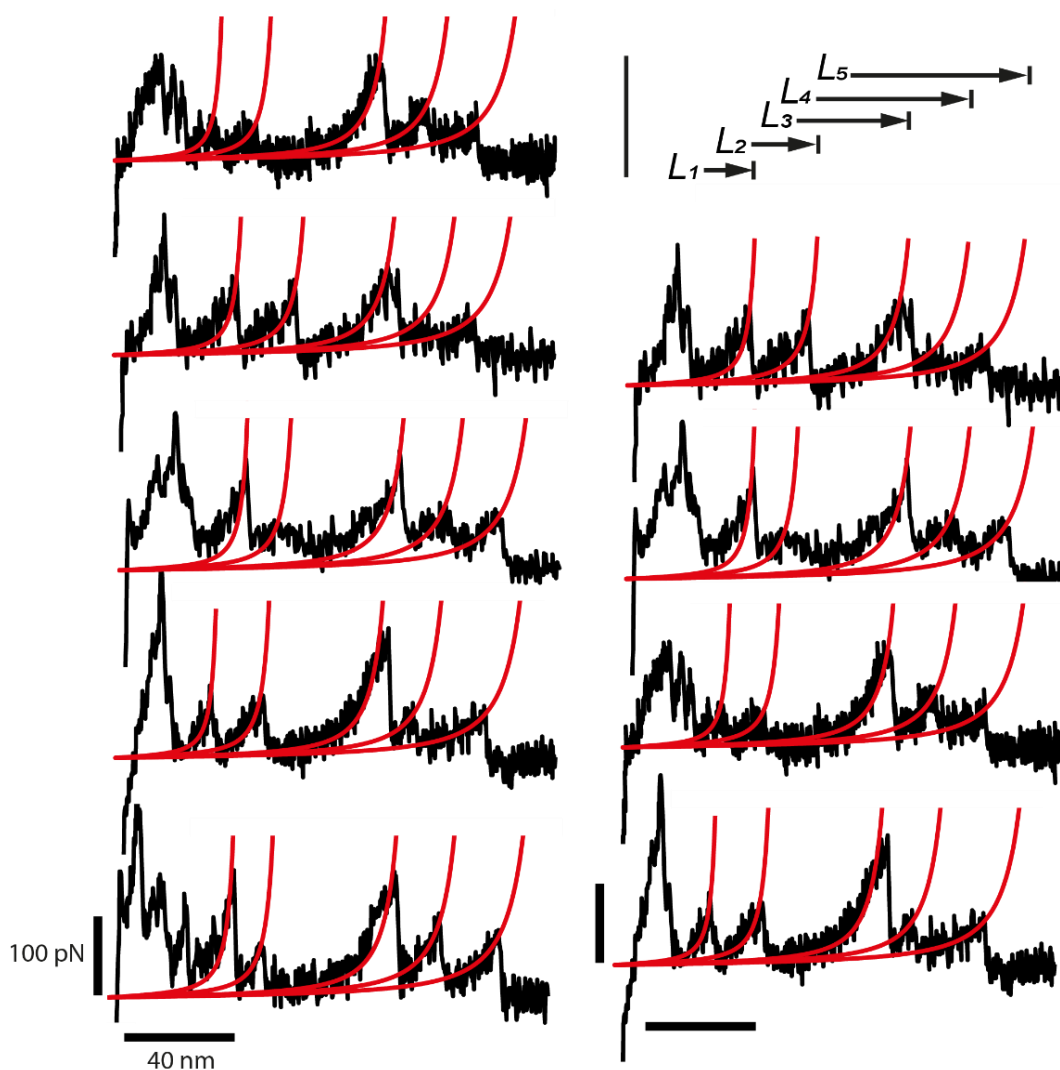


**Figure 6-8: Theoretical unfolding of WT XylE from C-terminal cantilever binding**

a) Schematic of experimental set-up, where XylE is reconstituted into a bilayer environment and deposited onto a cleaved mica substrate. We assume the AFM induces an upward force on the C-terminal of the protein, causing a concomitant unfolding of the protein in five distinct stages. We predict the five distinct stages of unfolding based on the sequence of XylE, and past work which suggests  $\alpha$ -helical proteins unfold pairwise. Each helix pair (coloured in orange, yellow, green, blue, purple, and gold) unfold as single structural segments and the number of amino acids unfolded can be determined from the structure and mapped onto a schematic of the protein. The final two helices (gold) cannot be detected as the protein either does not fully unfold and extract from the bilayer, or the helices are suspended freely suspended in buffer. WLC fits for the number of amino acids at each force peak can be converted into an  $L_C$  value, and used with persistence length of 0.6 nm to generate WLC fits. The  $L_C$  is also dependent on force applied. For 65 pN (the force used experimentally), the  $\Delta L_C$  values are calculated as the following:  $\Delta L_1 = 33$  nm,  $\Delta L_2 = 16$  nm,  $\Delta L_3 = 34$  nm,  $\Delta L_4 = 21$  nm and  $\Delta L_5 = 21$  nm which will aid in assignment of FC data in **section 6.4.7**.

#### 6.4.2. Defining the Xyle-tip anchor point using FX SMFS

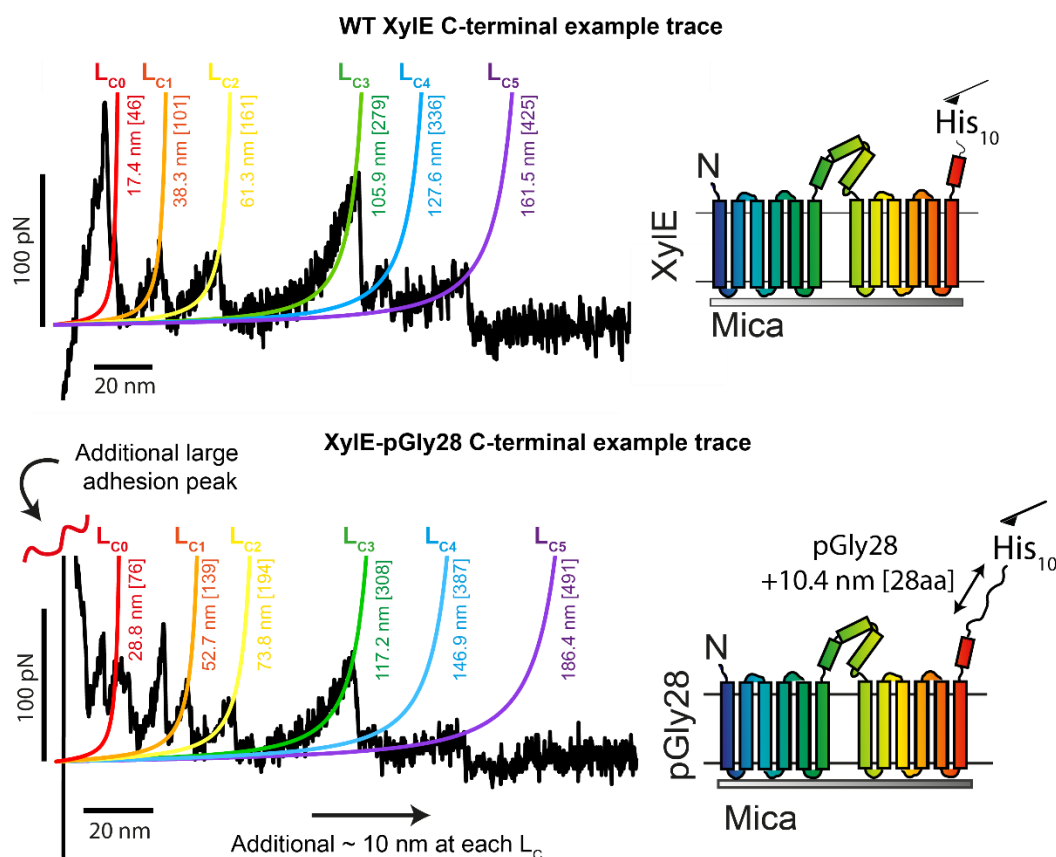
40:40:20 DOPE:DOPC:DOPG proteoliposomes with WT Xyle were deposited onto the mica substrate and imaged to find protein rich regions as shown in **section 6.3.2**. If the tip found a protein and non-specifically bound, the protein was extracted and displayed a characteristic trajectory of various force peaks. A selection of traces obtained are shown in **Figure 6-9**. Each trace is fitted with the WLC model at 5 of its major force peaks. As Xyle is a symmetrical protein, it was not clear at this stage which terminal the cantilever binds too, although the 4GBY crystal structure suggests that the N-terminal is barely exposed from the membrane, and therefore would likely not attach.



*Figure 6-9: Example traces from WT Xyle corresponding to N- or C-terminal binding*

*Each force peak is fit to WLC model using  $p = 0.6$  nm for the five major unfolding peaks. It is unclear at this stage which terminal was tethered to the AFM tip as Xyle is a symmetrical protein and shows a very similar unfolding trajectory when pulled from either side. Scale bars for 100 pN (vertical) and 40 nm (horizontal) scale are shown and is consistent for each trajectory.*

The pGly28 linker based on previous literature (Serdiuk et al., 2019, Serdiuk et al., 2015, Serdiuk et al., 2017, Serdiuk et al., 2014, Serdiuk et al., 2016) is used to improve pickup rate. However, it also yields an additional 10.36 nm extension which should be observed in all unfolding trajectories if C-terminal bound; all  $L_C$  values calculated were at least 10 nm larger when C-terminal bound than when N-terminal bound allowing us to assign each of the trajectories to one of three groups, C-terminal WT, C-term pGly28 or N-terminal. Two example traces for C-terminal extraction are annotated and  $L_C$  values for major force peaks identified are summarised in **Figure 6-10**.



**Figure 6-10: Summary of trajectory assignment for C-terminal Xyle binding using Xyle-pGly28**

Two characteristic traces were fit to WLC to determine the major unfolding  $L_C$  values. The number of amino acids for each  $L_C$  are shown in square brackets. Xyle pGly28 shows additional peaks which are lost in the noise of the initial adhesion peak resulting from the cantilever being near the membrane. In addition, the pGly28 samples show a much larger initial adhesion peak, omitted here for clarity. For this trajectory, the initial peak was more than 1 nN. The contour lengths shown increase by at least 10 nm which the cantilever attaches pGly28 confirming C-terminal binding allowing assignment of N- and C-terminal attachment. Additional peak fits for pGly28 which do not have corresponding WT peak are also omitted here for clarity, but values of these  $L_C$  values are found in **Table 6-2**. Any traces which did not show these characteristics were determined to be N-terminal bound, where N terminal binding WT and Xyle-pGly28 were almost identical in trajectory.

The major assignment arises from an additional large force adhesion event at the beginning of each pGly28 trace, likely due to the additional amino acids interacting with the surface of the membrane. pGly28 also showed additional peaks which were not present in the C-terminal WT binding trajectory and were likely lost in thermal noise which occurs when the tip is held at less than 15 nm from the membrane surface. The identifiable peaks corresponding to WT protein are consistent with subsequent force peaks unfolding with a  $\approx 10$  nm minimum increase at each  $L_C$ , compared to WT XylE. This shows that the cantilever binds to the end of the pGly28 linker on the C-terminal, rather than the N-terminal when compared to WT XylE. A summary of the calculated major  $L_C$  values for the example traces, including minor peaks are shown in **Table 6-2**, with comparison to the theoretical unfolding peaks from **section 6.4.1**. Any traces which did not show characteristic adhesions or extensions were determined to be N-terminal binding, where WT and XylE-pGly28 were almost identical in trajectory.

**Table 6-2: Comparison of major peak  $L_C$  between C-terminally bound WT and pGly28 XylE**

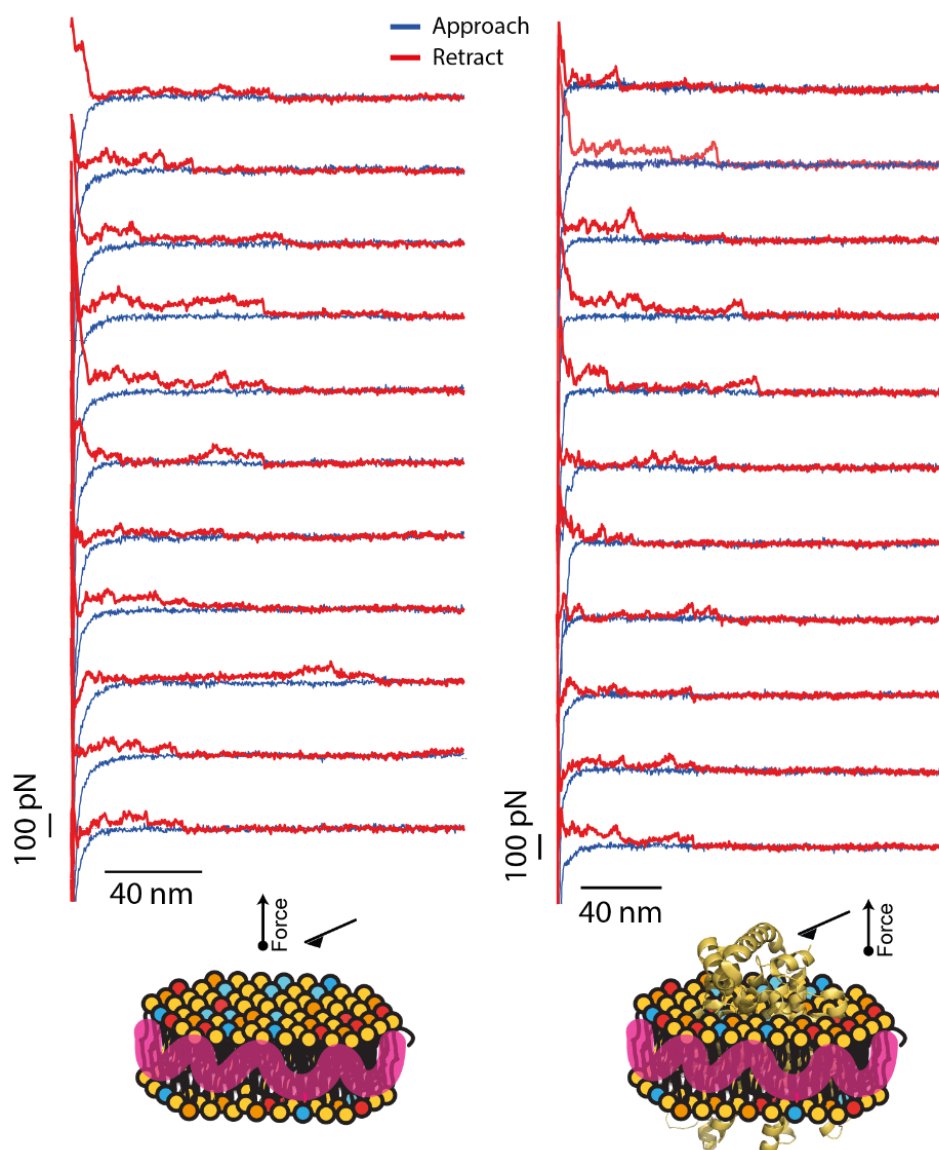
*\*equivalent peaks not observed, most likely hidden by thermal noise at the surface of the membrane*

Peaks identified	Theoretical Wt C-terminal bound $L_C$ [aa] (section 6.4.1)	Example WT trace C-terminal bound $L_C$ [aa] (nm)	Example pGly28 trace C-terminal bound $L_C$ [aa] (nm)	Difference in $L_C$ due to pGly28 ( $\approx 10.4$ nm)
pGly28 minor	-	N/A*	18.2 [48]	-
$L_{C0}$ (His <sub>10</sub> -ICH <sub>4</sub> )	-	17.4 [46]	28.8 [76]	11.4
minor	-	22.1 [58]	38.8 [102]	17.7
pGly28 minor	-	N/A	41.2 [108]	-
$L_{C1}$ (H12-11)	39 [103]	38.3 [101]	52.7 [139]	14.4
$L_{C2}$ (H10-9)	62 [163]	61.3 [161]	73.8 [194]	12.5
$L_{C3}$ (H8-7-ICH)	107 [282]	105.9 [279]	117.2 [308]	11.3
$L_{C4}$ (H6-5)	132 [347]	127.6 [336]	146.9 [387]	19.3
$L_{C5}$ (H4-3)	160 [421]	161.5 [425]	186.4 [491]	24.9

A lot of the time, the pick-ups from non-specific attachment yield only one or two unfolding force peaks which made it very difficult to assign to directionality of extraction or whether the cantilever was binding a contaminant or even lipids in the sample. This occurred in  $\approx 75$  % of all adhesions and the trajectories were omitted from analysis. Occasionally, trajectories that unfold past 200 nm are observed ( $\approx 5$  % of all adhesions) which must be due to a somewhat strong binding between protein oligomers or aggregates. The entire construct is only 503 AA equating to an extension of 191.14 nm as a maximum, and lipids will most likely not interact with the cantilever and therefore must correspond to unfolding proteinaceous material in the sample.

### 6.4.3. XylE and Halo-TMH-XylE in DIBMA nanodiscs

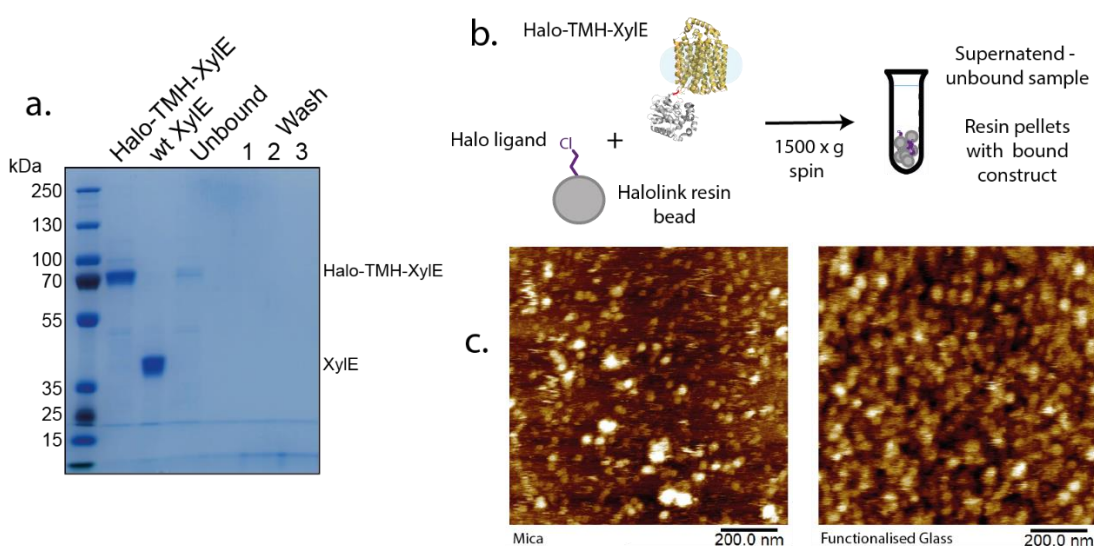
Subjecting both empty and XylE reconstituted DIBMA nanodiscs to FX spectroscopy yielded a similar trace summary. Although polymer use did increase ‘pick-up’, I could not be sure whether each trace corresponded to a non-specific interaction and unfolding of protein, or an interaction with the polymer itself, which was unexpected. A selection of trajectories observed for empty and reconstituted discs are shown in **Figure 6-11**. As there were obvious similarities between the protein-reconstituted discs and the empty discs, it was not possible to continue with data collection using native nanodiscs.



**Figure 6-11: Empty and WT XylE reconstituted DIBMA trace comparison**

An example 11 FX traces from empty DIBMA nanodiscs and XylE reconstituted nanodiscs are shown with approach (blue) and retract (red) curves of 400 nm/s. There are no clear XylE unfolding peaks in either set, suggesting that DIBMA also binds non-specifically with the  $\text{Si}_3\text{N}_4$  cantilever.

DIBMA purified samples of Halo-TMH-XylE were also checked for purity and function using AFM as a precursor for MT studies. **Figure 6-12a** shows an example Coomassie stained biochemical binding assay SDS-PAGE for the Halo-TMH-XylE construct with the experimental scheme shown in **Figure 6-12b**. The HaloLink resin beads, impregnated with O4 ligand, were incubated with the construct purified in DDM. This confirmed Halo can bind its ligand with a disappearance of protein from sample loaded to unbound, suggesting bead binding. Furthermore, this could not be eluted with buffer – mimicking the AFM stage and MT flow-cell washes. I next tested my sample on a mica substrate alongside a functionalised glass substrate **Figure 6-12c**. Both the mica and glass were able to bind my construct, this time purified in DIBMA native nanodiscs, the glass however is more uniform with fewer large patches where the DIBMA has not solubilised the native membrane correctly. I attempted some FX SMFS with non-specific attachment to the C-terminal end of the construct, however DIBMA interactions prevented this, as discussed above.



**Figure 6-12: Assaying the tethering of HaloTag to halo-ligand of DIBMA native nanodisc Halo-TMH-XylE**

a) Coomassie gel of Halo-TMH-XylE and WT XylE purified in DDM detergent. Binding the Halo construct to Halo-link resin beads removes the sample from the flow through – termed unbound here. The beads were then washed with 1 mL buffer 3 times. b) Schematic showing the construct binding the halo-link beads which can be harvested by a brief centrifugation step which in 100 % binding efficiency would remove the protein from the supernatant. c) Initial AFM imaging of the construct in DIBMA nanodiscs deposited on both a mica and a functionalised glass substrate. A 500 nm<sup>2</sup> area was scanned using HIRS-FB Si<sub>3</sub>N<sub>4</sub> cantilever ( $k_s$ : 0.05 N/m,  $f_{o\text{ fluid}}$ : 60 kHz). The glass (left) tethered shows a much cleaner sample with few large ‘dirty’ areas than the non-specific mica (right) interaction.



#### 6.4.4. Unfolding Xyle from the C-terminal

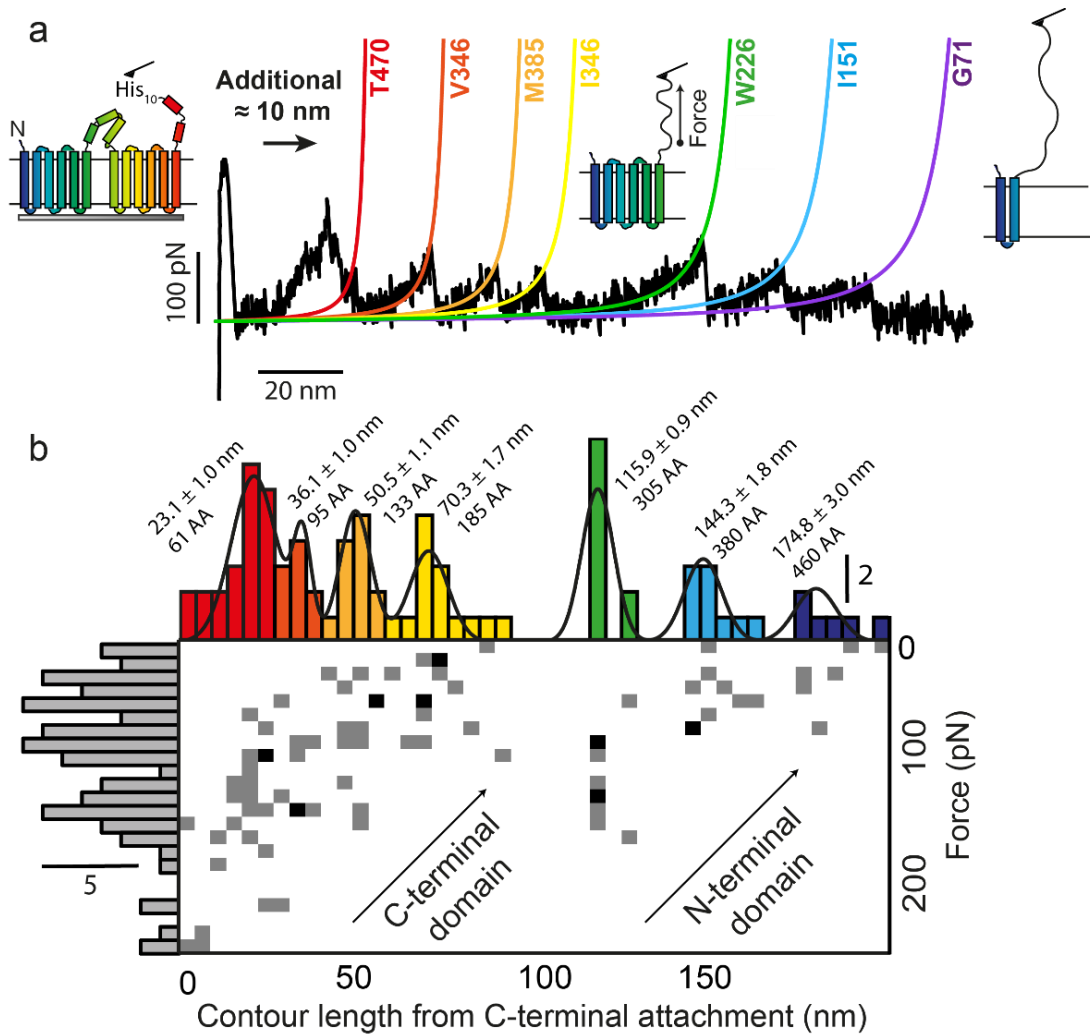
Looking at single trajectories, while informative when determining cantilever attachment sites and major force peaks, do not encompass every intermediate that may be found in a second trajectory for the same protein. Protein unfolding is a probabilistic process, meaning they many trajectories must be obtained, and summed to determine the central contour lengths for each major peak using Gaussian fits of histogram data. To each of the assigned traces, the  $L_C$  values of every determined peak, and corresponding rupture forces were used to produce bivariate histograms for each unfolding family; C-terminal binding WT and pGly28, and N-terminal binding. The resulting  $L_C$  values were subsequently used to calculate the stable structural segments of Xyle and map these onto the protein structure to determine the residue anchor points.

C-terminal binding histograms for Xyle-pGly28 ( $n = 10$ ) and WT Xyle ( $n = 38$ ) are shown in **Figure 6-13** and **Figure 6-14** respectively, each alongside a characteristic trace showing the major unfolding peaks. For Xyle-pGly28 averaging the data produced 7 major intermediates, unlike WT Xyle which produced 6 major intermediates, which are summarised in **Table 6-3**, with their resulting stable segments calculated, alongside the expected unfolding lengths determined in **section 6.4.1**.

For pGly28, the 7 unfolding peaks occur at  $23.1 \pm 1.0$  nm,  $36.1 \pm 1.0$  nm,  $50.5 \pm 1.1$  nm,  $70.3 \pm 1.7$  nm,  $115.9 \pm 0.9$  nm,  $144.3 \pm 1.8$  nm  $174.8 \pm 3.0$  nm, corresponding to anchor points at T470, V345, I346, M385, W226, I151 and G71. The corresponding rupture forces show that pGly28 unfold in two distinct parts, with mean forces of 145 pN at the beginning of each domain, reducing to around 80 pN for C-terminal unfolding. Mean force values were slightly smaller when pulled from the N-terminal, with 105 pN required to unfold the beginning of each domain, reducing to 45-65 pN as they unfold.

A complete assignment of  $L_C$  to Xyle structure is discussed in **section 6.5.1**.

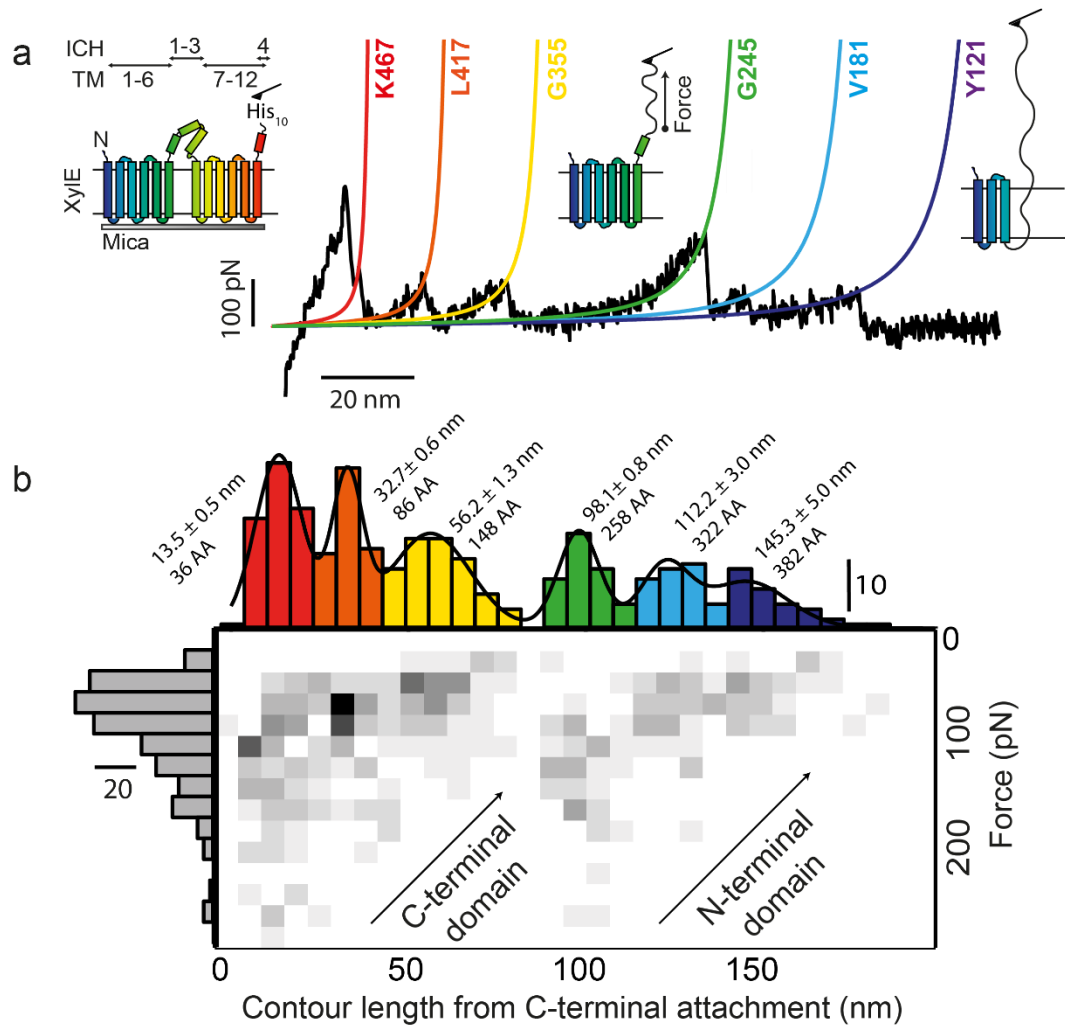




**Figure 6-13: 2D histogram summary of Xyle-pGly28 C-terminal-binding force peaks**

a) DNP-C (theoretical  $k_s = 0.24$  N/m)  $\text{Si}_3\text{N}_4$  AFM cantilever non-specifically binds to the C-terminal His<sub>10</sub>-tag of Xyle with additional pGly28 linker, within a 40:40:20 DOPE:DOPC:DOPG lipid bilayer deposited on mica. An example trajectory of C-terminal pGly28 pulling of Xyle at a constant velocity of 400 nm/s. Each peak in the force-extension trace is fit to the WLC with  $p = 0.6$  nm to determine contour length and coloured with the same scheme as above. c) 2D histogram, binning event rupture force (pN) against corresponding protein contour length ( $n = 10$ ). Force peaks were fit to multiple Gaussian peaks and corresponding contour lengths are displayed. Structural anchor points are labelled above each peak are calculated from the full length of the pGly28 construct of 531 amino acids. The force required to unfold the C- and N- domains decreases as each domain unfolds; the N- and C-terminal domains show independent structural stabilities.

Unfolding from the C-terminal of WT Xyle ( $n = 38$ ) (**Figure 6-14**) showed a similar unfolding strategy with central  $L_c$  values of  $13.5 \pm 0.5$  nm,  $32.7 \pm 0.6$  nm,  $56.2 \pm 1.3$  nm,  $98.1 \pm 0.8$  nm,  $122.2 \pm 3.0$  nm, and  $145.3 \pm 5.0$  nm corresponding anchor points at residues K467, L417, G355, G245, V181 and Y121. WT Xyle also unfolds with two distinct features corresponding to the N and C-domains independent stabilities with initial force peaks rupturing at  $\sim 160$  pN, reducing to 45 pN for C-domain, and 120 pN to 45 pN for the N-terminal domain. Full mapping is discussed in **section 6.5.1**.



**Figure 6-14: 2D histogram summary of WT Xyle C-terminal-binding force peaks**

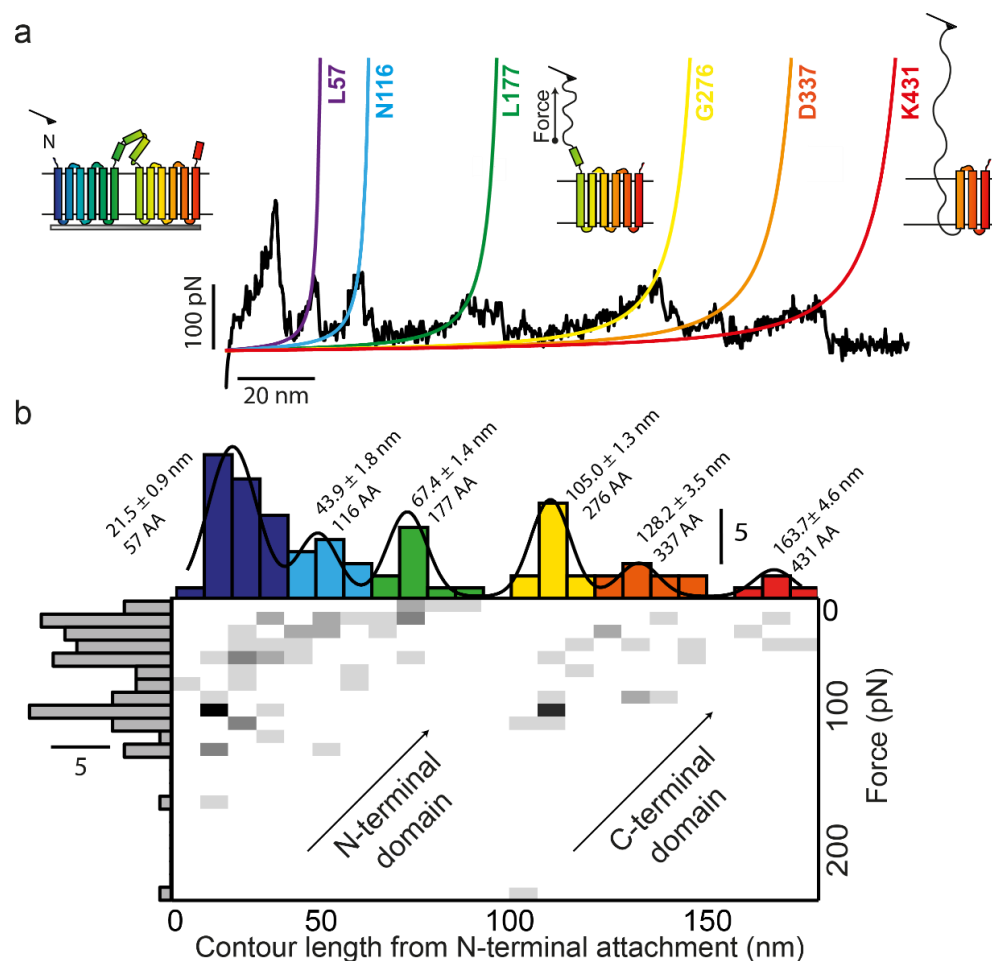
a) DNP-C, SNL-C or MLCT-C (theoretical  $k_s = 0.24$  N/m,  $0.24$  N/m and  $0.01$  N/m respectively)  $Si_3N_4$  AFM cantilevers non-specifically bind to the C-terminal His<sub>10</sub>-tag of Xyle within a 40:40:20 DOPE:DOPC:DOPG lipid bilayer deposited on mica. Example trajectory of C-terminal pulling of Xyle at a constant velocity of 400 nm/s. Each peak in the force-extension trace is fit to the WLC with  $p = 0.6$  nm to determine contour length of the unfolded protein. c) 2D histogram, binning peak rupture force (pN) against corresponding protein contour length ( $n = 38$ ). Peaks were fit to multiple Gaussian peaks and mean corresponding contour lengths are displayed with the number of amino acids unfolded. Structural anchor points at each  $L_c$  are labelled above each event. The force required to unfold the C- and N- domains decreases as each domain unfolds and the N- and C-terminal domains show independent structural stabilities when unfolded from the C-terminal.

Table 6-3: Central  $L_c$  values and anchor points for C-terminal binding XylE-pGly28 and WT XylE

Force Peak	XylE-pGly28 mean force peak (n = 10) (nm)[AA]	pGly28 anchor residue	WT XylE mean force peak (n = 38) (nm)[AA]	WT XylE anchor residue	Difference in $L_c$ due to pGly28 ( $\approx 10.4$ nm)	Theoretical WT C-terminal bound $L_c$ [aa] (section 6.4.1)
$L_{C0}$	$23.1 \pm 1.0$ [61]	T470	$13.5 \pm 0.5$ [36]	K467	25	-
$L_{C1}$	$36.1 \pm 1.0$ [95]	V346	$32.7 \pm 0.6$ [86]	L417	9	39 [103]
$L_{C1}'$	$50.5 \pm 1.1$ [146]	M385	Not observed			-
$L_{C2}$	$70.3 \pm 1.7$ [185]	I346	$56.2 \pm 1.3$ [148]	G355	37	62 [163]
$L_{C3}$	$115.9 \pm 0.9$ [305]	W226	$98.1 \pm 0.8$ [258]	G245	47	107 [281]
$L_{C4}$	$144.3 \pm 1.8$ [380]	I151	$122.2 \pm 3.0$ [322]	V181	58	132 [347]
$L_{C5}$	$174.8 \pm 3.0$ [460]	G71	$145.3 \pm 5.0$ [382]	Y121	78	160 [422]

### 6.4.5. Unfolding XyleE from the N-terminal

Traces that were analysable, and not mapped to C-terminal binding for XyleE or XyleE-pGly28 were determined to be N-terminal-tip binding ( $n = 10$ ) **Figure 6-15**. For this section, the colours of each contour length are inverted and unfold from purple to red. Six distinct intermediates are shown with central  $L_C$  values at  $21.5 \pm 0.9$  nm,  $43.9 \pm 1.8$  nm,  $67.4 \pm 1.4$  nm,  $105.0 \pm 1.3$  nm,  $128.2 \pm 1.3$  nm, and  $163.7 \pm 4.6$  nm, corresponding to anchor points at L57, N116, L177, G276, D337 and K431. Similarly to C-terminal extraction, N-terminal unfolds in two distinct domains, with the N domain unfolding at  $\sim 110$  pN, reducing to  $\sim 30$  pN, and the C domain at  $\sim 100$  pN to  $\sim 30$  pN.



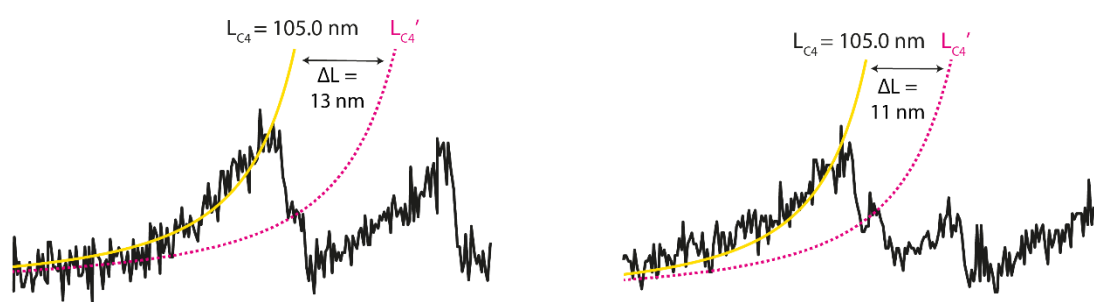
**Figure 6-15: 2D histogram summary of XyleE and XyleE-pGly28 N-terminal-binding force peaks**

a) DNP-C (theoretical  $k_s = 0.24$  N/m)  $Si_3N_4$  AFM cantilever non-specifically binds to the N-terminal of XyleE, or XyleE with additional pGly28 linker, reconstituted within a 40:40:20 DOPE:DOPC:DOPG lipid bilayer deposited on mica to produce an example trace of N-terminal pGly28 pulling of XyleE at a constant velocity. b) Each peak in the force-extension trace is fit to the WLC with  $p = 0.6$  nm to determine contour length. c) 2D histogram, binning force peak rupture force (pN) against corresponding protein contour length ( $n = 10$ ). Force peaks were fit to multiple Gaussian peaks and corresponding contour lengths and number of amino acids unfolded are displayed. Structural anchor points are labelled above each event. The force required to unfold the C- and N- domains decreases as each domain unfolds; the N- and C-terminal domains show independent structural stabilities when unfolded from the N-terminus.

20 % of N-terminal unfolding traces show a small intermediate after  $L_{C4}$  with a  $\Delta L$  of  $\approx +12$  nm **Figure 6-16**. Using Gaussian fits with an  $L_{C4}$  at 105.0 nm, I can calculate the new intermediate to be at 118.0 nm ( $L_{C4}'$ ). These results however are based on 2 trajectories of 10 - A much higher n-value is required to accurately deduce the precise number of amino acids unfolded from the histogram produced, to determine whether this force peak is a significant intermediate in the unfolding pathway.

All determined peaks for N-terminal extraction and residue anchor points are summarised in **Table 6-4**.

A complete assignment to structure is discussed in **section 6.5.1**.



**Figure 6-16: N-terminal binding traces showing  $L_{C4}'$  intermediate**

20 % of N-terminal traces showed an additional intermediate after the  $L_{C4}$  (yellow) force peak. The average  $\Delta L$  for this intermediate was calculated to be 12.0 nm. Using the central Gaussian value of the  $L_{C4}$  fit as 105.0 nm, I calculated the  $L_{C4}'$  (pink) to be 117 nm, corresponding to an additional anchor point at residue T311.

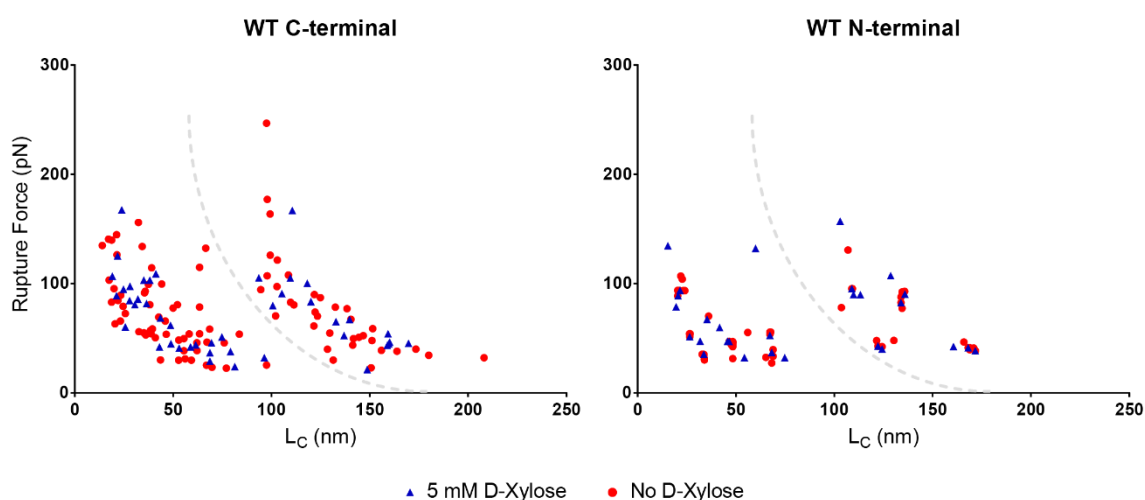
**Table 6-4: Summary of  $L_C$  values and anchor points for N-terminal extraction of XylE**

\* not observed in bivariate histogram, only 2 of 10 trajectories showed this intermediate

Identified force peak	N-terminal peak $L_C$ (n = 10) [AA]	N-terminal anchor residue
$L_{C1}$	$21.5 \pm 0.9$ [57]	L57
$L_{C2}$	$43.9 \pm 1.8$ [116]	N116
$L_{C3}$	$67.4 \pm 1.4$ [177]	L177
$L_{C4}$	$105.0 \pm 1.3$ [276]	G276
* $L_{C4}'$	$L_{C4}+12$ (117.0) [311]	T311
$L_{C5}$	$128.2 \pm 3.5$ [337]	D337
$L_{C6}$	$163.7 \pm 4.6$ [431]	K431

#### 6.4.6. D-xylose binding does not affect Xyle unfolding pathways

To determine whether the binding of the native transporter substrate D-xylose affected the unfolding pathway of Xyle from bilayers, SMFS experiments were repeated in buffer containing 5 mM D-xylose.  $L_C$  and corresponding rupture forces for each force peak were plotted in a scatter graph alongside no-xylose experiments for comparison. Bivariate histograms could not be plotted due to the low n number of xylose 'pick-ups'. The scatter graph showed no significant changes when extracted from either the N- (n = 3) or C-terminal (n = 5) of Xyle at this velocity (400 nm/s). The protein still unfolds in two distinct stages with comparable forces as apo-Xyle **Figure 6-17**.



**Figure 6-17: WT Xyle binding D-xylose plotting rupture force against  $L_C$**

*Plots of C-terminal and N-terminal binding of WT Xyle with and without the addition of D-xylose. Rupture force of each force peak has been plotted against its corresponding  $L_C$ . There appears to be no significant difference between the forces with (blue triangles) and without D-xylose (red circles). In each tethering position, the protein unfolds in two separate domains.*

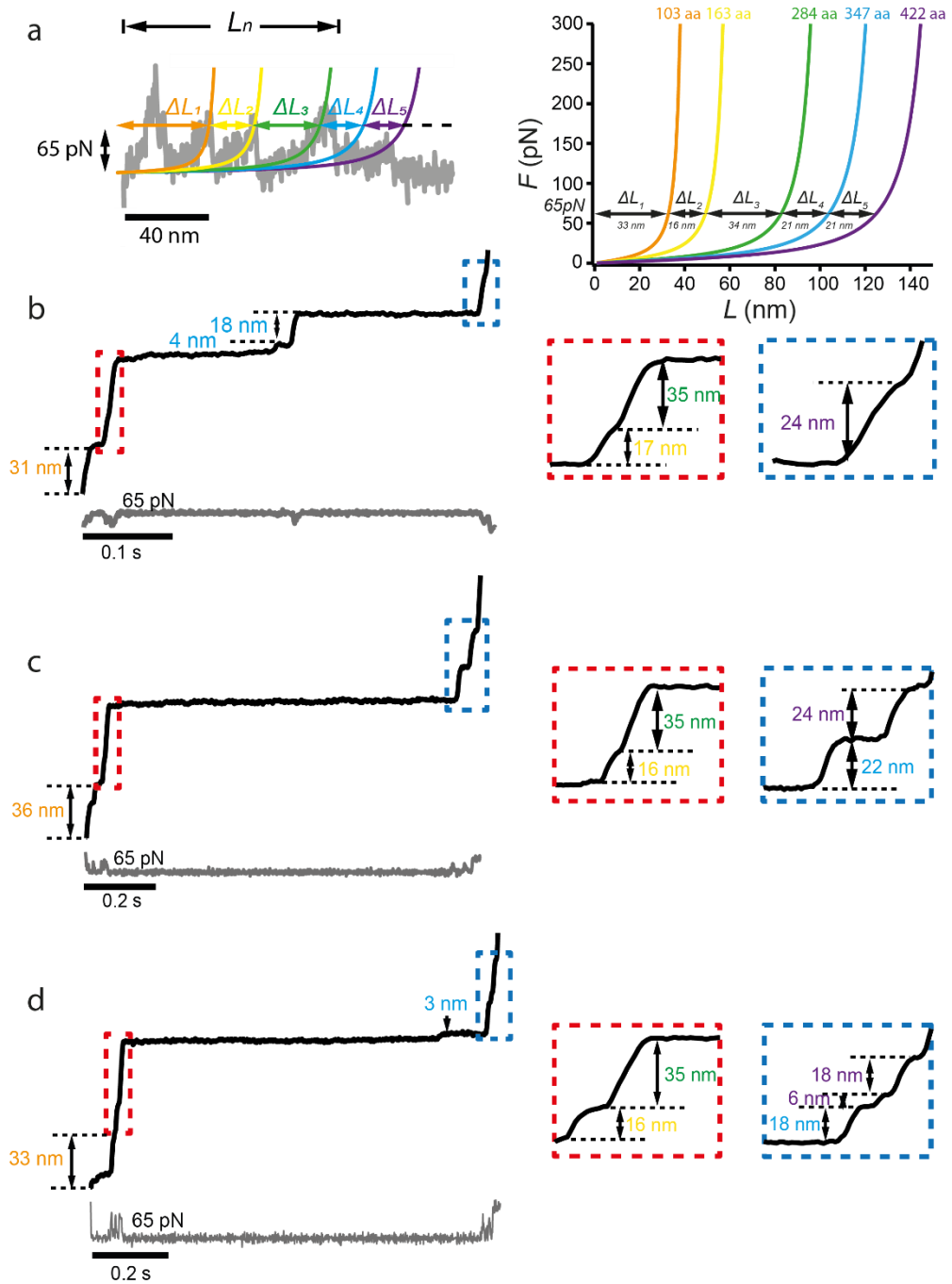
#### 6.4.7. Force-clamp unfolding of WT Xyle

Unlike FX data which can unfold proteins at a constant velocity, FC can maintain a constant force along the protein, unfolding it in a stepwise manner as a function of time. Each 'step' corresponds to the length of the unfolded intermediate at a particular force. Using an example FX trajectory, we decided to unfold at a force of 65 pN which would capture most intermediates that are shown in an example trace, as shown in **Figure 6-18a**. If the force is too large, intermediates may be missed in the FC data, if the force is too small, peaks may be lost in noise, due to the thermal fluctuation of the cantilever.

Using the generated WLC fits generated in **section 6.4.1**, we can work out the expected step sizes ( $\Delta L_c$ ) at 65 pN for the major expected contour lengths. The step sizes were calculated to be  $\Delta L_1 = 33$  nm,  $\Delta L_2 = 16$  nm,  $\Delta L_3 = 34$  nm,  $\Delta L_4 = 21$  nm and  $\Delta L_5 = 21$  nm when extracted from the C-terminus. Step sizes were consistent with experimental FX data (**sections 6.4.4 and 6.4.5**). **Figure 6-18bcd** shows three clamp traces which match the predicted unfolding pathway. The step size colours correspond to the  $\Delta L_c$ . Some  $\Delta L_c$  events occur in two stages during experimental conditions, these are denoted by the same colour and suggest additional 'hidden intermediates' in the stepwise unfolding pathway which FX at a single velocity of 400 nm/s could not identify. To describe the unfolding pathway in a FC experiment, a specific attachment of the tip to the protein is required to keep the starting position constant. The step-sizes from these initial trajectories are summarised in **Table 6-5**. However, assignments cannot be made due to the lack of knowledge on where the cantilever attachment point is, this unfolding trajectory could equally relate to N-terminal extraction.

When extracted at a constant force, it is often possible to observe 'hopping' between different intermediate states as shown in **Figure 6-19**. This transient unfolding intermediate was observed in three separate  $\Delta L_c$  force peaks ( $\Delta L_{C3}$ : pink box,  $\Delta L_{C4}$ : blue box and  $\Delta L_{C5}$ : red box), with the transient hopping increasing the total unfolding time of the protein to 8 s before early attachment which does not capture the complete final contour length, approximately 45 amino acids are unaccounted for.

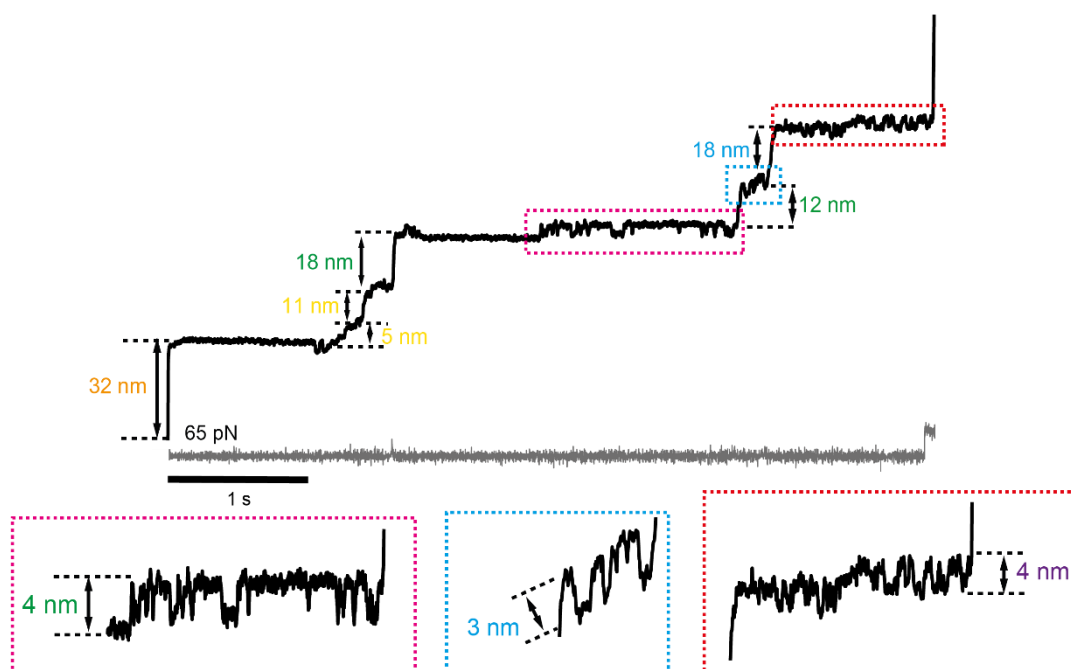
All step-heights and corresponding hopping intermediates are summarised in **Table 6-5**.



**Figure 6-18: Selection of clamp traces for WT XylE at 65 pN**

Selection of FC traces for WT XylE. These appear to be of C-terminal tethering based on the unfolding step-sizes, however at this stage it is not possible to determine this without a specific cantilever attachment. A) A force of 65 pN can unfold the entire protein whilst capturing each force peak detected in the FX trajectories. This force is mapped on the theoretical unfolding scheme from **Figure 6-8**, allowing estimated step sizes for the protein to be calculated at this force. b-d) Three traces each show the expected step sizes as a function of time. Lc<sub>1</sub> in orange, Lc<sub>2</sub> in yellow, Lc<sub>3</sub> in green, Lc<sub>4</sub> in blue and Lc<sub>5</sub> in purple. Some contour lengths occur in two separate steps which are not detected in FX traces. Areas in dashed boxes are expanded for a clearer view of unfolding intermediates. Each trajectory here shows complete unfolding of the protein within 1 s before tip detachment. The grey force trace shows 65 pN force is kept constant during unfolding.





**Figure 6-19: Example FC trace showing intermediate 'hopping' at 65 pN**

The colour scheme matches previous figures for  $L_c$  values: Orange  $\Delta L_1 = 32$  nm, yellow  $\Delta L_2 = 5$  nm + 11 nm, green  $\Delta L_3 = 18$  nm + 4 nm hopping + 12 nm, blue  $\Delta L_4 = 3$  nm hopping + 18 nm and purple  $\Delta L_5 = 4$  nm hopping + early detachment. The hopping steps are expanded for a clearer view. It is however still unclear which terminal unfolding this trajectory corresponds to without specific attachment. Intermediate hopping of 3 and 4 nm step heights correspond to around 10 amino acids which is roughly the length of half a TM helix. The complete unfolding trajectory occurs in roughly 8 s.

**Table 6-5: Step-Sizes for XylE unfolding at 65 pN**

\* We cannot explicitly say that these initial clamp traces correspond to C-terminal binding over N-terminal binding; these experiments should be repeated with pGly28, or more usefully a specific cantilever attachment site. For experimental data, step heights were calculated as follows:  $\Delta L_1 = L_{C_1}$ ,  $\Delta L_1 = L_{C_2} - L_{C_1}$ , etc. The additional  $L_c$  intermediate was omitted when calculating step heights. Generated WLC refers to data in **section 6.4.1**. H: hopping intermediate, ED: early detachment.

Step	*C-term extraction step height (nm)					
	Generated	Experimental	Figure 6-18a	Figure 6-18b	Figure 6-18c	Figure 6-19
$\Delta L_1$	33	32.7	31	36	33	32
$\Delta L_2$	16	23.5	17	16	18	5+11
$\Delta L_3$	34	41.9	35	35	35	18+4(H)+12
$\Delta L_4$	21	24.1	4+18	22	3+18	3(H)+18
$\Delta L_5$	21	23.1	24	24	6+18	4(H)+ED

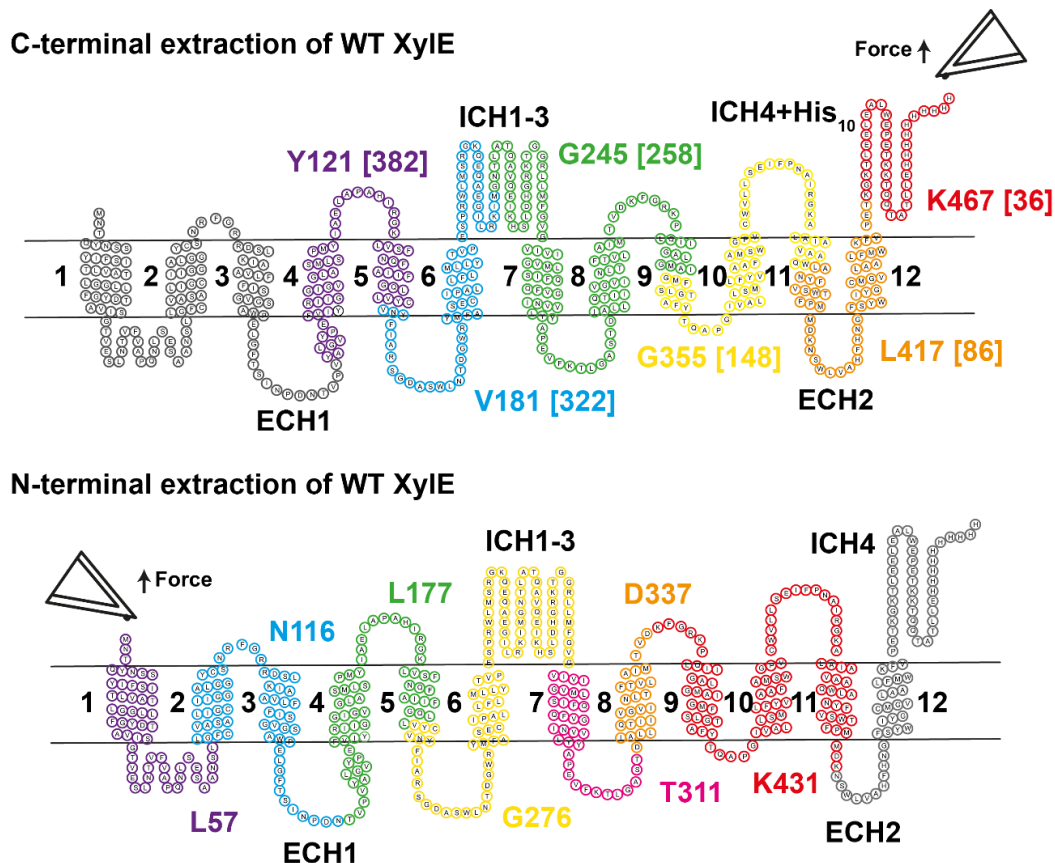
## 6.5. Discussion

### 6.5.1. XylE unfolding trajectory and force regimes

AFM SMFS reveals that WT XylE unfolds with six distinct unfolding intermediates when extracted from the C-terminal, and seven intermediates when extracted from the N-terminal. The position of binding was determined using an additional 28 amino acid poly-glycine based linker which was cloned onto the C-terminal of the protein based on existing literature on LacY (Serdiuk et al., 2015, Serdiuk et al., 2014).

The preliminary SMFS data obtained in this study can be used to map the unfolding trajectory of the XylE protein using averaged  $L_C$  values for the N- and C-domain extraction trajectories on WT protein. The anchor points determined from the average contour lengths are mapped onto a schematic structure of XylE **Figure 6-20**. This shows that when extracted from the C-terminal - firstly, His<sub>10</sub> and the ICH<sub>4</sub> helix unfold to K467, next helix 12 and half of helix 11 unfold with anchoring at L417. Next the remaining Helix 11 unfolds with helix 10, and half of helix 9, with the anchor position at G355. The fourth event completes the unfolding of the C-domain, and ICH domains 2 and 3, anchoring at G245. This suggests that ICH 1 remains bound to the membrane or the N-domain of the protein at this stage. Event 5 unfolds ICH 1 and helix 6, with the anchor point at V181 at the bottom of helix 5. The final observed event unfolds helices 5 and 4, with the final anchor point at Y121. The final three helices were not observed when averaging the contour lengths, and this is likely an artifact of data averaging or early detachment of the cantilever from the protein. The final unfolding event is never captured using AFM SMFS due to a lack of force associated when the protein is freely suspended in the aqueous environment.

When extracted from the N-terminal, which is perhaps more indicative of the cellular folding process which occurs co-translationally, with the N-domain folding first, the first 57 amino acids unfold, which include helix 1 and the loop region, with anchoring at L57. The second event unfolds helices 2 and 3, and half of ECH1 to position N166. Next, helices 4, and the majority of 5 unfold to L177. Event 4 sees the unfolding of helix 6 and ICH<sub>1-3</sub>, with anchoring at G276. The additional intermediate unfolds to the bottom of helix 8, T311, with event 6 unfolding helix 8 to D337. The final unfolding event unfolds the remaining observed segment, helices 9, 10 and 11 anchoring at K431, with the remaining helix 12 not observed in the unfolding trajectory. This is also depicted in **Figure 6-20**.



**Figure 6-20: Summary of FX SMFS anchor points for WT Xyle**

Residue anchor points were calculated using the mean  $L_C$  values from each bivariate histogram and the average amino acid length of 0.38 nm and mapped onto the 4GBY structure. The number of amino acids unfolded is shown in square brackets. I have assumed that the cantilever binds to the methionine at position 1 when extracted from N-terminal, and the final histidine at position 503 when extracted from the C-terminal. The colour scheme is consistent with WLC fits and derived  $L_C$  values throughout this chapter. Grey residues were not observed in the averaged unfolding trajectories likely due to early detachment of the cantilever.

Mapping the unfolding segments for Xyle shows that the protein unfolds with two distinct force families, corresponding to a high force (160 pN) unfolding of the first domain, which gradually reduces as it unfolds, with the final segment rupturing at 45 pN. This indicates a gradual loss of protein stability as the protein is unfolded. Stability is then reinstated as the ICH domains begin to unfold which appears to be characteristic of the second domain beginning to unravel (120 pN), this again is accompanied by a reduction in force as helices are extracted (45 pN) **Figure 6-14**. The same is observed for N-terminal unfolding, with slightly lower forces; 120-30 pN for N-domain and 100-30 pN for the C-domain **Figure 6-15**. Xyle appears to unfold similarly to other proteins which have been studied using AFM, primary showing that helices largely unfold as pairs as reviewed in (Jefferson et al., 2018). It is not explicitly clear in the literature whether other MFS transporters also unfold as two separate domains, it is likely that they do, however it is maybe more pronounced with the inclusion of

the highly stable ICH domains which aren't present in LacY or NhaA, and all major force peaks are of a similar magnitude for each protein unfolding event, often in excess of 100 pN (Kedrov et al., 2004, Serdiuk et al., 2014, Oesterhelt et al., 2000), which notably is higher than what we see for each domain of XyleE unfolding.

In XyleE, the ICH domains increase the stability of the protein, and ICH 4 and ICH 1-3 form a latch region which tightens the intracellular gate of XyleE during its transport cycle (Deng et al., 2014, Bai et al., 2017), likely eluding to the higher intrinsic stability when MFS family proteins are artificially denatured using urea and GuHCl (Harris et al., 2017a). This likely explains the higher ICH unfolding event compared to other XyleE unfolding events, which contrasts what has been shown for other proteins where the magnitude of rupture forces remain relatively similar for each event.

The increased force at the ICH region was also observed in the initial clamp traces shown in **Figure 6-18**, with  $L_{C4}$  taking considerably longer to unfold than flanking intermediates of the N or C domains unfolding. The  $\Delta L_C$  steps closely match the expected values theorized, and experimentally determined for both N- and C- FX unfolding, although due to the non-specific nature of attachment it is not possible to assign the direction of extraction for the FC traces. The longer dwell times for these stable intermediates was also captured in the example FC 'hopping' trajectory. The unfolding of the ICH domains (green force peak **Figure 6-19**) captures a 4 nm hopping intermediate with a long dwell time (>1 s), which roughly corresponds to 10/11 amino acids. This hopping event is likely a transient interaction between the ICH helices, which perhaps captures the dynamicism of the transport protein. Without a specific attachment point however, it is impossible to define the exact position of this intermediate using FC. When this 4 nm hopping phenomena occurs in a transmembrane region of the protein, such as the intermediates shown in  $\Delta L_4$  and  $\Delta L_5$ , it is likely due to a helix transiently hopping in and out of the membrane due to helix/helix or helix/lipid interaction when unfolded at 65 pN. Our data suggests this is happening once the second protein domain is unfolding, and perhaps due to lipids moving to fill in spaces left by prior extracted protein to maintain intrinsic structure of the protein. Unfolding at a lower force using a high-resolution MT set up will allow us to probe additional low force intermediates we are starting to uncover in these preliminary FC experiments for membrane proteins.

Binding ligand to proteins has been previously shown to alter the unfolding pathways of secondary transporters using mechanical force (Serdiuk et al., 2014, Kedrov et al., 2006b,

Zocher et al., 2012a). I attempted a similar experiment here, by including D-xylose in the SMFS buffer.  $L_C$  and rupture force values were plotted as a scatter graph. Bivariate histograms could not be plotted due to the low number of such traces, although in the raw data there appeared to be no discernible difference between the apo- and D-xylose bound XylE. This however may be due to its very low affinity (0.40 mM, **section 5.2.4**) - previous studies used artificial binding partner to proteins with higher affinities for LacY with  $\alpha$ NPG (10  $\mu$ M) (Serdiuk et al., 2014), or significant force peak and energetics changes when NhaA binds to inhibitor (Kedrov et al., 2006b, Kedrov et al., 2008), which is not possible to assay here without an extensive DFS study. Although no significant changes were shown here, it has been previously shown that the addition of 1 mM xylose was able to stabilise XylE on GuHCl unfolding (Harris et al., 2017a), and a more extensive DFS study may uncover a mechanism for this stabilisation. Xylose addition, however, did increase the pickup of the XylE, this was unexpected, and perhaps due to a conformational locking of the protein into a conformation which further exposes the protein C terminal, removing the dynamic nature of the protein in the bilayer.

### **6.5.2. Limitations of this study**

Our experiments rely on a non-specific attachment of the cantilever to the protein of interest. The non-specific nature of the  $\text{Si}_3\text{N}_4$  cantilever to the protein and the difficulties faced trying to incorporate high protein:lipid ratios into the bilayers are the primary drawbacks of the technique. This results in low numbers of successful data collection, and a substantial amount of well characterised data is required for a complete DFS study. DFS studies are the only way to achieve an in-depth study of the energetics of membrane protein folding and was unfortunately not possible to achieve during the time of my PhD, mainly due to lack of equipment hours.

The 28 amino linker is perhaps not long enough to promote a reproducible binding position and consistent 10 nm shift in every unfolding trajectory compared to unfolding protein without linker. To make doubly sure each of our extraction assignments are correct, a longer linker can be used, or the WT protein can be cleaved with trypsin prior to SMFS. This was shown with trypsin cleavage at the C-terminal end, and the loop region connecting helices 8 and 9 of NhaA - revealing a reproducible trajectory which could be assigned depending on direction of extraction (Kedrov et al., 2004). This would likely be difficult in these experiments due to relatively low protein:lipid ratios compared in liposomes compared to the 2D crystals of NhaA used in the study. Alternatively, a specific mode of interaction between the tip and the protein may be useful - a gold coated cantilever bound to an engineered cysteine residue

for example (Oesterhelt et al., 2000). This ensures a correct binding position for each protein, although altering the surface chemistry of the protein can have adverse effects on the protein. However non-specific attachment methods allow for higher throughput (Muller et al., 2002), and the same cantilever can be used for imaging.

Additionally its suggested that much higher lipid:protein ratios can be achieved by using purple membrane extracts of *H. salinarum* to study bR for example these yield very high lipid:protein ratios where the protein is essentially a 2D crystal (75 % of membrane), interspaced with very few lipid molecules (25 % of membrane) (Oesterhelt and Stoeckenius, 1971). This yielded around 15 % total pick-up corresponding to protein adhesion compared to the 0.28 % in this study. 2D crystallisation of NhaA for AFM SMFS has also been achieved (Kedrov et al., 2004, Kedrov et al., 2005, Kedrov et al., 2006a, Kedrov et al., 2006b, Kedrov et al., 2008), however the pickup rate is not quoted in the literature, and this was not possible for XylE experiments in timeframe of this PhD. Lipid nanodiscs used to improve pick-up, and to maintain the native lipid environment (Oluwole et al., 2017a, Barniol-Xicota and Verhelst, 2018) were also trialled for XylE. Unfortunately, the data revealed a considerable amount of non-specific binding to the polymer and we were unable to pursue this. MSP-based purple membrane nanodiscs are currently the only studies of AFM SMFS for bR (Zocher et al., 2012b) which showed no significant change in dynamics to when AFM SMFS was carried out in bilayers, but allowed for easier handling of the sample. This perhaps could be an alternative to the native-nanodisc using *E. coli* lipid extracts as the bilayer environment once the unfolding characteristics of the MSP protein itself are characterised.

### **6.5.3. Moving towards magnetic tweezer spectroscopy**

AFM extracts protein from the bilayer environment under a high force. The protein here transverses many unfavourable energy barriers as loops are pulled through membranes, and the high force used may not capture the smaller ‘hidden-intermediate’ steps which are the most interesting to study, such as the ICH hopping captured in FC experiments shown here. Although the 65 pN used in experiments shown in this chapter is still is a relatively high force, similar intermediates maybe highlighted at much lower forces, forces like those experienced in the physiological cell setting. AFM can often miss such intermediates due to thermal noise generated when the cantilever is held at constant low forces. Forces as low as 50 pN can be achieved before this noise starts to mask data. To counter this, XylE was modified for MT spectroscopy to unfold the protein within the bilayer environment (see **sections 5.2.7** and **6.4.3**), within a native nanodisc. MT can assess intrinsic protein forces whilst maintaining

these relatively high forces, with much less thermal noise which is absorbed by the large magnetic bead. MT also avoids the ‘hit-and-hope’ aspect of AFM experiments due to the specific tethering of the protein at each end.

The necessary modifications to tether the protein on either side of the membrane were achieved with the cloning of the N-terminal HaloTag with additional *de novo* designed helix (Cymer et al., 2014), as well as the C-terminal avi-tag for *in vivo* biotinylation. This allows for the tethering of the N-terminal end of the protein to a functionalised glass substrate, which I have shown to be functional with bead capture assays and AFM imaging on functionalised glass cover slides, and the tethering of the C-domain end to a streptavidin coated magnetic bead for low-force manipulation of the protein. The latter hasn’t been directly shown in this chapter for the TMH-Xyle construct due to PhD time constraints, but *in vivo* biotinylation, and subsequent binding to streptavidin species is possible and was shown using RNC samples in **Chapters 3** and **4**. This set-up allows for continuity on unfolding Xyle from the C-terminal domain, perpendicular to the membrane, whilst accessing lower force regimes to build on preliminary work shown in this chapter using MT. Of course, the advantage of MT over AFM experiments is a lateral pulling to the membrane. This could likely be achieved using this construct with a deletion of the additional TMH domain, which is currently being taken forward in the Booth and Garcia-Manyes labs.

## **6.6. Chapter summary**

All experiments shown in this chapter arise from extraction at a single velocity or force. To extract any meaningful energetics and conclusions from this data an extensive DFS study is required, with each condition containing more than 100 trajectories with high confidence of cantilever anchor point for data averaging. This is an incredibly time-consuming project, mainly limited by protein pick-up and machine hours available which was not possible to achieve in this PhD time.

Nevertheless, I have determined that Xyle unfolds in two distinct domains, with 6-8 independent structural segments with pulling from N or C-terminals. I have also shown it is possible to study transporter unfolding under constant force using AFM, which, to my knowledge has not been shown previously and may allow time-resolved studies of mechanical force unfolding allowing an increased knowledge of folding intermediates which can be studied with higher sensitivity using MT spectroscopy.

Overall, the work presented in this chapter is a minor stepping stone to an extensive single-molecule study of XylE, whilst developing methods to unfold of isolated membrane proteins by AFM or magnetic tweezers, to builds on the work of GlpG and the ClC transporter described the Bowie group (Min et al., 2015, Min et al., 2018), which could be developed to mechanically probe stability of RNC samples of membrane proteins whilst in a native lipid bilayer.



## **7. Summary and future directions**

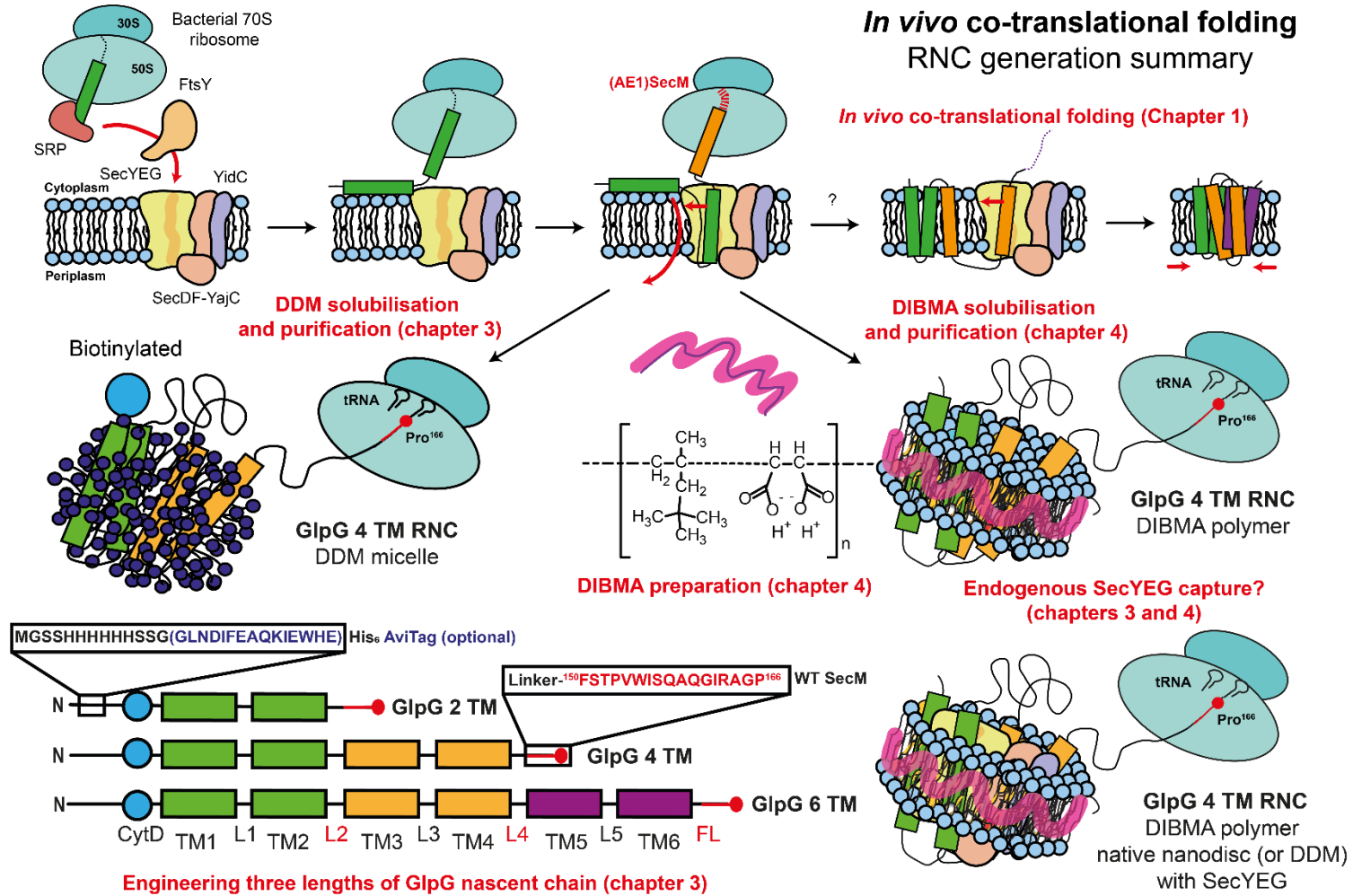
## 7.1. Summary of results

This chapter aims to summarise the work shown throughout this thesis, to highlight key outcomes, discuss their implications, and suggest possible improvements to the methods presented throughout. I also discuss possible future work based on the production of polytopic membrane protein RNCs to further our understanding of co-translational membrane protein folding.

Recent advances in protocol development shown throughout this thesis allow for in-depth biophysical studies of the co-translational folding pathways of polytopic  $\alpha$ -helical membrane proteins. Here I have used the SecM arrest-peptide sequence, which stalls translation while leaving the ribosome complex intact, engineered into the GlpG sequence to produce stalled-ribosome-nascent-chain complexes. RNCs capture a ‘snap-shot’ of the co-translational process at varying positions through the folding pathway as the protein inserts into its native membrane. This RNC purification method utilises a novel polymer-based system to capture not only the partially expressed protein, but also its surrounding lipids from the native *E. coli* membrane to allow a more physiological snapshot of the co-translational folding process, without the need to first purify using detergents as is common in the field – although preparation in DDM can also be achieved. I have shown GlpG RNCs in detergent and native-nanodiscs to be homogenous and stable; improved by substitution of an arrest enhanced SecM, and that they contain the correct composition of lipids, with additional evidence of co-purification of the SecYEG bacterial translocon. Overall, these methods open many avenues for future research from both biochemical/physical and structural avenues. A visual summary of RNC samples generated can be found in **section 7.1**, with the significance of RNC samples and discussions of the future work now possible in **section 7.2**.

Additionally, I have made important progress in ascertaining the (un)folding pathways and force regimes of a larger transmembrane-spanning transporter, XyleE, using AFM-based force spectroscopy. This preliminary data permits further research using more sophisticated AFM and MT experiments to probe smaller, and more physiologically relevant forces which exist within protein folding intermediates. This, ultimately, should allow for sensitive, time-resolved unfolding studies of such proteins in bilayers, or native nanodisc environments.

## 7.2. Visual summary of RNC samples generated



### **7.3. A note on the Coronavirus lab shutdown**

The 2020 Coronavirus pandemic unfortunately led to university-wide laboratory shutdown. This affected data collection at a crucial point of my PhD, and further studies into the structure using CryoEM, and biochemical analyses of GlpG RNCs with co-expression of the translocon could not be carried out. Any potential findings based on future work discussed in this chapter were ultimately cut short.

### **7.4. Significance and future work**

#### **7.4.1. Generation of GlpG RNCs captured within in DDM micelles**

GlpG, a 6 TM helix rhomboid protease was designed using MPEx to contain a SecM stalling sequence which actively stalls ribosomal translation when overexpressed *in vivo*. This allowed the production of three ribosome-bound nascent chain complexes at three different lengths of GlpG polypeptide. I was able to capture these RNCs when grown in cells in a DDM detergent, and purify them to a high quality, and yield, of around 25-100 pmol/mL (around 0.1 mg per litre of culture), which is ample for most biochemical and CryoEM experiments. Additionally, this yield can be scaled up by increasing cell-density, using either larger growth volumes, or increased aeration in the growth media for potential to study using NMR, which requires much higher yields for experimentation. I also showed that an incorporation of an enhanced stalling variant of the SecM sequence can drastically improve the quality of the generated complex, with quality checks carried out using Western blotting and SDS-PAGE analysis, as well as negative staining of the RNC samples. The method of purification is relatively straightforward and is based on a standard high yielding membrane protein purification protocols, with amendments made to the growth conditions which provide a maximal number of available ribosomes, as shown previously when generating high yields of soluble protein RNCs (Cassaignau et al., 2016). The affinity-chromatography/size exclusion-based purification methods provide a simpler, and less time-consuming method to produce RNCs than sucrose gradient methods used previously (Schaffitzel and Ban, 2007, Cassaignau et al., 2016, Rutkowska et al., 2009). The work presented here is the first time a polytopic  $\alpha$ -helical membrane protein RNC has been purified from *in vivo*.

Additionally, I can further tag each RNC sample with an avi-tag peptide, allowing for *in vivo* biotinylation when the cells are co-expressed with biotin-ligase and the growth media supplemented with D-biotin. This secondary tagging is particularly useful for downstream (or alternative) purification strategies, or biophysical analysis which requires a strong, non-covalent bond, which can withstand high-force (Evans and Ritchie, 1997), and low pH (Jensen

et al., 2013). Additionally, the *in vivo* biotinylated RNCs were run on a binding matrix and shown that sepharose-streptavidin beads can bind the biotinylated sample, and monomeric avidin can bind, and elute the RNC using excess biotin, proving success of a further method of purification of the RNCs. The intrinsic features of the biotin-streptavidin bond, and the additional purification methods open many doors for mechanical manipulation and studies of co-translational folding intermediates using MT spectroscopy with RNCs, or in-depth studies of nascent chain structure and dynamics using powerful HDX-MS techniques in the future and remain very interesting projects to follow up on.

RNC samples in DDM were shown to be stable and homogenous when probed with Western blot, which should allow for most biochemical/physical and structural analysis. However, to measure homogeneity directly, negative staining and 2D classification of the RNC samples were carried out by Dr Sara Alvira at the University of Bristol. Although intact RNC samples appeared to be of good quality, when using WT SecM, there was a significant release of the nascent chain for the 4<sup>TM</sup> sample which I would expect to be the most stable, based on the MPEx data – and this likely results from a combination of active cell-removal of the stalled nascent chains, and a force-sensor type release of the nascent chain as the protein folds into the membrane mimic (Ismail et al., 2012, Cymer and von Heijne, 2013). Stability was improved with the incorporation of an arrest enhanced SecM sequence (Cymer et al., 2015a, Kempf et al., 2017) which was shown here biochemically; however, we were unable to characterise arrest enhanced RNCs using negative stain due to the time constraints on this PhD. The well-defined negative stains, however, are a good precursor for a successful 3D reconstruction of structure when using CryoEM, which I have discussed further in **section 7.4.2**.

#### **7.4.2. Capture of GlpG RNCs in native nanodiscs with endogenous SecYEG**

I have also developed a method to capture the RNCs directly from their native membranes using the diisobutylene-maleic acid co-polymer – a variant of SMA (Oluwole et al., 2017a), which has been used very successfully to study membrane protein structure and function. I was able to solubilise the native membranes of RNC overexpressed cells to capture the samples in their lipid environment, before purifying the RNC nanodiscs using a similar method to DDM purification, with a few minor adjustments, as discussed in **section 4.3.1**. The resulting samples were not as clean as DDM samples, perhaps due to an increase in nascent chain release of the WT SecM sequence - this was also confirmed using negative stain and 2D classification, where although 6<sup>TM</sup> GlpG was very clean, only around 15 % of the ribosomes were occupied. Fortunately, occupancy for DIBMA RNCs was drastically improved with the

addition of the enhanced SecM sequence, however there was no time to classify DIBMA purified enhanced RNCs using negative stain during this PhD.

The preparation of membrane protein RNCs of this nature is a significant advance in the field. Not only have the membrane proteins been stalled to observe snapshots of the co-translational process, but the capture within a native lipid environment which is crucial for understanding the folding pathways *in vivo*. It has been previously shown that the lipid environment and composition crucial for correct function for GlpG, where there is a direct effect on the stability and likely folding pathways (Reading, 2018, Moin and Urban, 2012, Kreuzberger et al., 2019). Prior to the discovery of SMA technology, the only way to recapitulate this was through careful mixing and reconstitution into a synthetic lipid environment which may not accurately depict what happens *in vivo*. SMA/DIBMA technologies allow for direct extraction to the membrane which avoid any additional detergent steps which are usually required (Lee et al., 2016). This prevents multiple purification steps where protein degradation can often occur, and avoid the need for partially destabilising detergents, although their use has been incredibly insightful for past experiments of function, stability, and structure (Seddon et al., 2004).

It was possible to confirm the correct extraction of native lipids using HPTLC analysis when using DIBMA to capture RNCs. There was no discernible difference between each stall length and the standard, apart from a possible reduction in cardiolipin in the 2 TM GlpG RNC sample when grown in minimal media. This may reflect the local lipid environment at the early stages of GlpG biogenesis, with possible changes on interaction with the SecYEG translocon which requires cardiolipin for function (Ryabichko et al., 2020). Equally, this could be due to preferential solubilisation by DIBMA, or collisional lipid transfer in the discs (Danielczak and Keller, 2020), however the latter is relatively unexplored for the DIBMA co-polymer. Nevertheless, extraction of the native membrane is a significant advantage for folding studies and has not been shown before for a stalled RNC complex mid-way through its folding cycle.

It was also possible to capture endogenous translocon machinery in both detergent and native nanodisc purifications of RNC. It is assumed that like most helical membrane proteins, GlpG does in fact use the translocon to enter the membrane (Schibich et al., 2016), hence I used the translocon-assisted hydrophathy plots when designing these stable stalling sequences. With this knowledge, it is my ambition to co-express the SecYEG translocon with the RNC constructs to co-purify a complete snapshot of the translocon assisted insertion process at three different stalled lengths. At this stage, it has been possible to overexpress each component, and solubilise them together in DDM detergent, as confirmed by Western blotting. SecY was

identified using monoclonal SecY antibody, RNC using anti-poly-histidine, and SecE which was biotinylated *in vivo*, was detected using streptavidin HRP. The 70S ribosomes were identified by Coomassie stain. This confirms that overexpressing each of the components is possible; however, downstream purification procedures downstream are yet to be trialled. Nevertheless, this should lead to exciting structural project with the co-translational process capitulated within a DIBMA native nanodisc, and with a small degree of optimisation, and incorporation of the arrest enhanced sequence, it should be possible to increase the stability and occupancy of the ribosomes using the arrest-enhanced SecM sequence to improve on the quality of negative stains I have shown throughout this thesis, which is a prerequisite for detailed 3D CryoEM reconstruction of the nascent chain and translocon.

The above experiments should allow for a high-resolution study of the emerging nascent chain. It has been previously shown that obtaining dynamics of the SecY translocon channel on activation with a helix of a secretory protein (Kater et al., 2019, Bischoff et al., 2014), when reconstituted into a synthetic lipid nanodisc is possible. However, can we improve these studies using these new methods to determine the dynamic processes encompassing the formation of structure during the GlpG co-translational folding cycle into its native environment? This could advance our understating of co-translational folding in three ways, 1) does the native lipid environment alter the SecY lateral gate function, 2) are processes different when inserting a polytopic membrane protein – when are secondary and tertiary structure formed, and how do lipids influence these interactions? and 3) is the ribosome, or any other protein factor which are inadvertently purified using my method also in play, and can we confirm this with the overexpression of the translocon, which I have also shown could be achievable with some degree of optimisation.

In complement to structural analyses, my GlpG RNC samples are highly amenable to a great deal of biochemical analysis which observe the local dynamics of the emerging nascent chain, and their role in the folding pathway. Lots of work on soluble RNCs has already given significant insight into the co-translational processes occurring in the cell using FRET (Woolhead et al., 2004, Mercier et al., 2020), and biochemical gels, from partial proteolysis to radiation pulse-chase analysis for study. Labelling of specific residues in TM helices and loops for example, will enable the tagging of these positions using maleimide cysteine chemistry, or ester lysine chemistry to incorporate fluorophores, or membrane (im)permeable species for gel shift assays. This would give an indication into the exposure of the residue to the membrane environment, which could be further enhanced by a membrane active dye (NBD or

IANBD), or the ribosome. Crosslinking agents could additionally be utilised for the proximity analysis of the nascent chain to other helices, or even ribosomal components, which are known to interact with the nascent chain in soluble proteins (Waudby et al., 2013, Cassaignau et al., 2016, Cabrita et al., 2009). Each of these experiments present possibility for time-resolved analysis should the RNCs be produced using a cell-free system. This should allow an observation of the kinetics of any folding intermediates between dyes in two domains or helices, with necessary dyes can be incorporated using amber-stop codon suppression (Woolhead et al., 2004, Johnson, 2005, Alder et al., 2008).

It may also be possible to study the kinetics and thermodynamics of folding using pulse proteolysis experiments, with forced denaturant unfolding of the nascent chain and the subsequent proteolysis of the RNC using a protease and following the unfolding using a gel-based assay. If the ribosome maintains its structure on the addition of denaturant, this technique can be powerful in yielding unfolding and refolding data. Ribosomes can withstand 4 M urea when unfolding soluble proteins RNaseH (Jensen et al., 2020) and Halotag (Samelson et al., 2018), however isolated GlpG is somewhat resistant to urea and only loses a fraction of secondary structure at 8 M urea (Harris et al., 2017b, Paslawski et al., 2015), a partially formed and folded nascent chain, however, maybe significantly less stable and agreeable to these sorts of denaturant folding experiments.

Currently, the only detailed time-resolved co-translational study of GlpG has been captured using the novel SEIRAS approach (Harris et al., 2017b). Can RNCs also be produced in a bottom-up cell-free approach, for the comparison of the co-translational folding pathway of the isolated GlpG protein using SEIRAS? (Harris et al., 2017b). Is it possible that we may lose mechanistic detail when by trapping the protein into a low energy state during RNC generation, which is perhaps not observed in the native folding pathway? It should be possible to ascertain this using a cell-free expression system with folding into bilayers, or MSP-based nanodiscs, and although not an *in vivo* environment, does highlight time-resolved fundamental processes in a bottom-up system which relies on spontaneous insertion. It is possible, however, to reconstitute the SecYEG translocon to further see the differences in effect of protein insertion, where the system is held in a conformation for a prolonged period, and the effect of SecYEG on this.

Finally, a significant chunk of literature concerns itself with the mechanical unfolding of stalled RNC soluble proteins using optical tweezers (Kaiser et al., 2011, Goldman et al., 2015). It may be possible to eventually trial these methods on RNCs, but currently this is a distant



goal for membrane proteins, and perhaps the mechanical folding methods detailed in **Chapter 6**, and in **section 7.4.3** will act as a suitable stepping-stone to achieve this goal.

### **7.4.3. Towards in-depth biophysical studies of folding using AFM and MT**

Biophysical studies of membrane protein folding often rely on bulk ensemble methods which cannot accurately determine minor folding intermediates. Single-molecule AFM based force spectroscopy was used here to characterise the unfolding pathways and force regime of a 12 TM helix Xyle protein, which will enable a gauging of the necessary forces required to study this protein in detail using MT spectroscopy. Xyle, and additional constructs, were all engineered and could be successfully produced for use with AFM.

Preliminary AFM experiments allowed us to unfold Xyle from its bilayer by extraction from either the N- or C-terminal of the protein. Extraction direction was determined using a modified Xyle, with 28 additional amino acids at its C-terminal end. The protein was shown to unfold with 6 (C-terminal extraction) or 7 (N-terminal extraction) stable segments, which for the most part unfold consistently with published work (Jefferson et al., 2018), but in two defined stages, which correspond to the two transporter domains. This was characterised using the associated rupture forces of the protein, which start high and successively reduce as the domains unfold. Additionally, we tried to determine whether there were any changes in the unfolding pathway of Xyle when bound to its natural substrate. Unfortunately, due to the low pickup rate of the protein - even after optimisation of reconstitution to increase protein:lipid ratio, and hence the lack of an extensive DFS study, it was not possible to accurately say whether there was any effect. However, by eye and with no statistical analysis, it looked like D-xylose had no effect on the unfolding pathways of Xyle. Optimisation to increase protein/lipid ratios were successful in this study, but it was not possible to compare these values to published work as these papers do not calculate efficiencies, and only quote the desired ratio (Serdiuk et al., 2014). To overcome the intrinsic low pick-up rates associated with these experiments, competitors have many automatic AFMs running at once to obtain enough information for DFS studies.

Using the determined FX unfolding regime and force profile obtained for WT Xyle, we could determine a suitable pulling force for force-clamp experiments. Initial FC data revealed an expected trajectory based on both theorized, and experimental data. It was not possible to assign these intermediates to structure as Xyle is symmetrical, and clamp experiments were run on WT protein only. It was however shown that several 'hopping' intermediates were

observed in the clamp unfolding trajectory, which correspond to the intra-cytoplasmic helix region of XyleE when mapped. This region is known to contribute to the XyleE increased protein stability when compared to other MFS proteins (Harris et al., 2017a), and form a latch during the transport cycle (Deng et al., 2014, Bai et al., 2017).

It should however be said that while no solid conclusions on XyleE unfolding pathways can be drawn from these experiments, every experiment is feasible but requires many more experiments and a lot of time to attain enough trajectories to confirm any hypothesis. I have however achieved a preliminary force range and unfolding trajectory to allow for more detailed time resolved and physiological FC and MT experiments, which are intrinsically large projects in themselves. Nevertheless, I have successfully laid the groundwork for them, with successful construct design and preliminary measurements of force and unfolding trajectories. It should also be noted that, to my knowledge, this study is the first to show force-clamp unfolding of membrane proteins from bilayers.

The preliminary data here shows some interesting features in the unfolding pathway of XyleE, with particular regions of the structure, namely the ICH domains, which could be interesting to look at with higher force resolution. Therefore, necessary modifications to the XyleE protein were made for the study of this protein with a similar regime using magnetic tweezers. MT experiments require tethering on either end of the protein. This was achieved with the cloning of an additional *de novo* helix and N-terminal HaloTag for stage tethering, and an avi-tag incorporation for *in vivo* biotinylation, and subsequent magnetic bead binding at the C-terminus. This modified XyleE could be purified in DDM, and DIBMA which provides a membrane environment for MT spectroscopy. Unfortunately, when DIBMA was used in classic AFM experiments with XyleE, there was a significant degree of non-specific attachment of the cantilever to the polymer, and it was not possible to assign traces to protein unfolding, or non-specific adhesion, however these experiments did show that the HaloTag modification of XyleE was functional with surface chemical tethering to glass, which should allow in depth MT studies in the near future.

## **7.5. Closing remarks**

The overarching aim of this thesis was to produce a new sample of a polytopic membrane protein RNC which would allow a detailed study of the co-translational studies of membrane protein folding. The ease of genetic manipulation, coupled with the relatively straightforward methods shown here mean that these protocols should be applicable to many proteins, and

the resulting high yielding and homogeneous samples, while remaining physiologically relevant, are also highly amenable to many structural and biochemical techniques. RNCs of this nature should allow a detailed, combined experimental approach of biochemical/physical and structural nature to fully understand the mechanisms which underpin co-translational membrane protein folding, and for GlpG, this should be achievable in the near future.

The additional development of force-spectroscopy experiments and sample optimisation for low-force tweezer and clamp experiments on large, isolated proteins could one day be coupled to RNC membrane protein studies to understand co-translational pathways from a mechanical perspective. Although, initial time-resolved force studies of isolated proteins in the bilayer remain an achievement. For RNCs however, mechanical manipulation may yet require significant optimisation and data collection time.

## References

---

- ABDUL-GADER, A., MILES, A. J. & WALLACE, B. A. 2011. A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics*, 27, 1630-6.
- ABRISHAMI, V., ZALDIVAR-PERAZA, A., DE LA ROSA-TREVIN, J. M., VARGAS, J., OTON, J., MARABINI, R., SHKOLNISKY, Y., CARAZO, J. M. & SORZANO, C. O. 2013. A pattern matching approach to the automatic selection of particles from low-contrast electron micrographs. *Bioinformatics*, 29, 2460-8.
- ALDER, N. N., JENSEN, R. E. & JOHNSON, A. E. 2008. Fluorescence mapping of mitochondrial TIM23 complex reveals a water-facing, substrate-interacting helix surface. *Cell*, 134, 439-50.
- ALLEN, S. J., CURRAN, A. R., TEMPLER, R. H., MEIJBERG, W. & BOOTH, P. J. 2004. Folding Kinetics of an  $\alpha$  Helical Membrane Protein in Phospholipid Bilayer Vesicles. *Journal of Molecular Biology*, 342, 1279-1291.
- ANFINSEN, C. B. 1973. Principles that govern the folding of protein chains. *Science*, 181, 223-30.
- ASHRAF, S. S., BENSON, R. E., PAYNE, E. S., HALBLEIB, C. M. & GRON, H. 2004. A novel multi-affinity tag system to produce high levels of soluble and biotinylated proteins in *Escherichia coli*. *Protein Expr Purif*, 33, 238-45.
- ATAKA, K., STRIPP, S. T. & HEBERLE, J. 2013. Surface-enhanced infrared absorption spectroscopy (SEIRAS) to probe monolayers of membrane proteins. *Biochim Biophys Acta*, 1828, 2283-93.
- BADA JUAREZ, J. F., HARPER, A. J., JUDGE, P. J., TONGE, S. R. & WATTS, A. 2019. From polymer chemistry to structural biology: The development of SMA and related amphipathic polymers for membrane protein extraction and solubilisation. *Chem Phys Lipids*, 221, 167-175.
- BAI, X., MORAES, T. F. & REITHMEIER, R. A. F. 2017. Structural biology of solute carrier (SLC) membrane transport proteins. *Mol Membr Biol*, 34, 1-32.
- BAKER, R. P. & URBAN, S. 2012. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol*, 8, 759-68.
- BALCHIN, D., HAYER-HARTL, M. & HARTL, F. U. 2016. In vivo aspects of protein folding and quality control. *Science*, 353, aac4354.
- BARNIOL-XICOTA, M. & VERHELST, S. H. L. 2018. Stable and Functional Rhomboid Proteases in Lipid Nanodiscs by Using Diisobutylene/Maleic Acid Copolymers. *J Am Chem Soc*, 140, 14557-14561.
- BARTOSZEWSKI, R. A., JABLONSKY, M., BARTOSZEWSKA, S., STEVENSON, L., DAI, Q., KAPPES, J., COLLAWN, J. F. & BEBOK, Z. 2010. A synonymous single nucleotide polymorphism in DeltaF508 CFTR alters the secondary structure of the mRNA and the expression of the mutant protein. *J Biol Chem*, 285, 28741-8.
- BAUMANN, A., KERRUTH, S., FITTER, J., BULDT, G., HEBERLE, J., SCHLESINGER, R. & ATAKA, K. 2016. In-Situ Observation of Membrane Protein Folding during Cell-Free Expression. *PLoS One*, 11, e0151051.
- BAZZONE, A., MADEJ, M. G., KABACK, H. R. & FENDLER, K. 2016. pH Regulation of Electrogenic Sugar/H<sup>+</sup> Symport in MFS Sugar Permeases. *PLoS One*, 11, e0156392.
- BECKER, M., GZYL, K. E., ALTAMIRANO, A. M., VUONG, A., URBAN, K. & WIEDEN, H. J. 2012. The 70S ribosome modulates the ATPase activity of *Escherichia coli* YchF. *RNA Biol*, 9, 1288-301.
- BELL, G. I. 1978. Models for the specific adhesion of cells to cells. *Science*, 200, 618-27.
- BERCOVICH-KINORI, A. & BIBI, E. 2015. Co-translational membrane association of the *Escherichia coli* SRP receptor. *J Cell Sci*, 128, 1444-52.
- BERKOVICH, R., GARCIA-MANYES, S., KLAFTER, J., URBACH, M. & FERNANDEZ, J. M. 2010. Hopping around an entropic barrier created by force. *Biochem Biophys Res Commun*, 403, 133-7.
- BERSCH, B., DORR, J. M., HESSEL, A., KILLIAN, J. A. & SCHANDA, P. 2017. Proton-Detected Solid-State NMR Spectroscopy of a Zinc Diffusion Facilitator Protein in Native Nanodiscs. *Angew Chem Int Ed Engl*, 56, 2508-2512.
- BHUSHAN, S., HOFFMANN, T., SEIDELT, B., FRAUENFELD, J., MIELKE, T., BERNINGHAUSEN, O., WILSON, D. N. & BECKMANN, R. 2011. SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol*, 9, e1000581.
- BIBI, E. 2011. Early targeting events during membrane protein biogenesis in *Escherichia coli*. *Biochim Biophys Acta*, 1808, 841-50.
- BIBI, E. & KABACK, H. R. 1990. In vivo expression of the lacY gene in two segments leads to functional lac permease. *Proc Natl Acad Sci U S A*, 87, 4325-9.
- BIBOW, S. 2019. Opportunities and Challenges of Backbone, Sidechain, and RDC Experiments to Study Membrane Protein Dynamics in a Detergent-Free Lipid Environment Using Solution State NMR. *Front Mol Biosci*, 6, 103.
- BIPPES, C. A., GE, L., MEURY, M., HARDER, D., UCURUM, Z., DANIEL, H., FOTIADIS, D. & MULLER, D. J. 2013. Peptide transporter DtpA has two alternate conformations, one of which is promoted by inhibitor binding. *Proc Natl Acad Sci U S A*, 110, E3978-86.
- BIPPES, C. A. & MULLER, D. J. 2011. High-resolution atomic force microscopy and spectroscopy of native membrane proteins. *Reports on Progress in Physics*, 74.

- BIPPES, C. A., ZELTINA, A., CASAGRANDE, F., RATERA, M., PALACIN, M., MULLER, D. J. & FOTIADIS, D. 2009. Substrate binding tunes conformational flexibility and kinetic stability of an amino acid antiporter. *J Biol Chem*, 284, 18651-63.
- BISCHOFF, L., WICKLES, S., BERNINGHAUSEN, O., VAN DER SLUIS, E. O. & BECKMANN, R. 2014. Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nat Commun*, 5, 4103.
- BOGDANOV, M. & DOWHAN, W. 2012. Lipid-dependent generation of dual topology for a membrane protein. *J Biol Chem*, 287, 37939-48.
- BOGDANOV, M., DOWHAN, W. & VITRAC, H. 2014. Lipids and topological rules governing membrane protein assembly. *Biochim Biophys Acta*, 1843, 1475-88.
- BOGDANOV, M., HEACOCK, P. N. & DOWHAN, W. 2002. A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition. *EMBO J*, 21, 2107-16.
- BOOTH, P. J. & FAROOQ, A. 1997. Intermediates in the assembly of bacteriorhodopsin investigated by time-resolved absorption spectroscopy. *Eur J Biochem*, 246, 674-80.
- BOOTH, P. J., FLITSCH, S. L., STERN, L. J., GREENHALGH, D. A., KIM, P. S. & KHORANA, H. G. 1995. Intermediates in the folding of the membrane protein bacteriorhodopsin. *Nat Struct Biol*, 2, 139-43.
- BOOTH, P. J., TEMPLER, R. H., MEIJBERG, W., ALLEN, S. J., CURRAN, A. R. & LORCH, M. 2001. In vitro studies of membrane protein folding. *Crit Rev Biochem Mol Biol*, 36, 501-603.
- BORGIA, A., WILLIAMS, P. M. & CLARKE, J. 2008. Single-molecule studies of protein folding. *Annu Rev Biochem*, 77, 101-25.
- BOUCHIAT, C., WANG, M. D., ALLEMAND, J., STRICK, T., BLOCK, S. M. & CROQUETTE, V. 1999. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. *Biophys J*, 76, 409-13.
- BOWIE, J. U. 2004. Membrane proteins: a new method enters the fold. *Proc Natl Acad Sci U S A*, 101, 3995-6.
- BROCKMANN, K. 2009. The expanding phenotype of GLUT1-deficiency syndrome. *Brain Dev*, 31, 545-52.
- BROECKER, J., EGER, B. T. & ERNST, O. P. 2017. Crystallogenesis of Membrane Proteins Mediated by Polymer-Bounded Lipid Nanodiscs. *Structure*, 25, 384-392.
- BUSTAMANTE, C., MARKO, J. F., SIGGIA, E. D. & SMITH, S. 1994. Entropic elasticity of lambda-phage DNA. *Science*, 265, 1599-600.
- BUTT, H. J. & JASCHKE, M. 1995. Calculation of thermal noise in atomic force microscopy. *Nanotechnology*, 6, 1-7.
- CABRITA, L. D., CASSAGNAU, A. M. E., LAUNAY, H. M. M., WAUDBY, C. A., WLODARSKI, T., CAMILLONI, C., KARYADI, M. E., ROBERTSON, A. L., WANG, X., WENTINK, A. S., GOODSELL, L., WOOLHEAD, C. A., VENDRUSCOLO, M., DOBSON, C. M. & CHRISTODOULOU, J. 2016. A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding. *Nat Struct Mol Biol*, 23, 278-285.
- CABRITA, L. D., HSU, S. T., LAUNAY, H., DOBSON, C. M. & CHRISTODOULOU, J. 2009. Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc Natl Acad Sci U S A*, 106, 22239-44.
- CASSAGNAU, A. M., LAUNAY, H. M., KARYADI, M. E., WANG, X., WAUDBY, C. A., DECKERT, A., ROBERTSON, A. L., CHRISTODOULOU, J. & CABRITA, L. D. 2016. A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat Protoc*, 11, 1492-507.
- CASSAGNAU, A. M. E., CABRITA, L. D. & CHRISTODOULOU, J. 2020. How Does the Ribosome Fold the Proteome? *Annu Rev Biochem*, 89, 389-415.
- CECCONI, C., SHANK, E. A., BUSTAMANTE, C. & MARQUSEE, S. 2005. Direct observation of the three-state folding of a single protein molecule. *Science*, 309, 2057-60.
- CECCONI, C., SHANK, E. A., MARQUSEE, S. & BUSTAMANTE, C. 2011. DNA molecular handles for single-molecule protein-folding studies by optical tweezers. *Methods Mol Biol*, 749, 255-71.
- CELLITTI, J., BERNSTEIN, R. & MARQUSEE, S. 2007. Exploring subdomain cooperativity in T4 lysozyme II: uncovering the C-terminal subdomain as a hidden intermediate in the kinetic folding pathway. *Protein Sci*, 16, 852-62.
- CHANG, Y. C. & BOWIE, J. U. 2014. Measuring membrane protein stability under native conditions. *Proc Natl Acad Sci U S A*, 111, 219-24.
- CHARALAMBOUS, K., MILLER, D., CURNOW, P. & BOOTH, P. J. 2008. Lipid bilayer composition influences small multidrug transporters. *BMC Biochem*, 9, 31.
- CHOI, H. K., MIN, D., KANG, H., SHON, M. J., RAH, S. H., KIM, H. C., JEONG, H., CHOI, H. J., BOWIE, J. U. & YOON, T. Y. 2019. Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science*, 366, 1150-1156.
- CHURCHWARD, M. A., BRANDMAN, D. M., ROGASEVSKAIA, T. & COORSSEN, J. R. 2008. Copper (II) sulfate charring for high sensitivity on-plate fluorescent detection of lipids and sterols: quantitative analyses of the composition of functional secretory vesicles. *J Chem Biol*, 1, 79-87.
- CISNEROS, D. A., OESTERHELT, D. & MULLER, D. J. 2005. Probing origins of molecular interactions stabilizing the membrane proteins halorhodopsin and bacteriorhodopsin. *Structure*, 13, 235-42.
- COLBASEVICI, A., VOSKOBOYNIKOVA, N., OREKHOV, P. S., BOZDAGANYAN, M. E., KARLOVA, M. G., SOKOLOVA, O. S., KLARE, J. P., MULKIDJANIAN, A. Y., SHAITAN, K. V. & STEINHOFF, H. J. 2020. Lipid dynamics in nanoparticles formed by maleic acid-containing copolymers: EPR spectroscopy and molecular dynamics simulations. *Biochim Biophys Acta Biomembr*, 1862, 183207.

- CORIN, K. & BOWIE, J. U. 2020. How bilayer properties influence membrane protein folding. *Protein Sci*, 29, 2348-2362.
- COURNIA, Z., ALLEN, T. W., ANDRICOAEI, I., ANTONNY, B., BAUM, D., BRANNIGAN, G., BUCHETE, N. V., DECKMAN, J. T., DELEMOTTE, L., DEL VAL, C., FRIEDMAN, R., GKEKA, P., HEGE, H. C., HENIN, J., KASIMOVA, M. A., KOLOCOURIS, A., KLEIN, M. L., KHALID, S., LEMIEUX, M. J., LINDOW, N., ROY, M., SELENT, J., TAREK, M., TOFOLEANU, F., VANNI, S., URBAN, S., WALES, D. J., SMITH, J. C. & BONDAR, A. N. 2015. Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *J Membr Biol*, 248, 611-40.
- CRAIG, A. F., CLARK, E. E., SAHU, I. D., ZHANG, R., FRANTZ, N. D., AL-ABDUL-WAHID, M. S., DABNEY-SMITH, C., KONKOLEWICZ, D. & LORIGAN, G. A. 2016. Tuning the size of styrene-maleic acid copolymer-lipid nanoparticles (SMALPs) using RAFT polymerization for biophysical studies. *Biochim Biophys Acta*, 1858, 2931-2939.
- CRUZ-VERA, L. R. & YANOFSKY, C. 2008. Conserved residues Asp16 and Pro24 of TnaC-tRNA<sup>Pro</sup> participate in tryptophan induction of Tna operon expression. *J Bacteriol*, 190, 4791-7.
- CURNOW, P. & BOOTH, P. J. 2007. Combined kinetic and thermodynamic analysis of alpha-helical membrane protein unfolding. *Proc Natl Acad Sci U S A*, 104, 18970-5.
- CURNOW, P. & BOOTH, P. J. 2009. The transition state for integral membrane protein folding. *Proc Natl Acad Sci U S A*, 106, 773-8.
- CURNOW, P., DI BARTOLO, N. D., MORETON, K. M., AJOJE, O. O., SAGGESE, N. P. & BOOTH, P. J. 2011. Stable folding core in the folding transition state of an alpha-helical integral membrane protein. *Proc Natl Acad Sci U S A*, 108, 14133-8.
- CURRAN, A. R., TEMPLER, R. H. & BOOTH, P. J. 1999. Modulation of Folding and Assembly of the Membrane Protein Bacteriorhodopsin by Intermolecular Forces within the Lipid Bilayer†. *Biochemistry*, 38, 9328-9336.
- CYMER, F., HEDMAN, R., ISMAIL, N. & VON HEIJNE, G. 2015a. Exploration of the arrest peptide sequence space reveals arrest-enhanced variants. *J Biol Chem*, 290, 10208-15.
- CYMER, F., ISMAIL, N. & VON HEIJNE, G. 2014. Weak pulling forces exerted on Nin-orientated transmembrane segments during co-translational insertion into the inner membrane of Escherichia coli. *FEBS Lett*, 588, 1930-4.
- CYMER, F. & VON HEIJNE, G. 2013. Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements. *Proc Natl Acad Sci U S A*, 110, 14640-5.
- CYMER, F., VON HEIJNE, G. & WHITE, S. H. 2015b. Mechanisms of integral membrane protein insertion and folding. *J Mol Biol*, 427, 999-1022.
- CZERSKI, L. & SANDERS, C. R. 2000. Functionality of a membrane protein in bicelles. *Anal Biochem*, 284, 327-33.
- DALBEY, R. E., KUHN, A., ZHU, L. & KIEFER, D. 2014. The membrane insertase YidC. *Biochim Biophys Acta*, 1843, 1489-96.
- DALE, H., ANGEVINE, C. M. & KREBS, M. P. 2000. Ordered membrane insertion of an archaeal opsin in vivo. *Proc Natl Acad Sci U S A*, 97, 7847-52.
- DALE, H. & KREBS, M. P. 1999. Membrane insertion kinetics of a protein domain in vivo. The bacterioopsin n terminus inserts co-translationally. *J Biol Chem*, 274, 22693-8.
- DAMAGHI, M., BIPPES, C., KOSTER, S., YILDIZ, O., MARI, S. A., KUHLBRANDT, W. & MULLER, D. J. 2010. pH-dependent interactions guide the folding and gate the transmembrane pore of the beta-barrel membrane protein OmpG. *J Mol Biol*, 397, 878-82.
- DANIELCZAK, B. & KELLER, S. 2018. Collisional lipid exchange among DIBMA-encapsulated nanodiscs (DIBMALPs). *European Polymer Journal*, 109, 206-213.
- DANIELCZAK, B. & KELLER, S. 2020. Lipid exchange among polymer-encapsulated nanodiscs by time-resolved Forster resonance energy transfer. *Methods*, 180, 27-34.
- DANIELCZAK, B., MEISTER, A. & KELLER, S. 2019. Influence of Mg(2+) and Ca(2+) on nanodisc formation by diisobutylene/maleic acid (DIBMA) copolymer. *Chem Phys Lipids*, 221, 30-38.
- DE LA ROSA-TREVIN, J. M., OTON, J., MARABINI, R., ZALDIVAR, A., VARGAS, J., CARAZO, J. M. & SORZANO, C. O. 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J Struct Biol*, 184, 321-8.
- DE LA ROSA-TREVIN, J. M., QUINTANA, A., DEL CANO, L., ZALDIVAR, A., FOCHE, I., GUTIERREZ, J., GOMEZ-BLANCO, J., BURGNET-CASTELL, J., CUENCA-ALBA, J., ABRISHAMI, V., VARGAS, J., OTON, J., SHAROV, G., VILAS, J. L., NAVAS, J., CONESA, P., KAZEMI, M., MARABINI, R., SORZANO, C. O. & CARAZO, J. M. 2016. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. *J Struct Biol*, 195, 93-9.
- DEL RIO, A., DUTTA, K., CHAVEZ, J., UBARRETXENA-BELANDIA, I. & GHOSE, R. 2007. Solution structure and dynamics of the N-terminal cytosolic domain of rhomboid intramembrane protease from Pseudomonas aeruginosa: insights into a functional role in intramembrane proteolysis. *J Mol Biol*, 365, 109-22.
- DEL RIO, A., PEREZ-JIMENEZ, R., LIU, R., ROCA-CUSACHS, P., FERNANDEZ, J. M. & SHEETZ, M. P. 2009. Stretching single talin rod molecules activates vinculin binding. *Science*, 323, 638-41.
- DENG, D., XU, C., SUN, P., WU, J., YAN, C., HU, M. & YAN, N. 2014. Crystal structure of the human glucose transporter GLUT1. *Nature*, 510, 121-5.

- DENISOV, I. G., GRINKOVA, Y. V., LAZARIDES, A. A. & SLIGAR, S. G. 2004. Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size. *J Am Chem Soc*, 126, 3477-87.
- DENISOV, I. G. & SLIGAR, S. G. 2017. Nanodiscs in Membrane Biochemistry and Biophysics. *Chem Rev*, 117, 4669-4713.
- DESAI, V. P., FRANK, F., LEE, A., RIGHINI, M., LANCASTER, L., NOLLER, H. F., TINOCO, I., JR. & BUSTAMANTE, C. 2019. Co-temporal Force and Fluorescence Measurements Reveal a Ribosomal Gear Shift Mechanism of Translation Regulation by Structured mRNAs. *Mol Cell*, 75, 1007-1019 e5.
- DILL, K. A. & CHAN, H. S. 1997. From Levinthal to pathways to funnels. *Nat Struct Biol*, 4, 10-9.
- DORR, J. M., KOORENGEVEL, M. C., SCHAFER, M., PROKOFYEV, A. V., SCHEIDELAAR, S., VAN DER CRUIJSEN, E. A., DAFFORN, T. R., BALDUS, M. & KILLIAN, J. A. 2014. Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K<sup>+</sup> channel: the power of native nanodiscs. *Proc Natl Acad Sci U S A*, 111, 18607-12.
- DOWHAN, W. & BOGDANOV, M. 2011. Lipid-protein interactions as determinants of membrane protein structure and function. *Biochem Soc Trans*, 39, 767-74.
- DRAYCHEVA, A., LEE, S. & WINTERMEYER, W. 2018. Cotranslational protein targeting to the membrane: Nascent-chain transfer in a quaternary complex formed at the translocon. *Sci Rep*, 8, 9922.
- DYNSKA-KUKULSKA, K., CIESIELSKI, W. & ZAKRZEWSKI, R. 2013. The use of a new, modified Dittmer-Lester spray reagent for phospholipid determination by the TLC image analysis technique. *Biomed Chromatogr*, 27, 458-65.
- DZIONARA, M., KALTSCHMIDT, E. & WITTMANN, H. G. 1970. Ribosomal proteins. 8. Molecular weights of isolated ribosomal proteins of *Escherichia coli*. *Proc Natl Acad Sci U S A*, 67, 1909-13.
- EJIMA, D., YUMIOKA, R., ARAKAWA, T. & TSUMOTO, K. 2005. Arginine as an effective additive in gel permeation chromatography. *J Chromatogr A*, 1094, 49-55.
- ENGELMAN, D. M. 2005. Membranes are more mosaic than fluid. *Nature*, 438, 578-80.
- EVANS, E. & RITCHIE, K. 1997. Dynamic strength of molecular adhesion bonds. *Biophys J*, 72, 1541-55.
- EVANS, M. S., UGRINOV, K. G., FRESE, M. A. & CLARK, P. L. 2005. Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat Methods*, 2, 757-62.
- FEDYUKINA, D. V. & CAVAGNERO, S. 2011. Protein folding at the exit tunnel. *Annu Rev Biophys*, 40, 337-59.
- FINDLAY, H. E. & BOOTH, P. J. 2006. The biological significance of lipid-protein interactions. *J Phys Condens Matter*, 18, S1281-91.
- FINDLAY, H. E. & BOOTH, P. J. 2017. The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition. *Sci Rep*, 7, 13056.
- FINDLAY, H. E., RUTHERFORD, N. G., HENDERSON, P. J. & BOOTH, P. J. 2010. Unfolding free energy of a two-domain transmembrane sugar transport protein. *Proc Natl Acad Sci U S A*, 107, 18451-6.
- FIORI, M. C., JIANG, Y., ALTENBERG, G. A. & LIANG, H. 2017. Polymer-encased nanodiscs with improved buffer compatibility. *Sci Rep*, 7, 7432.
- FLORES, J. A., HADDAD, B. G., DOLAN, K. A., MYERS, J. B., YOSHIOKA, C. C., COPPERMAN, J., ZUCKERMAN, D. M. & REICHOW, S. L. 2020. Connexin-46/50 in a dynamic lipid environment resolved by CryoEM at 1.9 Å. *Nat Commun*, 11, 4331.
- FOLCH, J., LEES, M. & SLOANE STANLEY, G. H. 1957. A simple method for the isolation and purification of total lipides from animal tissues. *J Biol Chem*, 226, 497-509.
- FOO, A. C., HARVEY, B. G., METZ, J. J. & GOTO, N. K. 2015. Influence of hydrophobic mismatch on the catalytic activity of *Escherichia coli* GlpG rhomboid protease. *Protein Sci*, 24, 464-73.
- FOSTER, W., HELM, A., TURNBULL, I., GULATI, H., YANG, B., VERKMAN, A. S. & SKACH, W. R. 2000. Identification of sequence determinants that direct different intracellular folding pathways for aquaporin-1 and aquaporin-4. *J Biol Chem*, 275, 34157-65.
- FRYDMAN, J., ERDJUMENT-BROMAGE, H., TEMPST, P. & HARTL, F. U. 1999. Co-translational domain folding as the structural basis for the rapid de novo folding of firefly luciferase. *Nat Struct Biol*, 6, 697-705.
- GABASHVILI, I. S., AGRAWAL, R. K., SPAHN, C. M. T., GRASSUCCI, R. A., SVERGUN, D. I., FRANK, J. & PENCZEK, P. 2000. Solution Structure of the *E. coli* 70S Ribosome at 11.5 Å Resolution. *Cell*, 100, 537-549.
- GANAPATHY, S., OPDAM, L., HONTANI, Y., FREHAN, S., CHEN, Q., HELLINGWERF, K. J., DE GROOT, H. J. M., KENNIS, J. T. M. & DE GRIP, W. J. 2020. Membrane matters: The impact of a nanodisc-bilayer or a detergent microenvironment on the properties of two eubacterial rhodopsins. *Biochim Biophys Acta Biomembr*, 1862, 183113.
- GARCIA-MANYES, S., BRUJIC, J., BADILLA, C. L. & FERNANDEZ, J. M. 2007. Force-clamp spectroscopy of single-protein monomers reveals the individual unfolding and folding pathways of I27 and ubiquitin. *Biophys J*, 93, 2436-46.
- GHASRIANI, H., KWOK, J. K., SHERRATT, A. R., FOO, A. C., QURESHI, T. & GOTO, N. K. 2014. Micelle-catalyzed domain swapping in the GlpG rhomboid protease cytoplasmic domain. *Biochemistry*, 53, 5907-15.
- GOLD, V. A., DUONG, F. & COLLINSON, I. 2007. Structure and function of the bacterial Sec translocon. *Mol Membr Biol*, 24, 387-94.

- GOLDMAN, D. H., KAISER, C. M., MILIN, A., RIGHINI, M., TINOCO, I., JR. & BUSTAMANTE, C. 2015. Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science*, 348, 457-60.
- GONG, F., ITO, K., NAKAMURA, Y. & YANOFSKY, C. 2001. The mechanism of tryptophan induction of tryptophanase operon expression: tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNA(Pro). *Proc Natl Acad Sci U S A*, 98, 8997-9001.
- GONG, F. & YANOFSKY, C. 2002. Instruction of translating ribosome by nascent peptide. *Science*, 297, 1864-7.
- GRETHEN, A., OLUWOLE, A. O., DANIELCZAK, B., VARGAS, C. & KELLER, S. 2017. Thermodynamics of nanodisc formation mediated by styrene/maleic acid (2:1) copolymer. *Sci Rep*, 7, 11517.
- GULAMHUSSEIN, A. A., UDDIN, R., TIGHE, B. J., POYNER, D. R. & ROTHNIE, A. J. 2020. A comparison of SMA (styrene maleic acid) and DIBMA (di-isobutylene maleic acid) for membrane protein purification. *Biochim Biophys Acta Biomembr*, 1862, 183281.
- GULATI, S., JAMSHAD, M., KNOWLES, T. J., MORRISON, K. A., DOWNING, R., CANT, N., COLLINS, R., KOENDERINK, J. B., FORD, R. C., OVERDUIN, M., KERR, I. D., DAFFORN, T. R. & ROTHNIE, A. J. 2014. Detergent-free purification of ABC (ATP-binding-cassette) transporters. *Biochem J*, 461, 269-78.
- GUMBART, J., SCHREINER, E., WILSON, D. N., BECKMANN, R. & SCHULTEN, K. 2012. Mechanisms of SecM-mediated stalling in the ribosome. *Biophys J*, 103, 331-41.
- GUO, R., GAFFNEY, K., YANG, Z., KIM, M., SUNGSUWAN, S., HUANG, X., HUBBELL, W. L. & HONG, H. 2016. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol*, 12, 353-360.
- GUO, Y. 2020. Be Cautious with Crystal Structures of Membrane Proteins or Complexes Prepared in Detergents. *Crystals (Basel)*, 10.
- HA, Y., AKIYAMA, Y. & XUE, Y. 2013. Structure and mechanism of rhomboid protease. *J Biol Chem*, 288, 15430-6.
- HAFFKE, M., DUCKELY, M., BERGSDORF, C., JAAKOLA, V. P. & SHRESTHA, B. 2020. Development of a biochemical and biophysical suite for integral membrane protein targets: A review. *Protein Expr Purif*, 167, 105545.
- HARAYAMA, T. & RIEZMAN, H. 2018. Understanding the diversity of membrane lipid composition. *Nat Rev Mol Cell Biol*, 19, 281-296.
- HARDIN, N. Z., RAVULA, T., MAURO, G. D. & RAMAMOORTHY, A. 2019. Hydrophobic Functionalization of Polyacrylic Acid as a Versatile Platform for the Development of Polymer Lipid Nanodisks. *Small*, 15, e1804813.
- HARRIS, N. J., CHARALAMBOUS, K., FINDLAY, H. E. & BOOTH, P. J. 2018. Lipids modulate the insertion and folding of the nascent chains of alpha helical membrane proteins. *Biochem Soc Trans*, 46, 1355-1366.
- HARRIS, N. J., FINDLAY, H. E., SANDERS, M. R., KEDZIERSKI, M., DOS SANTOS, A. & BOOTH, P. J. 2017a. Comparative stability of Major Facilitator Superfamily transport proteins. *Eur Biophys J*, 46, 655-663.
- HARRIS, N. J., FINDLAY, H. E., SIMMS, J., LIU, X. & BOOTH, P. J. 2014. Relative domain folding and stability of a membrane transport protein. *J Mol Biol*, 426, 1812-25.
- HARRIS, N. J., PELLOWE, G. A. & BOOTH, P. J. 2020. Cell-free expression tools to study co-translational folding of alpha helical membrane transporters. *Sci Rep*, 10, 9125.
- HARRIS, N. J., READING, E., ATAKA, K., GRZEGORZEWSKI, L., CHARALAMBOUS, K., LIU, X., SCHLESINGER, R., HEBERLE, J. & BOOTH, P. J. 2017b. Structure formation during translocon-unassisted co-translational membrane protein folding. *Sci Rep*, 7, 8021.
- HEDIGER, M. A., CLEMENCON, B., BURRIER, R. E. & BRUFORD, E. A. 2013. The ABCs of membrane transporters in health and disease (SLC series): introduction. *Mol Aspects Med*, 34, 95-107.
- HELLWIG, N., PEETZ, O., AHDASH, Z., TASCAN, I., BOOTH, P. J., MIKUSEVIC, V., DISKOWSKI, M., POLITIS, A., HELLMICH, Y., HANELT, I., READING, E. & MORGNER, N. 2018. Native mass spectrometry goes more native: investigation of membrane protein complexes directly from SMALPs. *Chem Commun (Camb)*, 54, 13702-13705.
- HERMANSSON, M. & VON HEIJNE, G. 2003. Inter-helical hydrogen bond formation during membrane protein integration into the ER membrane. *J Mol Biol*, 334, 803-9.
- HESKETH, S. J., KLEBL, D. P., HIGGINS, A. J., THOMSEN, M., PICKLES, I. B., SOBOTT, F., SIVAPRASADARAO, A., POSTIS, V. L. G. & MUENCH, S. P. 2020. Styrene maleic-acid lipid particles (SMALPs) into detergent or amphipols: An exchange protocol for membrane protein characterisation. *Biochim Biophys Acta Biomembr*, 1862, 183192.
- HESSA, T., MEINDL-BEINKER, N. M., BERNSEL, A., KIM, H., SATO, Y., LERCH-BADER, M., NILSSON, I., WHITE, S. H. & VON HEIJNE, G. 2007. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature*, 450, 1026-30.
- HILL, W. E., ROSSETTI, G. P. & VAN HOLDE, K. E. 1969. Physical studies of ribosomes from Escherichia coli. *J Mol Biol*, 44, 263-77.
- HOFFMANN, A., BECKER, A. H., ZACHMANN-BRAND, B., DEUERLING, E., BUKAU, B. & KRAMER, G. 2012. Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding. *Mol Cell*, 48, 63-74.
- HOFFMANN, A., MERZ, F., RUTKOWSKA, A., ZACHMANN-BRAND, B., DEUERLING, E. & BUKAU, B. 2006. Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J Biol Chem*, 281, 6539-45.
- HOLTkamp, W., KOKIC, G., JAGER, M., MITTELSTAET, J., KOMAR, A. A. & RODNINA, M. V. 2015. Cotranslational protein folding on the ribosome monitored in real time. *Science*, 350, 1104-7.



- HONG, H., BLOIS, T. M., CAO, Z. & BOWIE, J. U. 2010. Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc Natl Acad Sci U S A*, 107, 19802-7.
- HUGHES, M. L. & DOUGAN, L. 2016. The physics of pulling polyproteins: a review of single molecule force spectroscopy using the AFM to study protein unfolding. *Rep Prog Phys*, 79, 076601.
- HUMMER, G. & SZABO, A. 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc Natl Acad Sci U S A*, 98, 3658-61.
- HUMMER, G. & SZABO, A. 2005. Free energy surfaces from single-molecule force spectroscopy. *Acc Chem Res*, 38, 504-13.
- HUTTER, J. L. & BECHHOEFER, J. 1993. Calibration of atomic-force microscope tips. *Review of Scientific Instruments*, 64, 1868-1873.
- ISMAIL, N., HEDMAN, R., LINDEN, M. & VON HEIJNE, G. 2015. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat Struct Mol Biol*, 22, 145-149.
- ISMAIL, N., HEDMAN, R., SCHILLER, N. & VON HEIJNE, G. 2012. A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat Struct Mol Biol*, 19, 1018-22.
- JAMSHAD, M., CHARLTON, J., LIN, Y. P., ROUTLEDGE, S. J., BAWA, Z., KNOWLES, T. J., OVERDUIN, M., DEKKER, N., DAFFORN, T. R., BILL, R. M., POYNER, D. R. & WHEATLEY, M. 2015. G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Biosci Rep*, 35.
- JANOVJAK, H., KESSLER, M., OESTERHELT, D., GAUB, H. & MULLER, D. J. 2003. Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO J*, 22, 5220-9.
- JANOVJAK, H., SAPRA, K. T., KEDROV, A. & MULLER, D. J. 2008. From valleys to ridges: exploring the dynamic energy landscape of single membrane proteins. *Chemphyschem*, 9, 954-66.
- JEFFERSON, R. E., MIN, D., CORIN, K., WANG, J. Y. & BOWIE, J. U. 2018. Applications of Single-Molecule Methods to Membrane Protein Folding Studies. *J Mol Biol*, 430, 424-437.
- JENSEN, M. K., SAMELSON, A. J., STEWARD, A., CLARKE, J. & MARQUSEE, S. 2020. The folding and unfolding behavior of ribonuclease H on the ribosome. *J Biol Chem*, 295, 11410-11417.
- JENSEN, P. F., JORGENSEN, T. J., KOEFOED, K., NYGAARD, F. & SEN, J. W. 2013. Affinity capture of biotinylated proteins at acidic conditions to facilitate hydrogen/deuterium exchange mass spectrometry analysis of multimeric protein complexes. *Anal Chem*, 85, 7052-9.
- JIANG, X., WU, J., KE, M., ZHANG, S., YUAN, Y., LIN, J. Y. & YAN, N. 2019. Engineered Xyle as a tool for mechanistic investigation and ligand discovery of the glucose transporters GLUTs. *Cell Discov*, 5, 14.
- JOHNSON, A. E. 2005. The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer. *FEBS Lett*, 579, 916-20.
- KAISER, C. M., GOLDMAN, D. H., CHODERA, J. D., TINOCO, I., JR. & BUSTAMANTE, C. 2011. The ribosome modulates nascent protein folding. *Science*, 334, 1723-7.
- KAISER, C. M. & LIU, K. 2018. Folding up and Moving on-Nascent Protein Folding on the Ribosome. *J Mol Biol*, 430, 4580-4591.
- KATER, L., FRIEG, B., BERNINGHAUSEN, O., GOHLKE, H., BECKMANN, R. & KEDROV, A. 2019. Partially inserted nascent chain unzips the lateral gate of the Sec translocon. *EMBO Rep*, 20, e48191.
- KAWAMURA, S., COLOZO, A. T., MULLER, D. J. & PARK, P. S. 2010. Conservation of molecular interactions stabilizing bovine and mouse rhodopsin. *Biochemistry*, 49, 10412-20.
- KAWAMURA, S., GERSTUNG, M., COLOZO, A. T., HELENIUS, J., MAEDA, A., BEERENWINKEL, N., PARK, P. S. & MULLER, D. J. 2013. Kinetic, energetic, and mechanical differences between dark-state rhodopsin and opsin. *Structure*, 21, 426-37.
- KEDROV, A., APPEL, M., BAUMANN, H., ZIEGLER, C. & MULLER, D. J. 2008. Examining the dynamic energy landscape of an antiporter upon inhibitor binding. *J Mol Biol*, 375, 1258-66.
- KEDROV, A., JANOVJAK, H., ZIEGLER, C., KUHLBRANDT, W. & MULLER, D. J. 2006a. Observing folding pathways and kinetics of a single sodium-proton antiporter from Escherichia coli. *J Mol Biol*, 355, 2-8.
- KEDROV, A., KRIEG, M., ZIEGLER, C., KUHLBRANDT, W. & MULLER, D. J. 2005. Locating ligand binding and activation of a single antiporter. *EMBO Rep*, 6, 668-74.
- KEDROV, A., SUSTARSIC, M., DE KEYZER, J., CAUMANN, J. J., WU, Z. C. & DRIESSEN, A. J. 2013. Elucidating the native architecture of the YidC: ribosome complex. *J Mol Biol*, 425, 4112-24.
- KEDROV, A., WEGMANN, S., SMITS, S. H., GOSWAMI, P., BAUMANN, H. & MULLER, D. J. 2007. Detecting molecular interactions that stabilize, activate and guide ligand-binding of the sodium/proton antiporter MjNhaP1 from Methanococcus jannaschii. *J Struct Biol*, 159, 290-301.
- KEDROV, A., ZIEGLER, C., JANOVJAK, H., KUHLBRANDT, W. & MULLER, D. J. 2004. Controlled unfolding and refolding of a single sodium-proton antiporter using atomic force microscopy. *J Mol Biol*, 340, 1143-52.
- KEDROV, A., ZIEGLER, C. & MULLER, D. J. 2006b. Differentiating ligand and inhibitor interactions of a single antiporter. *J Mol Biol*, 362, 925-32.
- KEMPF, N., REMES, C., LEDESCH, R., ZUCHNER, T., HOFIG, H., RITTER, I., KATRANIDIS, A. & FITTER, J. 2017. A Novel Method to Evaluate Ribosomal Performance in Cell-Free Protein Synthesis Systems. *Sci Rep*, 7, 46753.
- KHUSHOO, A., YANG, Z., JOHNSON, A. E. & SKACH, W. R. 2011. Ligand-driven vectorial folding of ribosome-bound human CFTR NBD1. *Mol Cell*, 41, 682-92.

- KIMANIUS, D., FORSBERG, B. O., SCHERES, S. H. & LINDAHL, E. 2016. Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *Elife*, 5.
- KNOWLES, T. J., FINKA, R., SMITH, C., LIN, Y. P., DAFFORN, T. & OVERDUIN, M. 2009. Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *J Am Chem Soc*, 131, 7484-5.
- KOMAR, A. A., KOMMER, A., KRASHENINNIKOV, I. A. & SPIRIN, A. S. 1997. Cotranslational folding of globin. *J Biol Chem*, 272, 10646-51.
- KOMAR, J., ALVIRA, S., SCHULZE, R. J., MARTIN, R., LYCKLAMA, A. N. J. A., LEE, S. C., DAFFORN, T. R., DECKERS-HEBESTREIT, G., BERGER, I., SCHAFFITZEL, C. & COLLINSON, I. 2016. Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Biochem J*, 473, 3341-54.
- KOPF, A. H., DORR, J. M., KOORENGEVEL, M. C., ANTONICIELLO, F., JAHN, H. & KILLIAN, J. A. 2020. Factors influencing the solubilization of membrane proteins from Escherichia coli membranes by styrene-maleic acid copolymers. *Biochim Biophys Acta Biomembr*, 1862, 183125.
- KREUTZBERGER, A. J. B., JI, M., AARON, J., MIHALJEVIC, L. & URBAN, S. 2019. Rhomboid distorts lipids to break the viscosity-imposed speed limit of membrane diffusion. *Science*, 363.
- KURUMA, Y. & UEDA, T. 2015. The PURE system for the cell-free synthesis of membrane proteins. *Nat Protoc*, 10, 1328-44.
- KUTLUBAEVA, Z. S., CHETVERINA, H. V. & CHETVERIN, A. B. 2017. The Contribution of Ribosomal Protein S1 to the Structure and Function of Qbeta Replicase. *Acta Naturae*, 9, 24-30.
- LE, S., LIU, R., LIM, C. T. & YAN, J. 2016. Uncovering mechanosensing mechanisms at the single protein level using magnetic tweezers. *Methods*, 94, 13-8.
- LEE, S. C., COLLINS, R., LIN, Y. P., JAMSHAD, M., BROUGHTON, C., HARRIS, S. A., HANSON, B. S., TOGNOLONI, C., PARSLow, R. A., TERRY, A. E., RODGER, A., SMITH, C. J., EDLER, K. J., FORD, R., ROPER, D. I. & DAFFORN, T. R. 2019. Nano-encapsulated Escherichia coli Divisome Anchor ZipA, and in Complex with FtsZ. *Sci Rep*, 9, 18712.
- LEE, S. C., KNOWLES, T. J., POSTIS, V. L., JAMSHAD, M., PARSLow, R. A., LIN, Y. P., GOLDMAN, A., SRIDHAR, P., OVERDUIN, M., MUENCH, S. P. & DAFFORN, T. R. 2016. A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nat Protoc*, 11, 1149-62.
- LEES, J. G., SMITH, B. R., WIEN, F., MILES, A. J. & WALLACE, B. A. 2004. CDtool-an integrated software package for circular dichroism spectroscopic data processing, analysis, and archiving. *Anal Biochem*, 332, 285-9.
- LEMBERG, M. K., MENENDEZ, J., MISIK, A., GARCIA, M., KOTH, C. M. & FREEMAN, M. 2005. Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J*, 24, 464-72.
- LEMIEUX, M. J., FISCHER, S. J., CHERNEY, M. M., BATEMAN, K. S. & JAMES, M. N. 2007. The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. *Proc Natl Acad Sci U S A*, 104, 750-4.
- LEMIEUX, M. J. & OVERDUIN, M. 2021. Structure and function of proteins in membranes and nanodiscs. *Biochim Biophys Acta Biomembr*, 1863, 183445.
- LEVINE, M. V., CUENDET, M. A., KHELASHVILI, G. & WEINSTEIN, H. 2016. Allosteric Mechanisms of Molecular Machines at the Membrane: Transport by Sodium-Coupled Symporters. *Chem Rev*, 116, 6552-87.
- LÉVY, R. & MAALOU, M. 2002. Measuring the spring constant of atomic force microscope cantilevers: thermal fluctuations and other methods. *Nanotechnology*, 13, 33-37.
- LI, D., LI, J., ZHUANG, Y., ZHANG, L., XIONG, Y., SHI, P. & TIAN, C. 2015. Nano-size uni-lamellar lipidisq improved in situ auto-phosphorylation analysis of E. coli tyrosine kinase using (19)F nuclear magnetic resonance. *Protein Cell*, 6, 229-33.
- LIAO, M., CAO, E., JULIUS, D. & CHENG, Y. 2013. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature*, 504, 107-12.
- LIND, T. K., WACKLIN, H., SCHILLER, J., MOULIN, M., HAERTLEIN, M., POMORSKI, T. G. & CARDENAS, M. 2015. Formation and Characterization of Supported Lipid Bilayers Composed of Hydrogenated and Deuterated Escherichia coli Lipids. *PLoS One*, 10, e0144671.
- LIU, K., MACIUBA, K. & KAISER, C. M. 2019. The Ribosome Cooperates with a Chaperone to Guide Multi-domain Protein Folding. *Mol Cell*, 74, 310-319 e7.
- LIU, K., REHFUS, J. E., MATTSON, E. & KAISER, C. M. 2017. The ribosome destabilizes native and non-native structures in a nascent multidomain protein. *Protein Sci*, 26, 1439-1451.
- LIUTKUTE, M., SAMATOVA, E. & RODNINA, M. V. 2020. Cotranslational Folding of Proteins on the Ribosome. *Biomolecules*, 10.
- LLORIS-GARCERA, P., KLINTER, S., CHEN, L., SKYNNER, M. J., LOVING, R. & FRAUENFELD, J. 2020. DirectMX - One-Step Reconstitution of Membrane Proteins From Crude Cell Membranes Into Salipro Nanoparticles. *Front Bioeng Biotechnol*, 8, 215.
- LOMIZE, M. A., POGOZHEVA, I. D., JOO, H., MOSBERG, H. I. & LOMIZE, A. L. 2012. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res*, 40, D370-6.
- LU, H. & BOOTH, P. J. 2000. The final stages of folding of the membrane protein bacteriorhodopsin occur by kinetically indistinguishable parallel folding paths that are mediated by pH. *J Mol Biol*, 299, 233-43.

- LU, W., SCHAFER, N. P. & WOLYNES, P. G. 2018. Energy landscape underlying spontaneous insertion and folding of an alpha-helical transmembrane protein into a bilayer. *Nat Commun*, 9, 4949.
- LU, Y., TURNBULL, I. R., BRAGIN, A., CARVETH, K., VERKMAN, A. S. & SKACH, W. R. 2000. Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum. *Mol Biol Cell*, 11, 2973-85.
- LUTHRA, A., GREGORY, M., GRINKOVA, Y. V., DENISOV, I. G. & SLIGAR, S. G. 2013. Nanodiscs in the studies of membrane-bound cytochrome P450 enzymes. *Methods Mol Biol*, 987, 115-27.
- MADEJ, M. G., SUN, L., YAN, N. & KABACK, H. R. 2014. Functional architecture of MFS D-glucose transporters. *Proc Natl Acad Sci U S A*, 111, E719-27.
- MAEGAWA, S., ITO, K. & AKIYAMA, Y. 2005. Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry*, 44, 13543-52.
- MARINKO, J. T., HUANG, H., PENN, W. D., CAPRA, J. A., SCHLEBACH, J. P. & SANDERS, C. R. 2019. Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem Rev*, 119, 5537-5606.
- MARKWELL, M. A., HAAS, S. M., BIEBER, L. L. & TOLBERT, N. E. 1978. A modification of the Lowry procedure to simplify protein determination in membrane and lipoprotein samples. *Anal Biochem*, 87, 206-10.
- MARTENS, C., SHEKHAR, M., BORYSIK, A. J., LAU, A. M., READING, E., TAJKHORSHID, E., BOOTH, P. J. & POLITIS, A. 2018. Direct protein-lipid interactions shape the conformational landscape of secondary transporters. *Nat Commun*, 9, 4151.
- MARTOGLIO, B., HOFMANN, M. W., BRUNNER, J. & DOBBERSTEIN, B. 1995. The protein-conducting channel in the membrane of the endoplasmic reticulum is open laterally toward the lipid bilayer. *Cell*, 81, 207-214.
- MERCIER, E. & RODNINA, M. V. 2018. Co-Translational Folding Trajectory of the HemK Helical Domain. *Biochemistry*, 57, 3460-3464.
- MERCIER, E., WINTERMEYER, W. & RODNINA, M. V. 2020. Co-translational insertion and topogenesis of bacterial membrane proteins monitored in real time. *EMBO J*, 39, e104054.
- MILLER, D. M., FINDLAY, H. E., CES, O., TEMPLER, R. H. & BOOTH, P. J. 2016. Light-activated control of protein channel assembly mediated by membrane mechanics. *Nanotechnology*, 27, 494004.
- MIN, D., ARBING, M. A., JEFFERSON, R. E. & BOWIE, J. U. 2016. A simple DNA handle attachment method for single molecule mechanical manipulation experiments. *Protein Sci*, 25, 1535-44.
- MIN, D., JEFFERSON, R. E., BOWIE, J. U. & YOON, T. Y. 2015. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat Chem Biol*, 11, 981-7.
- MIN, D., JEFFERSON, R. E., QI, Y., WANG, J. Y., ARBING, M. A., IM, W. & BOWIE, J. U. 2018. Unfolding of a CIC chloride transporter retains memory of its evolutionary history. *Nat Chem Biol*, 14, 489-496.
- MOIN, S. M. & URBAN, S. 2012. Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife*, 1, e00173.
- MONKEMEYER, L., KLAIPS, C. L., BALCHIN, D., KORNER, R., HARTL, F. U. & BRACHER, A. 2019. Chaperone Function of Hgh1 in the Biogenesis of Eukaryotic Elongation Factor 2. *Mol Cell*, 74, 88-100 e9.
- MORA, M., STANNARD, A. & GARCIA-MANYES, S. 2020. The nanomechanics of individual proteins. *Chem Soc Rev*, 49, 6816-6832.
- MOREIN, S., ANDERSSON, A., RILFORS, L. & LINDBLOM, G. 1996. Wild-type Escherichia coli cells regulate the membrane lipid composition in a "window" between gel and non-lamellar structures. *J Biol Chem*, 271, 6801-9.
- MOSS, K., HELM, A., LU, Y., BRAGIN, A. & SKACH, W. R. 1998. Coupled translocation events generate topological heterogeneity at the endoplasmic reticulum membrane. *Mol Biol Cell*, 9, 2681-97.
- MULLER, D. J., KESSLER, M., OESTERHELT, F., MOLLER, C., OESTERHELT, D. & GAUB, H. 2002. Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys J*, 83, 3578-88.
- MULLER, D. J., SAPRA, K. T., SCHEURING, S., KEDROV, A., FREDERIX, P. L., FOTIADIS, D. & ENGEL, A. 2006. Single-molecule studies of membrane proteins. *Curr Opin Struct Biol*, 16, 489-95.
- MUTO, H., NAKATOGAWA, H. & ITO, K. 2006. Genetically encoded but nonpolypeptide prolyl-tRNA functions in the A site for SecM-mediated ribosomal stall. *Mol Cell*, 22, 545-52.
- NAGAMORI, S., VAZQUEZ-IBAR, J. L., WEINGLASS, A. B. & KABACK, H. R. 2003. In vitro synthesis of lactose permease to probe the mechanism of membrane insertion and folding. *J Biol Chem*, 278, 14820-6.
- NAKATOGAWA, H. & ITO, K. 2001. Secretion monitor, SecM, undergoes self-translation arrest in the cytosol. *Mol Cell*, 7, 185-92.
- NAKATOGAWA, H. & ITO, K. 2002. The Ribosomal Exit Tunnel Functions as a Discriminating Gate. *Cell*, 108, 629-636.
- NEUMAN, K. C. & NAGY, A. 2008. Single-molecule force spectroscopy: optical tweezers, magnetic tweezers and atomic force microscopy. *Nat Methods*, 5, 491-505.
- NIERHAUS, K. H. 2014. Mg<sup>2+</sup>, K<sup>+</sup>, and the ribosome. *J Bacteriol*, 196, 3817-9.
- NIESEN, M. J. M., MULLER-LUCKS, A., HEDMAN, R., VON HEIJNE, G. & MILLER, T. F., 3RD 2018. Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon. *Biophys J*, 115, 1885-1894.
- OESTERHELT, D. & STOECKENIUS, W. 1971. Rhodopsin-like protein from the purple membrane of Halobacterium halobium. *Nat New Biol*, 233, 149-52.

- OESTERHELT, F., OESTERHELT, D., PFEIFFER, M., ENGEL, A., GAUB, H. E. & MULLER, D. J. 2000. Unfolding pathways of individual bacteriorhodopsins. *Science*, 288, 143-6.
- OLUWOLE, A. O., DANIELCZAK, B., MEISTER, A., BABALOLA, J. O., VARGAS, C. & KELLER, S. 2017a. Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angew Chem Int Ed Engl*, 56, 1919-1924.
- OLUWOLE, A. O., KLINGLER, J., DANIELCZAK, B., BABALOLA, J. O., VARGAS, C., PABST, G. & KELLER, S. 2017b. Formation of Lipid-Bilayer Nanodiscs by Diisobutylene/Maleic Acid (DIBMA) Copolymer. *Langmuir*, 33, 14378-14388.
- ORWICK, M. C., JUDGE, P. J., PROCEK, J., LINDHOLM, L., GRAZIADEI, A., ENGEL, A., GROBNER, G. & WATTS, A. 2012. Detergent-free formation and physicochemical characterization of nanosized lipid-polymer complexes: Lipodisq. *Angew Chem Int Ed Engl*, 51, 4653-7.
- OVERDUIN, M. & ESMAILI, M. 2019. Memtein: The fundamental unit of membrane-protein structure and function. *Chem Phys Lipids*, 218, 73-84.
- PARK, E., MENETRET, J. F., GUMBART, J. C., LUDTKE, S. J., LI, W., WHYNOT, A., RAPOPORT, T. A. & AKEY, C. W. 2014. Structure of the SecY channel during initiation of protein translocation. *Nature*, 506, 102-6.
- PARK, S. H., WU, J., YAO, Y., SINGH, C., TIAN, Y., MARASSI, F. M. & OPELLA, S. J. 2020. Membrane proteins in magnetically aligned phospholipid polymer discs for solid-state NMR spectroscopy. *Biochim Biophys Acta Biomembr*, 1862, 183333.
- PARMAR, M., RAWSON, S., SCARFF, C. A., GOLDMAN, A., DAFFORN, T. R., MUENCH, S. P. & POSTIS, V. L. G. 2018. Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim Biophys Acta Biomembr*, 1860, 378-383.
- PARMAR, M. J., LOUSA CDE, M., MUENCH, S. P., GOLDMAN, A. & POSTIS, V. L. 2016. Artificial membranes for membrane protein purification, functionality and structure studies. *Biochem Soc Trans*, 44, 877-82.
- PASLAWSKI, W., LILLELUND, O. K., KRISTENSEN, J. V., SCHAFER, N. P., BAKER, R. P., URBAN, S. & OTZEN, D. E. 2015. Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc Natl Acad Sci U S A*, 112, 7978-83.
- PAVLOV, M. Y., WATTS, R. E., TAN, Z., CORNISH, V. W., EHRENBERG, M. & FORSTER, A. C. 2009. Slow peptide bond formation by proline and other N-alkylamino acids in translation. *Proc Natl Acad Sci U S A*, 106, 50-4.
- PENMATSA, A. & GOUAUX, E. 2014. How LeuT shapes our understanding of the mechanisms of sodium-coupled neurotransmitter transporters. *J Physiol*, 592, 863-9.
- PITONZO, D. & SKACH, W. R. 2006. Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon. *Biochim Biophys Acta*, 1758, 976-88.
- POLLOCK, N. L., LEE, S. C., PATEL, J. H., GULAMHUSSEIN, A. A. & ROTHNIE, A. J. 2018. Structure and function of membrane proteins encapsulated in a polymer-bound lipid bilayer. *Biochim Biophys Acta Biomembr*, 1860, 809-817.
- POPA, I., BERKOVICH, R., ALEGRE-CEBOLLADA, J., BADILLA, C. L., RIVAS-PARDO, J. A., TANIGUCHI, Y., KAWAKAMI, M. & FERNANDEZ, J. M. 2013a. Nanomechanics of HaloTag tethers. *J Am Chem Soc*, 135, 12762-71.
- POPA, I., KOSURI, P., ALEGRE-CEBOLLADA, J., GARCIA-MANYES, S. & FERNANDEZ, J. M. 2013b. Force dependency of biochemical reactions measured by single-molecule force-clamp spectroscopy. *Nat Protoc*, 8, 1261-76.
- POPA, I., RIVAS-PARDO, J. A., ECKELS, E. C., ECHELMAN, D. J., BADILLA, C. L., VALLE-ORERO, J. & FERNANDEZ, J. M. 2016. A HaloTag Anchored Ruler for Week-Long Studies of Protein Dynamics. *J Am Chem Soc*, 138, 10546-53.
- POPOT, J. L. & ENGELMAN, D. M. 1990. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*, 29, 4031-7.
- POSTIS, V., RAWSON, S., MITCHELL, J. K., LEE, S. C., PARSLAW, R. A., DAFFORN, T. R., BALDWIN, S. A. & MUENCH, S. P. 2015. The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochim Biophys Acta*, 1848, 496-501.
- PRABUDIANSYAH, I., KUSTERS, I., CAFORIO, A. & DRIESSEN, A. J. 2015. Characterization of the annular lipid shell of the Sec translocon. *Biochim Biophys Acta*, 1848, 2050-6.
- PREINER, J., JANOVIK, H., RANKL, C., KNAUS, H., CISNEROS, D. A., KEDROV, A., KIENBERGER, F., MULLER, D. J. & HINTERDORFER, P. 2007. Free energy of membrane protein unfolding derived from single-molecule force measurements. *Biophys J*, 93, 930-7.
- PYLE, E., KALLI, A. C., AMILLIS, S., HALL, Z., LAU, A. M., HANYALOGLU, A. C., DIALLINAS, G., BYRNE, B. & POLITIS, A. 2018. Structural Lipids Enable the Formation of Functional Oligomers of the Eukaryotic Purine Symporter UapA. *Cell Chem Biol*, 25, 840-848 e4.
- QIU, W., FU, Z., XU, G. G., GRASSUCCI, R. A., ZHANG, Y., FRANK, J., HENDRICKSON, W. A. & GUO, Y. 2018. Structure and activity of lipid bilayer within a membrane-protein transporter. *Proc Natl Acad Sci U S A*, 115, 12985-12990.
- QUISTGAARD, E. M., LOW, C., MOBERG, P., TRESAUGUES, L. & NORDLUND, P. 2013. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nat Struct Mol Biol*, 20, 766-8.
- QURESHI, A. A., SUADES, A., MATSUOKA, R., BROCK, J., MCCOMAS, S. E., NJI, E., ORELLANA, L., CLAESSEON, M., DELEMOTTE, L. & DREW, D. 2020. The molecular basis for sugar import in malaria parasites. *Nature*, 578, 321-325.

- RAPOPORT, T. A. 2007. Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. *Nature*, 450, 663-9.
- RAVULA, T., HARDIN, N. Z., DI MAURO, G. M. & RAMAMOORTHY, A. 2018. Styrene maleic acid derivates to enhance the applications of bio-inspired polymer based lipid-nanodiscs. *Eur Polym J*, 108, 597-602.
- RAVULA, T., HARDIN, N. Z. & RAMAMOORTHY, A. 2019. Polymer nanodiscs: Advantages and limitations. *Chem Phys Lipids*, 219, 45-49.
- READING, E. 2018. Structural Mass Spectrometry of Membrane Proteins within Their Native Lipid Environments. *Chemistry*, 24, 13391-13398.
- READING, E., HALL, Z., MARTENS, C., HAGHIGHI, T., FINDLAY, H., AHDASH, Z., POLITIS, A. & BOOTH, P. J. 2017. Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions. *Angew Chem Int Ed Engl*, 56, 15654-15657.
- RILEY, M. L., WALLACE, B. A., FLITSCH, S. L. & BOOTH, P. J. 1997. Slow alpha helix formation during folding of a membrane protein. *Biochemistry*, 36, 192-6.
- RUTKOWSKA, A., BEERBAUM, M., RAJAGOPALAN, N., FIAUX, J., SCHMIEDER, P., KRAMER, G., OSCHKINAT, H. & BUKAU, B. 2009. Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Lett*, 583, 2407-13.
- RYABICHKO, S., FERREIRA, V. M., VITRAC, H., KIYAMOVA, R., DOWHAN, W. & BOGDANOV, M. 2020. Cardiolipin is required in vivo for the stability of bacterial translocon and optimal membrane protein translocation and insertion. *Sci Rep*, 10, 6296.
- SAHIN-TOTH, M., KABACK, H. R. & FRIEDLANDER, M. 1996. Association between the amino- and carboxyl-terminal halves of lactose permease is specific and mediated by multiple transmembrane domains. *Biochemistry*, 35, 2016-21.
- SAMELSON, A. J., BOLIN, E., COSTELLO, S. M., SHARMA, A. K., O'BRIEN, E. P. & MARQUSEE, S. 2018. Kinetic and structural comparison of a protein's cotranslational folding and refolding pathways. *Sci Adv*, 4, eaas9098.
- SAMELSON, A. J., JENSEN, M. K., SOTO, R. A., CATE, J. H. & MARQUSEE, S. 2016. Quantitative determination of ribosome nascent chain stability. *Proc Natl Acad Sci U S A*, 113, 13402-13407.
- SANCHEZ-MORENO, I., GARCIA-JUNCEDA, E., HERMIDA, C. & FERNANDEZ-MAYORALAS, A. 2016. Development of a new method for d-xylose detection and quantification in urine, based on the use of recombinant xylose dehydrogenase from *Caulobacter crescentus*. *J Biotechnol*, 234, 50-57.
- SANDERS, C. R. & PROSSER, R. S. 1998. Bicelles: a model membrane system for all seasons? *Structure*, 6, 1227-1234.
- SANDERS, M. R., FINDLAY, H. E. & BOOTH, P. J. 2018. Lipid bilayer composition modulates the unfolding free energy of a knotted alpha-helical membrane protein. *Proc Natl Acad Sci U S A*, 115, E1799-E1808.
- SAPRA, K. T., BESIR, H., OESTERHELT, D. & MULLER, D. J. 2006. Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. *J Mol Biol*, 355, 640-50.
- SCHAFFITZEL, C. & BAN, N. 2007. Generation of ribosome nascent chain complexes for structural and functional studies. *J Struct Biol*, 158, 463-71.
- SCHERES, S. H. 2012. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J Struct Biol*, 180, 519-30.
- SCHIBICH, D., GLOGE, F., POHNER, I., BJORKHOLM, P., WADE, R. C., VON HEIJNE, G., BUKAU, B. & KRAMER, G. 2016. Global profiling of SRP interaction with nascent polypeptides. *Nature*, 536, 219-23.
- SCHLEGEL, S., RUJAS, E., YTTTERBERG, A. J., ZUBAREV, R. A., LUIRINK, J. & DE GIER, J. W. 2013. Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microb Cell Fact*, 12, 24.
- SCHNEIDER, C. A., RASBAND, W. S. & ELICEIRI, K. W. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, 9, 671-5.
- SCHÖNFELDER, J., DE SANCHO, D., BERKOVICH, R., BEST, R. B., MUÑOZ, V. & PEREZ-JIMENEZ, R. 2018. Reversible two-state folding of the ultrafast protein gpW under mechanical force. *Communications Chemistry*, 1.
- SCHULTZ, C. P. 2000. Illuminating folding intermediates. *Nat Struct Biol*, 7, 7-10.
- SEDDON, A. M., CURNOW, P. & BOOTH, P. J. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta*, 1666, 105-17.
- SEIDELT, B., INNIS, C. A., WILSON, D. N., GARTMANN, M., ARMACHE, J. P., VILLA, E., TRABUCO, L. G., BECKER, T., MIELKE, T., SCHULTEN, K., STEITZ, T. A. & BECKMANN, R. 2009. Structural insight into nascent polypeptide chain-mediated translational stalling. *Science*, 326, 1412-5.
- SERDIUK, T., BALASUBRAMANIAM, D., SUGIHARA, J., MARI, S. A., KABACK, H. R. & MULLER, D. J. 2016. YidC assists the stepwise and stochastic folding of membrane proteins. *Nat Chem Biol*, 12, 911-917.
- SERDIUK, T., MADEJ, M. G., SUGIHARA, J., KAWAMURA, S., MARI, S. A., KABACK, H. R. & MULLER, D. J. 2014. Substrate-induced changes in the structural properties of LacY. *Proc Natl Acad Sci U S A*, 111, E1571-80.
- SERDIUK, T., MARI, S. A. & MULLER, D. J. 2017. Pull-and-Paste of Single Transmembrane Proteins. *Nano Lett*, 17, 4478-4488.
- SERDIUK, T., STEUDLE, A., MARI, S. A., MANIOGLU, S., KABACK, H. R., KUHN, A. & MULLER, D. J. 2019. Insertion and folding pathways of single membrane proteins guided by translocases and insertases. *Sci Adv*, 5, eaau6824.
- SERDIUK, T., SUGIHARA, J., MARI, S. A., KABACK, H. R. & MULLER, D. J. 2015. Observing a lipid-dependent alteration in single lactose permeases. *Structure*, 23, 754-61.

- SEREK, J., BAUER-MANZ, G., STRUHALLA, G., VAN DEN BERG, L., KIEFER, D., DALBEY, R. & KUHN, A. 2004. Escherichia coli YidC is a membrane insertase for Sec-independent proteins. *EMBO J*, 23, 294-301.
- SHERRATT, A. R., BLAIS, D. R., GHASRIANI, H., PEZACKI, J. P. & GOTO, N. K. 2012. Activity-based protein profiling of the Escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry*, 51, 7794-803.
- SHERRATT, A. R., BRAGANZA, M. V., NGUYEN, E., DUCAT, T. & GOTO, N. K. 2009. Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain. *Biochim Biophys Acta*, 1788, 2444-53.
- SIMON, K. S., POLLOCK, N. L. & LEE, S. C. 2018. Membrane protein nanoparticles: the shape of things to come. *Biochem Soc Trans*, 46, 1495-1504.
- SINGER, S. J. & NICOLSON, G. L. 1972. The fluid mosaic model of the structure of cell membranes. *Science*, 175, 720-31.
- SMIRNOVA, I. A., SJOSTRAND, D., LI, F., BJORCK, M., SCHAFFER, J., OSTBYE, H., HOGBOM, M., VON BALLMOOS, C., LANDER, G. C., ADELROTH, P. & BRZEZINSKI, P. 2016. Isolation of yeast complex IV in native lipid nanodiscs. *Biochim Biophys Acta*, 1858, 2984-2992.
- SNIDER, C., JAYASINGHE, S., HRISTOVA, K. & WHITE, S. H. 2009. MPEX: a tool for exploring membrane proteins. *Protein Sci*, 18, 2624-8.
- SORZANO, C. O., BILBAO-CASTRO, J. R., SHKOLNISKY, Y., ALCORLO, M., MELERO, R., CAFFARENA-FERNANDEZ, G., LI, M., XU, G., MARABINI, R. & CARAZO, J. M. 2010. A clustering approach to multireference alignment of single-particle projections in electron microscopy. *J Struct Biol*, 171, 197-206.
- SU, T. & PUROHIT, P. K. 2009. Mechanics of forced unfolding of proteins. *Acta Biomater*, 5, 1855-63.
- SUN, C., BENLEKBIR, S., VENKATAKRISHNAN, P., WANG, Y., HONG, S., HOSLER, J., TAJKHORSHID, E., RUBINSTEIN, J. L. & GENNIS, R. B. 2018. Structure of the alternative complex III in a supercomplex with cytochrome oxidase. *Nature*, 557, 123-126.
- SUN, L., ZENG, X., YAN, C., SUN, X., GONG, X., RAO, Y. & YAN, N. 2012. Crystal structure of a bacterial homologue of glucose transporters GLUT1-4. *Nature*, 490, 361-6.
- SWIECICKI, J. M., SANTANA, J. T. & IMPERIALI, B. 2020. A Strategic Approach for Fluorescence Imaging of Membrane Proteins in a Native-like Environment. *Cell Chem Biol*, 27, 245-251 e3.
- TASCON, I., SOUSA, J. S., COREY, R. A., MILLS, D. J., GRIWATZ, D., AUMULLER, N., MIKUSEVIC, V., STANSFELD, P. J., VONCK, J. & HANELT, I. 2020. Structural basis of proton-coupled potassium transport in the KUP family. *Nat Commun*, 11, 626.
- TAYLOR, K. A. C. 1995. A modification of the phenol/sulfuric acid assay for total carbohydrates giving more comparable absorbances. *Applied Biochemistry and Biotechnology*, 53, 207-214.
- TEO, A. C. K., LEE, S. C., POLLOCK, N. L., STROUD, Z., HALL, S., THAKKER, A., PITT, A. R., DAFFORN, T. R., SPICKETT, C. M. & ROPER, D. I. 2019. Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein. *Sci Rep*, 9, 1813.
- TRIBET, C., AUDEBERT, R. & POPOT, J. L. 1996. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc Natl Acad Sci U S A*, 93, 15047-50.
- URBAN, S. 2016. A guide to the rhomboid protein superfamily in development and disease. *Semin Cell Dev Biol*, 60, 1-4.
- URBAN, S., SCHLIEPER, D. & FREEMAN, M. 2002. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr Biol*, 12, 1507-12.
- URBANUS, M. L., SCOTTI, P. A., FRODERBERG, L., SAAF, A., DE GIER, J. W., BRUNNER, J., SAMUELSON, J. C., DALBEY, R. E., OUDEGA, B. & LUIRINK, J. 2001. Sec-dependent membrane protein insertion: sequential interaction of nascent FtsQ with SecY and YidC. *EMBO Rep*, 2, 524-9.
- VAN MEER, G., VOELKER, D. R. & FEIGENSON, G. W. 2008. Membrane lipids: where they are and how they behave. *Nat Rev Mol Cell Biol*, 9, 112-24.
- VINOTHKUMAR, K. R. 2011. Structure of rhomboid protease in a lipid environment. *J Mol Biol*, 407, 232-47.
- VON LOEFFELHOLZ, O., JIANG, Q., ARIOSIA, A., KARUPPASAMY, M., HUARD, K., BERGER, I., SHAN, S. O. & SCHAFFITZEL, C. 2015. Ribosome-SRP-FtsY cotranslational targeting complex in the closed state. *Proc Natl Acad Sci U S A*, 112, 3943-8.
- WANG, P. & DALBEY, R. E. 2011. Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts. *Biochim Biophys Acta*, 1808, 866-75.
- WANG, X., BOGDANOV, M. & DOWHAN, W. 2002. Topology of polytopic membrane protein subdomains is dictated by membrane phospholipid composition. *EMBO J*, 21, 5673-81.
- WANG, Y., ZHANG, Y. & HA, Y. 2006. Crystal structure of a rhomboid family intramembrane protease. *Nature*, 444, 179-80.
- WAUDBY, C. A., DOBSON, C. M. & CHRISTODOULOU, J. 2019. Nature and Regulation of Protein Folding on the Ribosome. *Trends Biochem Sci*, 44, 914-926.
- WAUDBY, C. A., LAUNAY, H., CABRITA, L. D. & CHRISTODOULOU, J. 2013. Protein folding on the ribosome studied using NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc*, 74, 57-75.
- WHITE, S. H. & WIMLEY, W. C. 1999. Membrane protein folding and stability: physical principles. *Annu Rev Biophys Biomol Struct*, 28, 319-65.

- WHITED, A. M. & PARK, P. S. 2014. Atomic force microscopy: a multifaceted tool to study membrane proteins and their interactions with ligands. *Biochim Biophys Acta*, 1838, 56-68.
- WHITEHOUSE, S., GOLD, V. A., ROBSON, A., ALLEN, W. J., SESSIONS, R. B. & COLLINSON, I. 2012. Mobility of the SecA 2-helix-finger is not essential for polypeptide translocation via the SecYEG complex. *J Cell Biol*, 199, 919-29.
- WIESLANDER, A., CHRISTIANSSON, A., RILFORS, L. & LINDBLÖM, G. 2002. Lipid bilayer stability in membranes. Regulation of lipid composition in *Acholeplasma laidlawii* as governed by molecular shape. *Biochemistry*, 19, 3650-3655.
- WIŚEDCHAIŚRI, G., PARK, M. S., IADANZA, M. G., ZHENG, H. & GONEN, T. 2014. Proton-coupled sugar transport in the prototypical major facilitator superfamily protein XylE. *Nat Commun*, 5, 4521.
- WOOLHEAD, C. A., JOHNSON, A. E. & BERNSTEIN, H. D. 2006. Translation arrest requires two-way communication between a nascent polypeptide and the ribosome. *Mol Cell*, 22, 587-98.
- WOOLHEAD, C. A., MCCORMICK, P. J. & JOHNSON, A. E. 2004. Nascent Membrane and Secretory Proteins Differ in FRET-Detected Folding Far inside the Ribosome and in Their Exposure to Ribosomal Proteins. *Cell*, 116, 725-736.
- WU, J., SUN, J. & KABACK, H. R. 1996. Purification and functional characterization of the C-terminal half of the lactose permease of *Escherichia coli*. *Biochemistry*, 35, 5213-9.
- XU, H., HILL, J. J., MICHELSEN, K., YAMANE, H., KURZEJA, R. J., TAM, T., ISAACS, R. J., SHEN, F. & TAGARI, P. 2015. Characterization of the direct interaction between KcsA-Kv1.3 and its inhibitors. *Biochim Biophys Acta*, 1848, 1974-80.
- XUE, Y. & HA, Y. 2012. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem*, 287, 3099-107.
- YAN, N. 2015. Structural Biology of the Major Facilitator Superfamily Transporters. *Annu Rev Biophys*, 44, 257-83.
- YAP, M. N. & BERNSTEIN, H. D. 2009. The plasticity of a translation arrest motif yields insights into nascent polypeptide recognition inside the ribosome tunnel. *Mol Cell*, 34, 201-11.
- YOUNG, R. & BREMER, H. 1976. Polypeptide-chain-elongation rate in *Escherichia coli* B/r as a function of growth rate. *Biochem J*, 160, 185-94.
- YU, H., SIEWNY, M. G., EDWARDS, D. T., SANDERS, A. W. & PERKINS, T. T. 2017. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science*, 355, 945-950.
- ZHANG, J., PAN, X., YAN, K., SUN, S., GAO, N. & SUI, S. F. 2015. Mechanisms of ribosome stalling by SecM at multiple elongation steps. *Elife*, 4.
- ZHAO, X., ZENG, X., LU, C. & YAN, J. 2017. Studying the mechanical responses of proteins using magnetic tweezers. *Nanotechnology*, 28, 414002.
- ZOCHER, M., FUNG, J. J., KOBILKA, B. K. & MULLER, D. J. 2012a. Ligand-specific interactions modulate kinetic, energetic, and mechanical properties of the human beta2 adrenergic receptor. *Structure*, 20, 1391-402.
- ZOCHER, M., ROOS, C., WEGMANN, S., BOSSHART, P. D., DOTTSCH, V., BERNHARD, F. & MULLER, D. J. 2012b. Single-molecule force spectroscopy from nanodiscs: an assay to quantify folding, stability, and interactions of native membrane proteins. *ACS Nano*, 6, 961-71.

# Appendices

---

## 1. N-His GlpG for pET28a

DNA:

```
ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCCTGGTGCCGCGCGGCAGCCATATGTGATGAT
TACCTCTTTTGGCTAACCCCGCGTGGCGCAGGCGTTTGTGATTACATGGCGACGCAGGGTGTATCCCTC
ACGATTCAACAACATAACCAAAGCGATGTCTGGCTGGCGGATGAGTCCCAGGCCGAGCGGTACGGGCGG
AGCTGGCGCGTTTTCTCGAAAACCCGGCAGATCCGCGTTATCTGGCGCGAGCTGGCAGGCAGGCCATAC
CGGCAGTGGCCTGCATTATCGCCGTTATCCTTTCTTTGCCGCTTGGCGTGAACCGCGCAGGTCCGGTAACC
TGGGTGATGATGATCGCCTGCGTGGTGGTGTATTGCCATGCAAATTCGCGGATCAGGAAGTGATGT
TATGGCTGGCCTGGCCATTGATCCAACACTGAAATTTGAGTTCTGGCGTTACTTCACCCACGCGTTAAT
GCATTCTCGCTGATGCATATCCTCTTTAACCTGCTCTGGTGGTGGTATCTCGGCGGTGCGGTGGAAAAA
CGCCTCGGTAGCGGTAAGCTAATTGTCATTACGCTTATCAGCGCCCTGTTAAGCGGCTATGTGCAGCAA
AATTCAGCGGGCCGTGGTTTGGCGGGCTTTCTGGCGTGGTGTATGCGCTGATGGGCTACGTCTGGCTACG
TGGCGAACCGGATCCGCAAAGTGGCATTACCTGCAACGTGGGTAAATTATCTTTGCGCTGATCTGGATT
GTCGCCGGATGGTTTGTATTGTTTGGGATGTCGATGGCGAACGGAGCACACATCGCCGGTTAGCCGTGG
GTTTAGCGATGGCTTTTGTGATTGCTCAATGCGCGAAAACGAAAATAA
```

Protein:

```
MGSSHHHHHHSSGLVPRGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWLADESQAERVRA
ELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWVMMIACVVVFIAMQIILGDQEV
LWLAWPFDPTLKFEFWRIFYTHALMHFSLMHILFNLLWWWYLGGAWEKRLGSGKLIVITLISALLSGYVQQ
KFSGPWFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLIIFALIWIWAGWFDLFGMSMANGAHIAGLAV
GLAMAFVDSL NARKRK
```

Molecular weight 33470.02 g/mol

Molar extinction coefficient 86400 M<sup>-1</sup>.cm<sup>-1</sup>

## 2. WT Xyle for pET28a

DNA:

```
tctagaaataatTTTTgttttaactttaagaaggagatataccATGAATACCCAGTATAATTCCAGTTATAT
ATTTTCGATTACCTTAGTCGCTACATTAGGTGGTATTATTATTTGGCTACGACACCGCCGTTATTTCCGGT
ACTGTTGAGTCACTCAATAACCGTCTTTGTTGCTCCACAAAACCTAAGTGAATCCGCTGCCAACTCCCTGT
TAGGGTTTTGCGTGGCCAGCGCTCTGATTGGTTCATCATCATCGGCGGTGCCCTCGTGGTTATTGCGATAA
CCGCTTCGGTCGTCGATGATTCACTTAAGATTGCTGCTGTCTGTTTTTTTATTTTGGTGTAGGTTCTGCC
TGGCCAGAACTTGGTTTTACCTCTATAAACCCGGACAACACTGTGCCTGTTTTATCTGGCAGGTTATGTCC
CGGAATTTGTTATTTATCGCATTATTGGCGGTATTGGCGTTGGTTTAGCCTCAATGCTCTCGCCAATGTA
TATTGCGGAAGTGGCTCCAGCTCATATTGCGGGAAAACCTGGTCTCTTTTAACCAGTTTGGCATTATTTTC
GGGCAACTTTTAGTTTACTGCGTAAACTATTTTATTGCCCGTTCCGGTGATGCCAGCTGGCTGAATACTG
ACGGCTGGCGTTATATGTTTGCCTCGGAATGTATCCCTGCACTGCTGTTCTTAATGCTGCTGTATAACCGT
GCCAGAAAGTCCTCGCTGGCTGATGTCGCGCGGCAAGCAAGAACAGGCGGAAGGTATCCTGCGCAAAATT
ATGGGCAACACGCTTGCAACTCAGGCAGTACAGGAAATTAACACTCCCTGGATCATGGCCGCAAAACCG
GTGGTCTGCTGCTGATGTTTGGCGTGGGCGTGATTGTAATCGGCGTAATGCTCTCCATCTTCCAGCAATT
TGTCGGCATCAATGTGGTGTGTACTACGCGCCGGAAGTGTTCAAAACGCTGGGGGCCAGCACGGATATC
GCGCTGTTGCGAGCATTATTGTGCGGAGTTATCAACCTCACCTTACCCTTCTGGCAATTATGACGGTGG
ATAAATTTGGTCGTAAGCCACTGCAAAATATCGGCGCACTCGGAATGGCAATCGGTATGTTTAGCCTCGG
TACCGCGTTTTACTCAGGCACCGGTTATTGTGGCGCTACTGTCGATGCTGTTCTATGTTGCCGCCCTTT
GCCATGCTCTGGGGTCCGGTATGCTGGGTACTGCTGTGCGAAATCTTCCCGAATGCTATTCGTGGTAAAG
CGCTGGCAATCGCGGTGGCGGCCAGTGGCTGGCGAACTACTTCGTCCTTGACCTTCCCGATGATGGA
CAAAAACCTCTGGCTGGTGGCCATTTCCACAACGGTTTCTCTACTGGATTTACGGTTGTATGGGCGTT
CTGGCAGCAGCTTTATGTGGAATTTGTCCCGGAAACCAAAGTAAAACCTTGAGGAGCTGGAAGCGC
TCTGGGAACCGAAACGAAGAAAACACAACAACTGCTACGCTGCTCGAGCATCATCATCACCACCA
CCACCACCACTGA
```



Protein:

MNTQYNSSYIFSITLVATLGGLLFGYDTAVISGTVESLNTVVFVAPQNLSESAANSLLGFCVASALIGCII  
GGALGGYCSNRFGRRDSLKIAAVLFFISGVGSAWPELGFTSINPDNTVPVYLAGYVPEFVIYRIIGGIGV  
GLASMLSPMYIAELAPAHIRGKLVSNQFAIIFGQLLVYCVNYFIARSGDASWLNTDGWRYMFASECIPA  
LLFLMLLYTVPEsprwLMSRGKQEQAEgILRKIMGNTLATQAVQEIKHSLDHGRKTGGRLLMFGVGVIVI  
GVMLSI FQQFVGINVVLYYAPEVFKTLGASTDIALLOTIIVGVINLTFVLAIMTVDKFGRKPLQIIGAL  
GMAIGMFSLGTAFYTTQAPGIVALLSMLFYVAAFAMSWGPVCWVLLSEIFPNAIRGKALAIAVAAQWLANY  
FVSWTFPMMDKNSWLVAHFHNGFSYWIYGCMGVLAALFMWKFVPEtKGKTLEELEALWEPETKKTQQTAT  
LLEHHHHHHHHHH

Molecular weight 55221.73 g/mol

Molar extinction coefficient 94310 M<sup>-1</sup>.cm<sup>-1</sup>

### 3. Xyle mechanical unfolding sequences and constructs

#### Xyle-pGly28-His<sub>10</sub>:

MNTQYNSSYIFSITLVATLGGLLFGYDTAVISGTVESLNTVVFVAPQNLSESAANSLLGFCVASALIGCII  
GGALGGYCSNRFGRRDSLKIAAVLFFISGVGSAWPELGFTSINPDNTVPVYLAGYVPEFVIYRIIGGIGV  
GLASMLSPMYIAELAPAHIRGKLVSNQFAIIFGQLLVYCVNYFIARSGDASWLNTDGWRYMFASECIPA  
LLFLMLLYTVPEsprwLMSRGKQEQAEgILRKIMGNTLATQAVQEIKHSLDHGRKTGGRLLMFGVGVIVI  
GVMLSI FQQFVGINVVLYYAPEVFKTLGASTDIALLOTIIVGVINLTFVLAIMTVDKFGRKPLQIIGAL  
GMAIGMFSLGTAFYTTQAPGIVALLSMLFYVAAFAMSWGPVCWVLLSEIFPNAIRGKALAIAVAAQWLANY  
FVSWTFPMMDKNSWLVAHFHNGFSYWIYGCMGVLAALFMWKFVPEtKGKTLEELEALWEPETKKTQQTAT  
LLETVPESPRWLMsRGKQEQAEgILRKIMGNTLATQAVQEIKHSLDHGRKTGGRLLMFGVGVIVIGVMLS  
IFQQFVGINVVLYYAPEVFKTLGASTDIALLOTIIVGVINLTFVLAIMTVDKFGRKPLQIIGALGMAIG  
MFSLGTAFYTTQAPGIVALLSMLFYVAAFAMSWGPVCWVLLSEIFPNAIRGKALAIAVAAQWLANYFVSWT  
FPMMDKNSWLVAHFHNGFSYWIYGCMGVLAALFMWKFVPEtKGKTLEELEALWEPETKKTQQTATLGAWR  
AARGARKVEEAGGGGGGGGGGSL EHHHHHHHHHH

#### Halo-TEV-TMH-Xyle-His<sub>10</sub>:

MAEIGTGFPFDPHYVEVLGERMHYVDVGP RDGTPVLFHGNPTSSYVWRNIIPHVAPTHRCIAPDLIGMG  
KSDKPD LGYFFDDHVR FMDAFIEALGLEEVV LVIHDWGSALGFHWAKRNP ERVKGI AFMEFIRPIPTWDE  
WPEFARETFQAFRTTDVGRKLIIDQNVFIEGTLPMGVVRPLTEVEMDHYREPF LN PV DREPLWRFPNELP  
IAGEPANIVALVEEYMDWLHQSPVPKLLFWGTPGVLI PPAAEARLAKSLPNCKAVDIGPGLNLLQEDNPD  
LIGSEIARWLSTLEISGEPTTEDLYFQSDNAIAGSMANMGPGGPGAAAALALALLLALAAAA GSSGMN  
TQYNSSYIFSITLVATLGGLLFGYDTAVISGTVESLNTVVFVAPQNLSESAANSLLGFCVASALIGCIIIGG  
ALGGYCSNRFGRRDSLKIAAVLFFISGVGSAWPELGFTSINPDNTVPVYLAGYVPEFVIYRIIGGIGVGL  
ASMLSPMYIAELAPAHIRGKLVSNQFAIIFGQLLVYCVNYFIARSGDASWLNTDGWRYMFASECIPALL  
FLMLLYTVPEsprwLMSRGKQEQAEgILRKIMGNTLATQAVQEIKHSLDHGRKTGGRLLMFGVGVIVIGV  
MLSI FQQFVGINVVLYYAPEVFKTLGASTDIALLOTIIVGVINLTFVLAIMTVDKFGRKPLQIIGALGM  
AIGMFSLGTAFYTTQAPGIVALLSMLFYVAAFAMSWGPVCWVLLSEIFPNAIRGKALAIAVAAQWLANYFV  
SWTFPMMDKNSWLVAHFHNGFSYWIYGCMGVLAALFMWKFVPEtKGKTLEELEALWEPETKKTQQTATLL  
EHHHHHHHHHH



## 6. RNC protein sequences and constructs

Key: His-tag- (AviTag) -GlpG (RNC) -WT SecM (Stalling proline)

Note: All WT SecM sequences were also swapped for AE1 SecM where FSTPVWISQAQGIRAGP was changed to FSTPVWIWWWPRIRGPP

### 2 TM GlpG RNC SecM:

MGSSHHHHHHSSGLVPRGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWLADESQAERVRA  
ELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWMMIACVVVFIAMQILGDQEV  
MLWLAWPFDPDLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAWEKRGSGGGSTPSEKGYRIDYAHF  
TPQAKFSTPVWISQAQGI RAGPQRLSFM MMMMH

### 4 TM GlpG RNC SecM:

MGSSHHHHHHSSGLVPRGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWLADESQAERVRA  
ELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWMMIACVVVFIAMQILGDQEV  
MLWLAWPFDPDLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAWEKRLGSGKLIVITLISALLSGYV  
QQKFSGPWFGGLSGVVYALMGYVWLRGERDPQSGSGGGSTPSEKGYRIDYAHFTPQAKFSTPVWISQAQGI  
RAGPQRLSFM MMMMH

### 6 TM GlpG RNC SecM:

MGSSHHHHHHSSGLVPRGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWLADESQAERVRA  
ELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWMMIACVVVFIAMQILGDQEV  
MLWLAWPFDPDLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAWEKRLGSGKLIVITLISALLSGYV  
QQKFSGPWFGGLSGVVYALMGYVWLRGERDPQSGIYLRGLIIFALIWIWVAGWFDLFGMSMANGAHIAGLAV  
GLAMAFVDSL NARKRKGSGGGSTPSEKGYRIDYAHFTPQAKFSTPVWISQAQGI RAGPQRLSFM MMMMH

### 2 TM Avi-GlpG RNC SecM:

MGSSHHHHHHSSGGLNDIFEAQKIEWHEGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWL  
ADESQAERVRAELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWMMIACVVVFI  
AMQILGDQEVMLWLAWPFDPDLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAWEKRGSGGGSTPSEK  
GYRIDYAHFTPQAKFSTPVWISQAQGI RAGPQRLSFM MMMMH

### 4 TM Avi-GlpG RNC SecM:

MGSSHHHHHHSSGGLNDIFEAQKIEWHEGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWL  
ADESQAERVRAELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWMMIACVVVFI  
AMQILGDQEVMLWLAWPFDPDLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAWEKRLGSGKLIVITL  
ISALLSGYVQKFSGPWFGGLSGVVYALMGYVWLRGERDPQSGSGGGSTPSEKGYRIDYAHFTPQAKFS  
TPVWISQAQGI RAGPQRLSFM MMMMH

### 6 TM Avi-GlpG RNC SecM:

MGSSHHHHHHSSGGLNDIFEAQKIEWHEGSHMLMITSFANPRVAQAFVDYMATQGVILTIQQHNQSDVWL  
ADESQAERVRAELARFLENPADPRYLAASWQAGHTGSGLYHRRYPFFAALRERAGPVTWMMIACVVVFI  
AMQILGDQEVMLWLAWPFDPDLKFEFWRYFTHALMHFSLMHILFNLLWWWYLGGAWEKRLGSGKLIVITL  
ISALLSGYVQKFSGPWFGGLSGVVYALMGYVWLRGERDPQSGIYLRGLIIFALIWIWVAGWFDLFGMSM  
ANGAHIAGLAVGLAMAFVDSL NARKRKGSGGGSTPSEKGYRIDYAHFTPQAKFSTPVWISQAQGI RAGPQ  
RLSFM MMMMH

## 7. Bell-Evans-Richie energy landscape example calculation

The force of unfolding depends on the loading rate used in the experiment. The loading rate **Equation A-1** is given as the force experienced by the protein when extracted at a constant velocity and is calculated from the spring constant multiplied by the pulling velocity (Oosterhelt et al., 2000).

*Equation A-1: Loading rate varies with pulling velocity*

$$r = k_s \cdot v$$

The force of unfolding of a protein, or a stable segment can be related to the unfolding rate, and height of the barrier which separates the folded and unfolded states of a proteins. The most common models are the Bell-Evans-Richie model , and the Hummer-Szabo model (Bell, 1978, Evans and Ritchie, 1997, Hummer and Szabo, 2001, Hummer and Szabo, 2005), which are used for the extraction of kinetic and thermodynamic properties of DFS experiments. The Bell-Evans-Richie model **Equation A-2** describes protein unfolding as a two-state process separated with a transition state with a high energy barrier. The energy barrier **Figure 6-2** reaction coordinate is defined as the end-to-end distance of the stretched polymer. Bell-Evans-Richie suggests that the rate of unfolding increases exponentially with force ( $F$ ):

*Equation A-2: Bell-Evans-Richie model*

$$F = \frac{k_B T}{x_u} \cdot \ln\left(\frac{x_u \cdot r}{k_B T \cdot k_0}\right)$$

Where  $k_B T$  is the product of the Boltzmann constant and the absolute temperature of the system,  $r$  is the most probable loading rate,  $x_u$  is the distance between the native and transition state, and  $k_0$  is the loading rate at zero fore (equilibrium).

The force (pN) for each structural segment unfolding can be plotted against each loading rate (pN/s) and the values for  $x_u$  and  $k_0$  can be calculated from the slope and the intercept of the Bell-Evans-Richie linear regression fit for each segment. The Arrhenius **Equation A-3** can then be used with the rate of unfolding at equilibrium to calculate Gibbs free energies of unfolding (height of the energy barrier) for each segment, or the protein as a whole:

*Equation A-3: Arrhenius equation*

$$\Delta G_U = -k_B T \cdot \ln(k_0)$$

Other models for DFS energy extraction, such as Hummer-Szabo and Friddle-de-Yoreo have also emerged which incorporate more detail and are extensively reviewed in (Hughes and Dougan, 2016).

### 8. Theoretical contour lengths for mapped Xyle PBD 4GBY

Xyle-(pGly28)	Residues	AA	$L_c$ (nm)	← N-terminal unfolding read down table ↓	Cumulative N-term		↑ C-terminal unfolding read up table →	Cumulative C-term		Cumulative pGly28 C-term		
					AA unfolded	$L_c$ (nm)		AA unfolded	$L_c$ (nm)	AA unfolded	$L_c$ (nm)	
Helix 1	1-32	32	12.2		32	12.2		503	191.3		531	201.9
Loop 1	33-56	26	9.9		58	22.1		471	179.1		499	189.7
Helix 2	57-79	23	8.7		81	30.8		445	169.2		473	179.8
Loop 2	80-84	5	1.9		86	32.7		422	160.5		450	171.1
Helix 3	85-105	21	8.0		107	40.7		417	158.6		445	169.2
ECH1	106-128	23	8.7		130	49.4		396	150.6		424	161.2
Helix 4	129-150	22	8.4		152	57.8		373	141.9		401	152.5
Loop 5	151-162	12	4.6		164	62.4		351	133.5		379	144.1
Helix 5	163-183	21	8.0		185	70.4		339	128.9		367	139.5
Loop 5	184-200	17	6.5		202	76.9		318	120.9		346	131.5
Helix 6	201-221	21	8.0		223	84.9		301	114.4		329	125.0
ICH 1-3	222-276	55	20.9		278	105.8		280	106.4		308	117.0
Helix 7	277-299	23	8.7		310	114.5		223	85.5		251	96.1
Loop 7	300-312	13	4.9		314	119.4		202	76.8		230	87.4
Helix 8	313-334	22	8.4		336	127.8		189	71.9		217	82.5
Loop 8	335-343	9	3.4		345	131.2		167	63.5		195	74.1
Helix 9	344-364	21	8.0		366	139.2		158	60.1		186	70.7
Loop 9	365-369	4	1.5		370	140.7		137	52.1		165	62.7
Helix 10	370-390	21	8.0		391	148.7		133	50.6		161	61.2
Loop 10	391-407	17	6.5		408	155.2		112	42.6		140	53.2
Helix 11	408-428	21	8.0		429	163.2		95	36.1		123	46.7
ECH2	429-442	14	5.3		443	168.5		74	28.1		102	38.7
Helix 12	443-463	21	8.0		464	176.5		60	22.8		88	33.4
ICH4-His <sub>10</sub>	464-503	39	14.8		503	191.3		39	14.8		67	25.4
ICH4-(pGly28)-His <sub>10</sub>	464-531	67	25.5		531	201.9						

## 9. *BBA Biomembranes*: Mini review

### Structural insight into co-translational membrane protein folding

Grant A. Pellowe, Paula J. Booth\*

Department of Chemistry, King's College London, 7 Trinity Street, SE1 1DB, London, UK

\* Corresponding author: [paula.booth@kcl.ac.uk](mailto:paula.booth@kcl.ac.uk)

#### *Abstract*

Membrane protein folding studies lag behind those of water-soluble proteins due to immense difficulties of experimental study, resulting from the need to provide a hydrophobic lipid-bilayer environment when investigated *in vitro*. A sound understanding of folding mechanisms is important for membrane proteins as they contribute to a third of the proteome and are frequently associated with disease when mutated and/or misfolded. Membrane proteins largely consist of  $\alpha$ -helical, hydrophobic transmembrane domains, which insert into the membrane, often using the SecYEG/Sec61 translocase system. This mini-review highlights recent advances in techniques that can further our understanding of co-translational folding and notably, the structure and insertion of nascent chains as they emerge from translating ribosomes.

#### *Abbreviations*

SRP: signal recognition particle, RNC: ribosome-bound nascent chain complex, HTL: holotranslocon, ATP: adenosine-triphosphate, Hsp: heat-shock protein, TRAM: translocon-associated membrane protein, Trap: translocon-associated protein, OST: oligosaccharyltransferase, FRET: Förster resonance energy transfer, TM: transmembrane, MP: membrane protein, PURE: protein synthesis using recombinant elements, NC: nascent chain, PTC: peptidyl transferase centre, CMC: critical micelle concentration, IMV: inner membrane vesicle, MSP: membrane scaffold protein, SMA: styrene/maleic acid, DIBMA: diisobutylene/maleic acid, UAA: unnatural amino acid, AMS: 4-Acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid disodium salt, AP: arrest-peptide, SEIRAS: surface-enhanced infrared spectroscopy, RRL: rabbit reticulocyte lysate, PE: phosphatidylethanolamine, ER: endoplasmic reticulum, PMF: proton motive force, DMPC: 1,2-dimyristoyl-sn-glycero-3-phosphocholine

#### *1. Membrane protein folding in vivo*

The majority of *Escherichia coli*  $\alpha$ -helical nascent chains follow the Sec-dependent pathway of insertion, and are captured by the signal recognition particle (SRP) and delivered towards its receptor FtsY for the binding of the ribosome nascent chain (RNC) to the SecYEG insertion apparatus [1]. SecYEG with associated chaperones is often termed the holotranslocon (HTL); this larger complex is composed of SecYEG-SecDFYajC-YidC [2], and acts as a protein channel, translocase and insertase and protects the nascent chain from aggregation as it begins to fold on the cytoplasmic side of the membrane [3, 4]. In eukaryotic organisms, a hetero-trimeric complex Sec61, of three subunits ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) is essential for both co-translational insertion of membrane proteins and translocation of secretory proteins into, and across the endoplasmic reticulum membrane for cellular trafficking. Sec62 and Sec63 bind during a post-translational event in yeast, and most likely other eukaryotes, to ratchet peptides through the Sec61 channel with the aid of BiP, an ATPase of the Hsp70 family [5]. This mechanism is similar to protein secretion mechanisms across the bacterial inner membrane, however SecA pushes the peptide through the channel in an ATP-dependent manner [1]. Sec61 also associates itself with chaperoning proteins in higher-order complexes like translocon-associated membrane protein (TRAM), translocon-associated protein (Trap) and oligosaccharyltransferase (OST), which together are required for the folding of glycosylated proteins in eukaryotes [5].

TMs of polytopic membrane proteins are generally assumed to insert into the membrane sequentially, however more complex scenarios suggest that insertion can also occur as two-helix hairpins which are



assembled close to the membrane or even in the vicinity of the Sec61/SecY channels before their coordinated release into the membrane [6-9]. The helical structure can form in the ribosome exit tunnel, as shown by Förster resonance energy transfer (FRET), gel shift assays and structural studies [10]. A long standing model of translocon function proposes that transmembrane TM helices insert into the bilayer via a channel and lateral gate in the translocon [11]. This model is supported by chemical-crosslinking studies whereby stalled nascent chains were found to crosslink to both the translocon and surrounding lipids, in isolated ER-microsomes [12]. A recent hypothesis suggests that TM helices do not necessarily enter the SecYEG channel, but slide down the outside of the translocon, inserting via lipid head groups before the hydrophobic interior of the membrane provides the driving for insertion of the TMs [11]. More polar regions of the protein may use the translocon channel to insert into, or cross the bilayer. The chemistry of the lipid head groups, packing of hydrophobic tails and overall lipid bilayer properties are important for the insertion of nascent TM helices in either model.

Smaller and less complex proteins have been observed to insert with only YidC-type insertases in a 'Sec-independent' manner. Substrates for YidC, or Oxa1 of mitochondria or Alb3/Alb4 of chloroplasts are likely to favour smaller proteins with one or two helices [13]. The precise function of the YidC-like proteins is currently unknown, with only a direct catalytic activity observed thus far when reconstituted *in vitro* [14]. Mitochondria do not contain the Sec machinery and only contain Oxa1, suggesting that direct contact with lipids is essential for MP insertion and spontaneous folding can occur with the insertase partitioning charged/highly polar residues across the membrane [11]. However, Oxa1 may cooperate with the TIM23 machinery and also Mba1 during membrane protein insertion, forming a larger insertion complex like HTL in *E. coli*, therefore it remains unclear whether YidC/Oxa1 act as standalone insertases, or remain as a small unknown part of a whole [15].

## 2. Tools required for co-translational study

### 2.1. Synthesis machinery

Cell-free translation systems (**figure 1a**) are often used to study co-translational folding due to the ease of manipulation of translation conditions and synthesis components, such as amino acids. Cell-free systems were originally derived from cell lysates of a specific expression organism. *E. coli* strains optimised for protein expression had their lysate harvested giving rise to the S30 expression system. Endogenous membranes have been removed from these extracts during their preparation, therefore for the expression of MPs a membrane mimic must be added. Lysate based expression systems have since been commercialised and used in combination with lipid-nanodiscs for functional protein and folding studies. More recently, a minimal system was developed that only contains translation components necessary for protein synthesis. This was termed PURE (protein synthesis using recombinant elements) [16]. These systems can produce large complex proteins for *in vitro* study when membrane mimics and coding DNA or mRNAs are added to the mix, and are open to a great deal of optimisation as certain components can be added or removed, for example allowing for the introduction of fluorescent labels or cross-linking reagents through unnatural-amino acid (UAA) technology. PURE is also a much cleaner approach than cytosolic extracts, facilitating the isolation and purification of affinity tagged proteins for biochemical analysis of only the protein of interest.

MP insertion has been probed using truncated or stalled RNC complexes which allow the early environmental effect on the nascent chain (NC) to be probed (**figure 1b**). RNCs are produced by incorporating a peptide stalling sequence into a coding DNA sequence. These stalling sequences are used in nature to regulate NC translation allowing structure segments to fold upstream of the sequence [17]. The *E. coli* wild type secretion monitor (SecM) sequence, which once stalled can alter mRNA secondary structure due to ribosome binding and in turn regulate the binding of SecA to, and thus translocation of the N-terminal NC to the periplasm. The stalling occurs in the final codon of the SecM sequence, where the resulting translated proline alters the geometry of the peptidyl-transferase centre (PTC) of the ribosome [17, 18] and halts translation until nascent chain folding produces a significant

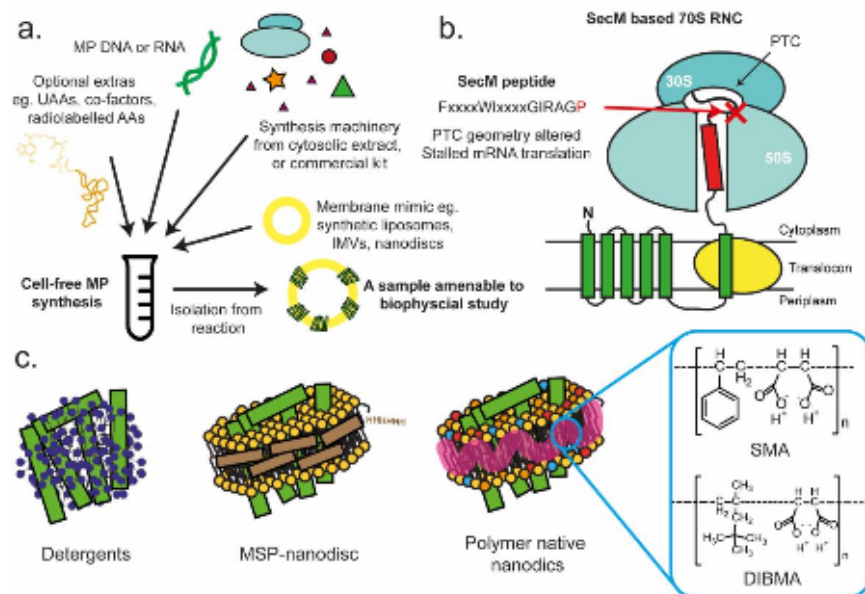


force to release the stalled sequence [19], making SecM a useful tool for RNC study [20]. The stall provided by SecM has also been enhanced based on stronger stalling sequences in other bacteria to supporting a new tool for biochemical and structural characterisation of RNCs where NC stalling release is not desired [11].

## 2.2. A membrane mimic

A membrane mimic with particular properties is also required for the correct folding of MPs (**figure 1b**). Each lipid molecule has a complex chemistry, with a variable number carbon chain tail length, of which the two can be asymmetric or have different degrees of saturation. Equally, the lipid head groups provide an added physicochemical specificity. Charged head groups and non-lamellar forming lipids can affect protein topology, folding and function [3, 21-23]. In addition, amphipathic detergent molecules, often used to solubilise MPs out of the native membranes at a concentration above their critical micelle concentration (CMC) are used to encapsulate MPs by forming a micelle with their hydrophobic hydrocarbon tails facing the protein, and hydrophilic head groups facing outwards towards the aqueous buffer environment. Detergents are often the first step in MP preparation and purification and are used as a vehicle for reconstitution into synthetic lipid mixes, and in the reconstitution process itself [24].

Most studies on aspects of co-translational folding have used extracts of inner membrane vesicles (IMVs), [25-27] harvested from cells, or microsomes [28, 29], isolated from dog pancreas. More recently, nanodiscs have been employed, where lipid bilayer discs are surrounded by a membrane scaffold protein (MSP). Nanodiscs can be used with cell-free systems to produce protein that is amenable to biochemical, biophysical and structural analysis [30-32], although the MSP limits some spectroscopic analyses. Naturally, the native membrane of an organism provides the optimal environment for MP folding and function. Novel co-polymer based chemicals styrene/maleic acid (SMA) and diisobutylene/maleic acid (DIBMA) (amongst others) have been developed to purify MPs directly from native membranes without detergent. These polymers act as a 'cookie cutter' to harvest the MP and surrounding lipids and any associated proteins, making these native-nanodiscs perfect for the study of MP dynamics or folding when coupled with stalled ribosome nascent chain (RNC) complexes. These native-nanodiscs, from production through to use have been reviewed extensively elsewhere [33].

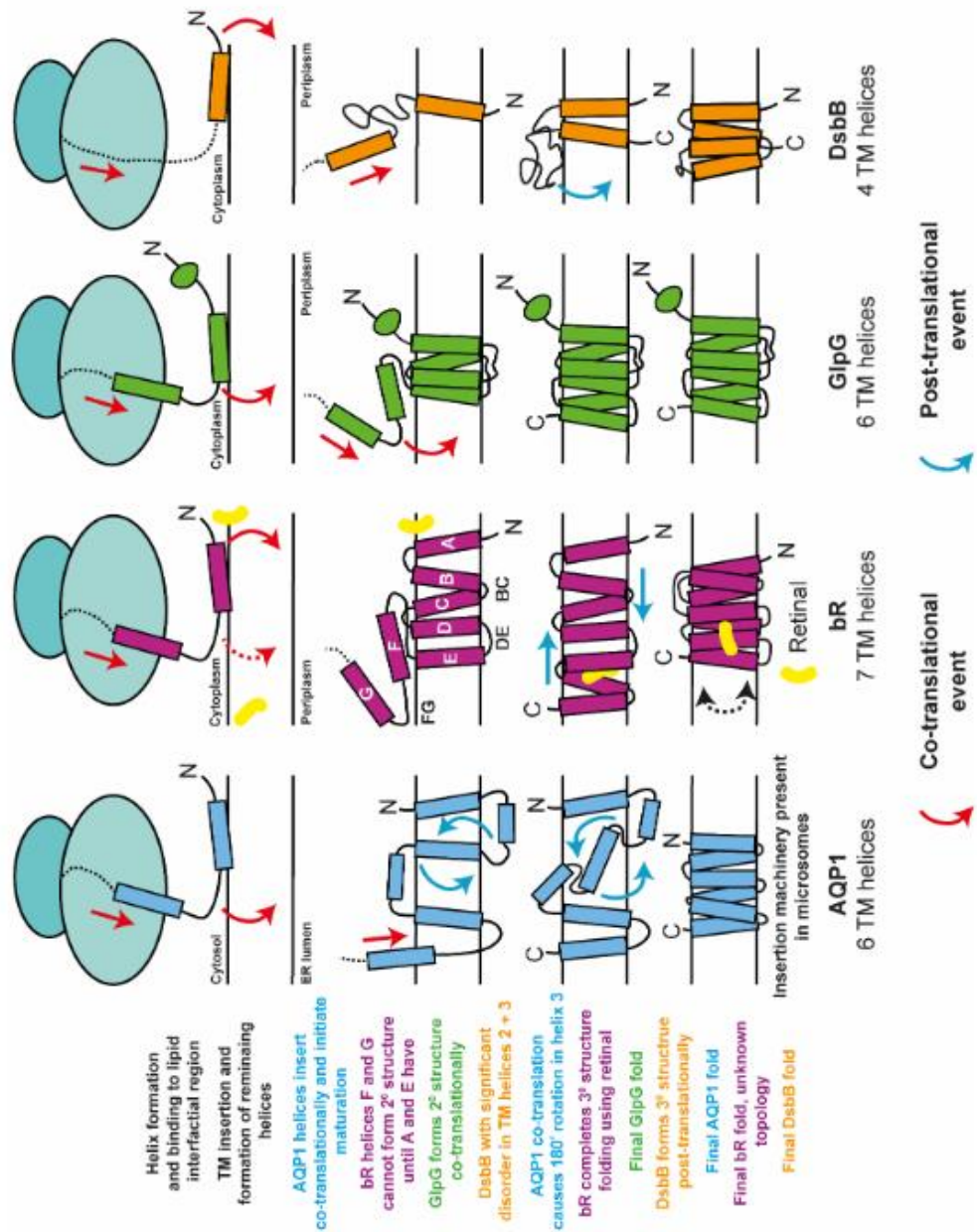


**Figure 1: Summary of the necessary tools for co-translational study of MPs**

a) Classic *in vitro* cell-free synthesis reaction scheme. Synthesis machinery from either cytosolic extracts of an organism, or recombinant components (PURE), containing transcription and translation components (ribosomes; amino acids, tRNA and synthetases; polymerases; initiation, elongation and release factors; energy replenishment systems), are mixed with DNA or RNA of desired membrane protein, and a membrane mimic of choice to produce a sample amenable to biophysical characterisation. Additional UAAs, co-factors specific to protein function, or radiolabelled tags/amino acids can be added to the reaction mixture to label the protein of choice. The sample can be isolated from the reaction and studied, or studied using novel biophysical techniques as the reaction proceeds (see section 4). b) Simplified schematic of a SecM induced stalled 70S ribosome to produce RNCs. The minimal SecM peptide stalls its mRNA translation once the final codon for proline enters the PTC. The geometry of the PTC is altered and peptide bond formation cannot occur. The expressed, truncated nascent chain still inserts into the membrane *in vivo*, most likely using the translocon. *In vitro* it is possible for proteins to insert directly into lipids using a minimal system without the translocon present. The effect of the ribosome, and the length of produced nascent chain provide insight into co-translational insertion and folding. This latter bottom up approach is amenable to lipid composition alteration which can increase spontaneous insertion of MPs [23]. c) A selection of membrane mimics, used in protein used in co-translational protein folding. Detergents, proteinaceous MSP-nanodiscs, which are often poly-his tagged for affinity purification or tethering to a surface, and novel polymer based chemicals (SMA and DIBMA) which can purify MPs with their native membrane intact and without the need for initial detergent purification.

**3. Biochemical methods to study co-translational folding and topology**

Early biochemical techniques to ascertain a time-courses for the order of polytopic TM helix insertion *in vivo*, used single-cysteine mutants amenable to labelling and were observed by a phosphorescent gel shift. Membrane impermeable 4-Acetamido-4'-maleimidylstilbene-2,2'-disulfonic acid disodium salt (AMS) was used alongside <sup>35</sup>S-Methionine pulse-chase labelling to determine the insertion rates of radiolabelled bacterioopsin (bO) (bacteriorhodopsin without retinal co-factor), a well-studied 7 TM  $\alpha$ -helical archaic proton pump of *Halobacterium salinarum* with each TM labelled A-G (figure 2). A cysteine located near the N-terminus showed a bilayer translocation rate faster than the translation of full length protein, as determined by AMS derivatisation at certain time points. This suggests that the insertion of the first TM helix must occur co-translationally [34] and acts as a folding 'nucleus' for the rest of the protein. This was later advanced, and each of the extracellular loops in bO were also labelled and pulse-chased. The BC and FG loops of bO are translocated in sequence order after the first helix [34, 35], but the FG loop, unlike the BC, completes its translocation after full length protein elongation is complete, and therefore must cross the membrane post-translationally. These results together suggest a sequential order of co-translational TM insertion for bO in *H. salinarum*. Direct structure formation however, cannot be characterised using this approach.



**Figure 2: Possible models for co-translational folding in the absence of translocon from SEIRAS**

Summary of possible models for protein folding and insertion using SEIRAS data for bR, and SEIRAS with crystal structure and mPEX predictions for GlpG and DsbB adapted from [4]. AQP1 comparison from protease reporter work in rough-ER microsomes [8]. Topology for bR in this cell-free system is unknown, the topologies of GlpG and DsbB depicted were confirmed by AMS labelling in liposomes



[4]. For each protein, helices are co-translationally (red arrows) produced and bind to the membrane interfacial region. AQP1 undergoes a co-translational maturation step where TM3 flips in the membrane 180° to produce the final folded structure. Insertion machinery may aid this process in the system used. Truncated proteins were used in the AQP1 study, this may have given the NC more time to rearrange their folding pathway. bR forms helices A-E co-translationally but F and G cannot fold until A and E are correctly inserted (loops labelled for section 3). A post-translational (blue arrows) event then forms the final 3D structure of bR using retinal included in the reaction. GlpG forms secondary and tertiary structure co-translationally. DsbB Regions corresponding to TMs 2 and 3 in DsbB are not predicted to insert suggesting that the unordered regions in SEIRA correspond to TMs 2 and 3, and they insert only insert post-translationally after TMs 1 and 4. Most likely, helices 2 and 3 favour a disordered state at the DMPC interface rather than partitioning through the bilayer in the absence of any translocon machinery.

Co-translational folding of eukaryotic polytopic proteins has been studied using protease digestion to assess proteins regions outside the membrane in the cytosol. Truncated polypeptides of the 6 TM human Aquaporin-1 (AQP1) water channel [8] were fused into protease sensitive reporter constructs at TM helix-connecting loops via epitopes, which allowed determination of topology for the truncated proteins at the endoplasmic reticulum in canine pancreas rough microsomal membranes with rabbit-reticulocyte lysate (RRL). The location of protease reporter in the cytosol or ER lumen indicated the orientation of TM helices. It was determined that 4 of 6 helices initially transverse the membrane, with helices 2 and 4 binding the lipid interfacial region (figure 2). A late stage maturation process appeared to flip TM helix 3 from an N-out orientation 180° through the membrane to N-in orientation, in turn pulling helices 2 and 4 into the bilayer. This is a co-translational process and rotation of TM3 increases as relative C-terminal helices are released from the ribosome [8]. Other AQP1 TM segments (omitted in the diagram for clarity) appear to transiently insert into the bilayer, these regions may remain in the translocon before integration adding yet another variable to the folding pathway. The pathway for AQP1 is much more complex than the sequential insertion studies of bO. However, truncated protein chains were used to ascertain topology in AQP1, which may have given the NC extra time to alter its folded conformation before measurements have taken place. In addition, bO studies were time resolved and therefore offer a more direct measurement of co-translational folding and topology that epitope labelling can. Yet, neither study can assess structure formation during co-translational insertion nor actual folding of the protein. In other work, lipid composition has been shown to influence topology with altered *E. coli* lipids [36]. Notably lack of the main lipid phosphatidylethanolamine (PE), causes an inversion of a 6 TM domain of the 12 TM lactose permease in the membrane, which can be rescued on addition of PE.

Epitope protease assays have also been used to show that the helices themselves contain topological information in the form of signal-anchoring sequences, which open the translocon allowing translocation of loops into the ER lumen and result in an N-out helix orientation, and stop-transfer sequences, which close the translocon keeping the connecting loops on the cytosolic side of the ER membrane [37]. The strength of stop-transfer activity, as a result of hydrophilic/phobic residues in helices, and loop length, can vary resulting in extensive topology and diversity among a range of proteins. The 6 TM helix AQP4 protein regulates the co-translational folding of its final structure in this way unlike AQP1 [9], where there is no co-translational helix rotation.

As NCs are elongated, the environments of the ribosome exit tunnel and the cytosol influence folding. By incorporating fluorescent or photocrosslinking labels into the nascent chain, changes in the peptide conformation as it is produced can be monitored. UAA technology uses modified aminoacyl-tRNAs to incorporate a non-natural amino acid directly into the nascent chain during translation [38]. FRET can occur between a donor and acceptor dye when in close proximity to produce an observable fluorescence, when fluorophore based UAAs are used in translation. The efficiency of this energy transfer is proportional to the distance between the two fluorophores. This can give spatial information between

nascent chain positions, and when photo-crosslinking agents are introduced, interactions between the unnatural amino acids and translocon components are observed [38, 39].

Single membrane-spanning TM peptides appear to fold near the ribosome PTC to produce compact  $\alpha$ -helices. This is shown by a high FRET efficiency. The compact nature is lost as the NC is released into the exit-tunnel, characterised by a reduced FRET efficiency. The final membrane inserted TM shows a similar FRET efficiency to when positioned near the PTC. Together, this shows that the ribosome itself can play a role in the initial folding strategy for TM spanning proteins. This pathway was proven to be orchestrated by specific ordered ribosome/TM interactions using photo-crosslinking reagents. The TM segment was shown to crosslink with three eukaryotic ribosomal proteins [39], whereas non-TM helix sequences crosslinked to only one. This suggests that the ribosome contains specific sequence selectors for TM helices, with the first acting as an initial sensor for TM segments, to provide a hydrophobic nucleation point for folding [39], indicating that folding pathways are dependent on sequence, and protein identity. These interactions further show that specific sequences in the TM-helix primary structure can control the entry of the helix into the translocon much like the epitope labelling work, however UAAs have only been used to study single-membrane spanning proteins with the overall picture of insertion of polytopic proteins still to be uncovered. Certain fluorescence dyes coupled with UAA technology can also give insightful information into its hydrophobic microenvironment due to a red or blue shift in its emission maxima. This was nicely used to uncover minor structural alterations in the TIM23 mitochondrial channel when subject to a proton motive force (PMF) using wheat-germ extract, an alternative cell-free expression system sometimes used for eukaryotic protein expression [40].

The SecM arrest-peptide (AP) has been used as an *in vivo* force-sensor by the von Heijne group, giving insight into forces acting on the co-translationally inserting and folding segments of the NC [41, 42]. The force of NC release is related to a measurement of the fraction of truncated protein per fully translated protein ( $f_{TL}$ ), and plotted against helix number, and a force-profile for a particular protein is determined. This methodology seeks to describe the force (or  $f_{TL}$ ) acting on a hydrophobic segment at varying distance from the AP, most likely describing interactions arising from NC insertion into the translocation machinery and TM helix partitioning into the membrane.

It was hypothesised that the detected force experienced by hydrophobic domains increases when the upstream TM helices fold and partition into the membrane, possibly signifying interactions between helices in the co-translational folding process [43]. Polytopic membrane proteins; CaiT, NhaA, EmrD, BtuC and GtpT, were cloned into the AP construct [44]. The apparent free-energies of protein insertion ( $\Delta G_{app}$ ) were predicted using  $\Delta G$ -predictor [45] and of each candidate, one TM with a positive predicted  $\Delta G_{app}$  was selected suggesting unfavourable helix insertion. The interactions between the helix and the upstream N-terminal helices were quantified by an  $f_{TL}$  value, and show that an increased hydrophobicity subsequently increased insertion, as did the presence of previously synthesised upstream helices. The shorter the loop length between the helix and the prior helices also increased  $f_{TL}$  signifying the reason for the short loop length in polytopic membrane proteins. The exact identity each of these interactions themselves however cannot be determined, nor can they give us a time resolved sequence of events of folding and tertiary structure formation on the growing polypeptide.

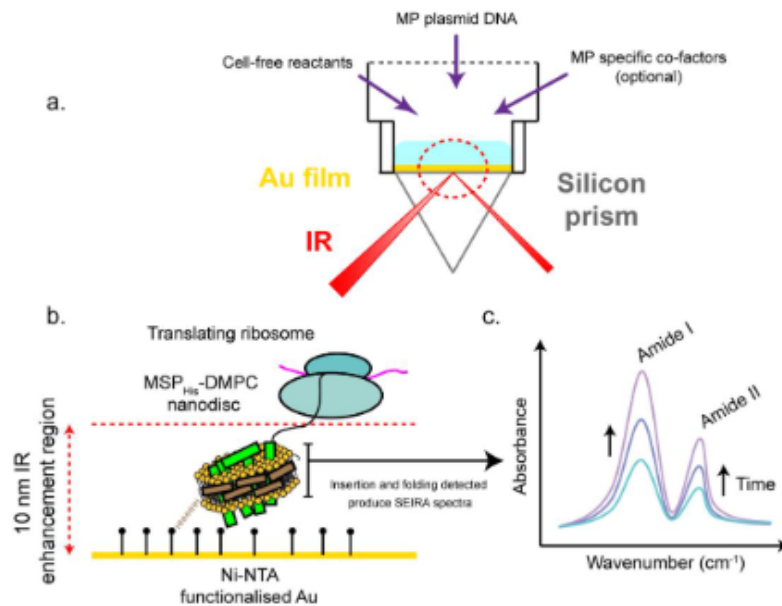
Furthermore, cryo-electron microscopy has been used to visualise static interactions between 7 TM proteorhodopsin (PR) [46], and the SecY translocon machinery in an RNC stalled state. A tryptophan dependent TnaC stalling peptide was used to generate SecY-bound RNCs with two complete TM helices and a hydrophilic region, necessary for stalling and producing tight interactions with SecY. These RNCs were co-purified in detergent with SecYEG and a RNC-SecY structure was obtained to 7.3 Å. The structure supports the idea that the TM helices may exit the SecY channel via the lateral gate for partition into the membrane as TMs 1 and 2 are directly in front of the lateral gate. It was proposed that YidC also positions itself outside of the lateral gate (although not detected in this study), suggesting

further chaperone regulation for TM partitioning, along with SecA, which cannot bind to the ribosome at the same time as SecY. Interactions between the positively charged cytoplasmic loop of the first two helices with a ribosomal rRNA helix were also observed. This suggests a role for the ribosome in retaining charged loops on the cytoplasmic side of the bilayer during TM integration into the membranes in this system.

Membrane protein co-translational folding studies currently lag behind that of soluble folding studies. Nascent chain studies of the latter, soluble proteins include HemK folding pathways by the Rodnina lab, using *in vitro* FRET based assays between fluorophores at different positions along the nascent chain [47]. The changes in FRET were time-resolved and suggested 4 independent intermediates for the co-translational folding of HemK which begin in the ribosome exit tunnel, and are only limited by translation rate. Translation rate and its effect on folding has been further explored harnessing rare codon technology [48]. Furthermore, much progress has been made in the structural dynamics of trapped nascent chains by the Christodoulou group using various NMR techniques with SecM mediated stalling [49, 50]. The globular FLN5 protein domain sampled transient interactions with the ribosomal RNA and the proteins surrounding the exit site.

#### 4. Time-resolving nascent chain structure formation during co-translational folding

A recent approach to directly observe temporal formation of NC structure, during co-translational folding of MPs into membranes is surface enhanced IR spectroscopy (SEIRAS). The SEIRA setup (figure 3) allows for IR spectra to be in the field of IR enhancement, resulting from plasmonic resonance at a thin gold surface deposited on a silicon prism. This gives a 10-100-fold increase in sensitivity compared to conventional IR spectroscopy [51]. There is a 10 nm enhancement region, which allows spectra to be obtained only of uniformly oriented lipid-nanodiscs and protein within the nanodisc, but not the translating ribosomes nor anything outside the membrane. The IR amide I (C=O stretching vibration) and II band (C-N stretching and N-H bending) intensities give fingerprints of protein structure, and particularly the position and width of the amide I band indicate helix formation packing during co-translational insertion.





### Figure 3: SEIRAS set-up and methodology

a) Reaction chamber and contents. The chamber sits on silicon prism with a thin-layer of gold. A 10 nm enhancement area above the gold surface results from plasmon resonance, due a reflection of the IR beam in the gold-prism interface. Plasmid DNA of the MP of interest is added, along with the cell-free kit containing ribosomes, T7 polymerase and 'feed' mixture with amino acids and energy replenishment components. Here, any co-factors required for MP folding (such as retinal for bR) are added. b) The IR beam illuminates a region up to 10 nm where a uniform layer of MSP-DMPC poly-histidine tethered nanodiscs sit on the Ni-NTA functionalised gold. The ribosomes translate the plasmid DNA producing NCs, which spontaneously insert into the nanodisc allowing in situ measurements of co-translational folding. c) Schematic SEIRA spectra output showing amide I and amide II peak intensity, which increase over time, and yielding kinetic information on folding.

The first MP co-translational study was carried out on bacteriorhodopsin (bR), using a commercial cell-free kit with bR DNA, retinal cofactor, and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) based MSP-nanodiscs. Both amide I and II bands increased as a function of time after the reaction started until a significant shift in peaks suggested that membrane insertion had occurred. In the late stages of folding, tertiary structure formation was observed with a sharpening of the amide I band, characteristic of  $\alpha$ -helical bundling [52]. No structure formation was observed when retinal was absent.

More recently, cell-free expression of the *E. coli* membrane proteins rhomboid protease GlpG and disulphide reductase DsbB were followed by SEIRAS. As with bR, no translocon was present and thus the observed folding was spontaneous and efficient, unassisted by translation apparatus [4]. For GlpG, the earliest detected IR bands corresponded to  $\alpha$ -helices and  $\beta$ -structure, and disordered structure. The latter then decreased as more  $\alpha$ -helical structure was produced.  $\alpha$ -helical bundling between TM helices occurred [4, 52] as GlpG folds into its polytopic structure. These structural changes all occurred within the time required to make the full length protein, strongly indicating that all folding occurred co-translationally.

DsbB co-translational folding was slower with more disordered structure initially, and helix formation and packing occurring both co- and post-translationally [4]. The observed structure formation during folding of both GlpG and DsbB was consistent with hydrophobicity plots; all 6 helices of GlpG are predicted to be stable by hydrophobicity scales for insertion into the lipid headgroups as well as via the translocon [53]. In contrast, TMs 2 and 3 of the 4 TM DsbB are predicted to be unstable, and thus could initially form disordered structure and require TMs 1 and 4 already in the membrane for the remaining helices to insert and fold to give the final protein structure summarised in (figure 2). DsbB however, contains a large unstructured periplasmic loop, which does not cross the membrane in this reductionist system. This results in an inverted topology as shown by AMS labelling in liposomes. *In vivo*, the translocon is likely to translocate the periplasmic loop across the membrane. SEIRAS itself, without labelling of the protein, cannot identify which regions of the protein are folding.

Each protein tested using SEIRAS so far has inserted into the bilayer in the absence of any translocation machinery suggesting, and confirming previous work [54-57] that the lipids themselves possess the necessary mechanical energy to modulate lipid binding independent of the translocon. The co-translational folding of GlpG for example was enhanced by certain lipids, including PE which alters the mechanical properties and phosphatidylglycerol (PG) that introduces negative headgroup charge. There is a degree of folded and misfolded structure for GlpG, DsbB and bR (figure 2) in their early stages, and it may be that addition of a translocon would reduce the amount of misfolded protein.

### 6. Conclusions

There are currently very few studies on the structure formation of  $\alpha$ -helical membrane proteins from a co-translational perspective. However, our ever expanding and refined arsenal of tools to study these systems, coupled with both biochemical and biophysical methods, is starting to paint a creditable picture

of the folding pathways of a handful of polytopic membrane proteins. Bacteriorhodopsin folding has been extensively studied through classic denaturant refolding experiments and biochemical assays to determine the order of TM insertion, and subsequently by time-resolved SEIRAS. This latter approach has considerable potential to reveal the temporal formation of structure during co-translational insertion and folding, particularly when combined with other approaches such as FRET, protease digestion and RNCs. The details of secondary structure formation for GlpG and DsbB has been observed and refined with input from predicted hydropathy plots. Additionally, biochemical principles of using *in vivo* force-mediated sensors can give insight into insertion, whilst cryoEM is beginning to deliver high resolution structural information of nascent chains and their local interactions with the insertion apparatus.

#### **Acknowledgements**

We thank Dr Nicola Harris, Dr Heather Findlay and Dr Eamonn Reading for proofreading the manuscript.

We thank the ERC and King's College, London for funding (Advanced Grant 294342 to PJB and PhD studentship to GP).

#### **References**

- [1] V.A. Gold, F. Duong, I. Collinson, Structure and function of the bacterial Sec translocon, *Mol Membr Biol*, 24 (2007) 387-394, DOI: 10.1080/09687680701416570
- [2] J. Komar, S. Alvira, R.J. Schulze, R. Martin, A.N.J.A. Lycklama, S.C. Lee, T.R. Dafforn, G. Deckers-Hebestreit, I. Berger, C. Schaffitzel, I. Collinson, Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC, *Biochem J*, 473 (2016) 3341-3354, DOI: 10.1042/BCJ20160545
- [3] W. Dowhan, M. Bogdanov, Lipid-protein interactions as determinants of membrane protein structure and function, *Biochem Soc Trans*, 39 (2011) 767-774, DOI: 10.1042/BST0390767
- [4] N.J. Harris, E. Reading, K. Ataka, L. Grzegorzewski, K. Charalambous, X. Liu, R. Schlesinger, J. Heberle, P.J. Booth, Structure formation during translocon-unassisted co-translational membrane protein folding, *Sci Rep-Uk*, 7 (2017), DOI: ARTN 8021 10.1038/s41598-017-08522-9
- [5] T.A. Rapoport, Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes, *Nature*, 450 (2007) 663-669, DOI: 10.1038/nature06384
- [6] M. Hermansson, G. von Heijne, Inter-helical hydrogen bond formation during membrane protein integration into the ER membrane, *Journal of Molecular Biology*, 334 (2003) 803-809, DOI: 10.1016/j.jmb.2003.10.019
- [7] D. Pitonzo, W.R. Skach, Molecular mechanisms of aquaporin biogenesis by the endoplasmic reticulum Sec61 translocon, *Biochim Biophys Acta*, 1758 (2006) 976-988, DOI: 10.1016/j.bbame.2006.04.021
- [8] Y. Lu, I.R. Turnbull, A. Bragin, K. Carveth, A.S. Verkman, W.R. Skach, Reorientation of aquaporin-1 topology during maturation in the endoplasmic reticulum, *Mol Biol Cell*, 11 (2000) 2973-2985, DOI: 10.1091/mbc.11.9.2973
- [9] W. Foster, A. Helm, I. Turnbull, H. Gulati, B.X. Yang, A.S. Verkman, W.R. Skach, Identification of sequence determinants that direct different intracellular folding pathways for aquaporin-1 and aquaporin-4, *Journal of Biological Chemistry*, 275 (2000) 34157-34165, DOI: 10.1074/jbc.M000165200
- [10] D.V. Fedyukina, S. Cavagnero, Protein Folding at the Exit Tunnel, *Annual Review of Biophysics*, Vol 40, 40 (2011) 337-359, DOI: 10.1146/annurev-biophys-042910-155338
- [11] F. Cymer, G. von Heijne, S.H. White, Mechanisms of integral membrane protein insertion and folding, *J Mol Biol*, 427 (2015) 999-1022, DOI: 10.1016/j.jmb.2014.09.014
- [12] B. Martoglio, M.W. Hofmann, J. Brunner, B. Dobberstein, The Protein-Conducting Channel in the Membrane of the Endoplasmic-Reticulum Is Open Laterally toward the Lipid Bilayer, *Cell*, 81 (1995) 207-214, DOI: 10.1016/0092-8674(95)90330-5



- [13] R.E. Dalbey, A. Kuhn, L. Zhu, D. Kiefer, The membrane insertase YidC, *Biochim Biophys Acta*, 1843 (2014) 1489-1496, DOI: 10.1016/j.bbamcr.2013.12.022
- [14] J. Serek, G. Bauer-Manz, G. Struhalla, L. van den Berg, D. Kiefer, R. Dalbey, A. Kuhn, *Escherichia coli* YidC is a membrane insertase for Sec-independent proteins, *EMBO J*, 23 (2004) 294-301, DOI: 10.1038/sj.emboj.7600063
- [15] P. Wang, R.E. Dalbey, Inserting membrane proteins: the YidC/Oxa1/Alb3 machinery in bacteria, mitochondria, and chloroplasts, *Biochim Biophys Acta*, 1808 (2011) 866-875, DOI: 10.1016/j.bbamem.2010.08.014
- [16] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, Cell-free translation reconstituted with purified components, *Nat Biotechnol*, 19 (2001) 751-755, DOI: 10.1038/90802
- [17] K. Ito, S. Chiba, K. Pogliano, Divergent stalling sequences sense and control cellular physiology, *Biochem Biophys Res Commun*, 393 (2010) 1-5, DOI: 10.1016/j.bbrc.2010.01.073
- [18] S. Bhushan, T. Hoffmann, B. Seidelt, J. Frauenfeld, T. Mielke, O. Berninghausen, D.N. Wilson, R. Beckmann, SecM-Stalled Ribosomes Adopt an Altered Geometry at the Peptidyl Transferase Center, *Plos Biol*, 9 (2011), DOI: ARTN e1000581 10.1371/journal.pbio.1000581
- [19] N. Ismail, R. Hedman, M. Linden, G. von Heijne, Charge-driven dynamics of nascent-chain movement through the SecYEG translocon, *Nat Struct Mol Biol*, 22 (2015) 145-149, DOI: 10.1038/nsmb.2940
- [20] M.S. Evans, K.G. Ugrinov, M.A. Frese, P.L. Clark, Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro, *Nat Methods*, 2 (2005) 757-762, DOI: 10.1038/nmeth790
- [21] P.J. Booth, P. Curnow, Folding scene investigation: membrane proteins, *Curr Opin Struct Biol*, 19 (2009) 8-13, DOI: 10.1016/j.sbi.2008.12.005
- [22] H.E. Findlay, P.J. Booth, The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition, *Sci Rep-Uk*, 7 (2017), DOI: ARTN 13056 10.1038/s41598-017-13290-7
- [23] N.J. Harris, K. Charalambous, H.E. Findlay, P.J. Booth, Lipids modulate the insertion and folding of the nascent chains of alpha helical membrane proteins, *Biochem Soc T*, 46 (2018) 1355-1366, DOI: 10.1042/Bst20170424
- [24] A.M. Seddon, P. Curnow, P.J. Booth, Membrane proteins, lipids and detergents: not just a soap opera, *Bba-Biomembranes*, 1666 (2004) 105-117, DOI: 10.1016/j.bbamem.2004.04.011
- [25] Y. Kuruma, K. Nishiyama, Y. Shimizu, M. Muller, T. Ueda, Development of a minimal cell-free translation system for the synthesis of presecretory and integral membrane proteins, *Biotechnol Prog*, 21 (2005) 1243-1251, DOI: 10.1021/bp049553u
- [26] J.J. Wu, J.R. Swartz, High yield cell-free production of integral membrane proteins without refolding or detergents, *Bba-Biomembranes*, 1778 (2008) 1237-1250, DOI: 10.1016/j.bbamem.2008.01.023
- [27] K. Nishiyama, M. Maeda, M. Abe, T. Kanamori, K. Shimamoto, S. Kusumoto, T. Ueda, H. Tokuda, A novel complete reconstitution system for membrane integration of the simplest membrane protein, *Biochem Biophys Res Commun*, 394 (2010) 733-736, DOI: 10.1016/j.bbrc.2010.03.061
- [28] R.B. Quast, O. Kortt, J. Henkel, S.K. Dondapati, D.A. Wustenhagen, M. Stech, S. Kubick, Automated production of functional membrane proteins using eukaryotic cell-free translation systems, *J Biotechnol*, 203 (2015) 45-53, DOI: 10.1016/j.jbiotec.2015.03.015
- [29] R.B. Quast, A. Sonnabend, M. Stech, D.A. Wustenhagen, S. Kubick, High-yield cell-free synthesis of human EGFR by IRES-mediated protein translation in a continuous exchange cell-free reaction format, *Sci Rep-Uk*, 6 (2016), DOI: ARTN 30399 10.1038/srep30399
- [30] T.H. Bayburt, Y.V. Grinkova, S.G. Sligar, Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins, *Nano Lett*, 2 (2002) 853-856, DOI: 10.1021/nl025623k

- [31] T.H. Bayburt, S.G. Sligar, Single-molecule height measurements on microsomal cytochrome P450 in nanometer-scale phospholipid bilayer disks, *P Natl Acad Sci USA*, 99 (2002) 6725-6730, DOI: 10.1073/pnas.062565599
- [32] I.G. Denisov, Y.V. Grinkova, A.A. Lazarides, S.G. Sligar, Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size, *J Am Chem Soc*, 126 (2004) 3477-3487, DOI: 10.1021/ja0393574
- [33] K.S. Simon, N.L. Pollock, S.C. Lee, Membrane protein nanoparticles: the shape of things to come, *Biochem Soc Trans*, 46 (2018) 1495-1504, DOI: 10.1042/BST20180139
- [34] H. Dale, M.P. Krebs, Membrane insertion kinetics of a protein domain in vivo. The bacterioopsin n terminus inserts co-translationally, *J Biol Chem*, 274 (1999) 22693-22698, DOI: 10.1074/jbc.274.32.22693
- [35] H. Dale, C.M. Angevine, M.P. Krebs, Ordered membrane insertion of an archaeal opsin in vivo, *Proc Natl Acad Sci U S A*, 97 (2000) 7847-7852, DOI: 10.1073/pnas.140216497
- [36] M. Bogdanov, P.N. Heacock, W. Dowhan, A polytopic membrane protein displays a reversible topology dependent on membrane lipid composition, *EMBO J*, 21 (2002) 2107-2116, DOI: 10.1093/emboj/21.9.2107
- [37] K. Moss, A. Helm, Y. Lu, A. Bragin, W.R. Skach, Coupled translocation events generate topological heterogeneity at the endoplasmic reticulum membrane, *Molecular Biology of the Cell*, 9 (1998) 2681-2697, DOI: 10.1091/mbc.9.9.2681
- [38] A.E. Johnson, The co-translational folding and interactions of nascent protein chains: a new approach using fluorescence resonance energy transfer, *Febs Lett*, 579 (2005) 916-920, DOI: 10.1016/j.febslet.2004.11.046
- [39] C.A. Woolhead, P.J. McCormick, A.E. Johnson, Nascent membrane and secretory proteins differ in FRET-detected folding far inside the ribosome and in their exposure to ribosomal proteins, *Cell*, 116 (2004) 725-736, DOI: 10.1016/S0092-8674(04)00169-2
- [40] N.N. Alder, R.E. Jensen, A.E. Johnson, Fluorescence mapping of mitochondrial TIM23 complex reveals a water-facing, substrate-interacting helix surface, *Cell*, 134 (2008) 439-450, DOI: 10.1016/j.cell.2008.06.007
- [41] M.J.M. Niesen, A. Muller-Lucks, R. Hedman, G. von Heijne, T.F. Miller, 3rd, Forces on Nascent Polypeptides during Membrane Insertion and Translocation via the Sec Translocon, *Biophys J*, 115 (2018) 1885-1894, DOI: 10.1016/j.bpj.2018.10.002
- [42] D.H. Goldman, C.M. Kaiser, A. Milin, M. Righini, I. Tinoco, Jr., C. Bustamante, Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo, *Science*, 348 (2015) 457-460, DOI: 10.1126/science.1261909
- [43] N. Ismail, R. Hedman, N. Schiller, G. von Heijne, A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration, *Nat Struct Mol Biol*, 19 (2012) 1018-1068, DOI: 10.1038/nsmb.2376
- [44] F. Cymer, G. von Heijne, Cotranslational folding of membrane proteins probed by arrest-peptide-mediated force measurements, *Proc Natl Acad Sci U S A*, 110 (2013) 14640-14645, DOI: 10.1073/pnas.1306787110
- [45] T. Hessa, N.M. Meindl-Beinker, A. Bernsel, H. Kim, Y. Sato, M. Lerch-Bader, I. Nilsson, S.H. White, G. von Heijne, Molecular code for transmembrane-helix recognition by the SecE1 translocon, *Nature*, 450 (2007) 1026-1030, DOI: 10.1038/nature06387
- [46] L. Bischoff, S. Wickles, O. Berninghausen, E.O. van der Sluis, R. Beckmann, Visualization of a polytopic membrane protein during SecY-mediated membrane insertion, *Nat Commun*, 5 (2014), DOI: ARTN 4103 10.1038/ncomms5103
- [47] W. Holtkamp, G. Kocic, M. Jager, J. Mittelstaet, A.A. Komar, M.V. Rodnina, Cotranslational protein folding on the ribosome monitored in real time, *Science*, 350 (2015) 1104-1107, DOI: 10.1126/science.1250344

- [48] F. Buhr, S. Jha, M. Thommen, J. Mittelstaet, F. Kutz, H. Schwalbe, M.V. Rodnina, A.A. Komar, Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations, *Mol Cell*, 61 (2016) 341-351, 10.1016/j.molcel.2016.01.008
- [49] A.M.E. Cassaignau, H.M.M. Launay, M.E. Karyadi, X.L. Wang, C.A. Waudby, A. Deckert, A.L. Robertson, J. Christodoulou, L.D. Cabrita, A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy, *Nat Protoc*, 11 (2016) 1492-1507, DOI: 10.1038/nprot.2016.101
- [50] L.D. Cabrita, A.M.E. Cassaignau, H.M.M. Launay, C.A. Waudby, T. Wlodarski, C. Camilloni, M.E. Karyadi, A.L. Robertson, X.L. Wang, A.S. Wentink, L.S. Goodsell, C.A. Woolhead, M. Vendruscolo, C.M. Dobson, J. Christodoulou, A structural ensemble of a ribosome-nascent chain complex during cotranslational protein folding, *Nat Struct Mol Biol*, 23 (2016) 278-285, 10.1038/nsmb.3182
- [51] K. Ataka, S.T. Stripp, J. Heberle, Surface-enhanced infrared absorption spectroscopy (SEIRAS) to probe monolayers of membrane proteins, *Biochim Biophys Acta*, 1828 (2013) 2283-2293, DOI: 10.1016/j.bbamem.2013.04.026
- [52] A. Baumann, S. Kerruth, J. Fitter, G. Buldt, J. Heberle, R. Schlesinger, K. Ataka, In-Situ Observation of Membrane Protein Folding during Cell-Free Expression, *PLoS One*, 11 (2016) e0151051, DOI: 10.1371/journal.pone.0151051
- [53] C. Snider, S. Jayasinghe, K. Hristova, S.H. White, MPEX: A tool for exploring membrane proteins, *Protein Sci*, 18 (2009) 2624-2628, DOI: 10.1002/pro.256
- [54] P.J. Booth, R.H. Templer, W. Meijberg, S.J. Allen, A.R. Curran, M. Lorch, In vitro studies of membrane protein folding, *Crit Rev Biochem Mol Biol*, 36 (2001) 501-603, 10.1080/20014091074246
- [55] P. Curnow, P.J. Booth, Combined kinetic and thermodynamic analysis of alpha-helical membrane protein unfolding, *Proc Natl Acad Sci U S A*, 104 (2007) 18970-18975, DOI: 10.1073/pnas.0705067104
- [56] H. Lu, P.J. Booth, The final stages of folding of the membrane protein bacteriorhodopsin occur by kinetically indistinguishable parallel folding paths that are mediated by pH, *J Mol Biol*, 299 (2000) 233-243, DOI: 10.1006/jmbi.2000.3735
- [57] J.U. Bowie, Membrane proteins: a new method enters the fold, *Proc Natl Acad Sci U S A*, 101 (2004) 3995-3996, DOI: 10.1073/pnas.0400671101



**Capturing Membrane Protein Ribosome Nascent Chain Complexes in a Native-like Environment for Co-translational Studies**

Grant A. Pellowe, Heather E. Findlay, Karen Lee, Tim M. Gemeinhardt, Laura R. Blackholly, Eamonn Reading,\* and Paula J. Booth\*

Cite This: *Biochemistry* 2020, 59, 2764–2775

Read Online

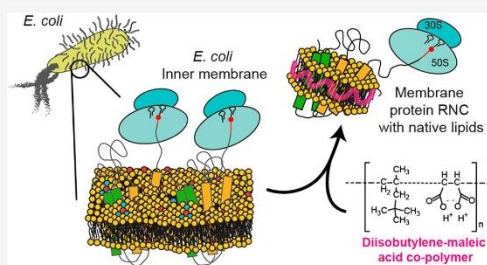
ACCESS |

Metrics &amp; More

Article Recommendations

Supporting Information

**ABSTRACT:** Co-translational folding studies of membrane proteins lag behind cytosolic protein investigations largely due to the technical difficulty in maintaining membrane lipid environments for correct protein folding. Stalled ribosome-bound nascent chain complexes (RNCs) can give snapshots of a nascent protein chain as it emerges from the ribosome during biosynthesis. Here, we demonstrate how SecM-facilitated nascent chain stalling and native nanodisc technologies can be exploited to capture *in vivo*-generated membrane protein RNCs within their native lipid compositions. We reveal that a polytopic membrane protein can be successfully stalled at various stages during its synthesis and the resulting RNC extracted within either detergent micelles or diisobutylene–maleic acid co-polymer native nanodiscs. Our approaches offer tractable solutions for the structural and biophysical interrogation of nascent membrane proteins of specified lengths, as the elongating nascent chain emerges from the ribosome and inserts into its native lipid milieu.



Most proteins begin to fold as they are translated unidirectionally from N- to C-termini by the ribosome. A significant drawback to much existing research into folding mechanisms is the study of full-length, isolated proteins that are unrepresentative of this co-translational cellular biosynthesis. A key advance has been the development of ribosome nascent chain (RNC) preparations in which elongating nascent chains are stalled at various points, so that biophysical methods can be used to probe the structures of specified length chains attached to ribosomes.<sup>1–9</sup> Typically, the *Escherichia coli* arrest sequence SecM is employed to halt translation.<sup>2,9,10</sup> Stalling translation, as well as truncating nascent chains, has proven to be instrumental in membrane protein insertion studies, but as yet, no membrane protein RNCs have been isolated directly from cells in their native lipid environments, which hinders detailed molecular mechanistic and structural investigations.

Co-translational folding of  $\alpha$ -helical transmembrane (TM) proteins is often assisted by a membrane insertase apparatus, such as the *E. coli* SecYEG translocon. SecYEG forms the basis of a holotranslocon (HTL), along with co-proteins SecDF and insertase YidC.<sup>11</sup> An emerging nascent chain is delivered to the SecYEG translocon by the signal recognition particle (SRP). Lipid interactions influence membrane insertion,<sup>12,13</sup> with lipid headgroup charge and interactions with the nascent chain affecting TM insertion efficiency and topology.<sup>14</sup> The bulk properties of the lipid bilayer, such as its lateral pressure and hydrophobic thickness, also likely impact membrane protein insertion and folding.<sup>12</sup> Thus, the cell membrane “arena”

required for membrane protein folding adds increasing complexity to the already intricate co-translational folding pathway of proteins.

Membrane protein co-translational folding and insertion have largely been studied using *in vitro* transcription/translation (IVTT). This *in vitro* approach enables control of protein translation within simplified cell and membrane extracts, or purified components such as the PURExpress system<sup>15</sup> and synthetic lipid mixtures. Assorted investigations of the inserted state have been undertaken using biochemical assays such as epitope binding, limited proteolysis, and chemical cross-linking.<sup>16–22</sup> Moreover, structure formation during co-translational folding has been probed by infrared spectroscopy.<sup>21</sup> These studies have led to exceptional insight into the co-translational insertion of membrane proteins; however, these translation systems often give low yields and are expensive.

The use of the *E. coli* SecM stalling sequence, found in nature to regulate the translation rate of higher-order proteins during co-translational folding,<sup>23,24</sup> is perhaps the most

Received: May 19, 2020

Revised: July 2, 2020

Published: July 6, 2020



common method for RNC generation. In wild-type (WT) SecM, a proline at position 166 alters the geometry of the peptidyl transferase center (PTC) preventing the formation of the ester linkage, resulting in the subsequent stalling of the ribosome along the mRNA.<sup>9,24,25</sup> Insertion of a SecM sequence downstream of a nascent chain allows RNCs to be generated both *in vitro* and *in vivo*,<sup>23</sup> thus trapping a polypeptide chain at a specified “length”. In addition, SecM has been used to give a measure of the force exerted on a TM helix during membrane insertion,<sup>16,17,26</sup> and an amino acid hydrophobicity scale has been derived from apparent free energies of TM insertion.<sup>27</sup> Other stalling sequences exist, such as the TnaC stalling peptide that makes extensive contacts with the majority of the ribosome exit tunnel,<sup>6</sup> rather than altering the PTC. Although these stalling methods have proven to be highly informative, we are still some way from understanding the molecular details of membrane protein co-translational folding, with our molecular and mechanistic knowledge currently deriving from refolding studies of full-length isolated nascent chains.<sup>28</sup>

For water-soluble proteins, the production and purification of high yields of homogeneous RNCs samples have enabled structural snapshots to be obtained at different nascent chain lengths, in particular by cryo-electron microscopy (cryoEM) analysis.<sup>1–3</sup> Moreover, it has been shown that RNCs are suited for advanced biophysical analysis, impressive examples being the use of solution-state nuclear magnetic resonance (NMR) to decipher a nascent chain structural ensemble<sup>1–3,9</sup> and mechanical force manipulation for stability assessment of the nascent chain fold.<sup>29,30</sup>

Membrane protein RNCs are considerably more challenging to isolate than those for water-soluble proteins, as they are complicated by the requirement of a lipid membrane. Nevertheless, notable achievements have been made in the cryoEM structural characterization of membrane protein RNCs. For example, Beckmann et al. found, at high resolution, that interaction of a TM nascent chain with SecY can cause opening of its channel, likely to facilitate its insertion into the membrane. To achieve this interaction, they used detergent-solubilized RNCs that were then subsequently reconstituted with SecY-containing membrane scaffold protein (MSP)-based nanodiscs consisting of non-native lipid DMPC.<sup>4,5</sup> Membrane mimetics such as these provide a lipid bilayer but lack the native lipid composition and, therefore, may not truly represent its functional environment.

To improve this, polymer scaffolds have recently been used as a method to directly solubilize MPs within nanodiscs of their native lipid surround (so-called “native” nanodiscs), thus removing the requirement for detergent purification and reconstitution. The most widely used scaffold is the styrene-maleic acid (SMA) co-polymer. However, when considering the preparation and analysis of ribosome-containing samples, the SMA co-polymer has its disadvantages: SMA contains a large styrene aromatic ring that contributes to high-absorption spectra in the low-ultraviolet region, which makes concentration and purity analysis of RNCs using  $A_{280}/A_{260}$  difficult.<sup>31,32</sup> SMA is also less tolerant to divalent cations, with the precipitation of the polymer occurring at  $Mg^{2+}$  concentrations of  $>5$  mM.<sup>33</sup> An *E. coli* 70S ribosome contains more than 170  $Mg^{2+}$  atoms, with the loss of bound  $Mg^{2+}$  leading to 70S dissociation and degradation.<sup>34</sup> Therefore, to maintain 70S integrity, an *in vitro*  $Mg^{2+}$  concentration of  $>5$  mM is typically required, making SMA unsuitable for RNC study. To circumvent these spectroscopic and divalent cation

issues, the recently utilized diisobutylene-maleic acid (DIBMA) co-polymer was selected.<sup>31</sup> Its polymer structure contains an aliphatic diisobutylene moiety instead of the aromatic styrene component found within SMA. This allows for spectroscopic study as the styrene component that absorbed strongly within the range of 260–280 nm has been removed. Importantly, DIBMA also has a proven tolerance to high concentrations of divalent cations ( $\leq 20$  mM for  $Mg^{2+}$  and  $\leq 35$  mM for  $Ca^{2+}$ ),<sup>31,35</sup> making it a suitable candidate for ribosome study. Both DIBMA and SMA co-polymers produce thermostable MP nanodiscs, but DIBMA has a weaker impact on the acyl chain order and lipid phase in solubilized membranes than SMA, maintaining a nearer physiological membrane environment.<sup>31</sup>

In view of the importance of lipid composition and membrane properties in defining the behavior of membrane proteins, we have developed an approach to extracting RNCs directly from their cellular membrane. We present a strategy for producing high yields of homogeneous polytopic membrane protein RNCs within their native lipid environments. Membrane protein RNCs are isolated from native membranes using DIBMA co-polymer native nanodisc technology.<sup>33,36,37</sup> Overall, our approach enables the structural and biochemical interrogation of cellular co-translational membrane protein translation and folding within a more physiological environment than is currently possible with existing *in vitro* or reconstituted systems.

## ■ MATERIALS AND METHODS

**Plasmid Construction.** His<sub>6</sub>-GlpG-SecM plasmids were constructed for GlpG RNC overexpression. pET28a was linearized with NdeI and XhoI restriction enzymes (New England Biolabs). The glpG gene (subcloned from Harris et al.<sup>21</sup>) and a 44-amino acid sequence motif encoding both the SecM arrest motif and additional amino acids (subcloned from Rutkowska et al.<sup>9</sup>) were amplified and cloned simultaneously into the linearized pET28a vector using In-Fusion HD cloning (Takara Bio). The resulting pET28a-His<sub>6</sub>-GlpG-SecM vector was then used in subsequent deletion cloning reactions to generate constructs of different GlpG “lengths” using the Q5 site-directed mutagenesis kit (New England Biolabs). A pET28a-His<sub>6</sub>-GlpG plasmid was also generated for overexpression of His<sub>6</sub>-GlpG without the linker-SecM sequence using the same procedures.

To substitute the WT SecM sequence for the arrest-enhanced (AE1) SecM sequence, primers were designed for the AE1 sequence based on Cymer et al.,<sup>26</sup> and polymerase chain reaction was carried out using the Q5 site-directed mutagenesis kit (New England Biolabs) on each RNC vector.

**GlpG-RNC Growth in BL21 (DE3).** N-Terminal His<sub>6</sub>-tagged GlpG constructs were truncated after two, four, and six TM helices using a WT SecM sequence, cloned into the pET28a expression vector, and transformed into *E. coli* strain BL21 (DE3). The following growth conditions are based on NMR protocols yielding large quantities of highly occupied RNCs,<sup>2</sup> which provide a maximal number of ribosomes ready for NC expression and are amenable to isotopic or selective labeling of the NCs.

Overnight LB cultures grown at 37 °C were used to seed 6 L of MDG {phosphate (M), aspartic acid (D), glucose (G) media [1× MDG salts (25 mM Na<sub>2</sub>HPO<sub>4</sub>, 25 mM KH<sub>2</sub>PO<sub>4</sub>, 50 mM NH<sub>4</sub>Cl, and 5 mM Na<sub>2</sub>SO<sub>4</sub>), 0.2% (w/v) L-aspartic acid (pH 7.0), 0.4% (w/v) D-glucose, 2 mM MgSO<sub>4</sub>, and 0.2×



trace metals, which were subsequently grown to saturation at 30 °C with shaking at 220 rpm]. Cells were harvested by centrifugation before suspension of an equal volume of enhanced M9 media [1× EM9 salts (7.1 g/L Na<sub>2</sub>HPO<sub>4</sub>, 3.4 g/L KH<sub>2</sub>PO<sub>4</sub>, and 0.58 g/L NaCl adjusted to pH 8.0–8.2), 0.4% (w/v) D-glucose, 5 mM MgSO<sub>4</sub>, 200 μM CaCl<sub>2</sub>, 0.25× trace metals, and 0.1% (w/v) NH<sub>4</sub>Cl] for induction. Trace metals were mixed as a 1000× stock containing 5 g/L EDTA, 0.83 g/L FeCl<sub>3</sub>·6H<sub>2</sub>O, 0.05 g/L ZnCl<sub>2</sub>, 0.01 g/L CuCl<sub>2</sub>, 0.01 g/L CoCl<sub>2</sub>·6H<sub>2</sub>O, 0.01 g/L H<sub>3</sub>BO<sub>3</sub>, and 1.6 g/L MnCl<sub>2</sub>·6H<sub>2</sub>O (pH 7.0). All media were supplemented with 30 μg/mL kanamycin. Cells were induced for protein production using 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 1.5–2 h at 30 °C before being washed with ice-cold PBS, harvested, and resuspended in lysis buffer [50 mM HEPES-KOH (pH 7.5), 1 M KOAc, 12 mM Mg(OAc)<sub>2</sub>, 5% (v/v) glycerol, 5 mM EDTA, 2 mM 2-mercaptoethanol, 1 mM phenylmethanesulfonyl (PMSF), 250 μg/mL chloramphenicol, and cOmplete EDTA-free protease inhibitor tablet (Roche)] with 1% lysozyme and RNase-free DNase I (NEB) and stored at –20 °C or flash-frozen and stored at –70 °C as “pellets” for cryo-milling.

For biotinylated RNC-GlpG, constructs with an N-terminal Avi tag were co-transformed with pBirAcm (Avidity), maintained by 10 μg/mL chloramphenicol, to express BirA. Cells were also grown to saturation in MDG media. However, to the enhanced M9 media was added 50 μM D-biotin dissolved in 10 mM bicine buffer (pH 8.3). The two plasmids were both induced using 1 mM IPTG and grown as described above, producing biotinylated RNC-GlpG constructs.

Alternatively, BL21-AI cells can be grown and RNCs expressed in Luria-Bertani or 2-YT medium. For these, 100 mL overnight cultures were seeded as described above, and cells grown at 37 °C until an OD<sub>600</sub> of 1.8 was reached. Cultures were cooled to 30 °C and induced with 1 mM IPTG and 0.1% L-arabinose for 1 h. Cells were washed and harvested as described above.

**DIBMA Preparation.** The diisobutylene/maleic acid copolymer (DIBMA) was precipitated out of a Sokalan CP9 solution (BASF) with 0.6 V of 4 M HCl. The solid was centrifuged at 17000g for 15 min and washed with doubly distilled H<sub>2</sub>O (ddH<sub>2</sub>O). This was repeated four times; 1.2 V of 4 M NaOH was used to solubilize the solid, before repeating the precipitation and washing steps described above. The wet pellet was then lyophilized for 72 h to produce the pure DIBMA polymer.<sup>38,39</sup>

The polymer was used to produce 20% (w/v) stocks in 2 M NaOH. Base was then added dropwise to carefully dissolve the solid and adjust the pH to 8.0. The concentration of the polymer solution was checked using a Reichert AR200 digital refractometer with a dn/dc of 1.346 M<sup>-1</sup>.<sup>31</sup> The 20% stock was used to produce 2.5% stocks in the necessary buffer for the solubilization of *E. coli* membranes.

Lyophilized DIBMA stocks were quality checked<sup>39</sup> using a Shimadzu IR Affinity-1s or a PerkinElmer Spectrum Two FTIR instrument, scanned across a complete wavenumber range of 4000–400 cm<sup>-1</sup>, with 16 scans to measure transmittance (Figure S3a). In particular, the carboxylate (1705 cm<sup>-1</sup>) and anhydride (1775 cm<sup>-1</sup>) bands were checked to ensure that our treatment of the polymer with strong acids did not result in a condensation of the anhydride ring, which would result in a significantly reduced solubilization efficiency.

**Purification of DDM and DIBMA GlpG-RNCs.** The cell lysis buffer described above and all subsequent RNC preparation buffers were prepared using RNase-free water to prevent the degradation of the rRNA content in our RNCs. This was prepared by incubating ddH<sub>2</sub>O with 0.1% diethyl pyrocarbonate (DEPC) to inactivate any RNase, prior to autoclaving that facilitated the removal of DEPC by decomposition into CO<sub>2</sub> and ethanol.

Cells were defrosted slowly and lysed using sonication, or cryo-milled in a Spex 6875 Freezer/Mill High Capacity Cryogenic Grinder using 15 cycles of 15 cps and allowed to thaw on ice. The cell debris was removed by centrifugation at 20000g for 45 min at 4 °C. This was repeated in fresh tubes if the resulting supernatant did not clear. The supernatant was then spun at 125000g for 30 min at 4 °C to harvest cell membranes with associated ribosomes. The crude membrane pellet was resuspended at concentrations of 25–50 mg/mL in washing buffer [50 mM HEPES-KOH (pH 7.5), 500 mM KOAc, 12 mM Mg(OAc)<sub>2</sub>, 5% (v/v) glycerol, 2 mM 2-mercaptoethanol, and 1 mM PMSF] aided by homogenization. To this suspension, 1% dodecyl maltoside (DDM) or 2.5% DIBMA (pH adjusted to 8.0) was added, and the mixture solubilized for 2 h at 4 °C or for 1 h at 25 °C, respectively. The solubilized membranes were spun at 125000g, and the supernatant was filtered before purification using an AKTA Pure system.

DDM-solubilized ribosomes carrying the nascent chain were purified using an AKTA Pure purification system. First, samples were separated by affinity chromatography using a HisTrap Nickel column. The column was washed with 5 column volumes of washing buffer containing 0.1% DDM. RNCs were passed over the column to allow binding. The column was washed with 30 mL of washing buffer with 20 mM imidazole with 0.1% DDM. RNCs were then eluted in washing buffer containing 500 mM imidazole and directly loaded onto a 16/60 HiPrep Sephacryl S-400 HR size-exclusion column pre-equilibrated with Tico size-exclusion buffer [10 mM HEPES-KOH (pH 7.5), 30 mM NH<sub>4</sub>Cl, 12 mM Mg(OAc)<sub>2</sub>, 5% (v/v) glycerol, 1 mM EDTA, 2 mM 2-mercaptoethanol, and 0.1 mM PMSF] containing 0.1% DDM. According to the manufacturer's specifications, the separation range of the Sephacryl S-400 column is 20 kDa to 8 MDa.

DIBMA-solubilized ribosomes with native membranes were first bound overnight at 4 °C to superaffinity Ni-NTA beads (Generon) pre-equilibrated with washing buffer before further cleanup using AKTA Pure. The supernatant containing unbound species was discarded, and beads were washed with 20 mM imidazole in washing buffer, before elution in 5 mL of washing buffer containing 500 mM imidazole. DDM was not present in any of these buffers. The eluate was subsequently concentrated to 2 mL using a 100 or 300 kDa concentrator and loaded directly onto the 16/60 HiPrep Sephacryl S-400 HR size-exclusion column pre-equilibrated with Tico size-exclusion buffer containing 0.2 M L-arginine. The left-hand side of the elution peak, previously shown to contain the highest proportion of 70S ribosomes with the nascent chain,<sup>40</sup> was taken and concentrated, and buffer exchange was carried out in a 100 or 300 kDa molecular weight cutoff spin concentrator and passed over a PD10 G25-superdex column, which facilitated the removal of arginine.

The RNC absorbance at 260 and 280 nm was monitored to evaluate the 70S content and its homogeneity using the OD<sub>260</sub>/OD<sub>280</sub> ratio (expected range of 1.9–2.0; 1 OD<sub>260</sub> = 24



pmol/mL<sup>41</sup>]. The samples were flash-frozen after purification; the samples were stable after freezing and thawing and observed to be homogeneous.

**Western Blotting Analysis.** To assess the RNC integrity, 1–2 pmol samples were run on 12% (w/v) NuPAGE gels at neutral pH and with a sample dye at pH 5.7 [30% glycerol, 0.25 M Bis-Tris (pH 5.7), 0.8% DTT, 8% SDS, and bromophenol blue] to maintain the ester bond between the tRNA and the nascent chain. DIBMA samples were boiled at 95 °C for 10 min prior to gel loading, and insoluble DIBMA material was spun out for 2 min at the maximum microfuge speed. To obtain the released forms of the nascent chain, the RNC samples were treated with 10 µg of RNase A to digest the rRNA (at room temperature for 5 min). The samples were analyzed by Western blotting using anti-histidine and streptavidin-HRP conjugate antibodies. For SecYEG Western blotting, an anti-SecY mouse antibody was used with an anti-mouse HRP secondary antibody. ImageJ<sup>42</sup> was used for densitometry analysis where necessary. Novex Sharp pre-stained protein ladder (Invitrogen) or PageRuler Plus prestained ladder (ThermoFisher) molecular weight markers were used.

It is worth noting that DIBMA, and other polymers, are prone to smearing when they are run on sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) gels and transferred to nitrocellulose. To allow the proteinaceous component to enter the acrylamide, it was necessary to boil the samples at 95 °C for 10 min to encourage denaturation before gel loading. The low-pH nature of the SDS loading dye precipitated out the DIBMA as expected.

RNC occupancy was achieved through the botting of full-length GlpG standards to produce a standard curve of blot intensity against picomoles of GlpG (Figure S5c), against a known concentration of RNC based on the rRNA-determined concentrations where 1 OD<sub>260</sub> = 24 pmol/mL.<sup>2</sup>

**Dynamic Light Scattering.** Dynamic light scattering for DIBMA RNCs was carried out in Tico buffer in particle size mode on a Malvern ZetaSizer machine.

**Sucrose Gradient Purifications.** Samples to be purified with a sucrose density gradient were solubilized and bound/eluted to superaffinity Ni-NTA resin as described above. Eluates were buffer exchanged out of imidazole using 100 kDa cutoff concentrators and PD10 columns equilibrated with washing buffer or pelleted at 117000g for 4 h and suspended in a suitable volume of washing buffer. Sucrose gradients were prepared with five steps of 5%, 10%, 20%, 30%, and 40% sucrose in washing buffer containing no glycerol. The sample was layered on top and centrifuged at 202000g in a Beckmann SW40 rotor for 16 h at 4 °C before 1 mL or 500 µL fractions were taken for analysis.

**Thin-Layer Chromatography Analysis of DIBMA-Solubilized Samples.** To extract lipids of the native membrane RNCs, a modified Folch method was used.<sup>43</sup> A 500 µL DIBMA-solubilized sample was added to 1.1 mL of a 10:23:1 (v/v) CHCl<sub>3</sub>/MeOH/1 M Tris (pH 8.0) solvent and mixed well for 1 h at 25 °C. Phase separation was then achieved by adding 1 mL of a 1:1 CHCl<sub>3</sub>/1 M Tris (pH 8.0) solvent; the mixture was vortexed, and each phase was allowed to separate at room temperature. The upper aqueous phase was removed, and the organic layer washed with ion switch buffer [50 mM Tris, 100 mM NaCl, and 100 mM EDTA (pH 8.0)] before being dried under a stream of nitrogen and suspended in CHCl<sub>3</sub> to the desired TLC concentration.

HPTLC silica gel TLC plates were run as described in ref 44 and washed in a 1:1 CHCl<sub>3</sub>/MeOH solvent before being air-dried. The plate was immersed in 2.3% boric acid in EtOH, air-dried, and baked at 100 °C for 15 min. One microgram of 18:1 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine (DOPE), 18:1 1,2-dioleoyl-*sn*-glycero-3-phospho-(1'-*rac*-glycerol) (DOPG), and *E. coli* cardiolipin (all from Avanti) were spotted along with the extract samples, and plates were run in a 65:25:4 CHCl<sub>3</sub>/MeOH/CH<sub>3</sub>COOH solvent to achieve good head-group separation of the lipids. The plate was immersed in a copper char solution (10% CuSO<sub>4</sub> and 8% H<sub>3</sub>PO<sub>4</sub>) before air drying and further charring at 145 °C for 15 min. Plates were imaged using blue fluorescence once they had cooled. Alternatively, the dried plate was stained with iodine vapor or sprayed with a molybdenum blue stain (commercial Dittmer-Lester stain<sup>45</sup>) to specifically detect phospholipid. ImageJ<sup>42</sup> was used for densitometry analysis, where the raw spot intensity was converted to mole fraction, and the percentage of total lipid composition was plotted with SD for at least three biological repeats (new growth and preparation of RNCs, lipid extraction, and running of TLC plates).

**Growth and Preparation of Full-Length WT N<sub>His</sub>-GlpG and GlpG-MycHis<sub>6</sub>.** pET28a-His<sub>6</sub>-GlpG and pET28a-GlpG-MycHis<sub>6</sub> plasmids were transformed into C43 *E. coli* cells (Sigma-Aldrich). Seven milliliters of an overnight-saturated culture was added to 1 L of LB culture, both with 30 µg/mL kanamycin. These cultures were grown at 37 °C until an OD of 0.6–0.8 was achieved. The cultures were then cooled to 16 °C, induced with 1 mM IPTG, and left overnight before the cells were harvested. The cells were washed with PBS and either stored at –80 °C or used immediately.

Cell pellets were thawed and resuspended in buffer A [300 mM sodium chloride and 20 mM 2-amino-2-hydroxymethylpropane-1,3-diol (Tris) (pH 7.4) at room temperature] supplemented with a cOmplete protease inhibitor tablet (Roche), PMSF, Benzonase nuclease, and 5 mM β-mercaptoethanol (β-ME).

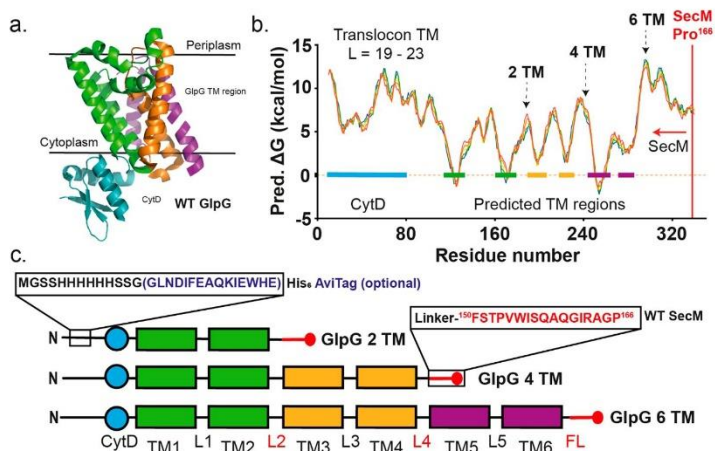
The cell suspension was passed twice through a microfluidizer (Constant Systems) at 25000 psi. Insoluble material was pelleted by centrifugation twice at 20000g for 25 min at 4 °C. Membranes were pelleted by centrifugation at 150000g for 1.5 h at 4 °C.

Membranes were resuspended at a concentration of 40 mg/mL in ice-cold buffer B [500 mM sodium chloride, 10% glycerol, and 50 mM Tris (pH 7.4) at room temperature] supplemented with a cOmplete protease inhibitor tablet (Roche) and PMSF and homogenized using a Potter-Elvehjem Teflon pestle and glass tube.

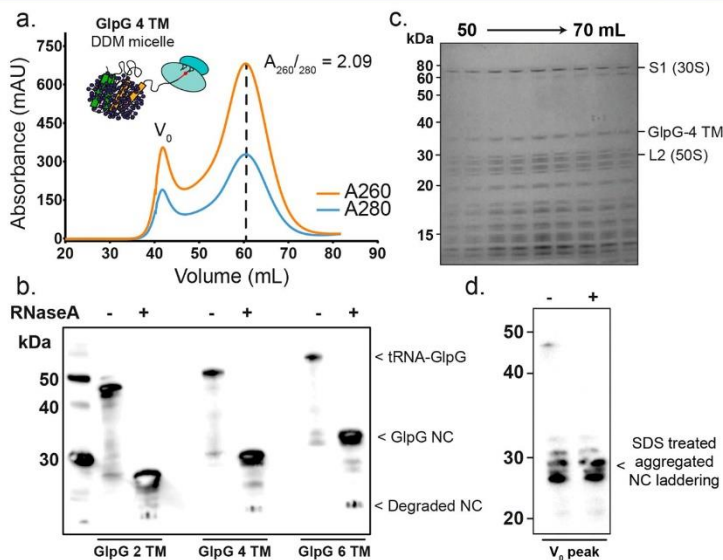
DDM was added to the suspension at a final concentration of 1% (w/v) to solubilize the membranes. The cell suspension with DDM was incubated overnight with gentle agitation at 4 °C, followed by centrifugation at 100000g for 1 h at 4 °C.

The supernatant was then filtered before being loaded onto a 1 mL HisTrap column (GE Healthcare) equilibrated in buffer B [500 mM sodium chloride, 20 mM imidazole, 10% (v/v) glycerol, 50 mM Tris (pH 7.4) at room temperature, and 0.025% (w/v) DDM].

The column was washed with 10 mL of buffer B and then 20 mL of 5% buffer C [500 mM sodium chloride, 500 mM imidazole, 10% (v/v) glycerol, 50 mM Tris (pH 7.4) at room temperature, and 0.025% (w/v) DDM].

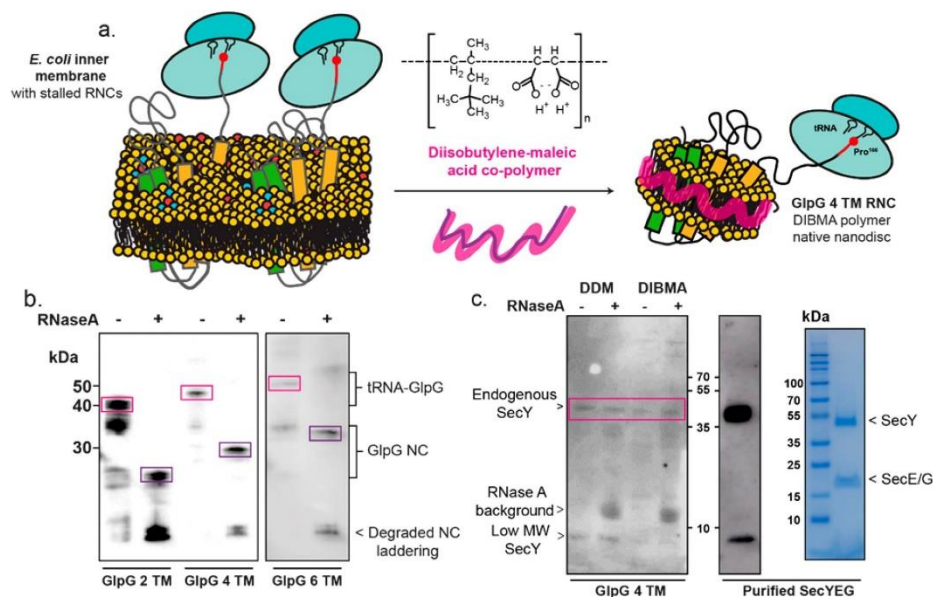


**Figure 1.** Construct design of GlpG RNCs at different co-translational “intermediates”. (a) PDB structure of WT GlpG: TM region, with helices 1 and 2 colored green, helices 3 and 4 colored orange, and helices 5 and 6 colored purple (PDB entry 2XTV<sup>19</sup>) and the cytoplasmic domain (CytD) colored blue (PDB entry 2LEP<sup>18</sup>). (b) MPEX<sup>27</sup> translocon TM predicted hydropathy plot for full-length GlpG with linker-WT SecM. A sliding scale length (*L*) of 19–23 amino acids was used for MPEX hydropathy prediction. Arrows depict the position chosen for the 3′ cloning of linker-WT SecM. This provides 44 amino acid residues between a stable TM helix and the SecM Pro<sup>166</sup> stalling residue. (c) Construct design for RNC truncations of GlpG. WT SecM was cloned to the C-terminus of each construct (red). A six-histidine tag was cloned to the N-terminus of the construct. An additional Avi tag (purple) could also be cloned and successfully utilized for *in vivo* biotinylation if desired (see the Supporting Information). The color scheme for GlpG constructs is consistent with panel a.



**Figure 2.** Purification of GlpG RNCs in DDM detergent micelles. (a) Representative SEC trace showing void and elution peaks at ~60 mL for four-TM GlpG RNC. (b) DDM-solubilized GlpG RNCs were purified using IMAC followed by SEC using a 16/60 HiPrep Sephacryl S-400 column. Low-pH SDS-PAGE and Western blotting analysis of fractions under major elution peak show stalled ribosome complex before (–) and after (+) RNase A treatment [tRNA-GlpG and GlpG nascent chain (NC)-labeled bands respectively]; the stall is confirmed by a decrease in molecular weight as tRNA is digested. The size of these bands increases as the protein increases by two TM helices with each stall. (c) SDS-PAGE of fractions under the elution peak in panel a, showing characteristic 70S ribosome protein bands: the 30S S1 protein at ~65 kDa, the 50S L2 protein at ~30 kDa, and a selection of other ribosomal proteins below 30 kDa after Coomassie staining.<sup>36</sup> (d) Western blot analysis of a void peak from a four-TM GlpG RNC preparation in the DDM detergent. This peak appears to contain aggregated, released nascent chain. All Western blots were probed using a polyhistidine-HRP antibody.





**Figure 3.** Preparation of GlpG RNCs in DIBMA nanodiscs shows evidence of co-purification with SecYEG. (a) Schematic of RNC DIBMA nanodisc preparation. *E. coli* inner membranes packed with membrane protein RNCs were subject to DIBMA solubilization to produce homogeneous RNC nanodiscs containing a native lipid environment. (b) GlpG RNC samples were solubilized with the DIBMA co-polymer and RNC DIBMA nanodiscs purified using IMAC followed by SEC. Arginine (0.2 M) was added to the SEC mobile phase to reduce the level of nonspecific aggregation of DIBMA nanodiscs with the resin of the 16/60 HiPrep Sephacryl S-400 size-exclusion column. Low-pH SDS-PAGE and Western blot analysis were used to assess the three RNC lengths from DIBMA preparations. Intact DIBMA-solubilized RNCs (pre-RNase A) for each stall length (two, four, and six TMs) are highlighted by pink boxes; released GlpG nascent chains due to RNase A treatment (+) are boxed in purple. Any unlabeled bands are identified as released, truncated, or aggregated GlpG caused by arginine-facilitated migration of the void peak into the SEC elution peak (see Figure S3c). All Western blots were probed using a polyhistidine-HRP antibody. (c) Evidence of endogenous SecYEG was detected within GlpG RNC samples prepared in the DDM detergent and DIBMA nanodiscs. Endogenous SecY was found to be present in samples before (–) and after (+) RNase A treatment (highlighted with a pink box, left panel). The low-molecular weight SecY band is characteristic of a C-terminal cleavage product.<sup>62</sup> Its 37 kDa band was confirmed by comparison with a Western blot (middle panel) and SDS-PAGE (right panel) of purified SecYEG, SecE and SecG co-migrating in this SDS-PAGE environment. All Western blots were probed using a monoclonal antibody for SecY.

GlpG was eluted with 100% buffer C and injected directly onto a Superdex 75 10/600 GL size-exclusion chromatography (SEC) column (GE Healthcare) equilibrated in buffer D [50 mM Tris (pH 7.4) at room temperature, 150 mM NaCl, 10% glycerol, and 0.025% (w/v) DDM]. Peak fractions that eluted from the SEC column containing GlpG were pooled and spin filtered before being flash-frozen and stored at  $-80$  or  $-20$  °C. SDS-PAGE and Western blotting, with an anti-histidine antibody, were used to assess GlpG purification.

**EnzCheck Protease Activity Assay for GlpG.** Detailed GlpG activity assay methods can be found in refs 21 and 46. Briefly, 5  $\mu\text{g/mL}$  EnzCheck BODIPY-Casein substrate (ThermoScientific) was cleaved by 1 mg/mL GlpG overnight and in the dark at 4 °C in GlpG purification buffer D (see above). The resulting fluorescence that was emitted at 510 nm upon excitation with 480 nm light was recorded on a Horiba Fluoromax 4 instrument. This was a measure of the reduction of the level of self-quenching of BODIPY as casein is cleaved, indicative of GlpG protease activity. GlpG's activity was inhibited with the addition of 100  $\mu\text{M}$  diisopropyl fluorophosphate (DFP).

**Growth and Preparation of SecYEG.** pBAD SecYEG was transformed into C43 cells and maintained with 100  $\mu\text{g/mL}$  ampicillin. Overnight cultures of transformed cells were grown at 37 °C and used to seed 3 L of LB medium containing 100  $\mu\text{g/mL}$  ampicillin. Cells were induced with 0.1% L-arabinose and grown until saturation was achieved. SecE (containing a C-terminal His<sub>6</sub> tag) is under the control of an arabinose promoter, and therefore, no IPTG is required for induction. Cells were harvested and suspended in 1× PBS with a cOmplete protease inhibitor tablet and stored at  $-20$  °C until the day of use.

Cells were defrosted slowly and incubated with Benzonase nuclease before lysis using a constant systems cell disruptor at 25000 psi. The membrane fraction was harvested at 150000g and solubilized in 20 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol, 0.1 mM PMSE, 2% DDM, 20 mM imidazole, and an EDTA-free protease inhibitor tablet for 2 h at 4 °C. Insoluble material was removed with a 100000g centrifugation step, and the supernatant was loaded directly onto an AKTA Pure purification system.

The sample was passed over a Ni-NTA column at a rate of 0.5 mL/min for affinity purification of the SecYEG using a six-

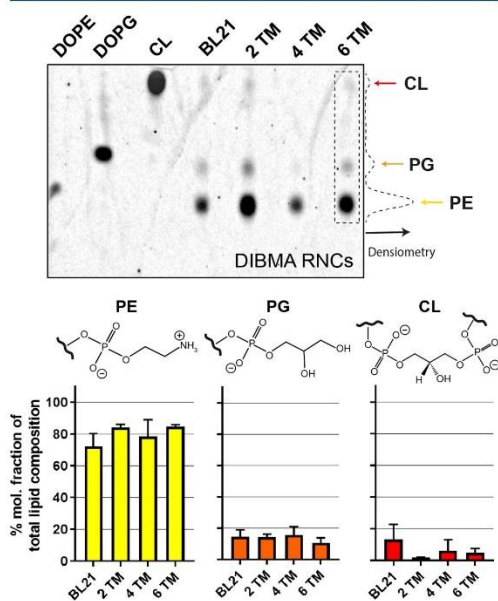
2769

<https://dx.doi.org/10.1021/acs.biochem.0c00423>  
Biochemistry 2020, 59, 2764–2775

His tag. The trap was washed with 30 mL of the buffer described above before elution in 20 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol, 0.2 mM PMSF, 0.1% DDM, 1 mM 2-mercaptoethanol, and 500 mM imidazole. The eluate was immediately passed over a Superdex 16/600 column in 20 mM Tris (pH 8.0), 300 mM NaCl, 10% glycerol, 0.1 mM PMSF, and 0.1% DDM. Peak fractions were collected and concentrated using a 100 kDa concentrator if necessary and subsequently snap-frozen and stored at  $-70^{\circ}\text{C}$  until use.

## RESULTS AND DISCUSSION

### A Strategy for the Preparation of Polytopic $\alpha$ -Helical Protein RNCs. The *E. coli* rhomboid protease GlpG was used



**Figure 4.** GlpG RNCs prepared in DIBMA nanodiscs contain native lipid compositions. DIBMA GlpG RNC samples and cell membranes, prepared from BL21 *E. coli* cells grown in minimal media, were subjected to Folch lipid extraction. Extractions were run alongside standards of DOPE, DOPG, and CL. A representative TLC plate from samples grown in MDG/EM9 media is shown. The three bar charts show a similar relative abundance of each lipid for each stall length (the 2 TM label denotes two-TM GlpG RNC extraction) as determined using densitometry analysis. The mean percent lipid composition and SD are plotted for three or more biological repeats for each stall length (see Materials and Methods). Phospholipid headgroups for phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and cardiolipin (CL) are shown.

to benchmark our strategies for RNC generation. GlpG has six transmembrane-spanning  $\alpha$ -helices and an N-terminal  $\alpha$ -helix/ $\beta$ -sheet structured domain housing its catalytic dyad (denoted CytD), separated from the TM domain by a flexible linker,  $L_{in}$ . High-resolution structures of each domain exist, but not that of the full-length protein.<sup>47–49</sup>

GlpG has become a model membrane protein for *in vitro* folding and stability work<sup>21,38</sup> and has been interrogated by chaotrope,<sup>50</sup> heating,<sup>51</sup> steric trapping,<sup>52</sup> or mechanical force.<sup>53</sup>

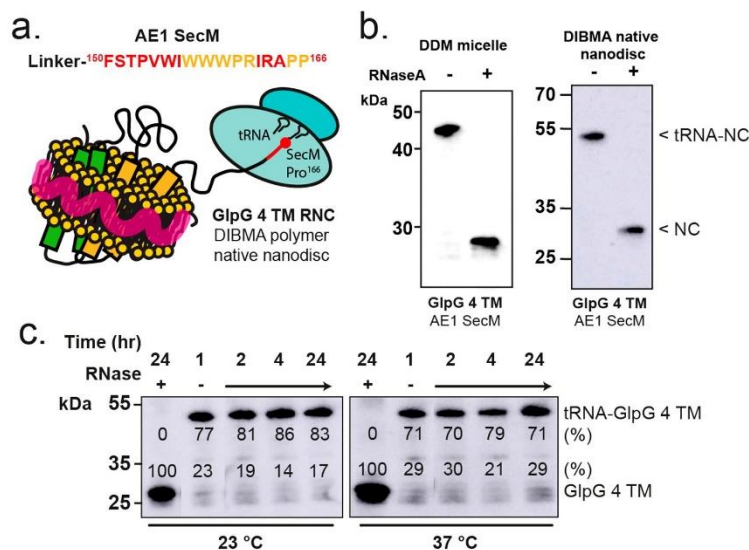
These systems, however, do not truly represent the *in vivo* folding situation whereby mRNA is translated by the ribosome and protein forms and folds during chain elongation in unison with translation. More recently, surface-enhanced infrared absorption spectroscopy (SEIRAS) and IVTT systems were used to determine secondary structure folding of the GlpG nascent chain as it was translated by the ribosome and inserted into translocon-free MSP-based lipid nanodiscs.<sup>20,21</sup> This allowed time-resolved analysis of co-translational folding of GlpG for the first time but nonetheless lacked a native membrane lipid environment.

GlpG RNCs of varying nascent chain lengths were generated *in vivo* using the SecM stalling motif (<sup>150</sup>FSTP<sup>166</sup>WISQA-QGIRAGP<sup>166</sup>). An *E. coli* SecM arrest sequence<sup>9,24</sup> was cloned downstream of two, four, and six predicted TM helices (see Materials and Methods). Stalling positions were chosen on the basis of translocon-mediated MPEX hydropathy analysis<sup>27</sup> (Figure 1a,b), which describes the free energy required for a stretch of amino acids to partition into the membrane environment, taking into account not only the identity of the amino acid but also the TM segment length and the influence of flanking amino acids. To allow all TM regions of interest to be fully translated, the length of the ribosome tunnel was also taken into account, which has been approximated to accommodate  $\sim 40$  fully extended amino acids.<sup>54</sup> Hence, the final Pro<sup>166</sup> stalling residue and the C-terminal end of the TM of interest were separated by 44 amino acids (including the SecM motif and linker residues).

To facilitate purification of *in vivo*-generated RNCs, a six-histidine affinity tag (His tag) was placed at the N-terminus of GlpG (Figure 1c). The introduction of the N-terminal tag did not affect protease activity when compared to a C-terminal construct (Figure S2a), or sorting to the membrane, and overexpression of these RNC constructs was found to be nontoxic to BL21 *E. coli* expression cells (Figure S2b). Additionally, we establish that *in vivo* biotinylation of the membrane protein nascent chain can be achieved through the incorporation of an Avi tag (see the Supporting Information and Figure S1). This expands the capability of our constructs for purification by biotin/avidin affinity chromatography and/or advanced biophysical interrogation.<sup>29,30</sup>

To assess the ability of the designed constructs to generate membrane protein RNCs, they were initially purified in detergent micelles following cell growth in both minimal and Luria-Bertani media. Cell membranes containing overexpressed GlpG RNCs were isolated and solubilized with *n*-dodecyl  $\beta$ -D-maltoside (DDM) detergent before purification with immobilized metal affinity chromatography (IMAC) and size-exclusion chromatography (SEC) (Figure 2a). Stalled nascent chains can be characterized by low-pH SDS-PAGE analysis, which preserves the nascent chain-tRNA ester bond supplying an increased shift in mass and a reduced rate of migration.<sup>2</sup> RNase A treatment digests the bound tRNA, leaving the stalled polypeptide. All three RNC constructs (two-, four-, and six-TM GlpG) eluted as two resolved peaks upon SEC (Figure 2a and Figure S2d); the void peak ( $V_0$ ) was found to contain aggregates of the released nascent chain (observed by Western blotting) and aggregated 70S ribosomes (observed by Coomassie staining) (Figure 2d and Figure S2c). The major elution peak at  $\sim 60$  mL contained homogeneous GlpG RNCs as confirmed by low-pH SDS-PAGE and Western blotting (Figure 2b,c), with an  $A_{260}/A_{280}$  ratio of  $\sim 2.1$  signifying the presence of homogeneous 70S ribosomes.<sup>41</sup>





**Figure 5.** Arrest-enhanced SecM provides yields MP RNCs with greater stability, occupancy, and purity. (a) AE1 SecM sequence, with residues differing from the WT SecM sequence colored orange. (b) Low-pH SDS-PAGE and Western blotting analysis show that the AE1 RNC constructs are consistent with WT SecM for four-TM GlpG in both DDM and DIBMA, but the degree of spontaneously released nascent chain is significantly reduced (see Figures 2a and 3b). (c) AE1 SecM GlpG four-TM RNCs purified in DIBMA nanodiscs were heated at 23 and 37 °C for 24 h. Samples were taken and blotted after 1, 2, 4, and 24 h before low-pH SDS-PAGE and Western blotting analysis. The ratio between intact RNC (~50 kDa band, tRNA-GlpG 4 TM) and spontaneously released nascent chain (~30 kDa band, GlpG 4 TM) was calculated for each lane using the band density and ImageJ.<sup>42</sup> Percentages of intact and released are quoted next to the corresponding band. There was no significant release for each temperature from 1 to 24 h.

This purification strategy is faster than sucrose gradient preparations commonly used for RNC preparations,<sup>2</sup> with the shorter purification time being preferable for maintaining the integrity of membrane protein samples in detergent micelles.<sup>53</sup> However, the RNC constructs were also amenable to sucrose gradient purification post-IMAC (Figure S2e).

Low-intensity bands underneath the major GlpG construct bands were observed by Western blotting and attributed to small populations of truncated GlpG. This is commonly observed within preparations of RNCs by SecM-based stalling,<sup>1,57</sup> caused by digestion with remaining proteases during purification. Notably, when the GlpG constructs were released from the ribosome, they became more prone to degradation, suggesting that the constructs are stabilized and/or protected by the ribosome to protease digestion (Figure 2b).

Yields of RNC varied between nascent chain lengths when overexpressed and purified under identical conditions. Purified yields for the two-, four-, and six-TM constructs were ~30, 100, and 25 pmol/mL, respectively, based on the rRNA ( $A_{260}$ ) content and assuming 100% ribosome/NC occupancy in DDM detergent. The individual yields may be linked to the extent of expected membrane force pulling of the different nascent chain lengths,<sup>16,17,26</sup> with an increased “insertion force” of some sequences reducing the stalling effect of SecM and in turn the quantity of stable RNCs obtained. MPEX analysis indicates that the four-TM construct is preceded by TM regions TM3 and TM4 that have the least favorable insertion energetics (Figure 1b). As a result, SecM stalling is effective and this is the highest-yield construct. On the

contrary, the two- and six-TM constructs are preceded by TMs with much more favorable insertion energetics (lower-energy minima), which may lead to decreased RNC stalling stability and, therefore, the reduced yields observed.

The yields achieved for GlpG RNCs are sufficient for further structural and biochemical analyses but are around 100-fold lower than those previously quoted for small globular protein RNCs by Cabrita et al.<sup>2</sup> This is likely due to the toxic nature of overexpressing membrane proteins, caused in part by saturation of available SecYEG translocons.<sup>58</sup> Indeed, GlpG has been shown to insert into the *E. coli* inner membrane using SecYEG and YidC,<sup>59</sup> however, details of these interactions are currently unknown.<sup>60</sup>

**Capturing Membrane Protein RNCs within DIBMA Nanodiscs.** To capture RNCs in a native lipid surrounding, GlpG RNCs with two, four, and six TMs were isolated with DIBMA using a protocol similar to that for DDM but with some important modifications (Figure 3a and Figure S3ab). Isolated membranes were solubilized using DIBMA, which has a lower solubilization efficiency compared to that of the DDM detergent and SMA co-polymer. Higher divalent cation concentrations (in the presence of 10 mM Mg<sup>2+</sup> or 7.5 mM Ca<sup>2+</sup>) have been shown to improve DIBMA-mediated solubilization of *E. coli* membranes by up to 2-fold, through association and neutralization of the polymer’s carboxylate groups.<sup>35</sup> To this end, the magnesium acetate concentration was increased in our membrane solubilization buffers from 6 to 12 mM. Solubilized membranes containing His-tagged GlpG RNCs were then batch bound to superaffinity Ni-NTA beads overnight at 4 °C to allow nickel–histidine coordination. This

2771

<https://dx.doi.org/10.1021/acs.biochem.0c00423>  
Biochemistry 2020, 59, 2764–2775

step was necessary as we found that DIBMA could interfere with nickel–histidine coordination,<sup>9,10</sup> reducing the level of the binding of the His tag to the Ni<sup>2+</sup> resin; in contrast, DDM detergent preparations did not require long incubation times for adequate binding to occur.

Dynamic light scattering was used to assess the size and homogeneity of solubilized, IMAC-purified four-TM GlpG RNCs and gave a mean peak size of 10.5 nm, which is close to the value of 12.7 nm previously reported for full-length GlpG in DIBMA.<sup>38</sup> The average particle size  $z$  (which is the intensity-weighted mean hydrodynamic size of the ensemble) was determined to be 33.0 nm, suggestive of populations of ribosome-bound and unbound discs causing sample polydispersity (Figure S3d).

Affinity-purified DIBMA-solubilized GlpG RNCs were then subjected to SEC for further purification from protein-free DIBMA discs, released nascent chains, and aggregated RNCs. However, during SEC, the DIBMA polymer was found to interact with the Sephacryl S400 resin, which prevented its elution from the column. To counter this, 0.2 M arginine was added to the SEC buffer mobile phase to mitigate interactions of the protein disc with the resin and eliminate possible aggregation<sup>61</sup> (Figure S3c). With arginine present, the DIBMA nanodisc samples could now be resolved by SEC but with an elution behavior different from that found within DDM detergent preparations: the void peak disappeared, and the major peak shifted to a larger elution volume (Figure S3c). 70S ribosomal controls in the presence and absence of arginine in SEC buffer showed that this shift was caused by arginine, while the  $A_{260}/A_{280}$  ratios were unaffected.

Low-pH SDS–PAGE and Western blotting analysis showed that two-, four-, and six-TM RNCs within DIBMA nanodiscs (pink box) were successfully obtained (Figure 3b). However, the Western blot shows an increased rate of nascent chain release (purple box), particularly in the two-TM GlpG RNC, even before treatment with RNase A. We suggest these bands are released nascent chain, previously found as aggregates in the void peak in DDM preparations, which migrate into the major elution peak during SEC, which is caused by the inclusion of arginine in the mobile phase (Figure 3b and Figure S3c).

Sucrose gradients, employing a continuous 5 to 50% gradient, were used to eradicate released nascent chain artifacts caused by SEC arginine preparation of RNC DIBMA nanodiscs following IMAC purification (see Materials and Methods). The increased density of the lipid discs results in DIBMA nanodisc samples residing in the bottom fifth of the sucrose layer (Figure S3e), as has been shown previously for membrane-associated RNCs,<sup>57</sup> providing pure and homogeneous samples. This supports the idea that sucrose gradient purification is also a viable purification strategy for RNC DIBMA nanodisc samples (DIBMA samples being more stable than those in DDM).

**The Endogenous SecYEG Translocon Associates with GlpG RNCs.** To explore whether any SecYEG remained associated with GlpG RNCs, we probed the four-TM GlpG RNC purified in DDM and DIBMA nanodiscs for the presence of any endogenous Sec translocase machinery using a monoclonal SecY antibody.<sup>62</sup> SDS–PAGE and Western blot analysis revealed the presence of endogenous SecY (which has a blot intensity that is lower than that of overexpressed RNC) within both DDM and DIBMA samples (Figure 3c). This copurification of SecY supports the idea that a GlpG RNC, Sec

translocase interaction *in vivo* is maintained upon isolation in DDM or DIBMA.

**DIBMA Captures Membrane Protein RNCs within a Native Lipid Environment.** To ensure that the DIBMA polymer had solubilized GlpG RNCs within a native lipid environment, we assessed the lipid compositions using high-performance thin-layer chromatography (HPTLC). Purified RNC DIBMA nanodisc samples and cell membranes from non-induced cells, grown at the same temperature, were subject to Folch lipid extraction. The lipid identity and relative mole ratio of each observed lipid were then assessed using HPTLC analysis (Figure 4). We found that all DIBMA GlpG RNC nanodiscs contained similar relative abundances of PE, PG, and CL and compared to the non-induced membrane extract when grown in MDG/EM9 media (Figure 4).

A small increase in the abundance of PE and a decrease in the abundance of CL were seen between uninduced BL21 cells and the GlpG RNC DIBMA preparations (especially for the two-TM GlpG construct). Lipid membrane solubilization by DIBMA is relatively unexplored, but these effects may arise from collision lipid transfer<sup>63</sup> or preferential solubilization by DIBMA or reflect the two-TM local lipid environment as previously shown with SMA.<sup>64</sup> When RNCs were grown in LB medium, neutral lipids were also found in addition to the expected phospholipid composition (Figure S4), as confirmed by iodine vapor and molybdenum blue staining.<sup>45</sup> This has also previously been observed in DIBMA nanodiscs of full-length GlpG.<sup>38</sup>

Taken together, we observed that DIBMA is capable of extracting MP RNCs within a composition akin to its native lipid bilayer.

**An Enhanced SecM Arrest Sequence Dramatically Improves Membrane Protein RNC Stability.** An enhanced arrest sequence (AE1) (<sup>150</sup>FSTPVVWVWVPRIRAPP<sup>166</sup>)<sup>26</sup> was employed to improve the stability of RNCs and mitigate the release of the nascent chain during and after purification, thus enhancing RNC quality and longevity. Compared to WT SecM, the AE1 SecM sequence contains an additional proline at position 165 and a substituted stretch of tryptophan amino acids (Figure 5a), which forms more interactions with the ribosome exit tunnel and further occludes peptide bond formation in the PTC.<sup>4</sup>

Four-TM GlpG RNCs were prepared with the AE1 SecM stalling sequence in both DDM micelles and DIBMA nanodiscs, yielding samples of a quality higher than that achieved with WT SecM, as evidenced by negligible nascent chain release and degradation observed by Western blotting (Figure 5b). Improved nascent chain retention was also found for AE1 SecM RNCs, with only 12% spontaneous release of nascent chain from AE1 SecM RNCs observed over 24 h at 23 °C, compared to 35% for WT SecM, for DDM purified samples under the same conditions (IMAC and SEC purification) (Figure S5a). These percentages were estimated by using densitometry analysis at each time point and calculating the relative ratios of tRNA-bound four-TM GlpG and naturally released four-TM nascent chain. Moreover, four-TM GlpG RNCs containing WT SecM released 48% of their nascent chain upon only one freeze–thaw cycle, whereas AE1 SecM released only 18% (Figure S5c).



Four-TM GlpG AE1 SecM RNCs, prepared in DIBMA (by IMAC and sucrose gradient purification), were incredibly stable; no significant release was found after 24 h at both 23 and 37 °C (Figure 5c). Furthermore, these samples were able to withstand at least six freeze–thaw cycles with no significant release of the nascent chain (Figure S5b).

We also estimated the fraction of ribosomes in the samples that contain bound nascent chain to understand purified RNC occupancy. Occupancy was calculated by blotting the known  $A_{260}$ -derived concentration of each RNC sample (where  $1 A_{260} = 24 \text{ pmol/mL}$ ) against full-length GlpG standards (Figure S5c) and calculating the picomoles of GlpG NC for the tRNA-bound band.<sup>2</sup> There was an increase in occupancy of AE1 SecM RNCs in DDM (85% occupancy) compared to that of WT SecM RNC (75% occupancy). An occupancy of 75% was determined for AE1 SecM RNC in DIBMA; however, there is no significant release of this DIBMA nascent chain compared to that of DDM detergent samples (Figure S5c). This signifies that incorporation of the AE1 sequence leads to a significant increase in RNC stability, which will aid future structural, biochemical, and biophysical analysis.

## CONCLUSIONS

We demonstrate that pure, stable RNCs of different nascent chain lengths can be prepared for a polytopic  $\alpha$ -helical integral membrane protein, in both detergent micelle and native lipid surroundings. This work advances membrane protein RNC preparations by establishing that (1) SecM stalling can successfully stall a polytopic  $\alpha$ -helical MP at various points throughout its synthesis to yield stable RNCs for structural interrogation, thus allowing folding of the full chain to be probed as opposed to single, short TM constructs previously investigated, (2) RNC constructs are tolerant to different purification tags, such as His and Avi tags, enabling diverse purification and experimental strategies to be devised, (3) the nascent chain is maintained within a lipid bilayer comprised of its cellular lipid mixtures rather than synthetic lipids, and (4) RNCs are extracted directly from native membranes, escaping reconstitution of isolated RNCs prepared in non-native environments and permitting co-purification of endogenous chaperones, such as the SecYEG translocon.

Overall, we reveal that homogeneous membrane protein RNCs can be captured within a native lipid environment using DIBMA native nanodisc technology. These samples are stable for the time frames and conditions required for future biochemical and structural studies.<sup>65,66</sup> *In vivo*-formed membrane protein RNCs generated here provide snapshots of co-translational folding, paused at various nascent chain lengths, and are a stepping stone toward studying the structure and dynamics of co-translational membrane protein folding in an increasingly more *in vivo* context.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.biochem.0c00423>.

Discussion of *in vivo* biotinylation procedures and Figures S1–S5 (PDF)

### Accession Codes

UniProtKB codes: GLPG, P09391; SECY, P0AGA2; SECE, P0AG96; SECG, P0AG99. Small ribosomal subunits (30S): RS1, P0AG67; RS2, P0A7V0; RS3, P0A7V3; RS4, P0A7V8;

RS5, P0A7W1; RS6, P02358; RS7, P02359; RS8, P0A7W7; RS9, P0A7X3; RS10, P0A7R5; RS11, P0A7R9; RS12, P0A7S3; RS13, P0A7S9; RS14, P0AG59; RS15, P0ADZ4; RS16, P0A7T3; RS17, P0AG63; RS18, P0A7T7; RS19, P0A7U3; RS20, P0A7U7; RS21, P68681; RS22, C8U8F3. Large ribosomal subunits (50S): RL1, P0A7L0; RL2, P60422; RL3, P60438; RL4, P60723; RL5, P62399; RL6, P0AG55; RL7, P0A7K2; RL9, P0A7R1; RL10, P0A7J3; RL11, P0A7J7; RL13, P0AA10; RL14, P0ADY3; RL15, P02413; RL16, P0ADY7; RL17, P0AG44; RL18, P0C018; RL19, B1LPB3; RL20, P0A7L3; RL21, P0AG48; RL22, P61175; RL23, P0ADZ0; RL24, P60624; RL25, P68919; RL27, P0A7M0; RL28, P0A7M2; RL29, P0A7M6; RL30, P0AG51; RL31, P0A7M9; RL32, C4ZS29; RL33, P0A7N9; RL34, P0A7P6; RL35, P0A7Q2; RL36, P0A7Q7.

## AUTHOR INFORMATION

### Corresponding Authors

Paula J. Booth – King's College London, Department of Chemistry, London SE1 1DB, U.K.; Email: paula.booth@kcl.ac.uk

Eamonn Reading – King's College London, Department of Chemistry, London SE1 1DB, U.K.; [orcid.org/0000-0001-8219-0052](https://orcid.org/0000-0001-8219-0052); Email: eamonn.reading@kcl.ac.uk

### Authors

Grant A. Pellowe – King's College London, Department of Chemistry, London SE1 1DB, U.K.; [orcid.org/0000-0003-4314-5261](https://orcid.org/0000-0003-4314-5261)

Heather E. Findlay – King's College London, Department of Chemistry, London SE1 1DB, U.K.

Karen Lee – King's College London, Department of Chemistry, London SE1 1DB, U.K.

Tim M. Gemeinhardt – King's College London, Department of Chemistry, London SE1 1DB, U.K.

Laura R. Blackholly – King's College London, Department of Chemistry, London SE1 1DB, U.K.

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.biochem.0c00423>

### Author Contributions

G.A.P., E.R., and P.J.B. designed the research. G.A.P., H.E.F., K.L., T.M.G., and E.R. performed all experiments and analyses, except for SecYEG preparation, which was performed by L.R.B. E.R. and P.J.B. supervised the project. G.A.P., E.R., and P.J.B. wrote the manuscript with input from the other authors.

### Funding

This work was supported by a BBSRC Future Leader Fellowship (BB/N011201/1) and a UKRI Future Leaders Fellowship (MR/S015426/1) to E.R., with support by the Erasmus+ programme of the European Union to T.M.G., an ERC Advanced Grant (294342) and a Wellcome Trust Investigator Award (214259/Z/18/Z) to P.J.B., and a King's College London Ph.D. studentship to G.A.P.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank Prof. Ian Collinson at the University of Bristol for the generous donation of the pBAD-SecYEG plasmid and the monoclonal mouse SecY antibody for endogenous SecYEG detection. The authors also thank Prof.

Bernd Bukau and Dr. Guenter Kramer for providing plasmids containing SecM genes.

## ■ REFERENCES

- (1) Cabrita, L. D., Hsu, S. T., Launay, H., Dobson, C. M., and Christodoulou, J. (2009) Probing ribosome-nascent chain complexes produced in vivo by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* 106, 22239–22244.
- (2) Cassaignau, A. M., Launay, H. M., Karyadi, M. E., Wang, X., Waudby, C. A., Deckert, A., Robertson, A. L., Christodoulou, J., and Cabrita, L. D. (2016) A strategy for co-translational folding studies of ribosome-bound nascent chain complexes using NMR spectroscopy. *Nat. Protoc.* 11, 1492–1507.
- (3) Waudby, C. A., Launay, H., Cabrita, L. D., and Christodoulou, J. (2013) Protein folding on the ribosome studied using NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* 74, 57–75.
- (4) Bischoff, L., Wickles, S., Berninghausen, O., van der Sluis, E. O., and Beckmann, R. (2014) Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. *Nat. Commun.* 5, 4103.
- (5) Kater, L., Frieg, B., Berninghausen, O., Gohlke, H., Beckmann, R., and Kedrov, A. (2019) Partially inserted nascent chain unzips the lateral gate of the Sec translocon. *EMBO Rep.* 20, No. e48191.
- (6) Seidelt, B., Innis, C. A., Wilson, D. N., Gartmann, M., Armache, J. P., Villa, E., Trabuco, L. G., Becker, T., Mielke, T., Schulten, K., Steitz, T. A., and Beckmann, R. (2009) Structural insight into nascent polypeptide chain-mediated translational stalling. *Science* 326, 1412–1415.
- (7) Woolhead, C. A., McCormick, P. J., and Johnson, A. E. (2004) Nascent Membrane and Secretory Proteins Differ in FRET-Detected Folding Far inside the Ribosome and in Their Exposure to Ribosomal Proteins. *Cell* 116, 725–736.
- (8) Hoffmann, A., Merz, F., Rutkowska, A., Zachmann-Brand, B., Deuerling, E., and Bukau, B. (2006) Trigger factor forms a protective shield for nascent polypeptides at the ribosome. *J. Biol. Chem.* 281, 6539–6545.
- (9) Rutkowska, A., Beerbaum, M., Rajagopalan, N., Fiaux, J., Schmieder, P., Kramer, G., Oschkinat, H., and Bukau, B. (2009) Large-scale purification of ribosome-nascent chain complexes for biochemical and structural studies. *FEBS Lett.* 583, 2407–2413.
- (10) Schaffitzel, C., and Ban, N. (2007) Generation of ribosome nascent chain complexes for structural and functional studies. *J. Struct. Biol.* 158, 463–471.
- (11) Schulze, R. J., Komar, J., Botte, M., Allen, W. J., Whitehouse, S., Gold, V. A., Lycklama a Nijeholt, J. A., Huard, K., Berger, I., Schaffitzel, C., and Collinson, I. (2014) Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proc. Natl. Acad. Sci. U. S. A.* 111, 4844–4849.
- (12) Harris, N. J., Charalambous, K., Findlay, H. E., and Booth, P. J. (2018) Lipids modulate the insertion and folding of the nascent chains of alpha helical membrane proteins. *Biochem. Soc. Trans.* 46, 1355–1366.
- (13) Findlay, H. E., and Booth, P. J. (2017) The folding, stability and function of lactose permease differ in their dependence on bilayer lipid composition. *Sci. Rep.* 7, 13056.
- (14) Bogdanov, M., Dowhan, W., and Vitrac, H. (2014) Lipids and topological rules governing membrane protein assembly. *Biochim. Biophys. Acta, Mol. Cell Res.* 1843, 1475–1488.
- (15) Kuruma, Y., and Ueda, T. (2015) The PURE system for the cell-free synthesis of membrane proteins. *Nat. Protoc.* 10, 1328–1344.
- (16) Ismail, N., Hedman, R., Linden, M., and von Heijne, G. (2015) Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat. Struct. Mol. Biol.* 22, 145–149.
- (17) Ismail, N., Hedman, R., Schiller, N., and von Heijne, G. (2012) A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat. Struct. Mol. Biol.* 19, 1018–1022.
- (18) Dale, H., Angevine, C. M., and Krebs, M. P. (2000) Ordered membrane insertion of an archaeal opsin in vivo. *Proc. Natl. Acad. Sci. U. S. A.* 97, 7847–7852.
- (19) Dale, H., and Krebs, M. P. (1999) Membrane insertion kinetics of a protein domain in vivo. The bacterioopsin n terminus inserts co-translationally. *J. Biol. Chem.* 274, 22693–22698.
- (20) Pellowe, G. A., and Booth, P. J. (2020) Structural insight into co-translational membrane protein folding. *Biochim. Biophys. Acta, Biomembr.* 1862, 183019.
- (21) Harris, N. J., Reading, E., Ataka, K., Grzegorzewski, L., Charalambous, K., Liu, X., Schlesinger, R., Heberle, J., and Booth, P. J. (2017) Structure formation during translocon-unassisted co-translational membrane protein folding. *Sci. Rep.* 7, 8021.
- (22) Ataka, K., Stripp, S. T., and Heberle, J. (2013) Surface-enhanced infrared absorption spectroscopy (SEIRAS) to probe monolayers of membrane proteins. *Biochim. Biophys. Acta, Biomembr.* 1828, 2283–2293.
- (23) Evans, M. S., Ugrinov, K. G., Frese, M. A., and Clark, P. L. (2005) Homogeneous stalled ribosome nascent chain complexes produced in vivo or in vitro. *Nat. Methods* 2, 757–762.
- (24) Nakatogawa, H., and Ito, K. (2002) The Ribosomal Exit Tunnel Functions as a Discriminating Gate. *Cell* 108, 629–636.
- (25) Bhushan, S., Hoffmann, T., Seidelt, B., Frauenfeld, J., Mielke, T., Berninghausen, O., Wilson, D. N., and Beckmann, R. (2011) SecM-stalled ribosomes adopt an altered geometry at the peptidyl transferase center. *PLoS Biol.* 9, No. e1000581.
- (26) Cymer, F., Hedman, R., Ismail, N., and von Heijne, G. (2015) Exploration of the arrest peptide sequence space reveals arrest-enhanced variants. *J. Biol. Chem.* 290, 10208–10215.
- (27) Snider, C., Jayasinghe, S., Hristova, K., and White, S. H. (2009) MPEx: a tool for exploring membrane proteins. *Protein Sci.* 18, 2624–2628.
- (28) Marinko, J. T., Huang, H., Penn, W. D., Capra, J. A., Schleich, J. P., and Sanders, C. R. (2019) Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem. Rev.* 119, 5537–5606.
- (29) Goldman, D. H., Kaiser, C. M., Milin, A., Righini, M., Tinoco, I., Jr., and Bustamante, C. (2015) Ribosome. Mechanical force releases nascent chain-mediated ribosome arrest in vitro and in vivo. *Science* 348, 457–460.
- (30) Kaiser, C. M., Goldman, D. H., Chodera, J. D., Tinoco, I., Jr., and Bustamante, C. (2011) The ribosome modulates nascent protein folding. *Science* 334, 1723–1727.
- (31) Oluwole, A. O., Danielczak, B., Meister, A., Babalola, J. O., Vargas, C., and Keller, S. (2017) Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angew. Chem., Int. Ed.* 56, 1919–1924.
- (32) Oluwole, A. O., Klingler, J., Danielczak, B., Babalola, J. O., Vargas, C., Pabst, G., and Keller, S. (2017) Formation of Lipid-Bilayer Nanodiscs by Diisobutylene/Maleic Acid (DIBMA) Copolymer. *Langmuir* 33, 14378–14388.
- (33) Simon, K. S., Pollock, N. L., and Lee, S. C. (2018) Membrane protein nanoparticles: the shape of things to come. *Biochem. Soc. Trans.* 46, 1495–1504.
- (34) Nierhaus, K. H. (2014) Mg<sup>2+</sup>, K<sup>+</sup>, and the ribosome. *J. Bacteriol.* 196, 3817–3819.
- (35) Danielczak, B., Meister, A., and Keller, S. (2019) Influence of Mg<sup>2+</sup> and Ca<sup>2+</sup> on nanodisc formation by diisobutylene/maleic acid (DIBMA) copolymer. *Chem. Phys. Lipids* 221, 30–38.
- (36) Denisov, I. G., and Sligar, S. G. (2017) Nanodiscs in Membrane Biochemistry and Biophysics. *Chem. Rev.* 117, 4669–4713.
- (37) Ravula, T., Hardin, N. Z., and Ramamoorthy, A. (2019) Polymer nanodiscs: Advantages and limitations. *Chem. Phys. Lipids* 219, 45–49.
- (38) Barniol-Xicot, M., and Verhelst, S. H. L. (2018) Stable and Functional Rhomboid Proteases in Lipid Nanodiscs by Using Diisobutylene/Maleic Acid Copolymers. *J. Am. Chem. Soc.* 140, 14557–14561.



- (39) Gulamhussein, A. A., Uddin, R., Tighe, B. J., Poyner, D. R., and Rothnie, A. J. (2020) A comparison of SMA (styrene maleic acid) and DIBMA (di-isobutylene maleic acid) for membrane protein purification. *Biochim. Biophys. Acta, Biomembr.* 1862, 183281.
- (40) Becker, M., Gzyl, K. E., Altamirano, A. M., Vuong, A., Urbahn, K., and Wieden, H. J. (2012) The 70S ribosome modulates the ATPase activity of Escherichia coli YchF. *RNA Biol.* 9, 1288–1301.
- (41) Hill, W. E., Rossetti, G. P., and Van Holde, K. E. (1969) Physical studies of ribosomes from Escherichia coli. *J. Mol. Biol.* 44, 263–277.
- (42) Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
- (43) Folch, J., Lees, M., and Sloane Stanley, G. H. (1957) A simple method for the isolation and purification of total lipides from animal tissues. *J. Biol. Chem.* 226, 497–509.
- (44) Churchward, M. A., Brandman, D. M., Rogasevskaja, T., and Coorsen, J. R. (2008) Copper (II) sulfate charring for high sensitivity on-plate fluorescent detection of lipids and sterols: quantitative analyses of the composition of functional secretory vesicles. *J. Chem. Biol.* 1, 79–87.
- (45) Dynska-Kukulka, K., Ciesielski, W., and Zakrzewski, R. (2013) The use of a new, modified Dittmer-Lester spray reagent for phospholipid determination by the TLC image analysis technique. *Biomol. Chromatogr.* 27, 458–465.
- (46) Reading, E., Hall, Z., Martens, C., Haghighi, T., Findlay, H., Ahdash, Z., Politis, A., and Booth, P. J. (2017) Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions. *Angew. Chem., Int. Ed.* 56, 15654–15657.
- (47) Brooks, C. L., and Lemieux, M. J. (2013) Untangling structure-function relationships in the rhomboid family of intramembrane proteases. *Biochim. Biophys. Acta, Biomembr.* 1828, 2862–2872.
- (48) Sherratt, A. R., Blais, D. R., Ghasiani, H., Pezacki, J. P., and Goto, N. K. (2012) Activity-based protein profiling of the Escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry* 51, 7794–7803.
- (49) Vinothkumar, K. R. (2011) Structure of rhomboid protease in a lipid environment. *J. Mol. Biol.* 407, 232–247.
- (50) Paslawski, W., Lillelund, O. K., Kristensen, J. V., Schafer, N. P., Baker, R. P., Urban, S., and Otzen, D. E. (2015) Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7978–7983.
- (51) Baker, R. P., and Urban, S. (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* 8, 759–768.
- (52) Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. (2016) Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* 12, 353–360.
- (53) Min, D., Jefferson, R. E., Bowie, J. U., and Yoon, T. Y. (2015) Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* 11, 981–987.
- (54) Jha, S., and Komar, A. A. (2011) Birth, life and death of nascent polypeptide chains. *Biotechnol. J.* 6, 623–640.
- (55) Seddon, A. M., Curnow, P., and Booth, P. J. (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta, Biomembr.* 1666, 105–117.
- (56) Dzionara, M., Kaltschmidt, E., and Wittmann, H. G. (1970) Ribosomal proteins. 8. Molecular weights of isolated ribosomal proteins of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 67, 1909–1913.
- (57) Bercovich-Kinori, A., and Bibi, E. (2015) Co-translational membrane association of the Escherichia coli SRP receptor. *J. Cell Sci.* 128, 1444–1452.
- (58) Schlegel, S., Rujas, E., Ytterberg, A. J., Zubarev, R. A., Luirink, J., and de Gier, J. W. (2013) Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels. *Microb. Cell Fact.* 12, 24.
- (59) Schibich, D., Gloge, F., Pohner, I., Bjorkholm, P., Wade, R. C., von Heijne, G., Bukau, B., and Kramer, G. (2016) Global profiling of SRP interaction with nascent polypeptides. *Nature* 536, 219–223.
- (60) Komar, J., Alvira, S., Schulze, R. J., Martin, R., Lycklama a Nijeholt, J. A., Lee, S. C., Dafforn, T. R., Deckers-Hebestreit, G., Berger, I., Schaffitzel, C., and Collinson, I. (2016) Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Biochem. J.* 473, 3341–3354.
- (61) Ejima, D., Yumioka, R., Arakawa, T., and Tsumoto, K. (2005) Arginine as an effective additive in gel permeation chromatography. *J. Chromatogr. A* 1094, 49–55.
- (62) Whitehouse, S., Gold, V. A., Robson, A., Allen, W. J., Sessions, R. B., and Collinson, I. (2012) Mobility of the SecA 2-helix-finger is not essential for polypeptide translocation via the SecYEG complex. *J. Cell Biol.* 199, 919–929.
- (63) Danielczak, B., and Keller, S. (2018) Collisional lipid exchange among DIBMA-encapsulated nanodiscs (DIBMALPs). *Eur. Polym. J.* 109, 206–213.
- (64) Teo, A. C. K., Lee, S. C., Pollock, N. L., Stroud, Z., Hall, S., Thakker, A., Pitt, A. R., Dafforn, T. R., Spickett, C. M., and Roper, D. I. (2019) Analysis of SMALP co-extracted phospholipids shows distinct membrane environments for three classes of bacterial membrane protein. *Sci. Rep.* 9, 1813.
- (65) Parmar, M., Rawson, S., Scarff, C. A., Goldman, A., Dafforn, T. R., Muench, S. P., and Postis, V. L. G. (2018) Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochim. Biophys. Acta, Biomembr.* 1860, 378–383.
- (66) Reading, E. (2019) Assessing Membrane Protein Structural Dynamics within Lipid Nanodiscs. *Trends Biochem. Sci.* 44, 989–990.

