

Manipulating sugar utilization in *Escherichia coli* through synthetic fusions to membrane transporters

Liam Neil Mike Chapman

PhD thesis

Biology

University of York

March 2021

Abstract

Substrate channelling describes the movement of intermediate between active sites without entering equilibrium with the bulk cell material. The fusion of membrane transport proteins to downstream enzymes may enhance substrate channelling of naturally occurring or synthetic pathways in microbes. This thesis describes the production of a system suitable for assembling libraries of protein fusions, with a focus on enhancing D-xylose metabolism, an industrially relevant metabolite.

The cloning technique Golden Gate assembly was selected to produce libraries of chimeric proteins, capable of producing fusions within a one-pot reaction. Low copy number plasmids suitable for expression of membrane proteins were developed and tested for this work, demonstrating high success rate and easy application. The D-xylose transporter XylE, alongside other MFS homologues, were then fused to GFP to confirm stability towards C-terminal fusion.

As XylE was a suitable target for fusion, attention turned towards the *E. coli* D-xylose isomerase, XylA. The crystal structure of XylA was obtained to understand oligomerization and organisation of the protein – finding that it formed a tetramer, which was distinct from early research suggesting a dimeric structure. However, direct fusion of the XylE C-terminal to the XylA N-terminal caused a loss of function for the D-xylose isomerase, possibly because of its oligomeric state or the location of the N-terminal with respect to the active site. Scaffolds are an alternative to protein fusion to enhance substrate channelling. The cohesin-dockerin scaffold system was therefore applied to reduce the steric hinderance seen in direct fusion – however while cohesin tagging increased XylA functionality, there was no evidence of scaffold formation, and therefore no substrate channelling occurred.

Overall, while this study did not demonstrate improved substrate channelling in the D-xylose utilization pathway, the systems produced are suitable for generating libraries of protein fusions and scaffolds, that, with time and troubleshooting, should facilitate substrate channelling.

Table of Contents

Abstract.....	2
List of Tables	9
List of Figures	11
Acknowledgements.....	17
Author's Declaration.....	19
List of Abbreviations.....	20
1. Introduction.....	24
1.1. Multi-enzyme reaction pathways - bringing reactions together	24
1.2. Protein Fusion.....	32
1.2.1. Protein fusions in nature	32
1.2.2. Engineering synthetic protein fusions.....	35
1.2.3. Design of synthetic protein fusion	38
1.3. Enzyme Scaffolds	43
1.3.1. Protein scaffolds	43
1.3.2. Nucleic acid scaffolds	48
1.3.3. The dynamics behind substrate channelling in scaffolds.....	51
1.4. Artificial Substrate Channelling Approaches.....	53
1.5. The Importance of Xylose Metabolism for Industrial Biotechnology.....	54
1.5.1 Xylose/H ⁺ symporter	60
1.5.2 Xylose isomerase	64
1.5.3 Xylulose kinase.....	65
1.6. Golden Gate Assembly as a Means for Producing Protein Fusion Libraries.....	66
1.7. Aims.....	70
2. Methods and Materials.....	72
2.1. Microbiology	72
2.1.1. Lysogeny broth and agar.....	72
2.1.2. Supplemented M9 minimal media	72
2.1.3. M9+X minimal media.....	72

2.1.4. M9+G minimal media	72
2.1.5. M9+S minimal media.....	72
2.1.6. Antibiotics.....	72
2.1.7. Growth assays	74
2.1.8. Calculating maximum growth rate	74
2.1.9. Producing Chemically Competent <i>E. coli</i>	74
2.1.10. Measuring the concentration of D-xylose in cell culture	75
2.2. Molecular biology.....	78
2.2.1. Strains and plasmids	78
2.2.2. Curing a Keio collection strain of the Kanamycin resistance cassette.....	78
2.2.3 Agarose gel electrophoresis.....	78
2.2.3. DNA preparation and extraction	80
2.2.4. Construction of Golden Gate expression plasmids	80
2.2.5. Gene amplification by Polymerase Chain Reaction	81
2.2.6. Site directed mutagenesis	81
2.2.7. Blunt End Mutagenesis	82
2.2.8. Duplex assembly from primers.....	82
2.2.9. Heat-shock transformation	82
2.2.10. Golden Gate Assembly	83
2.2.11. Designing of gBlocks.....	83
2.3. Microscopy	84
2.3.1 Sample preparation for Fluorescence Microscopy.....	84
2.3.2. Epifluorescence Microscopy.....	84
2.3.3 Total Internal Reflection Fluorescence Microscopy (TIRFM)	84
2.3.4 Processing of microscopy data.....	85
2.3.5 Median filtering of TIRFM data	85
2.3.6. Single particle tracking of TIRFM data.....	85
2.4. Recombinant protein production and purification.	87
2.4.1. Separating soluble and insoluble fractions	87

2.4.2. Nickel affinity chromatography	87
2.4.3. Size Exclusion Chromatography (SEC).....	87
2.4.4. Size Exclusion Chromatography – Multi-Angle Laser Light Scattering (SEC-MALLS).....	88
2.4.5. Protein crystallization	88
2.4.6. Sodium dodecyl sulphate polyacrylamide gel electrophoresis	89
2.4.7. Western blot.....	89
2.4.8. Coomassie blue staining	90
2.5. Bioinformatics	92
2.5.1. Determining Major Facilitator Superfamily protein fusions in nature.	92
2.5.2. Multiple sequence alignment.....	92
2.5.3. Maximum Likelihood Phylogenetic tree	92
3. Designing a Golden Gate assembly protocol to produce a library of membrane protein fusions to soluble proteins.....	119
3.1 Production of a Golden Gate expression vector suitable for membrane protein expression.....	119
3.2 Testing the function of Golden Gate cloning in novel expression vector pLC1	124
3.3 Naturally occurring Major Facilitator Superfamily protein fusions demonstrates a need for protein linkers.....	127
3.3.1 The MFS-PlsC fusions demonstrate significant sequence similarity in the linker region.....	129
3.3.2 The MFS-CBS protein fusions demonstrate greater variability in linker architecture in less abundant fusions.....	146
3.3.3. Summary of natural linkers.....	156
3.4 Designing peptide linkers suitable for Major Facilitator Superfamily protein fusion...	157
3.5 Using Golden Gate assembly to produce genetic fusions of XylE to a model GFP protein	161
3.6 Confirming function of the XylE-GFP fusion.....	164
3.6.1 Fluorescence analysis of C-terminal XylE-GFP demonstrates successful expression of fusion construct.....	164

3.6.2 Anti-GFP Western Blot further demonstrates production of the complete XylE-GFP fusion	167
3.6.3 Growth assay	170
3.7 Golden Gate cloning to produce additional MFS-GFP fusions	172
3.7.1 Producing a fusion of ExuT and GFP	172
3.7.2 Producing a fusion of NanT to GFP	175
3.8 Summary of chapter	178
4. The direct fusion of XylE to XylA introduces constraints to enzyme function.	180
4.1 Validating the oligomeric state of <i>E. coli</i> Xylose isomerase.....	180
4.1.1. Size Exclusion Chromatography – Multiple Angle Light Laser Scattering (SEC-MALLS) of <i>E. coli</i> Xylose Isomerase.	185
4.1.2. Crystallization of <i>E. coli</i> xylose isomerase.....	189
4.2. Producing fusions between <i>E. coli</i> Xylose/H ⁺ transporter and <i>E. coli</i> Xylose isomerase.	199
4.3. Fusion of XylE to XylA results in a decreased lag in cell growth over wild type XylE207	
4.4. Xylose isomerase activity is negatively impacted by its fusion to XylE.	211
4.5. Summary of Chapter	216
5. Assessment of the use of Scaffold components to overcome the enzyme activity issues associated with direct fusion.	219
5.1. Designing a scaffold system for localization of xylose isomerase at the XylE transporter.	219
5.2. Curing the Kanamycin resistance cassette of Keio collection BW25113 $\Delta xyIA::kan$ 228	
5.3. Anti-tetraHis western blots demonstrate expression of fused XylE-ctDoc and XylA-ctCoh/ctCoh-XylA fusions.....	234
5.3.1. The XylE-ctDoc shows comparable, if not greater, expression levels to XylE-XylA	234
5.3.2. High expression levels of the XylA-ctCoh and ctCoh-XylA fusions.	237
5.4. Determining the functionality of the scaffolds via a pull-down assay	243
5.4.1 Designing an <i>ex vivo</i> pull-down assay to confirm scaffold formation.....	243
5.4.2. The cohesin-dockerin scaffold confirmed to assemble in the pull-down assay .	249
5.4.3. Concluding remarks	258

5.5. The use of growth assays to determine the impact of scaffolds on D-xylose utilization	259
5.5.1. Fusion of cohesin to the N-terminal of XylA results in a loss of activity compared to C-terminal fusion.....	260
5.5.2. The transport of xylose across the membrane can be partially enhanced by coexpression of XylE-ctDoc and XylA-ctCoh.....	264
5.5.3. Coexpression of XylE-ctDoc and XylA-ctCoh fusions in $\Delta xylA$ results in an inverse of the effects seen for expression of XylA-ctCoh fusions alone.	275
5.6. Direct measure of D-xylose concentration in initial 13 hours of cell growth demonstrates no benefit with cohesin-dockerin system.	281
5.7. Total Internal Reflection Fluorescence Microscopy (TIRFM) to determine the localization of Dockerin tagged mCherry.	286
5.7.1. Designing TIRFM appropriate protein fusions.	286
5.7.2. Particle tracking of scaffold components did not demonstrate mobility that was comparable to membrane transporters.	295
5.8. Summary of Chapter	304
6. Discussion - Evaluating the role of scaffolds and protein fusions in substrate channelling.	306
6.1. Substrate channelling – a multifaceted challenge	306
6.2. Evaluating Golden Gate assembly as a tool for protein fusion libraries.....	310
6.3. Fusion – a question of sterics	314
6.4. Scaffolds as an alternative to direct protein fusion.....	321
6.5. To scaffold or fuse – what is the best choice for substrate channelling?	328
6.6. Concluding remarks.....	336
Appendices.....	337
Appendix A – plasmid maps of novel plasmids.....	337
Appendix B - Synthetic gene sequences.	340
Appendix C - Golden Gate assembly Code	341
Appendix D – Growth rates of D-xylose growth assays	344
Appendix E – SEC-MALLS data with time in X-axis.....	355
Appendix F – Data collection and refinement table of <i>E. coli</i> XylA.....	356

Appendix G – SDS-PAGE and western blots of a test of the coexpression pulldown assay.	357
References	359

List of Tables

Table 1.1. List of products of D-xylose fermentation.....	55
Table 2.1. Concentrations of antibiotics used in this work	73
Table 2.2. Reaction conditions of D-xylose assay kit	76
Table 2.3. List of <i>E. coli</i> strains used in this work	79
Table 2.4. List of antibodies used in this work.	91
Table 2.5. List of reagents used in this work.	93-95
Table 2.6. List of primers used in this work.	96-104
Table 2.7. List of plasmids used in this work.	105-117
Table 3.1. Components of the pLC series Golden Gate expression plasmids.	123
Table 3.2. Linker length and amino acid sequence of a random selection of natural MFS fusions within MFS-fusion group 2, belonging to one of the seven phyla with most recognised members.	130-132
Table 3.3 Linker length and amino acid sequence of a random selection of natural MFS-CBS(-CBS) fusions	147
Table 3.4 Amino acid sequence of a selection of linkers of natural MFS-CBS(-CBS) fusions.	149-151
Table 3.5. Synthetic protein linkers generated for Golden Gate assembly of Membrane Fusion proteins	158
Table 3.6. The components of XylE-GFP fusion constructs.	163
Table 3.7. The components of ExuT-GFP fusion constructs.	173
Table 3.8. The components of NanT-GFP fusion constructs.	176
Table 4.1. Oligomeric state of xylose isomerase according to online databases BRENDA and UniProt.....	181-182
Table 4.2. The components of XylE-XylA fusion constructs.	200
Table 5.1. Components of the pLD series Golden Gate expression plasmids.	220

Table 5.2. The components of XylE-ctDoc fusion constructs.	223
Table 5.3. The components of XylA-ctCoc and ctCoh-XylA fusion constructs.	224
Table 5.4. The components of the novel fusions produced for the pull-down assay.	244
Table 5.5. The diffusion coefficient of inner membrane proteins and cytoplasmic proteins.	292
Table 5.6. The components of novel fusions developed for TIRFM.	296
Table 6.1. Comparison of direct fusion and protein scaffolding for substrate channelling.	334-335

List of Figures

Figure 1.1. Schematics of representatives of different mechanisms of substrate channelling in prokaryotic cells.	27
Figure 1.2. Schematic demonstrating different methods of metabolite channelling in multidomain proteins or multienzyme complexes.	29
Figure 1.3. Schematics of naturally occurring multidomain protein fusions.	33
Figure 1.4. The three types of synthetic fusions.	36
Figure 1.5. The crystal structure of the resveratrol producing 4CL::STS fusion.	39
Figure 1.6. Synthetic linkers used within protein fusions.	41
Figure 1.7. Schematic of three different synthetic protein scaffolds produced in literature	45
Figure 1.8. Schematics of the nucleic acid scaffold assemblies.	49
Figure 1.9. Microbial methods of xylose metabolism from hemicellulose.	57
Figure 1.10. Schematic of the rocker switch motif of XylE.	61
Figure 1.11. Schematic diagrams of elements of Golden Gate Cloning.	67
Figure 1.12. Schematic diagram of the CIDAR MoClo assembly format including the fusion site overhangs.	69
Figure 3.1. The components of pLC1, a very low copy number Golden Gate expression vector.	120
Figure 3.2. Electrophoresis gels of PCR amplicons of desired regions of the plasmids pWKS30, pSB1C3 and DVK_AF.	122
Figure 3.3. Electrophoresis gel showing the enzyme digest pattern of pLC1 and pLC1_E1010m.	125
Figure 3.4. Confocal and epifluorescence microscopy images of transformed E. coli strain DH5 α	126
Figure 3.5. Schematic diagram of the eight most common groups of MFS fusion proteins in bacteria.	128

Figure 3.6. Multiple sequence alignment of the MFS-PlsC fusions focused on the C-terminal region of the MFS, whole linker sequence and N-terminal region of PlsC domain.....	134
Figure 3.7. Multiple sequence alignment of the C-terminal region of the MFS, linkers and N-terminal region of PlsC domain from group 2 Major facilitator Superfamily fusions.....	136
Figure 3.8. Maximum likelihood phylogenetic tree of MFS-PlsC fusions.....	139
Figure 3.9. Sequence of the <i>E. coli</i> genome focused on the 3- region of the <i>aas</i> gene and 5' region of the <i>lpIT</i> gene.	142
Figure 3.10. Hypothetical evolution pathway of the MFS-PlsC fusion architecture.	143
Figure 3.11. Maximum likelihood phylogenetic tree of a broad range of LpIT MFS homologues.	145
Figure 3.12. Multiple sequence alignment of the MFS-CBS fusions.	152
Figure 3.13. Multiple sequence alignment of the C-terminal region of the MFS, linkers and N-terminal region of CBS domain from Major facilitator Superfamily fusions with unfused homologues.	154
Figure 3.14. Electrophoresis gel of pLC1 fusion constructs following enzyme digest with XbaI.	162
Figure 3.15. Confocal and epifluorescence microscopy of <i>E. coli</i> expressing GFP and fusion proteins.....	165
Figure 3.16. Analysis of <i>E. coli</i> strain DH5 α cell lysate.	168
Figure 3.17. Recovery of TDXYlose growth on a low concentration xylose growth medium by expressing XylE-GFP fusions.	171
Figure 3.18. Growth of BW25113 wild-type and Δ <i>exuT</i> strains growth on a low concentration glucuronic acid.	174
Figure 3.19. Recovery of BW25113 Δ <i>nanT</i> growth on a low concentration sialic acid background.	177
Figure 4.1. The crystal structure of <i>Thermoanaerobacterium thermosulfurigenes</i> xylose isomerase	184
Figure 4.2. The elution profile of <i>E. coli</i> xylose isomerase during SEC-MALLS.	186

Figure 4.3. The molar mass of protein eluted over the duration of the SEC-MALLS procedure.	188
Figure 4.4. The solved structure of <i>E. coli</i> XylA suggests a tetrameric state.....	190
Figure 4.5. The missing loop of the <i>E. coli</i> XylA monomer	192
Figure 4.6. <i>E. coli</i> Xylose Isomerase shows strong homology with other class II Xylose Isomerase structures deposited on the PDB database, but does not share the two glycine domains predicted to be responsible for the PbXI flexible loop.....	193
Figure 4.7. The active site of <i>E. coli</i> XylA sits within the opening of the TIM-barrel like structure – with active sites of each monomer appearing to act independent of each other	196
Figure 4.8. The X-Ray Fluorescence Spectrum of <i>E. coli</i> Xylose Isomerase.	197
Figure 4.9. In absence of induction a XylE-XylA fusion is produced by transformed TDXYlose cells, however appears at lower levels than soluble proteins.....	201
Figure 4.10. Under induction conditions only XylA is detected in the total cell lysate, with a different ladder confirming correct protein mass.....	203
Figure 4.11. Variable amounts of the XylE-XylA fusion are detected by western blot depending on the fusion linker, with all samples showing high degrees of linker cleavage.	204
Figure 4.12. The polyproline, short Arai and poly(GlySer) linkers show strong detection by a western blot, however the rigid linkers appear more prone to linker cleavage..	206
Figure 4.13. Recovery of TDXYlose growth on a low concentration xylose growth medium by expressing XylE-XylA fusions.....	208
Figure 4.14. Maximum growth rate of TDXYlose growth on a low concentration xylose growth medium by expressing XylE-XylA fusions	210
Figure 4.15. Recovery of BW25113 $\Delta xyIA$ growth on a low concentration xylose growth medium by expressing XylE-XylA fusions.	212
Figure 4.16. Maximum growth rate of BW25113 strain $\Delta xyIA::Kan$ growth on a low concentration xylose growth medium expressing XylE-XylA fusions.	213
Figure 4.17. A cartoon representation of potential XylE-XylA fusion conformations.....	217
Figure 5.1. The components of pLD1, a low copy number Golden Gate expression vector compatible with pLC1.....	221

Figure 5.2. Cartoon representing the planned scaffold organisation.....	227
Figure 5.3. Spread plates of candidate xylA deletion mutants $\Delta xylA1$ and $\Delta xylA7$	229
Figure 5.4. Electrophoresis gels of colony PCR of candidate cured Keio collection strains $\Delta xylA1$ and $\Delta xylA7$	231
Figure 5.5. Growth of candidate $\Delta xylA$ cured strains on LB growth media over 24 hours	232
Figure 5.6. Anti-tetraHis western blot of the XylE-ctDoc fusions.....	235
Figure 5.7. Anti-tetraHis western blot of the XylA-ctCoh and ctCoh-XylA fusions.	238
Figure 5.8. Anti-tetraHis western blot of the XylA-ctCoh and ctCoh-XylA fusions with All Blue ladder.....	239
Figure 5.9. Anti-tetraHis western blot of XylA-ctCoh fusion with Super Signal ladder.	242
Figure 5.10. Cartoon representing the pulldown assay conditions.	245
Figure 5.11. SDS-PAGE and Western Blot of cells expressing mCherry-ctCoh following nickel affinity chromatography.....	247
Figure 5.12. Products of Nickel affinity chromatography of the ctCoh-mCherry expressed either in isolation or coexpressed with GFP-ctDoc-His.....	250
Figure 5.13. Anti-mCherry western blot of supernatant purified by nickel affinity chromatography.....	252
Figure 5.14. Products of Nickel affinity chromatography of cohesin-mCherry fusions expressed either in presence of GFP-XylA-His or GFP-ctDoc-His.	254
Figure 5.15. Anti-mCherry western blot of supernatant purified by nickel affinity chromatography.....	257
Figure 5.16. Fusion of cohesin to the XylA N-terminal reduces xylose isomerase activity	261
Figure 5.17. Fusion of cohesin to the C-terminal of XylA does not impact activity.	263
Figure 5.18. The impact of coexpression on a growth assay of TDXYlose cells on a low Xylose background.....	265

Figure 5.19. The impact of linkers in XylE-ctDoc on a growth assay of TDXYlose cells on a low Xylose background when cotransformed alongside XylA-waldo-ctCoh fusions.	268
Figure 5.20. The impact of linkers in XylA-ctCoh on recovery of growth of TDXYlose cells on a low Xylose background when cotransformed with both XylE-flex-ctDoc fusions.	271
Figure 5.21. Evaluating the impact of XylA-ctCoh linker variation on scaffold formation ..	275
Figure 5.22. The XylA-ctCoh/XylE-ctDoc coexpression shows the best cell growth amongst other scaffold and fusion conditions.	278
Figure 5.23. Concentration of D-xylose in cell cultures of $\Delta xylA$ co-expressing XylE-ctDoc and XylA-ctCoh.	283
Figure 5.24. The percentage of available D-xylose consumed by scaffold conditions and their controls.	284
Figure 5.25. Cartoon representing the different fusions used in the initial TIRFM study of protein mobility.	287
Figure 5.26. Box and whisker plot of Diffusion coefficients of each individual trajectory for the GFP fusions.	289
Figure 5.27. Histogram representing the distribution of particles across different diffusion coefficients for GFP fusions expressed in DH5 α	290
Figure 5.28. Epifluorescence microscopy of the BW25113 cells expressing XylE-mCherry fusions.	297
Figure 5.29. Anti-mCherry western blot of the mCherry and PAmCherry fusions.	298
Figure 5.30. Cartoon representing the different fusions used to study scaffold mobility via TIRFM.	299
Figure 5.31. A box and whisker plot of diffusion coefficient of the conditions used to study scaffold protein mobility.	302
Figure 5.32. Histogram representing the distribution of particles across different diffusion coefficients for mCherry fusions in BW25113.	303
Figure 6.1. A cartoon representing changes to local enzyme concentration in presence of scaffold/fusion.	307
Figure 6.2. Domain architecture of the ABC transporters.	309

Figure 6.3. High throughput Golden Gate assembly of 2 gene constructs composed of a transporter gene fused to downstream enzymes.....	312
Figure 6.4. A fusion between membrane bound CrtW and CrtZ to increase Astaxanthin production.....	315
Figure 6.5. Natural fusion from <i>Salinispora tropica</i> between an N-terminal aquaporin-derived arsenite channel and a C-terminal arsenate reductase domain.....	318
Figure 6.6. A cartoon representing the different scaffold assemblies seen both within this work and in literature.	323
Figure 6.7. Different scaffold compositions to achieve altered local enzyme stoichiometry.	330
Figure 6.8. Schematic of oligomeric proteins in a scaffold.....	332

Acknowledgements

Completing this thesis would not have been possible without the help and supervision I have received from others.

First and foremost, I would like to thank my supervisors Professor Gavin H Thomas and Dr Christoph Baumann for their support and guidance throughout my PhD. Gavin's guidance has been essential for getting this work completed, offering ideas and suggestions of where to take the project, and through his supervision he has fostered a strong and supportive lab group that has made my PhD the most comfortable time in my adult life. Christoph and his lab group have assisted me with planning and execution of microscopy techniques and has given me his time to help me understand the techniques involved. In addition, I would like to thank my original co-supervisor Professor Jennifer R. Potts for her supervision during the planning stages of this work and her suggestions throughout the first three years of my PhD.

Furthermore, I would also like to thank my Thesis Advisory Panel: Gideon Grogan and Jamie Blaza, who have provided me advice and ask tough questions in TAP meetings to structure my work and spot shortcomings. In addition, Christoph Baumann was originally part of this TAP panel, for which I would like to thank him once more.

This work was funded through the UKRI-sponsored White Rose DTP program, and this work would not be possible without their funding. Furthermore, I would like to thank UKRI for extending my funding to account for lost time due to the coronavirus pandemic. I want to thank the University of York for being such a great place to complete my research for the past four and a half years. The biology department at the University of York was an excellent place to complete research as it facilitates strong communication between biochemists, structural biologists, microbiologists and biophysicists which has been essential for completing this work – in addition, they have also sponsored events such as Gradshare, that have really helped during rough weeks.

I would also like to thank the Thomas lab as a whole for assisting both my personal and professional development. The lab was vital in helping me learn new techniques, talking me through difficulties so that I could identify shortcomings and for creating an environment that felt welcoming. I would like to thank Ivan Gyulev and Sophie Rugg for their guidance in the initial stages of my PhD and familiarising me within the lab, Michelle Rudden for assisting me with phylogenetic work, Matt "Mad-dog" Rose for taking over the hosting duties from me for Gradshare in my final year, Reyme Herman for all his work on the XylA crystal and responding to questions about the AKTA that I always seemed to have, and Ben Willson for providing his feedback for my work and working with me on the MFS fusion papers. More than all that,

however, I also would like to thank them for putting up with me for four and a half years and being such good friends throughout. The emotional support was invaluable.

The Baumann group have also assisted me greatly in fluorescent microscopy work for which I greatly appreciate. I would like to thank Joe Nabarro for taking time out of his epifluorescence microscopy booking slots to check my mCherry and PAmCherry samples, Rosie Leaman for talking me through particle tracking methods and median filtering, and Alex Scott for producing the particle tracking code that saved me an immense amount of time and stress by automating the procedure.

Finally, I would like to thank the network of support I have received from friends and family. I would like to thank Lydia, Ryan, Tom, Daniel, Martyn, Emma and Emilia for their friendship – it really helped during the hardest parts. I would like to thank my family for their belief in me and the time spent together. And finally, I would like to thank my parents, Janice and Robert, for their love and support throughout my PhD.

Thank you all so much.

Author's Declaration

I declare that this thesis, except where otherwise stated, is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Work presented in chapter 3 about MFS fusions in nature contains extracts from work previously published as:

Willson, B. J., Dalzell, L., Chapman, L. N. M., & Thomas, G. H. (2019). Enhanced functionalisation of major facilitator superfamily transporters via fusion of C-terminal protein domains is both extensive and varied in bacteria. *Microbiology*, *165*(4), 419–424. <https://doi.org/10.1099/mic.0.000771>

Willson, B. J., Chapman, L. N., & Thomas, G. H. (2019). Evolutionary dynamics of membrane transporters and channels: enhancing function through fusion. *Current Opinion in Genetics & Development*, 58–59, 76–86. <https://doi.org/10.1016/j.gde.2019.07.017>

List of Abbreviations

<i>Abbreviation</i>	<i>Meaning</i>
<i>ABC</i>	ATP-binding cassette
<i>ACS</i>	AMP-dependant Synthetase/Ligase
<i>ADH</i>	Alcohol Dehydrogenase
<i>ALDH</i>	Aldehyde Dehydrogenase
<i>ATP</i>	Adenosine Triphosphate
<i>AMP</i>	Adenosine Monophosphate
<i>BLAST</i>	Basic Local Alignment Search Tool
<i>bp</i>	Base Pair
<i>BPL</i>	Biotin Protein Ligases
<i>BRENDA</i>	BRaunschweig ENzyme DAtabase
<i>BSA</i>	Bovine Serum Albumin
<i>cAMP</i>	cyclic AMP
<i>CIDAR</i>	Cross-disciplinary Integration of Design Automation Research lab
<i>CpXI</i>	<i>Clostridium phytofermentans</i> Xylose Isomerase
<i>CRP</i>	cAMP Receptor Protein
<i>DNA</i>	Deoxyribonucleic acid
<i>FDH</i>	Formate Dehydrogenase
<i>GABA</i>	Gamma-Aminobutyric Acid
<i>GAPDH</i>	D-Glyceraldehyde-3-Phosphate Dehydrogenase (
<i>GBD</i>	GTPase binding domain (GBD)
<i>IPTG</i>	Isopropyl β -D-1-thiogalactopyranoside
<i>kb</i>	Kilobase

<i>kDa</i>	Kilodalton
<i>KSR</i>	Kinase Suppressor of Ras
<i>LB</i>	Lysogeny Broth
<i>LDH</i>	L-Lactate-Dehydrogenase
<i>MAP</i>	Mitogen-Activated Protein
<i>MFS</i>	Major Facilitatory Superfamily
<i>MoClo</i>	Modular Cloning
<i>MSD</i>	Mean Squared Displacement
<i>NRPS</i>	Non-Ribosomal Peptide Synthetase
<i>OD</i>	Optical Density
<i>PAGE</i>	Polyacrylamide Gel Electrophoresis
<i>PAmCherry</i>	Photo-activatable mCherry
<i>PbXI</i>	<i>Paenibacillus</i> sp. Xylose Isomerase
<i>PCR</i>	Polymerase Chain Reaction
<i>PDB</i>	Protein Data Bank
<i>PDZ</i>	PSD95/DlgA/Zo-1
<i>PHB</i>	Polyhydroxybutyrate
<i>PKS</i>	Polyketide Synthase
<i>RMSD</i>	Root Mean Square Deviation
<i>RNA</i>	Ribonucleic acid
<i>RuBisCO</i>	Ribulose 1,5-Bisphosphate Carboxylase/Oxygenase
<i>SDS</i>	Sodium Dodecyl Sulphate
<i>SEC-MALLS</i>	Size Exclusion Chromatography - Multiple Angle Laser Light Scattering
<i>SH3</i>	Src homology 3

<i>ssDNA</i>	Single Stranded DNA
<i>ssRNA</i>	Single Stranded RNA
<i>TatXI</i>	<i>Thermoanaerobacterium thermosulfurigenes</i> Xylose Isomerase
<i>TIRFM</i>	Total Internal Reflection Fluorescence Microscopy
<i>TMH</i>	Transmembrane Helix
<i>WH1</i>	WASP-homology 1
<i>XDH</i>	Xylitol Dehydrogenase
<i>XI</i>	Xylose Isomerase
<i>XK</i>	Xylulose Kinase
<i>XMR</i>	Xylose Mutarotase
<i>XR</i>	Xylose Reductase

Chapter One

Introduction

1. Introduction

1.1. Multi-enzyme reaction pathways - bringing reactions together

Complex biological reaction pathways rely on multiple enzymes to convert readily available compounds to important products (de Kok *et al.*, 1998; Lass *et al.*, 2011). However, a major limitation of multi-enzyme reaction pathways is the loss of intermediates. Reaction intermediates may instead be used by a competing enzyme (Huang, Holden and Raushel, 2001), cause inhibition of pathway enzymes (Berry *et al.*, 2002), be highly labile (Sibbesen *et al.*, 1994) or cause toxicity within the cell (Kizer *et al.*, 2008; Dueber *et al.*, 2009), ultimately reducing the product yields available (M. E. Lee *et al.*, 2013). The problems associated with the loss of intermediates, and subsequent reduced product yield, can challenge synthesis of the most complex molecules (Wheeldon *et al.*, 2016).

Metabolic channelling, also sometimes termed substrate channelling, is an important process in the functioning of pathways featuring multiple enzymes. Metabolic channelling is defined as the process by which the product of one enzyme is used by another enzyme, without entering into equilibrium with the bulk solution (Spivey and Ovádi, 1999). By limiting the free diffusion of substrate and intermediates, it is possible to accelerate multi-enzyme reaction pathways (Miles, Rhee and Davies, 1999). Furthermore, this can increase yields without necessitating high protein expression, reducing metabolic load (Dueber *et al.*, 2009).

Within eukaryotic cells, high substrate channelling may be achieved by localizing reactions to specific organelles. Eukaryotic cells will use these organelles to increase local substrate and enzyme concentration, block competitive reactions, provide specific cofactor, and to separate the bulk cell media from a toxic intermediate (Hammer and Avalos, 2017; Huttanus and Feng, 2017). One such example of a sequestered reaction is fatty acid metabolism, which is completed within the peroxisome and is associated with the release of hydrogen peroxide. The organelle is designed to tolerate the reactive H_2O_2 , preventing it from escaping into the cytoplasm and reducing risk to the health of the cell. In addition, enzymes such as peroxidases can then use the produced H_2O_2 to oxidise target molecules (Gabaldón, Ginger and Michels, 2016; Gao and Zhou, 2019).

Prokaryotic cells generally lack classical organelles therefore must rely on other methods to organise reactions and achieve metabolic channelling. In some prokaryotic cells large complexes akin to pseudo-organelles may be present (Greening and Lithgow, 2020). However in others, channelling is achieved by producing multiprotein complexes or multidomain enzymes (Zhang, 2011).

Gene fusion can facilitate end-to-end covalent binding of two or more enzymes so to produce a multifunctional protein (Argos, 1990; Gokhale and Khosla, 2000). It is possible that such a fusion can bring the active sites within close proximity to one another. By bringing two active sites within range of one another, the intermediate is less likely to diffuse into the bulk cell media, and instead is used by the next active site, thereby achieving substrate channelling (Srivastava and Bernhard, 1986). The human pyrroline-5-carboxylate enzyme is one such fusion, a bifunctional enzyme where the active site of the glutamate kinase moiety interacts with the catalytic domain of the γ -glutamyl phosphate reductase moiety (Fig. 1.1.A) (Pérez-Arellano *et al.*, 2010). It is important to note that proximity based substrate channelling requires close interaction, with models predicting rapid loss in the probability of reaction once distances exceed 10 Å, and benefits from proximity being lost over 25 Å (Fig. 1.2.A) (Bauler *et al.*, 2010).

Not all protein fusions position active sites within proximity of each other, necessitating more elaborate kinds of complex that can facilitate channelling over these larger distances. One such complex is characterised by intramolecular tunnels which connect two active sites (Fig. 1.2.B). These tunnels can physically separate the intermediate from the bulk cell material and limit its ability to diffuse so that it moves directly to the downstream active site (Thoden *et al.*, 1997). This system necessitates synchronising reaction rates so that the intermediate arrives at a ready proximal enzyme, which can be achieved through the use of protein gates (Gora, Brezovsky and Damborsky, 2013). The use of tunnels can facilitate transfer of intermediate over distances greater than proximity channelling would permit, such as ammonia diffusing 45 Å within the carbamoyl phosphate synthetase (Fan *et al.*, 2009). This is not the only step within the carbamoyl pathway, as ultimately the tunnels stretch to almost 100 Å distances from substrate to final product (Thoden *et al.*, 1997). Furthermore, these tunnels can occur both in direct fusions, such as the Eukaryotic imidazole glycerol-phosphate synthase/glutamine amidotransferase protein (Brilli and Fani, 2004), and complexes composed of multiple separate enzymes, such as the tryptophan synthase $\alpha_2\beta_2$ complex (Hyde and Miles, 1990).

In some fusions, an intramolecular tunnel is not needed, as instead a series of charged residues carry intermediates through electrostatic guidance from one active site to the next on the polypeptide (Fig. 1.2.C) (Wheeldon *et al.*, 2016). The Bifunctional Thymidylate Synthase-Dihydrofolate Reductase fusion exists within protozoa and some plant families, containing a 40-Å “electrostatic highway,” in which positively charged domains facilitate directional substrate channelling of negatively charged dihydrofolate (Knighton *et al.*, 1994; Stroud, 1994). The Aldehyde-alcohol dehydrogenase (AdhE) spiroosome is an interesting example of a substrate channel through electrostatic guiding. The AdhE protein alters conformation, from a compact state in which only the aldehyde dehydrogenase (ALDH) domain is solvent accessible, to an extended state that facilitates substrate channelling from the ALDH to the

alcohol dehydrogenase (ADH) domains. This occurs as the conformation change allows solvent access to the ADH, and due to the presence of NAD⁺ in the binding pockets, which guide the acetaldehyde from ALDH to ADH. This substrate channelling aids in protecting the cell from the cytotoxic effects of intermediate acetaldehyde (Kim *et al.*, 2020).

The previously mentioned multienzyme complexes are an alternative method to protein fusion. As opposed to covalent linkage, these complexes form due to strong protein-protein interaction (Nagradova, 2001), which on top of enhancing metabolic channelling (Youjun Zhang *et al.*, 2017) can also alter the activity of the enzymes involved (Roderer *et al.*, 2014). The tryptophan synthase is a classic example of a multienzyme complex, forming an $\alpha_2\beta_2$ arrangement. This arrangement produces a tunnel that protects the α -aminoacrylate intermediate from water preventing formation of any side reactions while the hydrophobic indole molecule travels from the α -domain to the β -domain (Fig. 1.1.B) (Hilario *et al.*, 2016), substantially increasing rate constants of intermediate formation and hydrolysis (Miles, Banzon and Raushel, 1998).

Some complexes may only form under specific conditions. Within the reaction of D-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and L-lactate-dehydrogenase (LDH), substrate channelling is observed within the enzyme under NADH limiting conditions. Under these conditions, the presence of a NADH-GAPDH interaction results in a transient GAPDH-LDH complex forming to channel the NADH to LDH (Svedružić *et al.*, 2020).

The protein scaffolds are an example of a multienzyme complex in which clustering takes place to localize many enzymes to a non-catalytic structure or tag (Wieczorek and Martin, 2012). The most famous example is the cellulosome, an extracellular structure that is used to localize a variety of enzymes associated with hemicellulose and cellulose degradation, controlling local enzyme concentration, as well as the composition of those enzymes, at the cellulose surface (Bayer, Morag and Lamed, 1994; Artzi, Bayer and Moraïs, 2016). As energy levels in anaerobic bacteria limit the production of enzymes, the purpose of this structure is to organise the enzymes in a way to maximise efficiency (Bayer *et al.*, 2004). The cellulosome is composed of a “scaffoldin” molecule derived of many cohesin motifs, which integrates the cellulase and xylanase enzymes by interacting with a complementary dockerin-tag. The scaffoldin is then either localized to the membrane by a second cohesin-dockerin tag system, or exist as a cell free system (Fig 1.1.C). In order to further aid enzyme

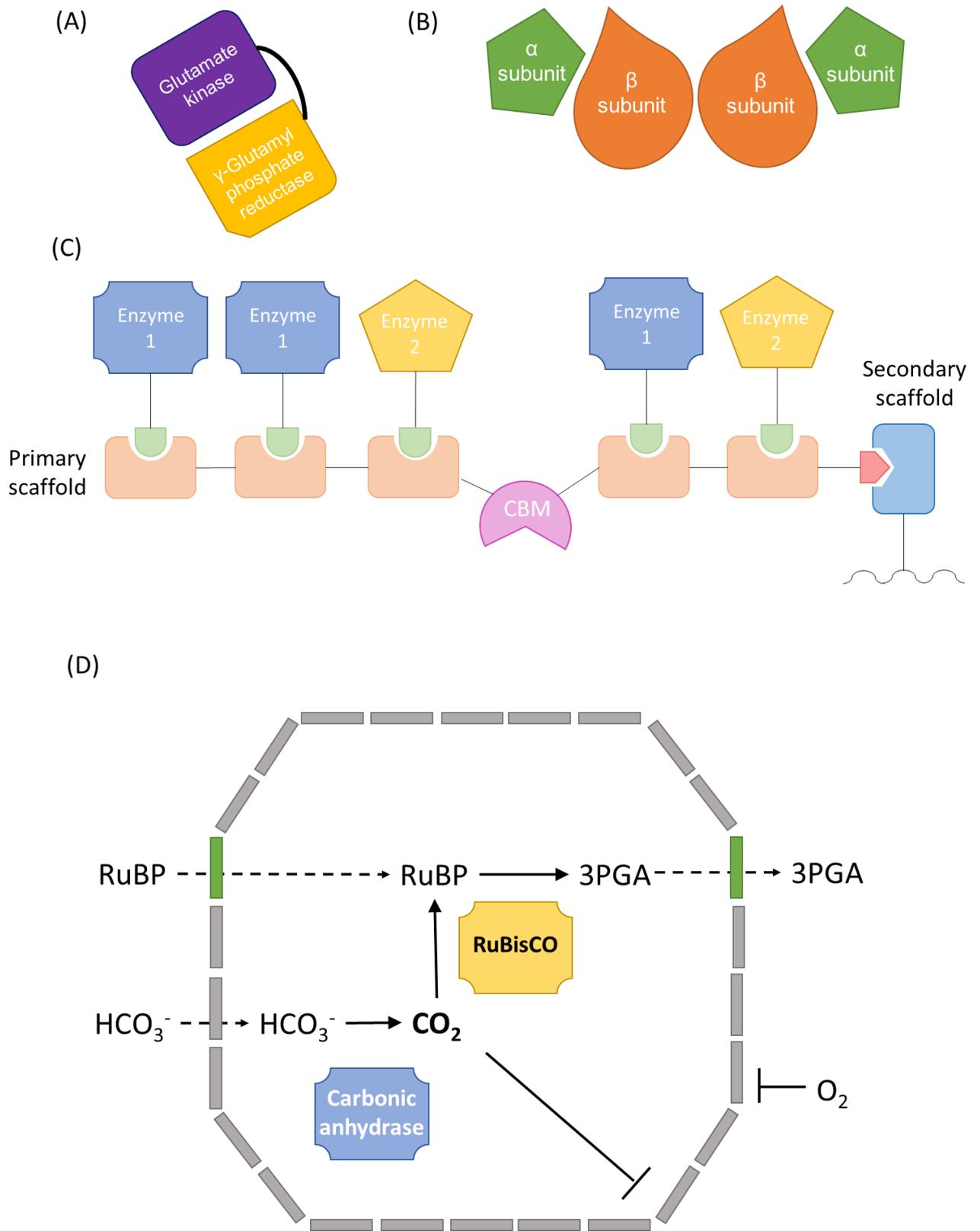


Figure 1.1. Schematics of representatives of different mechanisms of substrate channelling in prokaryotic cells. (A) Direct fusion between glutamate kinase and γ -glutamyl phosphate reductase as seen in the multidomain human pyrroline-5-carboxylate **(B)** The non-covalent tryptophan synthase multiprotein complex is formed between two α -domains and two

β -domains. **(C)** the primary scaffold of the cellulosome complex. The scaffold is composed of many repeats of cohesin onto which multiple cellulase or xylanase enzymes bind through interaction between their dockerin tag and the cohesin. Primary scaffold is bound to a second scaffold by a second type of cohesin-dockerin pair - figure derived from (Artzi, Bayer and Morais, 2016) **(D)** The carboxysome, a protein shell used to sequester enzymes associated with carbon-concentrating mechanisms from the bulk cell material. Reactants and products are able to enter the protein shell whereas inhibitors and intermediates are not – figure derived from (Turmo, Gonzalez-Esquer and Kerfeld, 2017).

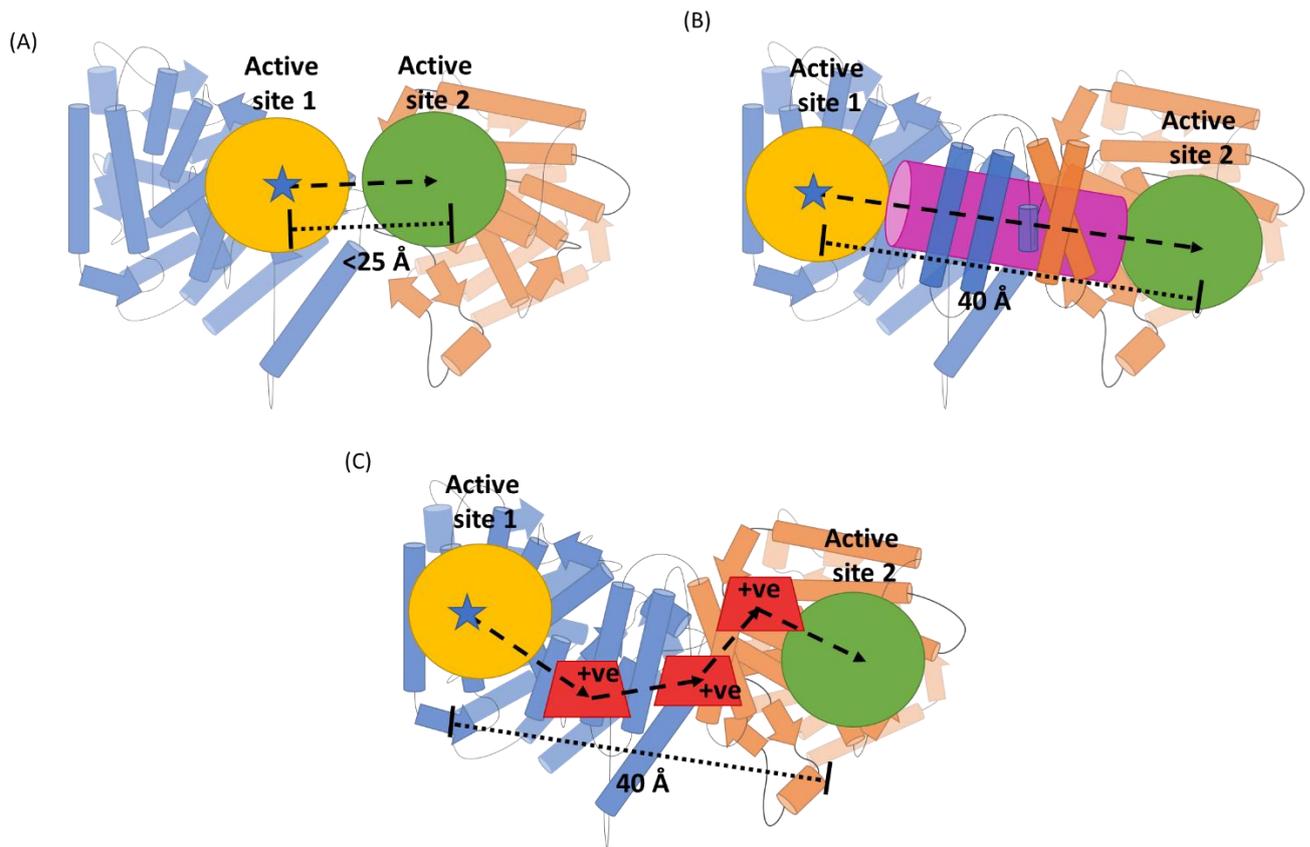


Figure 1.2. Schematic demonstrating different methods of metabolite channelling in multidomain proteins or multienzyme complexes. A cartoon showing the potential methods by which metabolic channelling is achieved between two domains. The first active site is represented by the yellow circle and the second by the green circle. An intermediate is represented as a blue star, and the direction travelled is shown by the dashed arrow. A hypothetical distance is also demonstrated by the dashed line. **(A)** Proximity channelling – the intermediate diffuses directly to the next active site due to the close proximity of the two active sites. Best performance is $<10 \text{ \AA}$ but some channelling may be seen $<25 \text{ \AA}$. **(B)** Intramolecular tunnel – an intermolecular tunnel (represented by the purple cylinder) physically connects the two active sites so that they do not interact with the outside bulk cell material. **(C)** Electrostatic guiding – a series of positively charged domains (represented by the red trapezoids) carry the negatively charged intermediate from active site 1 to 2 to facilitate channelling over large distances.

activity, a cellulose binding domain is also present within the scaffoldin, which interacts with target cellulose substrates (Artzi, Bayer and Moraïs, 2016; Anandharaj *et al.*, 2020).

The non-ribosomal peptide synthetase (NRPS) enzymes form a protein complex could be argued to be another example of a naturally occurring scaffold. The NRPS complex contains a series of donor and acceptor communication-mediating domains which facilitate interaction between enzyme modules responsible for each step of the reaction (Hahn and Stachelhaus, 2006; Huang, Stephan and Kries, 2020). By bringing the modules together, the substrate can be moved from one catalytic motif to the next in an assembly line through a series of covalent bonds with adjacent thioester groups (Fischbach and Walsh, 2006). The thioester bonds are therefore held in close proximity so that the next motif can attack the substrate while it is bound to the previous, overcoming the need for the substrate to diffuse.

Beyond fusion and complexes, some bacteria have demonstrated the ability to separate reactions or intermediates within large constructs. Membrane compartmentalization is one such method to achieve this – within the genus *Pirellula* a lipid-bilayer separates the chromosome and ribosomes from the rest of the cell, producing a structure termed the pirellosome and sequestering the cell in a manner akin to organelles in eukaryotes (Murat, Byrne and Komeili, 2010). Planctomycete cells can further separate out the pirellosome into smaller compartments to achieve greater control over structural organisation of the cell (Lindsay *et al.*, 2001). Intracytoplasmic membranes may also develop within *Nitrosomonas* and *Nitrobacter* species to localize high concentrations of membrane bound metal proteins (Wallace and Nicholas, 1969; Fiencke and Bock, 2006).

Other organisms can form pseudo-organelles through the expression of protein shells, bacterial microcompartments composed of many repeated units of protein and containing enzymes that are functionally related (Fig. 1.1.D) (Yeates *et al.*, 2008). Through this system, the prokaryotes can approximate the substrate channelling benefits of organelles – with a common trait seen of a small metabolite being produced by one enzyme within the protein shell that is either cytotoxic or volatile and is consumed by a downstream enzyme (Yeates, Crowley and Tanaka, 2010). The protein shell is itself composed of multiple repeats of “encapsulin” which come together to form the nanostructure. For example, *Thermotoga maritima* expresses a protein shell composed of 60 encapsulin monomers, producing an icosahedral shell with a 240 Å diameter (Sutter *et al.*, 2008).

Cyanobacteria and chemoautolithotrophs use a protein shell, termed the carboxysome, to protect the ribulose 1,5-bisphosphate carboxylase/oxygenase (RuBisCO) catalysed carbon concentrating reaction from inhibition in low CO₂ environments (Kaplan and Reinhold, 1999;

Murat, Byrne and Komeili, 2010). The carboxysome is a protein shell between 800 and 1400 Å in diameter that is rich in RuBisCo and carbonic anhydrase. It is produced from the interaction of hexameric proteins, that form the face of the icosahedral shell, and pentameric proteins, which form the vertices (Tanaka *et al.*, 2008). The cellulosome facilitates the reaction under low CO₂ conditions by controlling access to enzymes contained. The cellulosome will permit both ribulose 1,5-bisphosphate and HCO₃⁻ to enter the structure while preventing O₂ diffusing across the shell, preventing it from inhibiting RuBisCO activity. When CO₂ is produced by carbonic anhydrase-catalysed dehydration of bicarbonate, it is limited in its ability to exit the protein shell, increasing local concentration and saturating RuBisCO (Price *et al.*, 2008; De Araujo *et al.*, 2014). Through this system metabolic channelling is achieved, increasing production of 3-phosphoglycerate (Fig. 1.1.D) (Turmo, Gonzalez-Esquer and Kerfeld, 2017).

Synthetic biology is a growing field in the overlap of biology and engineering, in which artificial pathways are encoded within a host cell to generate new products (Benner and Sismour, 2005). Synthetic biology can enable cells to produce complex organic compounds which would be expensive or challenging to synthesise by traditional organic chemistry (Paddon and Keasling, 2014). However, the pathways introduced by synthetic biology can be hindered in practice by poor yields of product, as the new enzymes may be less active *in vivo* or intermediates within reaction pathways could be easily lost (Khalil and Collins, 2010).

Because of limitations in yields of vital products, substrate channelling is of growing interest in synthetic biology (Wheeldon *et al.*, 2016). If enhanced substrate channelling is achieved, multi-enzyme synthetic complexes could overcome the limitations of synthetic biology pathways, in order to increase product yield. This work is particularly interested in the role of metabolic channelling between membrane transport proteins and downstream enzymes and whether it can enhance product yields. In order to achieve such improvements, researchers look to natural examples of metabolic channelling to inspire design.

1.2. Protein Fusion

1.2.1. Protein fusions in nature

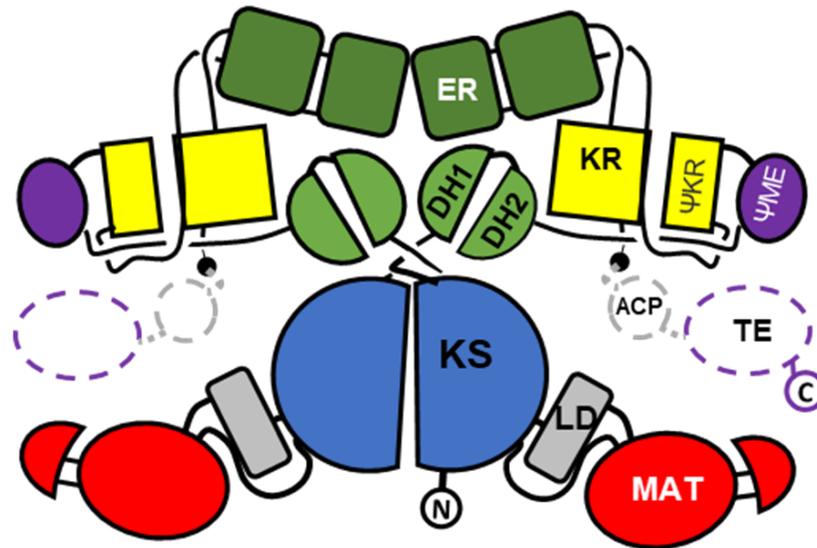
In nature, large polypeptide structures can be found containing multiple enzymes or active sites, notable examples being the Type I fatty acid synthase (Fig. 1.3.A) (Schweizer and Hofmann, 2004) and the multifunctional polyketide synthase (PKS) (Fig 1.3.B) (Beedessee *et al.*, 2015). These are examples of natural protein fusion, in which two or more disparate genetic elements are joined together so that they are transcribed as a single, large protein. A variety of fused proteins can be found throughout a diverse range of organisms, even within eukaryotes such as green algae (Madsen *et al.*, 2018), the protozoan parasite *Plasmodium falciparum* (Mamoun *et al.*, 1998), and humans (Pérez-Arellano *et al.*, 2010). Fusions can also form between multiple transporters (Bolhuis *et al.*, 1998; Willson *et al.*, 2019) and between enzymes and non-enzymatic proteins (Laskowski, Gerick and Thornton, 2009; Liang and Blumenthal, 2013).

Gene fusion has been identified as a key factor for the production of multidomain proteins (Doolittle, 1995). Through insertion of one protein domain within another, organisms can evolve additional functionality in absence of an external gene source (Willson, Chapman and Thomas, 2019). To facilitate the functional fusion, translation must be able to proceed through the entire sequence, necessitating that all components of the fusion are in frame, with no stop codons to interrupt the sequence (Long, 2000).

There are a number of mechanisms by which the organism may achieve such a fusion. Point mutation can remove the stop codon and cause downstream coding sequences to be brought into frame. Conversely, duplication within a gene can cause domains, or even the whole gene, to be copied to form a larger protein with increased functionality. On the other hand, recombinant gene shuffling can facilitate fusions of two or more different domains. Shuffling may be achieved through domain exchange, in which two multidomain protein-encoding genes swap a domain to create a novel fusion, or through the direct insertion of one dissimilar gene or domain within another gene (Conant and Wagner, 2005; Pasek, Risler and Brezellec, 2006). Within more complex organisms, fusions may be the result of gene shuffling between different chromosomes (Wu *et al.*, 2013).

Natural protein fusion can introduce a greater kinetic efficiency and improve control over coordinated syntheses by multienzyme reactions (Schweizer and Hofmann, 2004). Furthermore, fusion of regulatory domains to functional enzymes has allowed for post-translational control of enzyme activity (Laskowski, Gerick and Thornton, 2009). Fusion proteins are not always desired products, however, as they have been implicated in disease,

(A)



(B)

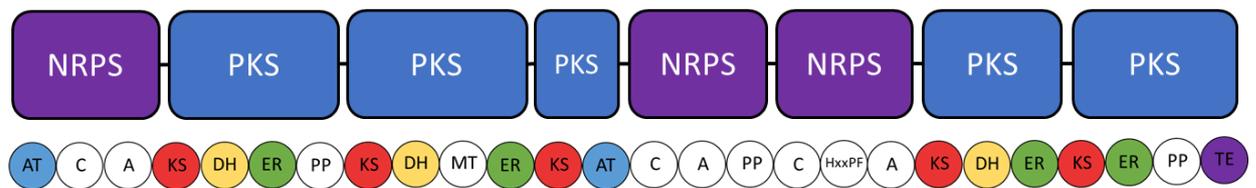


Figure 1.3. Schematics of naturally occurring multidomain protein fusions. (A) A cartoon showing the fusion of domains in the mammalian type I fatty acid synthase, derived from the crystal structure (Maier, Leibundgut and Ban, 2008). The fusion is composed of the Acyl carrier protein (ACP), β -ketoacyl synthase (KS), Malonyl transferase (MT), NADPH-dependent β -ketoreductase (KR), a dehydratase (DH), the NADPH-dependent enoyl reductase (ER), the malonyl-acetyl transferase (MAT), and the thioesterase (TE) domain. **(B)** The consensus schematic of the large NRPS-PKS fusion protein produced by the *Symbiodinium minutum* genome. Domains represented are the phosphopantetheine (PP), acyl transferase (AT), condensation (C), adenylation (A), ketosynthase (KS), dehydratase (DH), ketoreductase (KR), methyltransferase (MT), thioesterase (TE). Derived from (Beedessee et al., 2015).

such as the oncogenic fusions. Here, proteins associated with cell functionality become fused and develop disadvantageous activities, resulting in over expression or novel protein behaviours (Latysheva and Babu, 2019). For example, within acute myeloid leukaemia, transcriptional or epigenetic regulation proteins are fused, risking deregulation of transcription and protein signalling (Melo, 1996; Rapin and Porse, 2014).

Fusion has also had a role in the evolution of large membrane proteins, such as the major facilitator superfamily (MFS), which have been demonstrated to be the result of fusion and intergenic duplication of genes encoding smaller proteins (Saier, 2003, 2016). Furthermore, a diverse range of membrane transporter-soluble domain fusions are observed in nature, such as substrate binding domains fused to ABC domains (Bouvier *et al.*, 2000; Heide and Poolman, 2002; Fulyani *et al.*, 2013), or the promiscuous partners such as the cystathionine-beta-synthase domain, which are found fused to multiple membrane proteins (Meyer and Dutzler, 2006; Chen and Beattie, 2007; Tomita *et al.*, 2017). This has yet to be systemically studied, however.

There is a relative scarcity of fusions between membrane transporters and enzymes (Willson, Chapman and Thomas, 2019). A key example is found in bacterial NukT-like ABC transporters involved in lantibiotic synthesis (Severi and Thomas, 2019), where a peptidase may be found fused to the transmembrane domain to assist with the final stage of synthesis (Håvarstein, Diep and Nes, 1995; Zheng *et al.*, 2018), showing an example of crosstalk between domains. Three examples of transporter-enzyme fusion can be seen within the MFS superfamily: the LpIT flippase (Harvat *et al.*, 2005), *Magnetospirillum gryphiswaldense* MamZ (Raschdorf *et al.*, 2013), and the human PepT1 and PepT2 transporters (Beale *et al.*, 2015). Curiously, the fusion architecture of the LpIT flippase is the most abundant MFS-soluble protein fusion (Willson *et al.*, 2019), so while incidents of transmembrane domain fusion to enzymes are uncommon, this does not reflect how abundant the individual fusions will be within nature.

1.2.2. Engineering synthetic protein fusions

Synthetic direct fusion attempts to replicate multi-enzyme complexes by covalently linking two or more enzymes that are not otherwise linked (Chen, Zaro and Shen, 2013). Three key types of protein fusion can be achieved synthetically. The first is tandem fusion, which describes the genetic fusion of two genes end-to-end (Fig. 1.4.A). Conversely, domain insertion involves the addition of a single domain in the middle of the host gene (Fig. 1.4.B). Finally, post-translational conjugation involves the chemical or enzymatic cross-linking of two genetically separate proteins via a defined peptide tag (Fig. 1.4.C) (Yu *et al.*, 2015).

Synthetic fusion has many applications and is particularly useful for *in vivo* labelling studies. By fusing a fluorescent protein to the protein of interest, it is possible to study its function within a cell and elucidate information difficult to otherwise obtain (Keppler *et al.*, 2004). Fluorescence studies can be useful for identifying the location and topology of proteins in cells (Kenri *et al.*, 2004; Belardinelli and Jackson, 2017), measuring mobility of proteins *in vivo* (Reits and Neefjes, 2001; Rassam *et al.*, 2018), and determining whether protein-protein interactions are taking place (Kenworthy, 2001; Kleanthous, Rassam and Baumann, 2015).

It is also possible for proteins to be fused to tags to aid in the stability or localization of the protein. It has been shown that fusion of mCherry to thermostable trehalose transferase (TreT) has aided in increasing protein solubility and stability when expressed in an *E. coli* host (Mestrom *et al.*, 2019). The Maltose-Binding protein has been used in a comparable fashion, fused as a tag to mammalian proteins to localize them to the periplasm and increase stability (Malik, 2016; Reuten *et al.*, 2016).

In a pharmaceutical context, synthetic fusion can prolong the biological half-life of therapeutic peptides (Kim *et al.*, 2010). In addition, therapeutic peptides can be fused to suitable monoclonal antibodies to facilitate crossing of the blood-brain barrier to achieve enzyme replacement therapy or drug delivery in the central nervous system (Boado *et al.*, 2008, 2014; Lu *et al.*, 2010). Nanostructures produced through fusion have also been implicated in medical roles, able to facilitate drug and small molecule delivery within the body (Truffi *et al.*, 2016; Butterfield *et al.*, 2017; Reuter *et al.*, 2017), vaccine production (Kanekiyo *et al.*, 2013; Z. Wang *et al.*, 2019) and bio-sensing applications (Hwang *et al.*, 2013; Y. Wang *et al.*, 2019), the latter of which have additional applications outside pharmaceuticals.

Protein supramolecules are produced by fusing oligomeric peptides to one another to produce new structures (Kobayashi and Arai, 2017), altering the shape and size of the

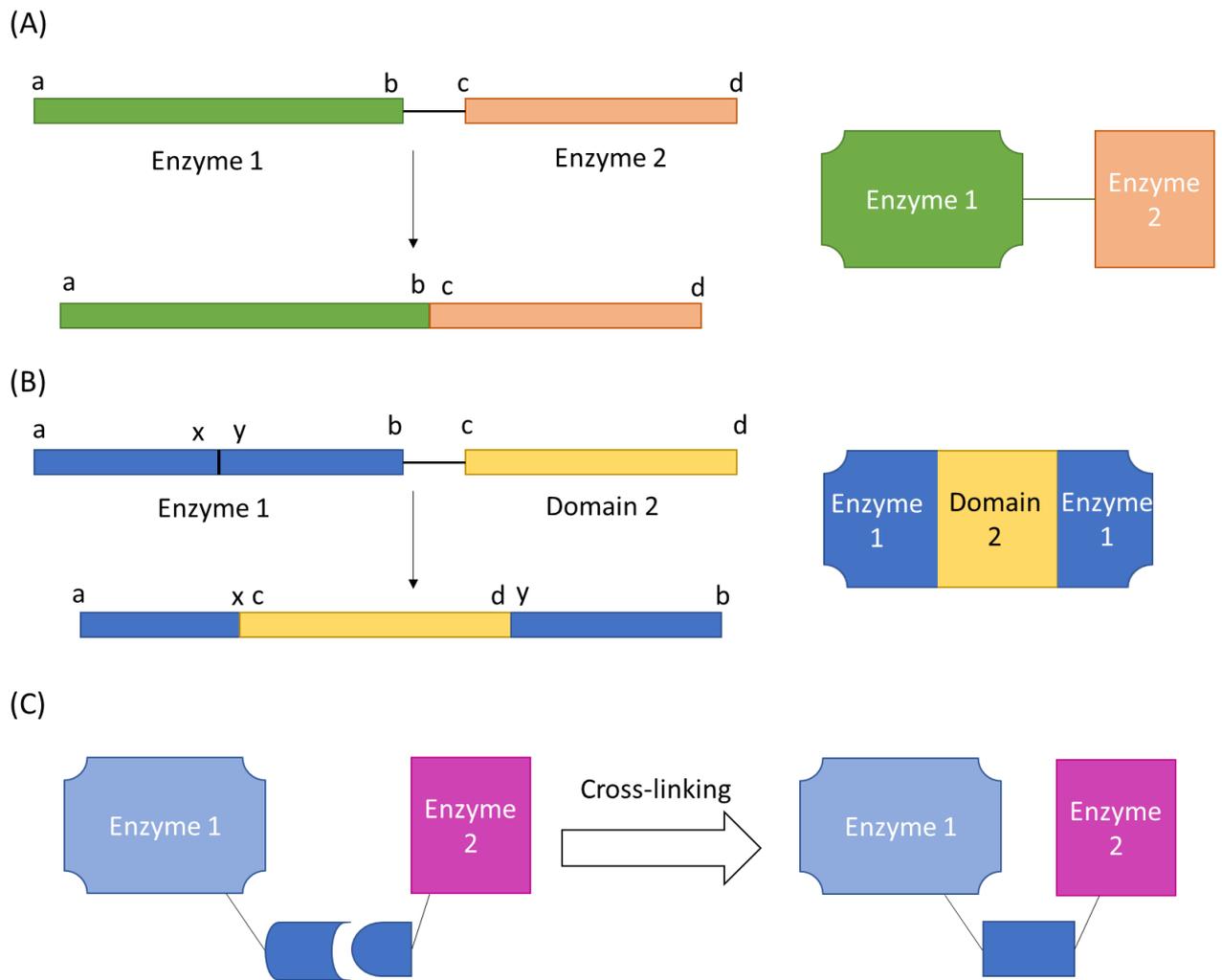


Figure 1.4. The three types of synthetic fusions. A schematic representing the fusion of two different enzymes **(A) Tandem fusion** – genes encoding enzyme 1 is fused to enzyme 2 by the end-to-end attachment at site b and c. **(B) Domain insertion** – the whole gene encoding domain 2 is inserted at site x and y of domain 1. **(C) Post-translational conjugation** – enzyme 1 and enzyme 2 are expressed independently, followed by chemical or enzymatic cross-linking at specific amino acids or sequences.

supramolecule and its cavity by varying the number of and composition of fused monomers (Lai, Cascio and Yeates, 2012; Cannon *et al.*, 2020).

Synthetic fusions can also be used to improve product yields and enzyme expression (Aalbers and Fraaije, 2019); however, using fusion to enhance such yields is limited when compared to the biotechnological methods discussed above (Iturrate *et al.*, 2009). Cells can be used as biological factories to produce high-value organic compounds from sugars or low-cost feedstocks (Chandel and Singh, 2011; Moon *et al.*, 2016; Liu, Ding and Jiang, 2017). However, there are considerations for yield and productivity that have limited large-scale adoption of biologically produced compounds (Gustavsson and Lee, 2016). Enhancing yields is a key application for protein fusion, with targets such as alcohols (Anderlund, Rådström and Hahn-Hägerdal, 2001; Lewicka *et al.*, 2014; Nebel *et al.*, 2014; Kim *et al.*, 2015), esters (Torres Pazmiño *et al.*, 2008; Torres Pazmiño *et al.*, 2009; Jeon *et al.*, 2015) or terpenes (Formighieri and Melis, 2015, 2016; Chaves *et al.*, 2017; Hu *et al.*, 2020). Fusion of two or more proteins can be applied to enhance channelling through metabolic pathways (Lewicka *et al.*, 2014), increase degradation of lignocellulosic feedstocks (Kim *et al.*, 2015; Xu *et al.*, 2018), remove toxic by-products (Chin *et al.*, 2013), or facilitate introduction of novel pathways through fused enzymes from different organisms (Jeon *et al.*, 2015).

1.2.3. Design of synthetic protein fusion

The increase in yield can be highly variable between examples of synthetic fusion, depending on the associated enzymes and pathways chosen. Resveratrol yield was increased by 15-fold through fusion of enzymes from unrelated species, compared to the two enzymes co-expressed but unfused (Zhang *et al.*, 2006; Wang *et al.*, 2011). In contrast, muconic acid saw a 4.44-fold gain in yield compared to its unfused homologues (Fujiwara *et al.*, 2018) whereas sesquiterpene yield was only increased 2-fold by fusion (Albertsen *et al.*, 2011). All of these were simple tandem fusions with flexible linkers, however differences in the proteins and pathways involved can affect the extent to which an improvement in yield is seen. An improvement seen in one pathway will not always be comparable to one seen in another. Furthermore, expressing systems in non-native hosts may require expression of additional enzymes beyond the studied pathway before a yield increase is observed (Formighieri and Melis, 2016). Therefore, when considering successful substrate channelling the structural features of the two fused proteins, as well as the nature of the host organism, are important to consider.

Looking in greater detail at the 15-fold increased resveratrol production in yeast, structural and kinetic studies have been used to observe the resveratrol synthetic pathway to determine what benefits fusion achieved (Wang *et al.*, 2011). The fusion exists in a dimeric form and is produced from *Arabidopsis thaliana* 4-coumaroyl-CoA ligase (At4CL1) and *Vitis vinifera* stilbene synthase (VvSTS) (Fig. 1.5). These studies suggested that a change in enzyme activity was partially responsible for improved yield, however the activity was no greater 3-fold that of the unfused enzymes, other factors must be responsible for the 15-fold increase in yield. This therefore implies that the remaining increase in *in vivo* activity was due to colocalization of the active sites (Wang *et al.*, 2011). Curiously, the active sites of the two domains were separated by a distance of ~ 70 Å, greater than what would be within the expected range of diffusion-mediated substrate channelling (Bauler *et al.*, 2010). It is therefore likely that other benefits were obtained by this enzyme localization – perhaps the dimerization resulted in an increased local enzyme concentration, for example. The nature of distance was obtained from crystal structure, and it is important to consider that in some cases the crystal structure does not accurately represent the *in vivo* state of the protein (Postle *et al.*, 2010; Fukasawa and Tomii, 2019).

A key concern for synthetic fusion is steric hindrance. The close proximity of two proteins when covalently bound can introduce steric pressure to the system or even result in misfolding. In order to alleviate the strain, short peptide linkers are introduced (Chen, Zaro and Shen, 2013). Linkers can vary from highly flexible to rigid, depending on the amino acid

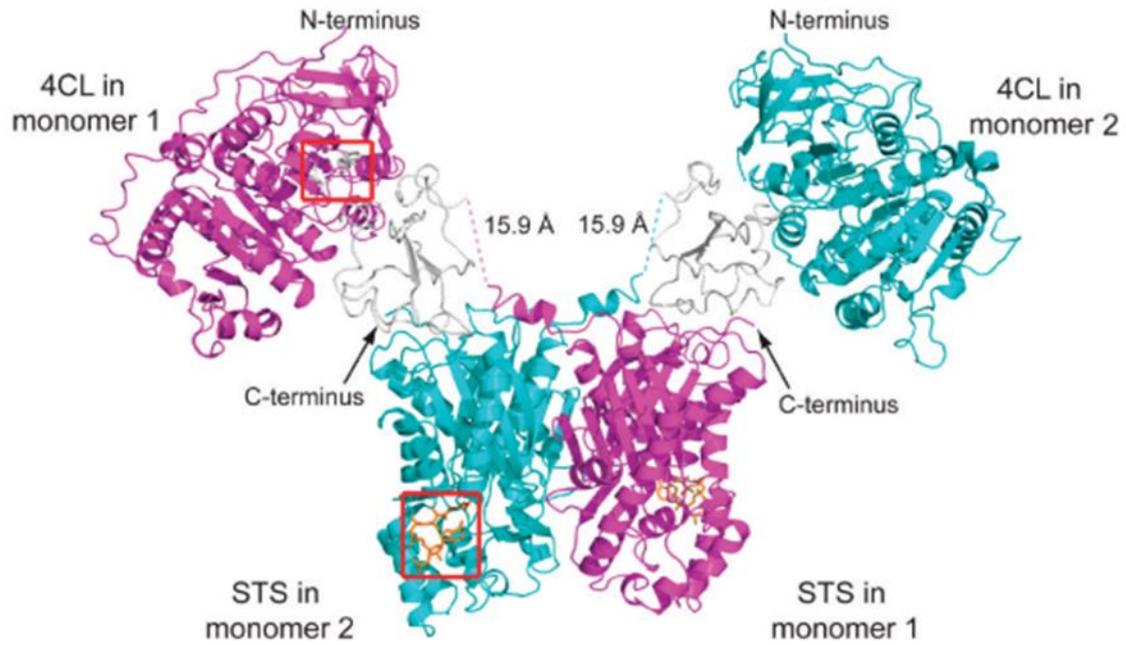


Figure 1.5. The crystal structure of the resveratrol producing 4CL::STS fusion. Produced by Wang *et al.* this figure demonstrates a 2-component fusion between *Arabidopsis thaliana* 4-coumaroyl-CoA ligase (4CL1) and *Vitis vinifera* stilbene synthase (STS). The active sites are highlighted by the red boxes and the 15.9 Å flexible linker represented by the dashed line. (Wang *et al.*, 2011).

composition, as well as differ in length, all of which may alter the stability and activity of the fusion protein (Fig. 1.6.A) (Arai *et al.*, 2001; Agapakis *et al.*, 2010).

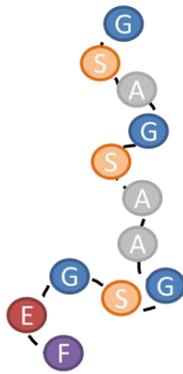
The more flexible linkers allow the components of the protein fusion a greater degree of movement (Fig. 1.6.B) (Chen, Zaro and Shen, 2013), allowing the linker to loop back on itself and permit interdomain interaction between active sites (Papaleo *et al.*, 2016); however, it can also introduce unfavourable interactions between the two domains (Maeda *et al.*, 1997; Shamriz, Ofoghi and Moazami, 2016). Comparatively, rigid linkers maintain distance between two domains, which can reduce interdomain interaction when unfavourable to increase stability and bioactivity (Fig 1.6.C) (Amet, Lee and Shen, 2009; Shamriz, Ofoghi and Moazami, 2016).

The optimal extent of flexibility within the linker can vary depending on the proteins, making it an important consideration when designing fusions (Robinson and Sauer, 1998). Furthermore, beyond just the consideration of whether linkers are rigid or flexible, the secondary structures of the linker peptides can also impact activity, with helix-forming rigid linkers stabilizing fusions to different amounts than alternative polyproline-rich rigid linkers (Zhao *et al.*, 2008).

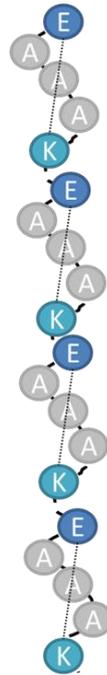
A key requirement of linkers is to ensure the optimal accessibility of active sites for the channelling of intermediate from one enzyme to the next (Yu *et al.*, 2015). Fusions with correct orientation between active sites increase product yields, compared to suboptimal orientations (Lim *et al.*, 2016). Using different linkers to change the distance between active sites is also a way of improving yields (Guo *et al.*, 2017). Orientation of the active sites is therefore another consideration when accounting for the flexibility of linkers, as too much flexibility means that an advantageous orientation is not maintained, reducing activity, compared to more rigid linkers in spite of the increased distance, and is increasingly susceptible to proteolytic cleavage (Papaleo *et al.*, 2016; Guo *et al.*, 2017).

Even beyond the linkers used, other features of the fusion can impact stability and the activity of the enzymes, such as the terminal at which the enzyme is fused (Aalbers and Fraaije, 2017). A key example is the fusion of formate dehydrogenase (FDH) to NADPH-dependant oxidoreductase enzymes, which can enhance cofactor regeneration. When fusion of FDH occurs at the C-terminal of the oxidoreductase, a severe loss of FDH enzyme activity is seen, compared to the fusion at the N-terminal (Hölsch and Weuster-Botz, 2010; Yonghui Zhang *et al.*, 2017). As FDH catalyses the rate limiting step, such a rate loss reduced the impact of the fusion. The importance of this consideration, however, can vary between

(A)



Flexible



Rigid helical



Rigid ruler

(B)



(C)

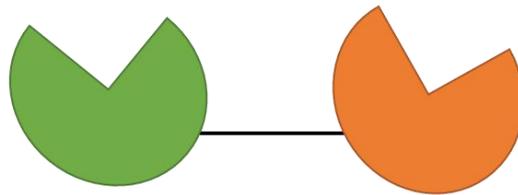


Figure 1.6. Synthetic linkers used within protein fusions. (A) a schematic of three hypothetical linkers, the first a flexible linker contains multiple amino acids that facilitate high degrees of freedom of movement, the second a rigid helical linker contains repeated EAAK motifs that form salt bridges between the glutamic acid and lysine amino acids creating the rigid structure and finally the third linker, the rigid ruler, contains repeats of proline which heavily limits the degrees of freedom to create a straight, ruler like linker. **(B)** A fusion between two enzymes, held together by a flexible linker allow close interaction between the two domains. **(C)** A fusion between two enzymes held together by a rigid linker, enforcing separation between the two domains.

systems and may not play a factor depending on the enzymes fused (Torres Pazmiño *et al.*, 2008).

It is also important to consider the enzymes used within the fusion. Depending on the catalysed reaction, it can be preferential to select enzymes or proteins of similar pH optima (Ljungcrantz *et al.*, 1989; Colpa *et al.*, 2017) and temperature optima (Mangala *et al.*, 2003; Adlakha *et al.*, 2012) as otherwise the fused protein may exhibit reduced bioactivity or stability within the cell due to changing optimal conditions (Yu *et al.*, 2015; Care *et al.*, 2017). However, in some cases changing optima can be less consequential or even the target of the fusion, such as in *in vitro* conditions (Lee *et al.*, 2011; Stressler *et al.*, 2016).

The oligomeric state of the enzyme is also an important consideration – some proteins are less active in their monomeric form, however fusion can impact the protein-protein interaction within the oligomer (Aalbers and Fraaije, 2019; Oreb, 2020). The β -galactosidase /galactose dehydrogenase fusion, for example, is derived from a tetrameric and dimeric protein respectively. This fusion was found to exist in complexes containing either four and six subunits – while the activity of the fusion was high, the reduced oligomeric state is a cause for concern (Ljungcrantz *et al.*, 1989). With a suitable linker design it is possible to maintain oligomeric state of the proteins involved, however (Liao *et al.*, 2021). Alternatively, it is also possible for fusions between oligomeric enzymes to increase subunit stoichiometry or multimerize, creating large assemblies of enzymes and increasing local enzyme concentration (Bülow and Mosbach, 1991; Conrado, Varner and DeLisa, 2008).

An increase in substrate channelling or electrostatic effects due to fusion can still facilitate improvements in the overall pathway activity, even when specific activity of the fused enzyme is lower than the individual components (Iturrate *et al.*, 2009; Hölsch and Weuster-Botz, 2010). This is demonstrated in the fusion of *Gloeophyllum trabeum* endoglucanase and xylanase to hydrolyse lignocellulose, where all feedstocks showed 1.3 to 1.4-fold increase in degradation when treated with the lower activity fusion protein than the individual components co-expressed (Kim *et al.*, 2015). The specific activity of the enzyme may also be influenced by altering the linker in some fusions (Adlakha *et al.*, 2011; Huang *et al.*, 2016).

1.3. Enzyme Scaffolds

In some examples, the increased substrate channelling facilitated by protein fusion may not overcome the reduced activity or misfolding resulting from fusing the two proteins (Bülow, Ljungcrantz and Mosbach, 1985; Dueber *et al.*, 2009). This is particularly true within multimeric components, where oligomerization can be impacted by the fusion (Conrado, Varner and DeLisa, 2008; Lee, DeLoache and Dueber, 2012). In such incidents, alternative methods of synthetic substrate channelling must be considered.

Attaching enzymes to a scaffold is a method to increase metabolic channelling within a reaction pathway while reducing the risk of steric hindrance between enzymes (Dueber *et al.*, 2009). As opposed to fusing two or more bulky enzymes to one another, each enzyme within the scaffold system is instead modified to contain only a small peptide tag (Bayer *et al.*, 2004).

Synthetic enzyme scaffolds are elongated polymeric structures to which enzymes bind via non-covalent interactions in order to form a complex. The close proximity of the enzymes within the complex then increases substrate channelling (Siu *et al.*, 2015). The use of scaffolds to localize enzymes to one another is notable for the transient interactions that form, which have a smaller impact on the structure and stability of the target enzymes, compared to protein fusion (Fierobe *et al.*, 2001).

1.3.1. Protein scaffolds

In nature, scaffolds are often proteinaceous and bind enzymes due to a complementary protein-binding site (Okamoto *et al.*, 1998; McDonald *et al.*, 2000). Enzyme scaffolds in nature modulate multiple roles within cells, such as in signalling in MAP kinases via Kinase Suppressor of Ras (KSR) scaffolds or the $\alpha\beta/\beta\alpha$ sandwich scaffold of metallohydrolase enzymes (Park, 2006; Brown and Sacks, 2009). Synthetic protein scaffolds have focused upon replicating systems where the scaffold interacts with a defined, easily replicated feature of the target protein (Dueber *et al.*, 2009; You and Zhang, 2014). Synthetic scaffolds either use a “scaffoldin” inspired separate structure (Fig. 1.1.C), or fuse the complimentary tags directly to the enzymes involved so that they localize to each other directly. Linker peptides will often be introduced in the new complexes to space peptide tags and the enzymes (Dueber *et al.*, 2009; Chen, Zaro and Shen, 2013).

An artificial protein scaffold was produced, composed of domains derived from different signalling proteins from various metazoan species, to enhance the mevalonate pathway, achieving a 77-fold increase in production in the non-native organism *E. coli*. Mevalonate itself is a precursor of the anti-malarial drug artemisinin (Dueber *et al.*, 2009). The signalling

proteins, such as mouse SH3, form strong protein-protein interactions with ligands composed of short peptide sequences (Phane Richard *et al.*, 1995); therefore, by fusing corresponding ligands to the target enzymes, the mevalonate pathway enzymes could be localized to the scaffold (Fig. 1.7.A). However, a limitation of scaffold size exists here due to the dependency of different signal domains from different animal species (Dueber *et al.*, 2009). While this signal protein scaffold was able to achieve a significant improvement in the yields of mevalonate, the improvements in yields of glucaric acid and resveratrol were a comparatively small 5-fold increase when the scaffold was used (Moon *et al.*, 2010; Wang and Yu, 2012). A review by DeLisa *et al.* suggested these signalling domains would not be feasible for all systems as the functions of the enzymes could be perturbed depending on how the enzymes and ligands interact (DeLisa and Conrado, 2009).

The result of ligand binding to an enzyme could interfere with the stability of the attached enzyme (Siu *et al.*, 2015). Therefore, when considering protein scaffolds, it is important to be able to change ligands to overcome changes in enzyme activity. The protein-protein interaction between dockerins and cohesins of the cellulosome are one alternative (Bayer, Morag and Lamed, 1994). Dockerins are small peptide tags found on enzymes associated with cellulose degradation, which complement a binding site on a cohesin molecule forming a strong protein-protein interaction (K_D as low as 10^{-11} M) (Mechaly *et al.*, 2001; Handelsman *et al.*, 2004). By adding dockerin tags to enzymes, it is possible to localize them to an assembled cohesin scaffold (Fig 1.7.B) (Liu *et al.*, 2013; You and Zhang, 2014). Cohesin-dockerin systems have been used as scaffolds in order to increase yields of fructose-6-phosphate, 2,3-butanediol and bioethylene (Jindou *et al.*, 2014; Kim and Hahn, 2014; You and Zhang, 2014). During the production of fructose 6-phosphate, cohesin-dockerin domains from three separate organisms, *Clostridium thermocellum*, *Clostridium cellulovorans*, and *Ruminococcus flavefaciens*, were obtained (You and Zhang, 2014). A key strength of the cohesin-dockerin system for synthetic scaffold purposes, which this experiment applied, is the species specificity of cohesins and dockerins. A dockerin from one species will not interact with the cohesin from a different organism.

In order to increase the range of protein scaffolds available for substrate channelling, scaffolds using affibody complexes have also been developed. Affibodies are proteins derived from the *Staphylococcus* protein A, and form non-covalent bonds with pre-defined molecules (Löfblom *et al.*, 2010). In the context of a protein scaffold, two types of affibody are used. The first is a standard affibody to which reaction enzymes are fused to, and the second is an anti-idiotypic affibody which binds the complementary, enzyme bound affibody. The scaffold is then assembled, capable of keeping two proteins within 80 Å of one another

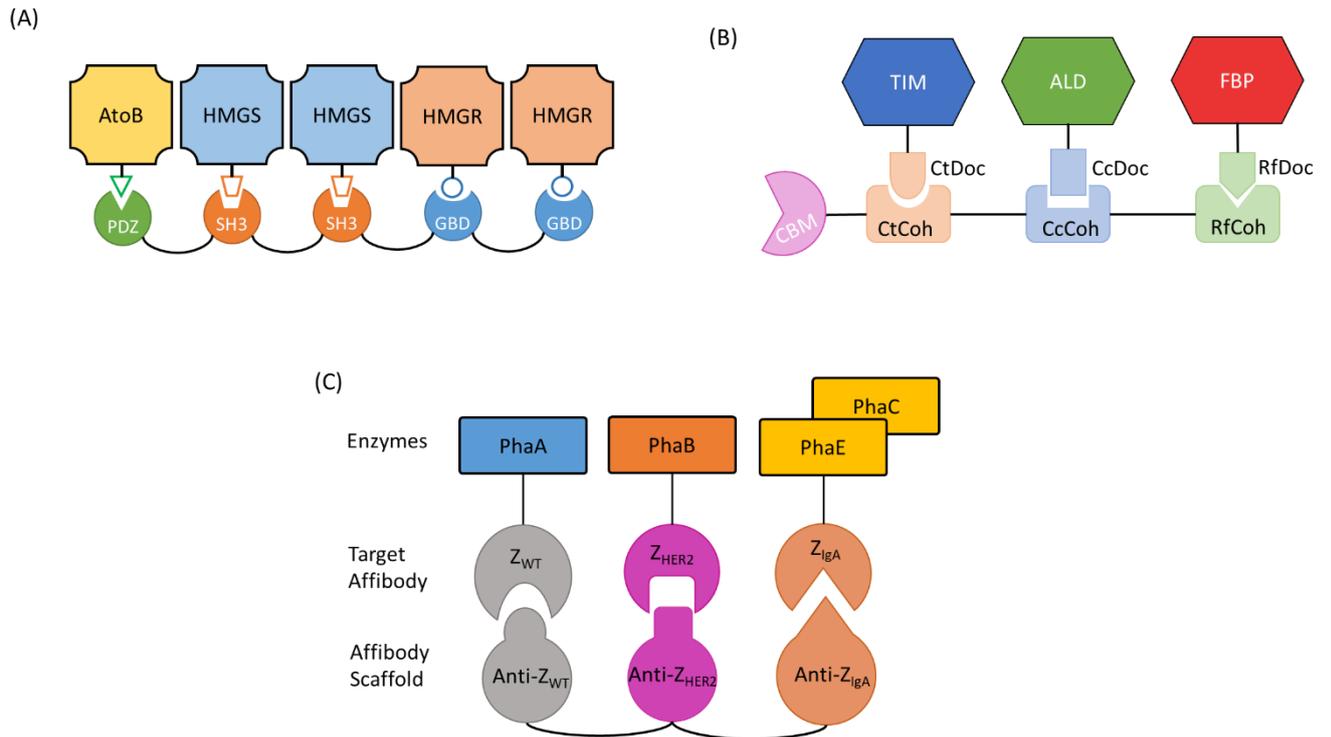


Figure 1.7. Schematic of three different synthetic protein scaffolds produced in literature **(A)** The metazoan signalling pathway ligand scaffold – a scaffold is produced from three separate ligand-domain pairs associated with strong protein-protein interaction. PDZ - the PSD95/DlgA/Zo-1 domain from the adaptor protein syntrophin, SH3 - Src homology 3 domain from the adaptor protein CRK, and GBD - GTPase binding domain (GBD) from the actin polymerization switch N-WASP. In order to increase melvanoate titre, the PDZ domain is fused to *E. coli* acetoacetyl-CoA thiolase (AtoB), the SH3 domain is fused to *Saccharomyces cerevisiae* hydroxy-methylglutaryl-CoA synthase (HMGS) and the GDB domain is fused to *S. cerevisiae* hydroxy-methylglutaryl-CoA reductase (HMGR). The scaffold is composed of the ligands, and displayed orientation is P1S2G2, the best performing stoichiometry. Figure derived from (Dueber *et al.*, 2009) **(B)** The cohesin-dockerin scaffold - Cohesin-dockerin pairs are obtained, derived from cellulosome proteins *Clostridium thermocellum* ATCC 27405 CipA (CtCoh-ctDoc), *Clostridium cellulovorans* ATCC 35296 CbpA (CcCoh-CcDoc), and *Ruminococcus flavefaciens* ScaB (RfCoh-RfDoc). The scaffold is produced from fusion between CtCoh, CcCoh and RfCoh. CtDoc is fused to triosephosphate isomerase (TIM), CcDoc is fused to aldolase (ALD), and RfDoc is fused to fructose 1,6-bisphosphatase (FBP). Each cohesin will only interact from the dockerin of the same organism. Figure derived from (You and Zhang, 2014) **(C)** The affibody scaffold – A scaffold is produced from 3 anti-affibody peptides, which are themselves affibodies that have high affinity to a specific target affibody.

The complimentary affibodies are ZWT a wildtype affibody, ZlgA designed to bind to IgA and ZHER2 designed to bind to tyrosine kinase receptor HER2. Each of these affibodies were fused to an enzyme of the polyhydroxybutyrate (PHB) pathway in *E. coli*. ZWT was fused to β -ketothiolase (PhaA), ZHER2 was fused to an acetoacetyl-CoA reductase (PhaB) and ZlgA was fused to the PhaEC complex, a PHA synthase. The affibody scaffold then interacts with each tag to localize the fused enzymes. Figure derived from (Tippmann *et al.*, 2016).

(Tippmann *et al.*, 2016). The affibody scaffold was used for production of polyhydroxybutyrate (PHB), increasing yield by 7.5-fold. The application of the affibody scaffold to the PHB pathway also provides direct evidence of protein scaffolds facilitating substrate channelling when a protein fusion fails (Fig 1.7.C) (Tippmann *et al.*, 2016).

A key advantage of the ligand-tag scaffold system is the control over stoichiometry of reactions. Within cohesin-dockerin scaffolds, for example, repeated units of one cohesin in the scaffold will increase the amount of its complementary enzyme bound to the scaffold, therefore changing the ratio of each enzyme connected (You, Myung and Zhang, 2012). This can alter local enzyme concentration and reduce the impact of the rate limiting step, therefore lessening the amount of released intermediate (Chen and Silver, 2012).

Protein scaffolds are not always a fusion of multiple small binding proteins, however, and in some instances, scaffolds may be present as large macrostructures. A method for increasing yield of ethanol from pyruvate involved producing large fibrous scaffold proteins. PduA, a single shell protein associated with micro-compartments, was mutated to produce cell length hollow filaments. By introducing a heterodimeric coiled-coil domain to the C-terminus of the mutants, much shorter filaments are produced, ranging from around 150 nm to 300 nm, which are abundant within the cell. Fusion of the complementary coiled-coil to enzymes of the pyruvate dehydrogenase pathway then results in localization, increasing ethanol yield 2-fold (Lee *et al.*, 2017). Compared to the other scaffold systems, however, stoichiometry appears to be more random in this system, with both enzymes of the pyruvate dehydrogenase pathway using the same coiled-coil, meaning they would compete with one another for binding to the scaffold.

In a paper by Myhrvold *et al.*, a new scaffold assembly was produced to offer greater protection to protein components by embedding them within a lipid membrane. The scaffold system used protein machinery derived from bacteriophage $\phi 6$, expressing two intermembrane proteins, one essential for structure, and a second that could be attached to another protein by fusion. By fusing the second intermembrane protein to an enzyme, it is possible to localize the enzyme to the surface of an internal membrane envelope (Myhrvold, Polka and Silver, 2016). The use of a lipid envelope to generate a scaffold is a new technology, and for this reason must be evaluated for all potential applications. Attaching the enzyme to the large transmembrane protein could limit the sterics of the system, for example. Furthermore, the effect of oligomerization when attached to a transmembrane protein could be a concern, either bringing together multiple envelopes or being limited by the sterics of the transmembrane protein within the same envelope.

1.3.2. Nucleic acid scaffolds

While proteins are important in the production of artificial scaffold systems, there is an interest in moving beyond the native structures seen in cells and producing more complex scaffold systems which offer greater control over structure. Protein scaffolds can be limited in terms of malleability, a problem when distance between proteins or orientation must be altered. Of increasing interest is the application of nucleic acids, DNA and RNA, as scaffolds (Siu *et al.*, 2015). Nucleic acids have been used to produce complex nanostructures *in vitro*, such as DNA origami and linking duplex constructs (Fu *et al.*, 2012; Liu *et al.*, 2016). As nucleic acids can be used to create more physically complex structures compared to proteins, scaffolds with a greater range of applications could also be produced.

A simple method for using DNA as a scaffold is to localize enzymes to a plasmid. To localize pathway enzymes to a designed plasmid, a DNA binding domain is fused to each enzyme. The “Artificial DNA Binding Domains” are comprised of 4 zinc finger binding motifs. As each zinc finger motif interacts with a specific codon, 2.56×10^6 options are available for controlling the binding sequence (Fig. 1.8.A) (J. H. Lee *et al.*, 2013). DNA plasmid scaffolds have been used to increase the production of N-acetylglucosamine, fumarate and L-threonine (J. H. Lee *et al.*, 2013; Liu *et al.*, 2014; Chen, Zhu and Liu, 2016). Furthermore, expression of oligomeric proteins does not impact functionality of the scaffold complex.

Using plasmids as a scaffold offers greater control over enzyme placement when compared to protein scaffolds. This is because DNA is sequence independent, offering greater stability with respect to protein and RNA, reducing risk of aggregation (Conrado *et al.*, 2012). The plasmid can be easily modified, such as increasing or decreasing the number of base pairs between the targeted DNA to change the distance between binding sites or changing the stoichiometry of enzymes by manipulating the number of DNA binding sites. However, a potential flaw of this system is the size of the DNA binding domain. The DNA binding domain is over 120 amino acids in length, larger than the 65-70 amino acid dockerin protein, and therefore is more likely to impact the activity of the attached enzymes (Handelsman *et al.*, 2004; Lee, DeLoache and Dueber, 2012).

As well as using DNA binding domains to localize a protein to a scaffold, single stranded DNA (ssDNA) can be used directly. By fusing ssDNA to the terminus of an enzyme through enzymatic crosslinking, it is possible to bring it into association with a second enzyme fused to a complementary ssDNA strand. This system could be compared to that of coiled-coils, although the length of the DNA can be altered to change orientation and sterics

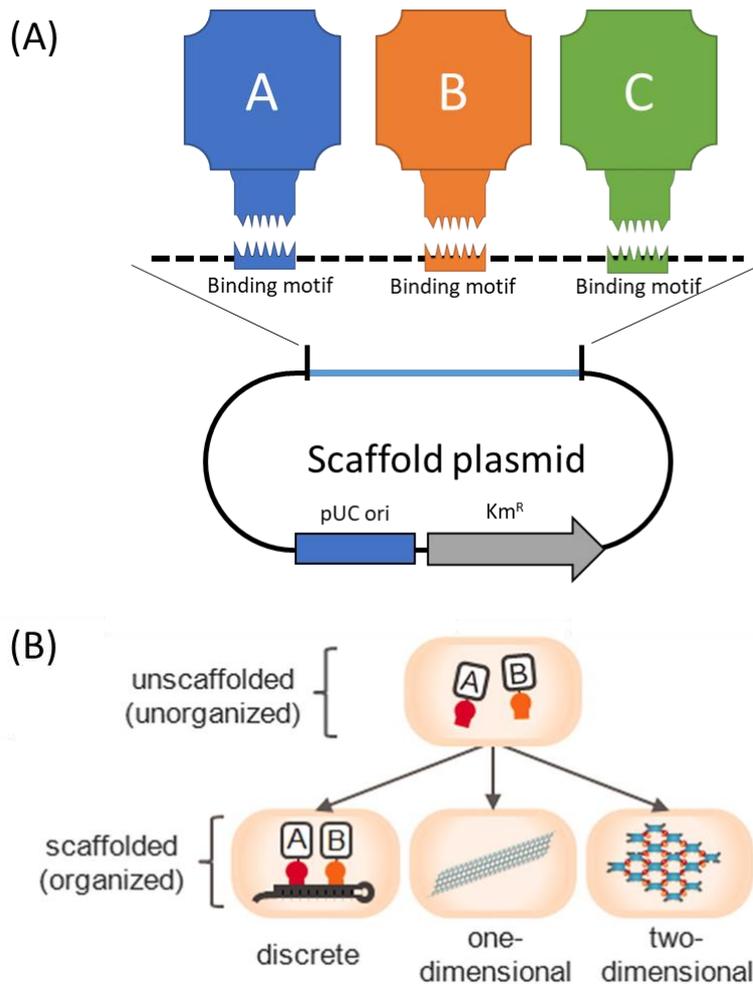


Figure 1.8. Schematics of the nucleic acid scaffold assemblies. Cartoons demonstrating the scaffold structures of DNA and RNA scaffolds **(A)** The plasmid DNA scaffold contains a coding region in which the binding sites for the Zinc-finger motifs are present. This region can be modified to alter the number of binding domains present and the distance between them. Figure derived from (J. H. Lee *et al.*, 2013) **(B)** Design of RNA modules to organise Proteins A and B. Proteins are either scaffolded onto discrete, 1D, and 2D RNA assemblies by interaction of RNA aptamers with protein adapters. In this image the 2D assembly is represented by a sheet, however this scaffold can also be assembled into a nanotube or other structures. Figure obtained from (Delebecque *et al.*, 2011).

(Wang *et al.*, 2015). Larger ssDNA strands have been used to produce DNA origami scaffolds, more complex 2D structures derived from the folding of the long ssDNA (Rothemund, 2006; Idan and Hess, 2013), which has enhanced the activity of glucose oxidase/horseradish peroxidase enzymes (Fu *et al.*, 2012). Within the scaffold, distance between the strands can be altered in order to change the distance between enzymes from 10 to 65 nm (Fu *et al.*, 2012).

The use of a DNA scaffold can be limited *in vivo*. A reduced activity seen when using DNA plasmids as scaffolds within a cell, compared to *in vitro* expression was believed to be the result of a limited plasmid copy number (Lee, DeLoache and Dueber, 2012). Furthermore, the expression of a ssDNA fused to an enzyme *in vivo* for DNA origami structures is a significant challenge, as it is achieved *in vitro* by chemical modification (Siu *et al.*, 2015). A final limitation is found in eukaryotic cells, where DNA scaffolds may be transferred to the nucleus; making DNA scaffolds unsuited for expression in these cell lines (Smolke, 2018).

Like DNA, it is hypothesised that RNA molecules could act as a scaffold and could allow production of highly complex scaffold structures. RNA scaffolds are based on the use of aptamers, regions of single stranded RNA (ssRNA) which select for specific peptide sequences (Delebecque *et al.*, 2011; Germer, Leonard and Zhang, 2013). The RNA aptamer itself contains both 2 polymerization and 2 dimerization domains. The dimerization domains facilitate interaction with other molecules of ssRNA, while the polymerization domain prevents the collapse of the ssRNA molecule (Delebecque, Silver and Lindner, 2012). Notably, the bound peptide sequences are smaller than the DNA binding domain, meaning that the target enzyme will require only small modifications to achieve RNA contact.

The RNA-aptamer scaffolds function in one of three forms: a series of distinct RNA monomers, a polymeric structure produced from one dimensional linear overlaps of RNA, or a complex two-dimensional structure where RNA overlaps in both horizontal and vertical directions (Fig. 1.8.B). The proteins then bind to the surface of the scaffold, with scaffolds of greater complexity resulting in increased substrate channelling (Delebecque *et al.*, 2011). The 2D nanostructure generated by RNA scaffolds have increased hydrogen production by up to 48-fold, and increase pentadecane production by up to 1.4-fold, both in *E. coli* (Delebecque *et al.*, 2011; Sachdeva *et al.*, 2014). By using FRET, it was demonstrated that the distance of enzymes bound to either the 1D or 2D scaffolds is almost identical, yet nevertheless the 2D scaffold caused increased production of product (Delebecque *et al.*, 2011). This could be explained by the greater concentration of enzymes in the immediate vicinity. Conversely the use of discrete RNA scaffolds resulted only in minimal changes in the production of the target

molecule, even though a scaffold was present (Siu *et al.*, 2015). However, the complex RNA scaffolds do contain the risk of aggregating beyond desired levels (Conrado *et al.*, 2012).

1.3.3. The dynamics behind substrate channelling in scaffolds

The role of scaffolds in increasing substrate channelling is often debated; however, there are three key theories: reduced distance between enzymes increasing channelling of intermediates between active sites, the aggregation of scaffolds due to polymeric enzymes increasing local enzyme concentration, or the scaffold components stabilizing the pathway enzymes to increase activity (Fu *et al.*, 2012; Lee, DeLoache and Dueber, 2012; Zhang, Tsitkov and Hess, 2016). Of these theories, Brownian motion models have suggested that any distance greater than 1 nm would not result in an increased enzyme activity. As scaffolds produce greater distances than 1 nm between enzymes, it is likely that scaffold aggregation or enzyme stabilization causes improved yields in the synthetic pathway (Fu *et al.*, 2012).

It was hypothesised by Lee *et al.* that, as with 2D RNA aptamers, all scaffolds increase the rate of reaction by increasing local enzyme concentration. Substrate channelling is achieved across not one scaffold-bound complex, but many, interacting as a result of the oligomeric state of the bound enzymes. For a dimeric enzyme, one scaffold-bound enzyme monomer forms protein-protein interactions with a sister monomer on a separate scaffold, causing both scaffolds to associate with the enzyme. The scaffolds are therefore brought into close proximity with one another, and will increase local downstream enzyme concentration (Lee, DeLoache and Dueber, 2012).

On the other hand, Zhang *et al.* suggested that scaffolds can also increase yields by stabilizing the attached enzymes. Using monomeric enzymes of the glucose oxidase-horseradish peroxidase system, it was shown that a small distance generated by protein fusion did not result in the improved yields comparatively measured in DNA origami scaffolds. Instead, it was hypothesized that the negative charge of the DNA molecules resulted in an increased acidity at the surface of the scaffold, in turn stabilizing the enzymes (Zhang, Tsitkov and Hess, 2016). Similarly, fusion of an affibody to proteins involved in PHB production resulted in a 3-fold increase in yield, even in absence of a scaffold, although yield was improved further when a scaffold was present (Tippmann *et al.*, 2016).

It is also important to note that the enzymes selected can play a role in the activity of the scaffold. When a cohesin-dockerin system was used to enhance fructose-6-phosphate production, initial rates of reaction were as high as 13.4-fold greater than the free enzyme control (You, Myung and Zhang, 2012). However, it was found that exchanging a high activity variant of an enzyme with a lower activity caused a 5 fold reduction in the degree of substrate

channelling, a calculation of initial rate of scaffold-bound system vs non-scaffold bound system (You and Zhang, 2014). This either suggests that the scaffold may stabilize the different enzymes to varying degrees, or that substrate channelling faces a cap based on the activity of the rate-limiting step.

1.4 Artificial Substrate Channelling Approaches

Synthetic protein fusion can facilitate the channelling of metabolites from two enzymes in order to enhance product yield (Lewicka *et al.*, 2014; Xu *et al.*, 2018). Protein fusion has been used for a broad range of applications; however, the use of it to improve activity of reactions has been less researched (Iturrate *et al.*, 2009). While it can be effective, it is important to note that linker variation should be considered (Chen, Zaro and Shen, 2013; Chen *et al.*, 2017), alongside the fact that for oligomeric enzymes, there is a greater risk of misfolding (Conrado *et al.*, 2012; Lee, DeLoache and Dueber, 2012).

Artificial scaffolds are an important tool in accelerating pathways generated by synthetic chemistry, which may not be optimal within the host organism (Lee, DeLoache and Dueber, 2012; Siu *et al.*, 2015). Unlike fusion proteins, these systems are considered less likely to cause changes to protein folding or steric strain (Tippmann *et al.*, 2016). The use of enzyme scaffolds can enhance substrate channelling through multienzyme reactions, in turn resulting in an improved yield, as high as 77-fold higher for the mevalonate pathway (Dueber *et al.*, 2009). Either protein or nucleic acids may be used as a scaffold, localizing enzymes by suitable peptide tags or binding sites (J. H. Lee *et al.*, 2013; Kim and Hahn, 2014; Siu *et al.*, 2015). Protein scaffolds have been used for many reactions, exploiting strong protein-protein interactions to achieve localization (Dueber *et al.*, 2009; Kim and Hahn, 2014; Tippmann *et al.*, 2016); however, they are susceptible to issues with respect to controlling the spacing between enzymes, requiring the testing of a number of different scaffolds to find the optimal conditions (Siu *et al.*, 2015). On the other hand, nucleic acid scaffolds can facilitate greater complexity, producing 2-dimensional structures with control over enzyme spacing and orientation (Fu *et al.*, 2012; Sachdeva *et al.*, 2014); however, they may introduce changes to the local chemical environment (Zhang, Tsitkov and Hess, 2016) and could be susceptible to nucleases in industrial *in vitro* conditions (Quin *et al.*, 2017).

In order to test substrate channelling at the membrane, use of both fusion and scaffolding should be considered. Examples of natural protein fusion between transporters and downstream enzymes, while rare, have been recorded (Willson, Chapman and Thomas, 2019). Comparatively, scaffold proteins reduce any risks of steric hindrance; however, considerations about the function of the scaffolds and the impact of the tags should be considered (Dueber *et al.*, 2009; Sachdeva *et al.*, 2014; Tippmann *et al.*, 2016). Finally, it is important to consider the impact that oligomeric states of downstream enzymes may have on the success of substrate channelling (Lee, DeLoache and Dueber, 2012).

1.5. The Importance of Xylose Metabolism for Industrial Biotechnology

In order to demonstrate the potential of metabolic channelling at the membrane, it is important to study a system with academic and/or industrial relevance. One of the most important systems in the public eye at this time is the metabolism of sugars, particularly pentose sugars, due to the implication in industrial fermentation. Through the engineering and optimization of microorganisms capable of metabolising such sugars, biotechnology can facilitate production of industrially relevant biochemicals from an increasing diversity of renewable feedstocks (Table 1.1) (Hassan, Williams and Jaiswal, 2019; Singh *et al.*, 2020).

The metabolism of free sugars to useful products is the foundation of industrial fermentation (Naik *et al.*, 2010; Lee and Lavoie, 2013; Singh *et al.*, 2017). Biofuels are a key industrial fermentation product, with the most common type seen being first generation biofuels, produced from free glucose derived from starch-rich agricultural crops (Guo, Song and Buhain, 2015). The use of agricultural crops as a source of biofuels is a source of ethical concern, risking food shortage due to either limited crop amounts or limited space for growing food crops (Vertès, Inui and Yukawa, 2006; Boretti, 2012). Second generation biofuels are derived from lignocellulosic feedstocks, such as waste crop material, and are rich in a range of sugars with respect to primary biofuels (Rodionova *et al.*, 2017). In addition to the high volume biofuel products, lignocellulose may also be used to produce high value organic compounds such as pharmaceuticals (Isikgor and Becer, 2015; Wijaya, Ismadji and Gunawan, 2021), feed ingredients (Jazini, Fereydouni and Karimi, 2017), organic acids (Francois, Alkim and Morin, 2020), biosurfactants, biolubricants, and bioplastics (Kawaguchi *et al.*, 2016). The combination of both high volume and high value products means that improving lignocellulose fermentation could aid in reaching UN sustainability goals (Hassan, Williams and Jaiswal, 2019).

Lignocellulosic feedstocks are largely inedible, and include grasses, jatropha and camelina (Lin and Lu, 2021). Lignocellulose makes up the plant cell wall in these crops, and is a polysaccharide composed of three separate polymers – lignin, cellulose and hemicellulose (Williams *et al.*, 2016). Hemicellulose makes up 30% of the lignocellulose composition, and while cellulose is a glucose rich polysaccharide linked by β -1,4 glycosidic bond, hemicellulose contains a mix of sugars including D-glucose, D-galactose, D-mannose, L-arabinose and D-xylose, with various types of linkage between them. Hemicellulose itself describes the range of non-cellulosic polysaccharides and can therefore be categorised as a combination of smaller polymers such as heteromannans, xyloglucan, heteroxylans, and mixed-linkage glucans (Pauly *et al.*, 2013).

Product	Reference
Bioethanol	(Huffer <i>et al.</i> , 2012)
Biodiesel	(Uthandi <i>et al.</i> , 2021)
Succinate	(Jo <i>et al.</i> , 2017)
Lactate	(Boer <i>et al.</i> , 2019)
Ethylene glycol	(Boer <i>et al.</i> , 2019)
Glycolic acid	(Boer <i>et al.</i> , 2019)
Citric acid	(Jagtap and Rao, 2018)
Single Cell Oil (Biolubricant)	(Bandhu <i>et al.</i> , 2018)
Xanthan (food thickening agent)	(Zhang and Chen, 2010)
Sophorolipids (Biosurfactant)	(Marcelino <i>et al.</i> , 2019)
Xylitol (Food and pharmaceutical use)	(Martins <i>et al.</i> , 2018)
Non-biodegradable bioplastics	(Rahman and Bhoi, 2021)
Isoprenoids (Cosmetic, Pharmaceutical and food)	(Jagtap and Rao, 2018)

Table 1.1. List of products of D-xylose fermentation. A list of various products produced from fermentation of D-xylose either derived from lignocellulosic biomass or in lab conditions using pure D-xylose – These products can be produced directly through fermentation, or may require further chemical reactions following production of a suitable intermediate.

The biological degradation of hemicellulose therefore completed by a diverse variety of enzymes to overcome the similar variety of the hemicellulose polysaccharides (Sánchez, 2009). This is clearly seen for the most abundant hemicellulose polymers, the xylan polysaccharides. Xylan is a heteropolysaccharide with homopolymeric backbone chains of 1,4-linked β -D-xylopyranose units. The xylans may then also contain branches containing a mix of xylose, arabinose, glucuronic acid or its 4-O-methyl ether, and acetic acid, ferulic acid, and p-coumaric acids. Therefore, to achieve degradation of xylan, a mix of enzymes are required. Endo- and exo-xylanases β -1,4-xylose linkages releasing smaller xylooligosaccharides or disaccharide xylobiose respectively, which is then degraded by xylosidase enzymes. Conversely, where side chains are present other enzymes such as α -arabinofuranosidase, which releases terminal nonreducing α -arabinofuranose from arabinoxylans, must be present to release all the sugars present (Fig. 1.9.A) (Saha, 2003).

Metabolism of these sugars can be achieved within microorganisms, taking advantage of a host of enzymatic pathways (Gronenberg, Marcheschi and Liao, 2013). The model organism *Escherichia coli* is one such species (Fig. 1.9.B and C) (Inokuma *et al.*, 2010; Bond-Watts, Bellerose and Chang, 2011), containing multiple pathways for sugar metabolism, focused around monosaccharides and some disaccharides (Ammar, Wang and Rao, 2018), with ethanologenic strains capable of converting the metabolised sugars into ethanol (Ingram *et al.*, 1987; Zaldivar, Martinez and Ingram, 1999).

High activity xylose metabolism is vital from an industrial standpoint in order to make cellulosic ethanol commercially feasible (Hahn-Hägerdal *et al.*, 2007; Jojima *et al.*, 2010). It is a major component of hemicellulose (Saha, 2003); however, few organisms are able to metabolise xylose to ethanol at yields suitable for the biofuel industry (Teunissen *et al.*, 1991; Dien, Cotta and Jeffries, 2003). Increasing the metabolism of xylose is therefore a key target for synthetic biology (Lee *et al.*, 2008; Kang *et al.*, 2014; Cam *et al.*, 2016).

Organisms capable of metabolising xylose and other lignocellulosic sugars to useful organic products such as ethanol, methane or butanol have been obtained from nature or produced by genetic modification, but can face challenges in industrial use (Bhatia *et al.*, 2017). Suboptimal yields, productivity and/or titre are common with these organisms (Sims *et al.*, 2010; Carriquiry, Du and Timilsina, 2011; Gustavsson and Lee, 2016), as enzyme activity may be poor (Sarthi *et al.*, 1987), substrates and intermediates may be sequestered by side reactions (Kötter and Ciriacy, 1993; Yazdani and Gonzalez, 2007), products or contaminants may be toxic to the cell (Oliva *et al.*, 2006), or scale up may be inaccessible (Chandel *et al.*, 2018). Such considerations necessitate further modification of the organism or enzyme pathways.

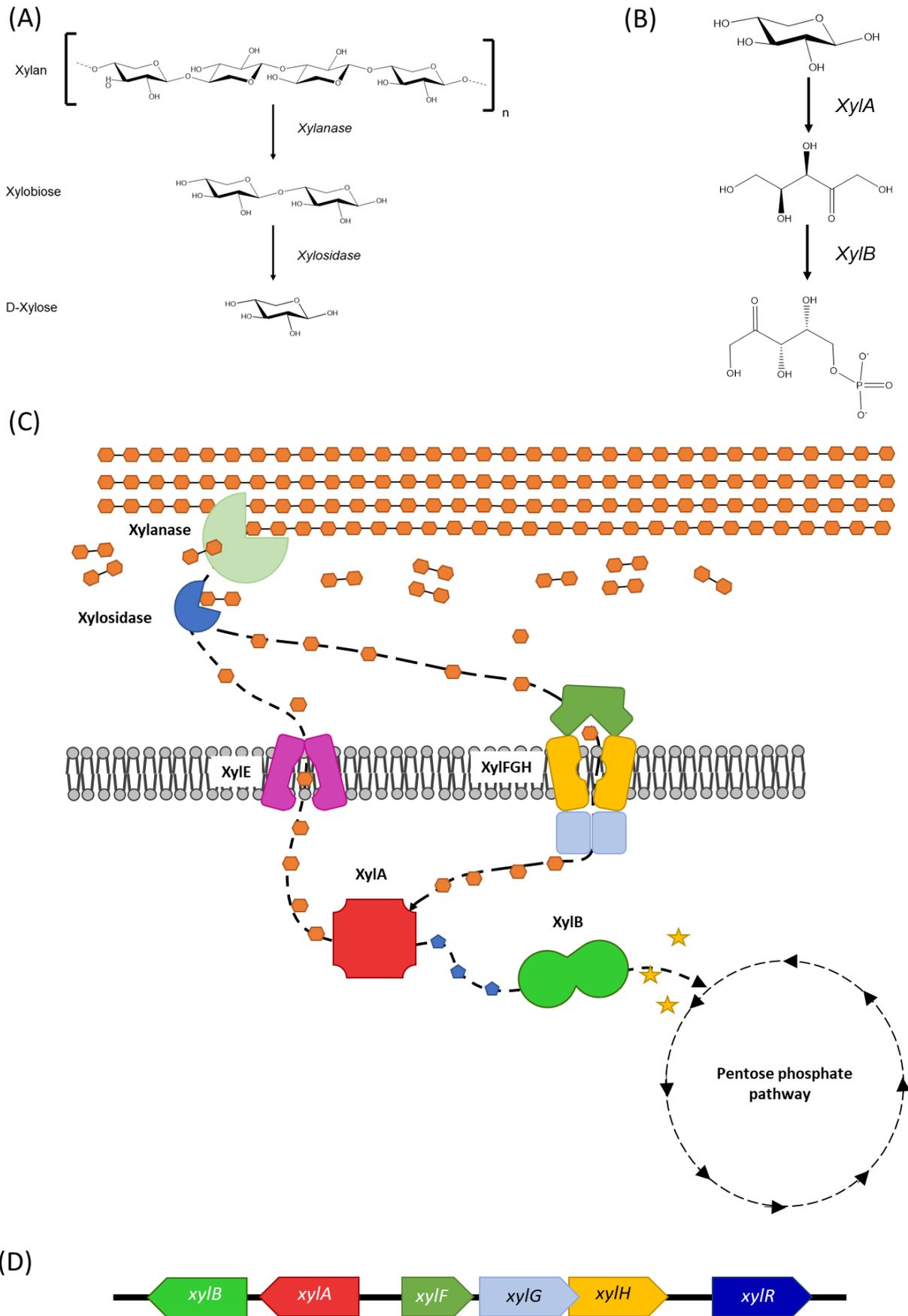


Figure 1.9. Microbial methods of xylose metabolism from hemicellulose (A) The chemical structures of the xylan to xylose degradation pathway - The enzymatic hydrolysis of xylan is achieved through the degradation of Xylan to xylobiose, catalysed by the exo- and endo-xylanase, and xylobiose to D-xylose by the xylosidase enzymes. **(B)** The two-step conversion of D-xylose to D-xylulose-5-phosphate catalysed by *E. coli* XylA and XylB enzymes – The isomerisation of D-xylose to D-xylulose is catalysed by XylA and phosphorylation of D-xylulose catalysed by XylB. **(C)** A schematic of the xylose utilization in *E. coli*, converting hemicellulose polysaccharide xylan to xylose and feeding into the pentose phosphate pathway. D-xylose is represented as the orange hexagons, D-xylulose is represented by the blue pentagons and D-xylulose-5-phosphate is represented by the yellow stars. The movement of the substrate and products is shown by the dashed line. D-xylose can enter the cell either through the action of MFS XylE or ABC transporter XylFGH. *E. coli* cannot directly utilize xylan (Zheng *et al.*, 2012), therefore the enzymes displayed here would either be from an industrial enzyme broth or organism capable to producing such enzymes in nature. **(D)** The xyl operon, activated by XylR – figure derived from (Song and Park, 1997)

This work aims to target enhanced substrate channelling within *E. coli*. As previously mentioned, strains of *E. coli* have been produced that are capable of metabolising xylose to ethanol (Ingram *et al.*, 1987; Zaldivar, Martinez and Ingram, 1999), and it is a target organism for industrial biofuel production (Koppolu and Vasigala, 2016).

Demonstrating enhanced substrate channelling in *E. coli* could also act as a model for other organisms. The enzymes of the prokaryotic xylose utilization pathway have been used to facilitate xylose metabolism in other organisms (Moes, Pretorius and Van Zyl, 1996; Walfridsson *et al.*, 1996; Voronovsky *et al.*, 2005); however, rates are below industrial standards, necessitating further refinement (Liu *et al.*, 2017; Feng *et al.*, 2018; Guamán *et al.*, 2018).

An important element to take into account when metabolising sugars is that *E. coli* relies on a hierarchy of sugar selection when growing, so to optimize growth by selecting the most energy rapidly metabolised sugars available (Desai and Rao, 2010). A notable example of this carbon catabolite repression is the glucose-lactose diauxic shift, in which bacteria in a mixed solution of glucose and lactose will first utilize the glucose alone, only switching to lactose once all of the other sugar is exhausted (Loomis and Magasanik, 1967). For non-glucose sugars, *E. coli* has a hierarchy of lactose > arabinose > xylose > sorbitol > rhamnose > ribose. The hierarchy is maintained by the global regulator CRP-cAMP, which activates each promoter at a different level (Aidelberg *et al.*, 2014).

The xylose metabolism pathway is controlled by the activator protein XylR. In absence of glucose and L-arabinose, XylR will activate the expression of genes encoding the high affinity ABC transporter XylFGH, alongside xylose metabolism proteins XylA and XylB (Ni *et al.*, 2013). Additionally, xylose is transported into the *E. coli* cell by a xylose/proton symporter, XylE, which is a transporter with lower affinity than XylFGH and is not part of the *xylAB* operon (Fig. 1.9.D), although still predicted to be regulated by XylR (Song and Park, 1997). In addition, the AraE and AraFGH, the L-arabinose transporters, are associated with strong transport of D-xylose and demonstrates control over the *xyl* operon (Desai and Rao, 2010).

Following the transport of xylose across the membrane to enter the cytoplasm, the sugar is metabolised through conversion to xylulose-5-phosphate (Vieille *et al.*, 1995). Xylulose-5-phosphate is subsequently catalysed within the pentose phosphate pathway to produce glyceraldehyde-3-phosphate (Jeppsson *et al.*, 2002; Liu *et al.*, 2018), which can then be converted into useful products.

1.5.1 Xylose/H⁺ symporter

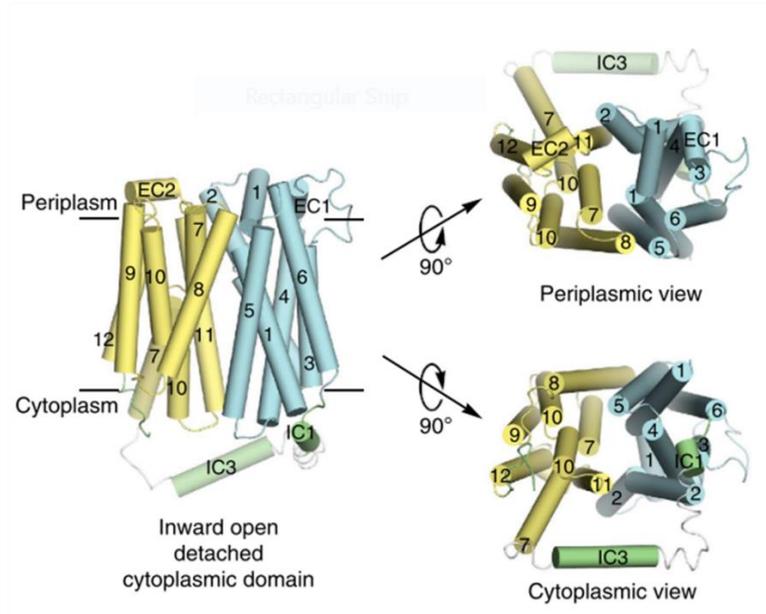
E. coli XylE is a 491 amino acid XylE/H⁺ symporter and a member of the MFS superfamily, demonstrating strong homology to the human glucose transporter GLUT1, another MFS transporter (Henderson, 1990; Ke *et al.*, 2017). While XylFGH is the preferred D-xylose transporter under low xylose conditions (Koirala, Wang and Rao, 2016), there is evidence that under high D-xylose concentration the MFS is more important to the cell than the ABC transporter (unpublished data – personal communication from Henrique Neves). MFS transporters may act as substrate/proton symporters, antiporters or uniporters (Reddy *et al.*, 2012) and transport a broad range of ligands across the membrane, including monosaccharides, organic acids, peptides, drugs, and inorganic ions (Lee, Sands and Biggin, 2016). With this diverse range of function, proteins of the MFS superfamily make up 25% of all transporter proteins (Saier *et al.*, 1999), containing 105 transporter families (Wang *et al.*, 2020).

As previously discussed and described later in this thesis, MFS transporters have shown a history of C-terminal fusion to soluble domains (Willson *et al.*, 2019). Furthermore, the most abundant MFS fusion to a soluble domain is a functional flippase fusion (Harvat *et al.*, 2005) – an example of the sparse few membrane protein to enzyme fusions (Willson, Chapman and Thomas, 2019). This suggests that XylE, an MFS transporter, may be better suited to study for substrate channelling in comparison to the XylFGH system.

XylE, as with most members of the MFS superfamily, is characterized by 12 transmembrane helices (TMH), separated into two halves, the N-terminal (covering TMH 1-6) and the C-terminal (TMH 7-12), with both termini present in the cytoplasm (Fig. 1.10.A) (Xuejun C. Zhang *et al.*, 2015). Furthermore, these helices form four pseudo-symmetrical three-helix repeats, each composed of the cavity helix (TMH 1,4,7 and 10), rocker helix (TMH 2, 5, 8 and 11) and the support helix (TMH 3, 6,9 and 12). The cavity helix forms the cavity in which xylose sits, whereas the rocker helix forms long curved structures that act as the side-walls of the cavity and interact with xylose itself. Finally, the support helices do not act on the cavity, but instead form contacts with the lipid bilayer, and so are involved with the hydrophobic mismatch within the cytoplasmic membrane (Xuejun C Zhang *et al.*, 2015).

Transport of xylose is achieved through a rocker switch mechanism, in which C- and N-domains maintain structures, but their relative positions change (Xuejun C. Zhang *et al.*, 2015). In the case of XylE, an asymmetric conformational change is measured, in which the N-domain movement is minimal (root-mean square deviation (rmsd) of 1.1 Å) whereas the

(A)



(B)

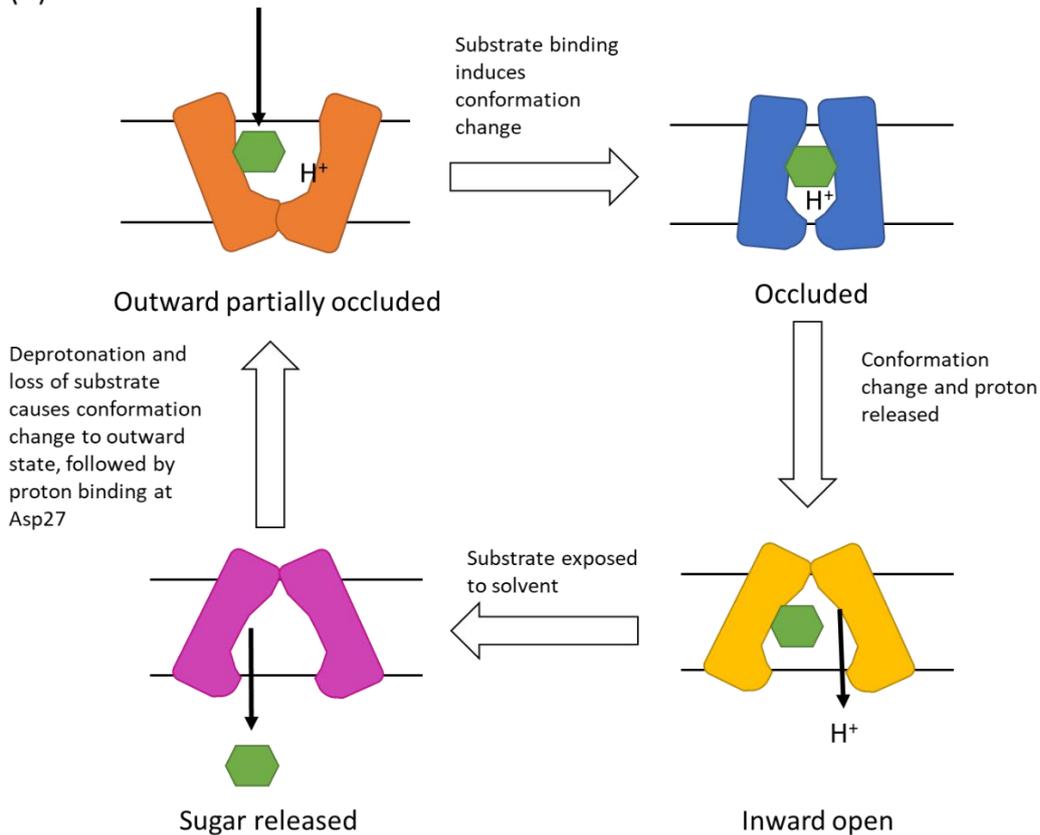


Figure 1.10. Schematic of the rocker switch motif of XylE – (A) the crystal structure of *E. coli* XylA in the inward open state. Crystal structure directly lifted from (Wisedchaisri *et al.*, 2014). **(B)** A cartoon representing the rocker switch motif of the D-xylose/proton symporter facilitates transport of both D-xylose and a proton from the periplasm to the cytoplasm.

Cartoon derived from (Du *et al.*, 2015), adapted towards the structural features discussed in (Wisedchaisri *et al.*, 2014).

C-terminal movement is more significant (rmsd of 2.9 Å). In particular, the helix-turn-helix structures of TMH 7-8, TMH 10-11 and TMG 11-12 have the most significant conformational changes, of greater than 3 Å (Wisedchaisri *et al.*, 2014).

The rocker switch mechanism requires three states of the transporter: an outward, partially occluded conformation, an intermediate occluded conformation, and an inward open conformation. The ground state is the outward facing state, in which the protein cavity is open to the periplasm (Fig. 1.10.B) (Wisedchaisri *et al.*, 2014; Du *et al.*, 2015). Xylose is bound to the cavity by 8 hydrogen bonds, while the proton binds to Asp27. The Asp27 protonation triggers a conformational change which causes Tyr298 to act as a periplasmic gate, preventing release of bound xylose. In the intermediate state, xylose is blocked by both Tyr298 and Trp392, the cytoplasmic gate which is one of the hydrogen-bond forming amino acids. As the conformational change is complete, Trp392 swings 6 Å away from adjacent Gln168 and creates a channel to the centre of the cavity, exposing the sugar to the cytoplasm, the inward facing motif. Once the sugar and proton are released from the xylose, the transporter returns to its ground state (Wisedchaisri *et al.*, 2014).

1.5.2 Xylose isomerase

Converting D-xylose to xylulose-5-phosphate is considered the rate limiting step within the reaction pathway in multiple organisms (Di Luccio *et al.*, 2007; Agrawal, Mao and Chen, 2011; Xin, Wu and He, 2014; Qi *et al.*, 2015), although there is some debate about the impact of rate limiting steps, with alternative explanations relating to flux control through a pathway used to explain improved product yields (Fell, 1998). The D-xylose to xylulose-5-phosphate reaction can be achieved through one of two different methods. *E. coli* and other prokaryotic organisms exploit two key enzymes, xylose isomerase (XI), which converts xylose to xylulose, and xylulose kinase (XK), which phosphorylates the xylulose (David and Weismeyer, 1970; Wilhelm and Hollenberg, 1985). Yeasts, such as *Hansenula polymorpha* and *Saccharomyces cerevisiae*, lack XI and must instead use two other enzymes to produce xylulose, xylose reductase (XR) and xylitol dehydrogenase (XDH). For the majority of XR-XDH systems, however, a redox imbalance between NADH and NADPH cofactors causes reduced ethanol yields and increased xylitol production (Nidetzky *et al.*, 2003; Voronovsky *et al.*, 2005; Qi *et al.*, 2015). As a result of this imbalance, the xylose isomerase pathway is preferable for metabolic engineering of xylose catabolism (Kuyper *et al.*, 2005; Hou *et al.*, 2016; Cunha *et al.*, 2019).

The xylose isomerase enzymes are a family of metalloenzymes containing two divalent metal ions, which can either be Mg^{2+} , Mn^{2+} or Co^{2+} , and can be characterized as one of two classes, depending upon the number of amino acids present in the monomer (Van Bastelaere *et al.*, 1992). Class I enzymes are the smaller of the two, with a size of around 390 amino acids, whereas class II XI contain a 50 amino acid N-terminal insert, resulting in a size of ~440 amino acids (Bhosale, Rao and Deshpande, 1996; Hartley *et al.*, 2000).

Of the two xylose isomerase groups, class I enzymes have been better studied, with application in industry; they mostly originate from mesophilic organisms (Epting *et al.*, 2005; Fatima and Javed, 2020). The 3D-structures of class I XI show great similarity to one another and correspondingly, the biochemical and thermostability of the class I XI are also similar (Hartley *et al.*, 2000). Conversely, class II XI have had less application in industry, but originate from a more diverse range of organisms, including mesophilic, thermophilic and hyperthermophilic bacteria (Vieille *et al.*, 1995; Epting *et al.*, 2005; Peng *et al.*, 2015; Fatima, Aftab and Haq, 2016). This in turn results in significant differences in the biochemical and thermostability of the class II enzymes (Hartley *et al.*, 2000; Fatima, Aftab and Haq, 2016).

The *E. coli* XylA is a 440 amino acid metalloenzyme and with a mass of 49.742 KDa (Schellenberg *et al.*, 1984; Tucker *et al.*, 1988). The size and sequence of the protein identifies

XylA as a member of the somewhat less-well studied class II XI (Epting *et al.*, 2005; Fatima and Javed, 2020). Knowledge of the structure of XylA is limited due to a lack of a crystal structure, however it is believed contain Mg^{2+} within the divalent core (Farber *et al.*, 1989; Black, Huang and Cowan, 1994) with experimental evidence showing the importance of two histidine side chains (Batt, Jamieson and Vandeyar, 1990).

1.5.3 Xylulose kinase

Following the isomerisation of D-xylose to D-xylulose, the sugar molecule must then be phosphorylated in order to be incorporated into the pentose phosphate pathway. In *E. coli*, the protein xylulose kinase (XylB) is used to achieve ATP-dependent phosphorylation (Wungsintaweeikul *et al.*, 2001; Tritsch *et al.*, 2004; Meijnen, De Winde and Ruijssenaars, 2008). XK can also be found within other bacteria species (Neuberger, Hartley and Walker, 1981; Ahmad and Scopes, 2002), as well as yeast (Rodriguez-Peña *et al.*, 1998; Jin *et al.*, 2002) and higher eukaryotes (Hemmerlin *et al.*, 2006; Bunker *et al.*, 2013).

E. coli XylB exists as a homodimeric protein, in which each monomer has a size of 484 amino acids and a mass of 53 KDa and requires Mg^{2+} to function (Di Luccio *et al.*, 2007). The enzyme has been shown to be somewhat promiscuous for 5-carbon sugars, capable of phosphorylating D-1-deoxy-xylulose, D-ribulose, D-arabitol and D-xylitol, although at much slower rates than xylulose (Wungsintaweeikul *et al.*, 2001; Di Luccio *et al.*, 2007).

Structural studies have determined the XylB monomer is comprised of two domains, the first domain acting as a sugar binding region and the second domain containing the majority of ATP binding motifs. A recognisable ATPase fingerprint region, required for binding and hydrolysis of the ATP cofactor can be seen within this domain. The two domains are linked by a hinge segment which is responsible for moving the first domain into close proximity with the ATPase domain upon sugar binding (Di Luccio *et al.*, 2007). Within these studies, the measured K_m for D-xylulose was 0.29 mM, with a k_{cat} of 244 s^{-1} .

1.6. Golden Gate Assembly as a Means for Producing Protein Fusion Libraries

Taking into account the need to produce both direct fusion between XylE and XylA/B and scaffolds, involving fusion between the xylose utilization pathway components and scaffold molecules, it will be important to account for changing both components of the fusion and the short peptide linkers to stabilize fusion. A potential limitation for producing functional protein fusions is the ability to introduce different linker peptide sequences to join the proteins of interest, while maintaining functionality and expression of the proteins (Chen, Zaro and Shen, 2013; Schuurman-Wolters *et al.*, 2018). Linkers with different properties can have differing impacts of the functionality of the fusion (Zhao *et al.*, 2008; Li *et al.*, 2016), consequently a method to easily alter the linkers used in the fusion of a sugar transporter to its downstream partner, be that enzyme or peptide tag, is needed.

The use of defined regions of DNA to engineer novel proteins is a key focus of synthetic biology, aiming for a rational design process that can predictably modify biological approaches (Flores Bueso and Tangney, 2017). A synthetic DNA library can be generated from multiple DNA fragments, each with with defined sequence variability (Oldham, Hall and Burton, 2012; Kosuri and Church, 2014), such that when they are assembled the impact of sampling a large space in the genetic diversity on the resulting protein can be tested (Harikumar *et al.*, 2017; Zhao *et al.*, 2019). In the context of protein fusions, this can include altering the types of proteins used in the fusion, such as by substituting homologous proteins from the same or different organisms, in order to maximise activity of the reaction (Yu *et al.*, 2015), introducing transgenic components, or altering the linkers used to increase fusion stability and reduce steric strain (Li *et al.*, 2016).

In order to facilitate production of libraries of fusions, differing only in linker structure, it is important to use a system for assembling multiple DNA fragments in a rapid and replicable method. Traditional cloning methods would necessitate the production of a plasmid with multiple restriction enzyme digest sites and would therefore require multiple steps of enzyme-facilitated digest, gel extraction and enzyme-facilitated ligation to achieve the final product (Marillonnet and Grützner, 2020). For this purpose, Golden Gate assembly was considered. Golden Gate assembly is a cloning method in which type IIS restriction enzymes, such as BsaI or BpiI, are used in order to cut DNA outside their recognition sites (Fig. 1.11.A), creating a cut region with a defined overhang (Engler *et al.*, 2009). By complementing the overhang with a matching overhang from a different DNA sample that was also cut, it is possible to ligate multiple DNA fragments together using a DNA ligase enzyme (Fig. 1.11.B).

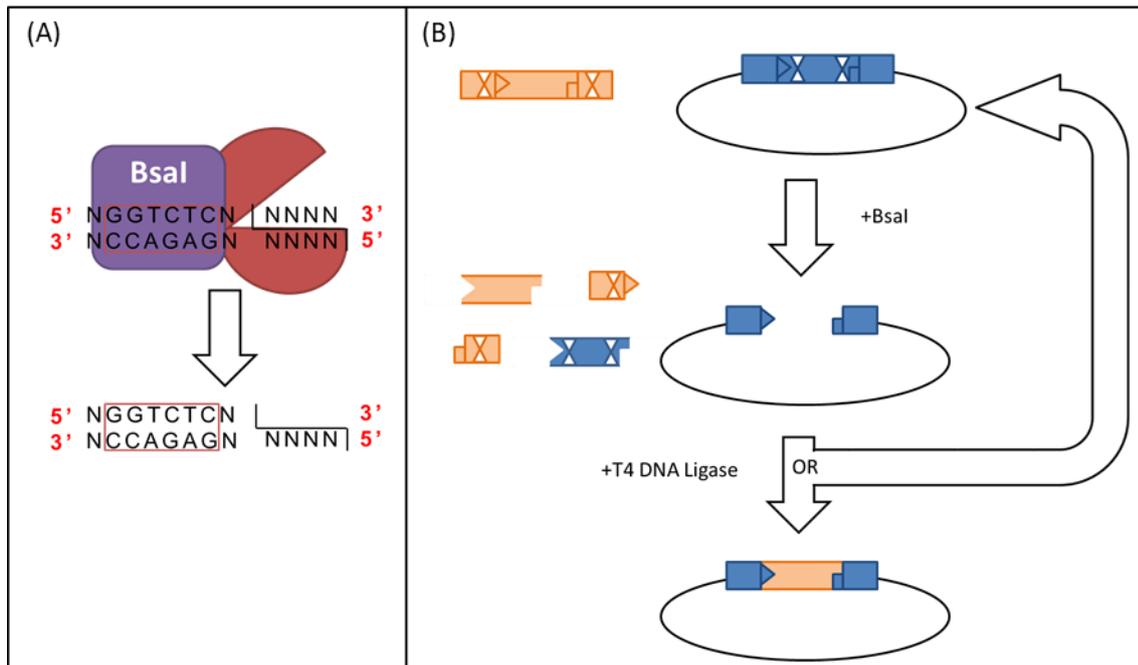


Figure 1.11. Schematic diagrams of elements of Golden Gate Cloning (A) The method of Type IIS restriction enzyme digest used by Golden Gate to create 4 bp overhangs. The purple rectangle represents the binding site of restriction digest enzyme BsaI, and the red semicircles represent the cleaving region of BsaI, which cuts downstream of binding site. **(B)** BsaI cuts a linear and circular molecule of DNA. Both the linear and circular DNA are shown in the uncut form in the first step, with BsaI binding sites represented as the colourless triangles. Introduction of BsaI results in the production of multiple DNA fragments in the second step with different pairs of complementary overhangs. T4 DNA Ligase will then ligate complimentary fragments, either producing the desired product, a circular DNA molecule containing the gene of interest from the linear DNA, which lacks the BsaI binding sites, or reproducing the original DNA molecules of stage 1, which can be cut again.

Using a series of predefined overhangs, Iverson *et al.* demonstrated the ability to assemble 6 fragments of DNA within a holding plasmid (Iverson *et al.*, 2016). Golden Gate assembly is a one-pot reaction protocol (Engler *et al.*, 2009), that has been used to produce phage display Fab libraries for antibody production (Chockalingam *et al.*, 2020), test the functionality of various antibiotic resistance markers and expression genes in newly isolated bacteria (Leonard *et al.*, 2018) and develop a more robust 5-hydroxymethylfurfural oxidase through gene shuffling of beneficial mutants (Martin *et al.*, 2018). An alternative method considered was Gibson assembly, which uses overlapping regions of linear DNA to achieve seamless cloning (Gibson *et al.*, 2009), however this methodology requires an additional stage of linearization for each component, is more cost expensive due to the enzymes associated and is not by nature modular (Terfrüchte *et al.*, 2014; Chiasson *et al.*, 2019).

Following the Iverson protocol, termed the CIDAR MoClo method (Fig. 1.12), this work used Golden Gate cloning in order to assemble between five and six fragments, facilitating production of an expression system of a fusion protein with or without a polyHis-tag, within suitable holding vectors, encoding libraries of novel protein fusions.

CIDAR MoClo Assembly Format

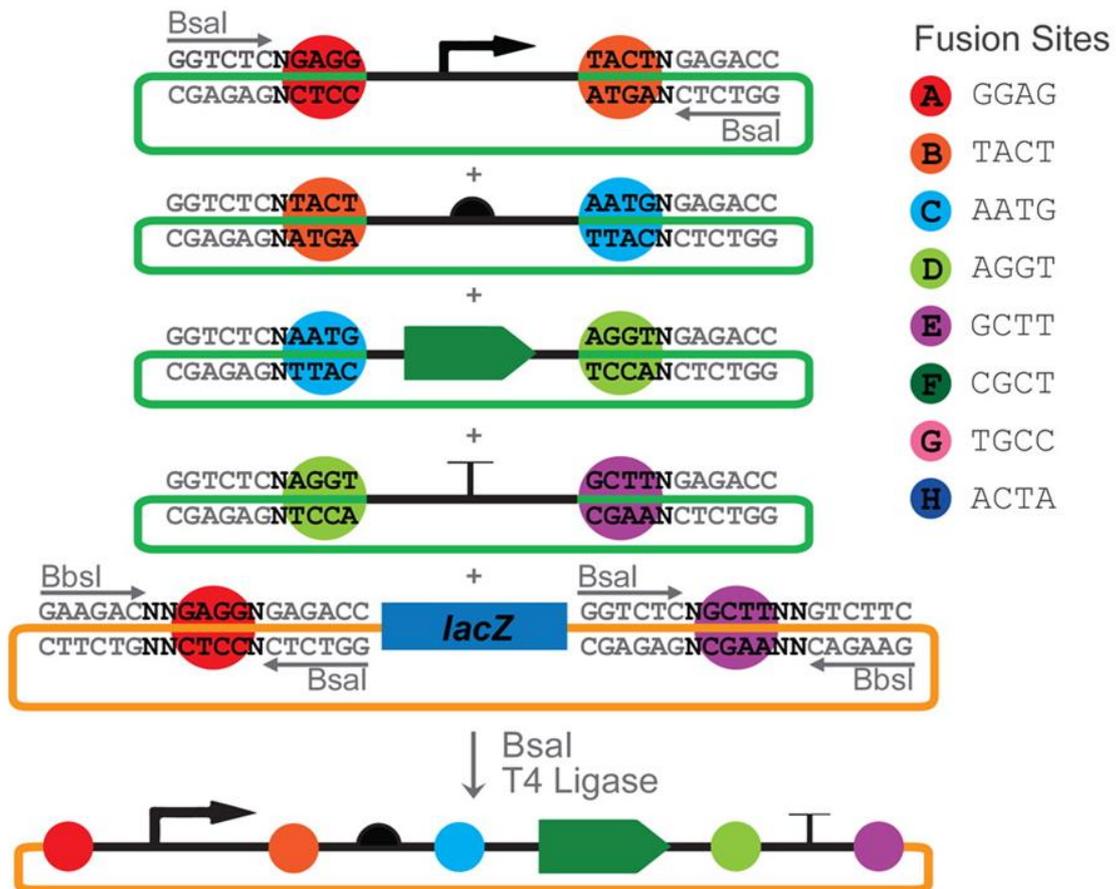


Figure 1.12. Schematic diagram of the CIDAR MoClo assembly format including the fusion site overhangs. The schematic diagram of the CIDAR MoClo assembly of components from holding DVA plasmids into an expression DVK plasmid, using Bsal and T4 DNA Ligase as detailed in Figure 1.11B. The 'fusion sites' section defines the sequences of the overhangs of the CIDAR MoClo assembly method produced by digestion with a type IIS endonuclease. Image taken from (Iverson *et al.*, 2016).

1.7. Aims

This work aims to enhance xylose utilization within *E. coli*, a model prokaryote with potential for producing biofuels and other valuable organic compounds (Koppolu and Vasigala, 2016; Hassan, Williams and Jaiswal, 2019). To achieve xylose utilization, the principles of substrate channelling will be applied as it does not require increased enzyme concentrations nor altering of the pathway expression. Within *E. coli* there are two key xylose transporters, XylE and XylFGH; however, XylE is of greater interest for substrate channelling as it belongs to the MFS superfamily, to which functional C-terminal fusions are prevalent (Willson *et al.*, 2019; Willson, Chapman and Thomas, 2019). In order to facilitate increased substrate channelling, a fusion of XylE and XylA will be produced, alongside a scaffold system featuring protein-protein interaction to take into account any steric stress that could impact the fusion system.

Overall, there are 6 key aims within this work:

- To generate plasmids of low copy number suitable for the expression of membrane protein fusions.
- To produce and validate a system capable of generating libraries of both protein fusions and scaffolds within this plasmid.
- To test the impact of direct fusion between membrane protein XylE and cytoplasmic enzyme XylA.
- To determine whether there is a different impact when scaffold proteins are used.
- To demonstrate the formation of a protein scaffold at the membrane.
- Finally, to demonstrate enhanced substrate channelling through an increased xylose metabolism through either the scaffold or direct fusion condition.

We hypothesize that both direct fusion and protein scaffolds can be used to increase substrate channelling through the xylose utilization pathway. However, due to the potential of steric hinderance, we also expect a high likelihood that enzyme or transport activity will be reduced in direct fusion compared to scaffold proteins. For this reason, we hypothesize only the scaffold will show increased metabolic flux, and therefore increased xylose utilization.

Chapter Two

Materials and Methods

2. Methods and Materials

2.1. Microbiology

2.1.1. Lysogeny broth and agar

Lysogeny broth (LB) was prepared in distilled water using 10 g/L tryptone, 5 g/L yeast extract and 10 g/L of sodium chloride. As the primary solution for culturing *E. coli*, transformations, purifications and western blot experiments were performed in LB at 37 °C unless otherwise noted. To produce LB agar, 15 g/L agar was added to LB broth. Media components were sourced from Oxoid.

2.1.2. Supplemented M9 minimal media

Fluorescent microscopy experiments were performed in supplemented M9. Supplemented M9 was adapted from Knight Lab. It is comprised of 1 x M9 (7 g/L Na₂HPO₄•7H₂O, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 1 g/L NH₄Cl), 0.34 mg/mL thiamine, 0.4% v/v Glycerol, 0.2% v/v casamino acids, 2 mM MgSO₄ (Fisher), 0.1 mM CaCl₂ (Fisher).

2.1.3. M9+X minimal media

M9+X minimal media was made up to the final concentration of 1 x M9 (7 g/L Na₂HPO₄•7H₂O, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 1 g/L NH₄Cl), 2 mM MgSO₄ and 10 mM xylose.

2.1.4. M9+G minimal media

M9+G minimal media was made up to the final concentration of 1 x M9 (7 g/L Na₂HPO₄•7H₂O, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 1 g/L NH₄Cl), 2 mM MgSO₄ and 10 mM glucuronic acid.

2.1.5. M9+S minimal media

M9+S minimal media was made up to the final concentration of 1 x M9 (7 g/L Na₂HPO₄•7H₂O, 3 g/L KH₂PO₄, 0.5 g/L NaCl, 1 g/L NH₄Cl), 2 mM MgSO₄ and 10 mM sialic acid.

2.1.6. Antibiotics

Antibiotics were used at concentrations specified in Table 2.1.

Antibiotic	Final $\mu\text{g/mL}$	Source
Chloramphenicol	34	Duchefa
Ampicillin	100	Sigma-Aldrich
Kanamycin	60	TCI Chemicals

Table 2.1. Concentrations of antibiotics used in this work

2.1.7. Growth assays

Each sample to be tested was grown in LB media with relevant antibiotic overnight at 37 °C, 200 rpm. The relevant buffer for limited growth assay was then prepared, either M9+X, M9+G or M9+S, from here on referred to as “growth buffer,” filter sterilized, and antibiotics added for selection. Following overnight growth, each sample is then normalized to the lowest OD₆₀₀ of the samples, followed by washing with 1 mL growth buffer and finally re-suspending in growth buffer.

A microplate is then obtained, and each well to be measured are filled with 200 µL of growth buffer, to which 5 repeats of each sample are loaded at an OD of 0.05. The sample is then read by an Epoch 2 microplate spectrophotometer, growing at 37 °C with double-orbital shaking for 48.5 h.

2.1.8. Calculating maximum growth rate

Growth rate calculation was adapted from (Hall *et al.*, 2014) The growth rate between two time points on the growth curve was calculated by the equation:

$$\mu = \left(\frac{\sum(t_n - \bar{t})(\text{LN}(C_n/\bar{C}))}{\sum(t_n - \bar{t})^2} \right) \left(\frac{\sum(t_n - \bar{t})(\text{LN}(C_n/\bar{C}))}{\sqrt{\sum(t_n - \bar{t})^2 \sum(\text{LN}(C_n/\bar{C})^2)}} \right)$$

Where μ is growth rate in h^{-1} , C is Optical Density at 600nm, and t is time in h. \bar{C} represents mean Optical Density, \bar{t} represents mean time Within this work Growth rate is calculated as the rate of the slope over 5 time points multiplied by the absolute value of the Pearson correlation over the same five time points, calculated throughout the reaction for each condition.

Subsequently the mean average of 5 consecutive growth rates were taken. From the mean average growth rate, the highest growth rate was found for each repeat. A second mean average was then calculated between highest growth rate of each repeat. By looking at the mean average this method aimed to reduce the impact of erroneous growth rates. The standard deviation was also calculated for variance between each repeat.

2.1.9. Producing Chemically Competent *E. coli*

Chemically competent *E. coli* was prepared by first incubating 5 mL of LB media with the desired strain and growing overnight at 37 °C, 160 rpm. 500 µL of the overnight culture of the desired strain was then used to inoculate 50 mL LBA. The culture was incubated at 37 °C, 180 rpm until OD₆₅₀ is between 0.5-0.6. The culture was transferred to 50mL Falcon tubes and chilled on ice for 5 min. The tube was then centrifuged at 2600 RCF, 4 °C for 10 min. The

supernatant was decanted, and pellet resuspended in 10 mL ice cold 0.1 M CaCl₂ solution and chilled on ice for 20 min. The solution was then centrifuged at 2600 RCF, 4 °C for 10 min, the supernatant discarded, and pellet resuspended in 2 mL of 0.1M CaCl₂, 15% (v/v) glycerol. The solution was then aliquot into Eppendorf's (between 100-150 µL each) and frozen at -80 °C.

2.1.10. Measuring the concentration of D-xylose in cell culture

Cultures of each sample were grown overnight in LB broth, before measuring the optical density at 600 nm. An aliquot of cells from each condition were taken and centrifuged at 12000 xg for 5 minutes, the supernatant discarded and the pellet resuspended in M9+X media. The volume of M9+X added was calculated to ensure all conditions were at the same OD. Each sample was then washed by centrifugation and resuspension in fresh M9+X, before being used to inoculate 2x or 3x 20 mL M9+X, each in a 50 mL falcon tube, at an OD of 0.05.

Samples were then grown for 13 hours, with 500 µL aliquots taken at regular time intervals. In accordance with a protocol by Megazyme, each aliquot then heat treated at 90°C for 10 minutes, followed by centrifugation at 12000 xg for 10 minutes. An aliquot of the supernatant for each sample was then used in the Megazyme D-xylose assay kit test to measure the concentration of D-xylose present. This kit uses a xylose dehydrogenase to convert xylose and NAD⁺ to D-xylonic acid and NADH. The amount of NADH present can then be inferred from colorimetric analysis of the sample with a spectrophotometer at a wavelength of 340 nm, with this absorbance giving insight into the volume of xylose present.

Within this work the protocol had to be adapted to 1 mL cuvettes, as the larger 3 mL cuvettes were not available, meaning the final volume in the cuvette was 0.99 mL as opposed to 2.97 mL. In addition, the concentration of D-xylose contained within M9+X was found to cause the reading at 340 nm to exceed instrument limit, meaning that the concentration of sample added was reduced. The conditions used for this experiment are shown in Table 2.2.

The concentration of xylose was then calculated by:

$$c = \frac{V \times MW}{\epsilon \times d \times v} \times (\Delta_{A2} - \Delta_{A1})$$

Where c = the concentration of xylose in g/L, V = the final volume (0.99 mL), MW is molecular weight (150.1 g/mol), ϵ = extinction coefficient of NADH at 340 nm (6300 L.mol⁻¹.cm⁻¹),

Pipette into cuvettes	Blank	Sample
Distilled water	0.700 mL	0.693 mL
Sample	-	0.007 mL
Solution 1 (buffer)	0.133 mL	0.133 mL
Solution 2 (NAD ⁺ /ATP)	0.133 mL	0.133 mL
Suspension 3 (Hexokinase)	0.007 mL	0.007 mL

Mixed by gentle vortexing, with the absorbances of the solutions (A1) read after at least 5 min

The reaction continued by addition of:

Solution 4 (XDH/XMR)	0.017 mL	0.017 mL
----------------------	----------	----------

Mixed by gentled vortexing, with the absorbance of the solutions (A2) read at the end of the reaction (~ 6 min).

Table 2.2. Reaction conditions of D-xylose assay kit

d = light path (1 cm), v = sample volume (0.007 mL), Δ_{A1} = absorbance before addition of solution 4 and Δ_{A2} = absorbance after addition of solution 4.

The percentage of available D-xylose consumed was then calculated as:

$$\frac{([Xyl]_i - [Xyl]_n)}{[Xyl]_i} \times 100\%$$

Where $[Xyl]_i$ is the initial concentration of xylose and $[Xyl]_n$ is the concentration of xylose at a particular time.

2.2. Molecular biology

2.2.1. Strains and plasmids

Strains used are listed in Table 2.3.

DH5 α is a polygenic derivative of *Escherichia coli* K-12 produced as a laboratory strain, designated W945. BW25113 is a derivative of *Escherichia coli* K-12 BD792, itself a derivative of W1485. Isogenic and polygenic derivatives of *E. coli* BW25113 were obtained from the KO collection (Baba *et al.*, 2006), with the kanamycin cassette cured in previous projects.

2.2.2. Curing a Keio collection strain of the Kanamycin resistance cassette

The isogenic derivative of *E. coli* BW25113 strain $\Delta xyIA::kan$ contains an undesired kanamycin antibiotic resistance cassette. Chemically competent $\Delta xyIA$ cells were produced as detailed above, and two aliquots were obtained, with 3 μ L purified pCP20 added to each. The sample was incubated on ice for 30 min, before heated to 42°C for 90 s on a heat block. The sample was then chilled on ice for 2 minutes, followed by addition of 900 μ L of pre-warmed LB solution to each aliquot. The sample was then incubated at 30°C for 1 hour to allow expression of plamid encoded antibiotic resistance marker. 150 μ L of each aliquot was then spread on a separate LB + ampicillin agar plate and incubated overnight at 30°C.

2 colonies were then selected, one from each plate, to subculture into separate 5 mL LB media without antibiotic. The cultures were then incubate overnight at 43°C to induce pCP20 flippase activity. The following morning, a loopful of each culture was streaked onto an LB agar plate, and was incubated overnight at 37°C. Several colonies from each plate was patched onto an LB agar plate, LB ampicillin (100 μ g/mL) plate and LB kanamycin (50 μ g/mL) plate. Colonies that grew on the LB agar plate, but not the LB ampicillin nor LB kanamycin plate, therefore were antibiotic sensitive, were then selected.

2.2.3 Agarose gel electrophoresis

Agarose gel electrophoresis was completed using 1% (w/v) agarose gel in TBE buffer. TBE buffer was composed of 1.62 g/L Tris (Invitrogen), 2.75 boric acid (Fisher scientific) and 0.95 g/L EDTA (Fisher scientific). To dissolve granulated agarose, the gel was brought to molten temperature. 1 x SYBR safe stain (Invitrogen) was added with gentle mixing to the agarose gel while molten, before the solution was poured into a mould and left to set for up to an hour.

For each gel, a DNA ladder was introduced to at least one terminal well (Bioline 1Kb+ or NEB 2-log ladder). 500ng of Plasmid was mixed with 1 μ L enzyme buffer and 5 U enzyme, with

Strain	Genotype	Source
DH5α	<i>endA1, recA1, gyrA96, thi, hsdR17, relA1, supE44ΔlacU169,80 lacZΔM15</i>	Invitrogen
BW25113	$\Delta(\text{araD-araB})567 \Delta(\text{rhaD-rhaB})568 \Delta\text{lacZ4787} (::\text{rrnB-3}) \text{ hsdR514 rph-1}$	(Datsenko and Wanner, 2000)
TDXylose	$\Delta\text{xylE}\Delta\text{xylH}\Delta\text{araH}$	Thomas lab
ΔxylE	ΔxylE	Thomas lab
$\Delta\text{xylE}\Delta\text{xylG}$	$\Delta\text{xylE}\Delta\text{xylG}::\text{KmR}$	Thomas lab
$\Delta\text{xylE}\Delta\text{xylH}$	$\Delta\text{xylE}\Delta\text{xylH}$	Thomas lab
$\Delta\text{xylA}::\text{kan}$	$\Delta\text{xylA}::\text{KmR}$	(Baba <i>et al.</i> , 2006)
ΔexuT	ΔexuT	Thomas lab
ΔnanT	ΔnanT	Thomas lab
ΔxylA	ΔxylA	This work

Table 2.3. List of E. coli strains used in this work

sample brought to 10 μ L by addition of Milli-Q water. The solution was incubated between 1 and 2 h at 37 $^{\circ}$ C to achieve plasmid cutting, before samples were prepared in 1x loading dye (Bioline) and gel electrophoresis followed (70V, 60 min).

2.2.3. DNA preparation and extraction

DNA from plasmids were prepared by using commercially available “mini prep” kits (Biobasic, Sigma) according to the manufacturer’s instructions. DNA from PCR reactions and agarose gel were extracted and cleaned up using commercial PCR product purification (Biobasic) and gel extraction kits (Biobasic) respectively.

2.2.4. Construction of Golden Gate expression plasmids

The Golden Gate compatible plasmid, pLC1, was produced by Quickfusion, assembling the vector from polymerase chain reaction (PCR) amplicons derived from three plasmids: pSB1C3, DVK_AF and pWKS30 (Wang *et al.*, 1991; Iverson *et al.*, 2016). Quickfusion ligates two or more amplicons when the terminal regions of one amplicon complement the terminal region of another amplicon. Primers FwOri and RevOri (10mM) were used to amplify the chloramphenicol resistance region of pSB1C3 (958bp), VR and VF2 amplified the LacZ α region alongside two flanking BsaI sites of DVK_AF (841bp), and primers VCR and VCFw amplified the backbone of pWKS30, including the pSC101 origin of replication (2984bp). The amplicons were then combined in a 0.015: 0.03: 0.03 pmol mix, pWKS30: pSB1C3: DVK_AF, alongside QuickFusion buffer (2 μ L) and QuickFusion ligating enzyme (1 μ L), obtained from Biotool, and brought to 10 μ L with Milli-Q water. The solution was incubated at 37 $^{\circ}$ C for 30 min to produce the pLC1 vector.

Application of Quickfusion was also used for the production of pLC2, a temperature sensitive variant of pLC1, as a result of the oriR101 of pKD46. Primers pKD46 Ori Fw and pKD46 Ori Rev amplify the temperature sensitive pKD46 origin of replication (1886 bp), VR and VF2 amplified the LacZ α region of DVK_AF (841bp), and primers FwOri and RevOri (10mM) amplified the chloramphenicol resistance region of pSB1C3 (958bp).

Plasmid pLC3 was constructed through the process of blunt end mutagenesis. Two primers were produced, one containing the required 2bp frame shift mutations followed by a 20bp region matching the plasmid (VpLC3 blunt end fw). On the other side of the mutation, the second primer, 20bp in length (VpLC3 blunt end Rv), annealed. After a phase of PCR amplification, 1 μ L of PCR product was mixed with 2 μ L NEB T4 DNA Ligase Buffer, 1 μ L NEB T4 DNA Ligase, 1 μ L NEB PNK and 15 μ L autoclaved MilliQ water, and left overnight at 16 $^{\circ}$ C to achieve phosphorylation and ligation.

Finally, Quickfusion was applied to produce Golden Gate compatible plasmids pLD1 and pLD3 from pLC1 or pLC3 respectively, alongside pET28a and pBBR1MCS-2. Primers VpLC3QF Fw and VpLC3QF Rev2 amplified the lacZ α region of either pLC1 or pLC3 (852bp), VpBBR1 rep Fw and VpBBR1 rep rv amplified the origin of replication of pBBR1MCS-2 (1495bp) and VKanpET28 QF Fw2 and VKanpET28 QF rev amplified the Kanamycin resistance genes of pET28a (1052bp).

2.2.5. Gene amplification by Polymerase Chain Reaction

A list of primers used for PCR are specified in Table 2.6. Amplification of desired genes or DNA regions was completed by PCR using Q5 DNA polymerase (New England Biolabs), with the reaction completed within a thermocycler. Where a restriction enzyme digest is required of the amplicon, an addition of several extra base pairs was required to enable effective digestion. PCR primers were purchased either from IDT or Sigma.

PCR was performed according to the manufacturer's protocol provided New England Biolabs, accounting for the T_m of the primers and the length of the sequence to be amplified. A typical reaction mixture of 50 μ L comprised of 1x Q5 reaction buffer, 0.2 U/ μ L Q5 polymerase, <1000 ng template DNA, 0.5 μ M forward primer, 0.5 μ M reverse primer, 200 μ M dNTPs and nuclease-free water to bring total volume up to 50 μ L. This solution was then inserted into the thermocycler, with an initial step of 30 s at 98 $^{\circ}$ C. Following this, 35 cycles followed of: 10 s at 98 $^{\circ}$ C, 30 s at T_m (50-72 $^{\circ}$ C) and 30 s/kb at 72 $^{\circ}$ C. A final extension followed for 2 min at 72 $^{\circ}$ C and temperature was brought to 4 $^{\circ}$ C for holding.

2.2.6. Site directed mutagenesis

The methodology for site directed mutagenesis was adapted from a protocol by (Liu and Naismith, 2008).

Pairs of primers were designed of 30-40 bases in length. These would match the target region, however only the half of each primer would overlap with the complimentary primer. The primers were designed so that they were composed of a 15-20 base 5' region which would overlap with the other primer (hereafter termed the overlapping sequence), and within the region was the mutation. At the 3' end the primer extended for an additional 15-20 bases, matching only the upstream region of the DNA (hereafter termed the non-overlapping sequence). The overlapping sequence was designed so that the T_m ($T_{m_{no}}$) was 5 to 10 $^{\circ}$ C higher than the T_m of the non-overlapping sequence of the primer ($T_{m_{pp}}$).

A reaction mixture was prepared using 50-100 ng of plasmid, 10 mM of each primer, 1x Q5 reaction buffer, 0.2 U/ μ L Q5 polymerase, 200 μ M dNTPs and nuclease-free water to bring

total volume up to 50 μL . This solution was then inserted into the thermocycler, with an initial step of 5 min at 98 $^{\circ}\text{C}$. Following this, 12 cycles followed of: 1 min at 98 $^{\circ}\text{C}$, 1 min at $T_{m_{no}}-5$ $^{\circ}\text{C}$ and 10 min at 72 $^{\circ}\text{C}$. A final step of annealing at $T_{m_{pp}}-5$ $^{\circ}\text{C}$ for 1 min, followed for 30 s at 72 $^{\circ}\text{C}$ and temperature was brought to 4 $^{\circ}\text{C}$ for holding.

The PCR product was then purified using a Bio Basic PCR clean up kit. Purified amplicon was then digested with restriction enzyme DpnI to remove methylated DNA for 1 hour at 37 $^{\circ}\text{C}$, the enzyme then inactivated by heating to 80 $^{\circ}\text{C}$ for 20 minutes. The product was then used to transform *E. coli* strain DH5 α , before extracting plasmid DNA of 2-4 colonies and sequencing to confirm the desired mutation is present.

2.2.7. Blunt End Mutagenesis

Two primers were developed between 15 and 30bp in length, with complementary annealing temperatures. The primers were designed to anneal in opposing direction, with the 5' region of each primer back to back. Only one primer contained the desired mutation(s). The sample was amplified by PCR, and the product was purified using a PCR clean up kit (Bio Basic Inc).

A 20 μL reaction mixture of PCR product (20-100 ng), T4 DNA Ligase buffer (1x), T4 DNA Ligase (20 U) and T4 Polynucleotide Kinase (10 U) was produced, with the remainder being autoclaved milliQ water. The mixture was allowed to react overnight at 16 $^{\circ}\text{C}$, followed by restriction enzyme digest with DpnI to remove methylated DNA.

2.2.8. Duplex assembly from primers

Two complimentary primers were produced, overlapping with the exception of a 4bp single stranded overhang suitable for insertion into Golden Gate arrangement at either end. The two primers were added to a reaction vessel (10 μM), each with a final concentration of 1.6 μM . Both T4 PNK and T4 Ligase buffer, provided by New England Biolabs, were added, and the solution was brought up to 20 μL with nuclease-free water.

The reaction mixture was incubated at 37 $^{\circ}\text{C}$, 30 min before removing and adding an equal volume of 6x saline-sodium citrate (SSC) (20 μL) and half volume of 100 μM EDTA (10 μL). This solution was heated at 95 $^{\circ}\text{C}$, 5 min and cooled to 12 $^{\circ}\text{C}$ at a gradient of -0.1 $^{\circ}\text{C}/\text{s}$, holding at 12 $^{\circ}\text{C}$ for 10 min before cooling to 4 $^{\circ}\text{C}$ to be put on hold. 1.3 μL of resultant product was then used in Golden Gate assembly as insert DNA.

2.2.9. Heat-shock transformation

Transformation was performed with 100-200 ng of plasmid pipetted into a 150 μL sample of thawed chemically competent *E. coli*, mixed by pipetting and incubated for 30 min on ice. The

cells were heated to 42°C for 1 min, then cooled on ice for a further 2 min. Once cooled, 500 mL LB media was added to the cells and an overgrowth media was grown at 37 °C, 220 rpm for 1 h. The transformed cells were then grown overnight on LB agar with appropriate antibiotic. Blue-White screening was performed to select for desired colonies by adding 20 µg/mL X-Gal and 1 mM IPTG to the plate (Bebenek and Kunkel, 1995). Desired colonies were then expressed in overnight cultures and the plasmid was separated by miniprep.

2.2.10. Golden Gate Assembly

For producing component plasmids, the level 0 holding vector was introduced to the system at 80-90 fmol. The PCR amplicon containing desired insert was added at a 2:1 ratio of insert to backbone. 2 µL of NEB T4 DNA buffer (10x), 5 U thermo scientific Bpil and 200U NEB T4 DNA ligase were added to the reaction vessel, which was then filled to 20 µL with milliQ water. The reaction solution was heated to 37°C for 20 s, followed by 20 cycles of 37°C for 3 min, then 4 min at 16 °C. Once the cycle was complete, the solution was heated to 50°C for 5 min, followed by 80°C for 10 min.

For producing an expression vector, the level 1 holding vector is introduced to the system at 10-60 fmol. Vectors containing inserts were added at a 2:1 ratio of insert to backbone, therefore each component vector would require double the fmol value of pLC1 (20-120 fmol). 2 µL of NEB T4 DNA buffer (10x), 10 U NEB Bsal-HF and 200 U NEB T4 DNA ligase were added to the reaction vessel, with solution brought to 20 µL by addition of MilliQ water. Using a thermocycler, the reaction solution was heated to 37 °C for 20 s, followed by 25 cycles of 37 °C for 3 min, then 4 min at 16 °C. Once the cycling was complete, the solution was heated to 50 °C for 5 min, followed by 80 °C for 10 min, and then brought to 4 °C.

2.2.11. Designing of gBlocks

In order to produce proteins that were not native to *Escherichia coli*, such as the cohesin dockerin system, gBlocks were designed and ordered from Integrated DNA Technologies (IDT). The amino acid sequence of the target proteins were obtained from the Protein Data Bank (PDB) (Berman *et al.*, 2000) and codon optimised using the IDT codon optimization tool, optimizing against *E. coli* strain K12. The gBlock contained multiple domains from the same organism, with the domains of interest cloned from the gBlock with Bpil restriction enzyme binding sites, and then inserted into a DVA plasmid through Golden Gate Assembly. Multiple colonies were then selected and the plasmid purified and sequenced to confirm no mutations were present from the gBlock.

2.3. Microscopy

2.3.1 Sample preparation for Fluorescence Microscopy

A fresh solution of supplemented M9 was prepared to induce overnight samples of desired expression vectors expressed in either BW25113 or DH5 α . These overnight samples were diluted to an OD₆₀₀ of 0.2 and grown for up to 5 h at 37 °C, 180 rpm. Following this, cells were either diluted or concentrated, depending on the rate of growth, to an OD₆₀₀ of 1.0 at a volume of 1 mL in supplemented M9. 96 μ L of each sample was then mixed in solution with 4 μ L of 25x stock of silica beads (diameter = 5 μ m) in M9. 30 μ L of each mixture was then pipetted along the length of a 10 μ g/mL poly-D-lysine coated quartz slide, and a no. 1 thickness cover slip was applied, using nail varnish to hold it in place, and the varnish was dried under a small fan for 10 min.

2.3.2. Epifluorescence Microscopy

Epi-fluorescence imaging was done using either an Axioskop 40 microscope (Zeiss) for work completed in chapter 3, or a 780 inverted fluorescence confocal microscope (Zeiss) for mCherry work in chapter 5, the latter of which was completed by Joe Nabarro (PhD student from Baumann group). Following the protocol covered in 2.3.1. a small drop of immersion oil (Zeiss) was placed on the coverslip and the slide was loaded onto the mechanical stage and held by the clasp. Fluorescence imaging was achieved by different filters. For recording Green Fluorescence Protein, a filter with an emission spectrum of 530/43 was used. For recording mCherry a filter of 605/70 was used.

Images were recorded using the imaging software, Micro-Manager (Stuurman, Amdodaj and Vale, 2007; Edelstein *et al.*, 2010), an opensource software package that acts as a plug-in for ImageJ in order to collect images from an interfaced camera on a microscope. Epi-fluorescence images were recorded with CoolSnap CCD camera (Photometrics) at a resolution of 0.102 μ m/px.

2.3.3 Total Internal Reflection Fluorescence Microscopy (TIRFM)

Following protocol covered in 2.3.1, once the slides were dry a drop of refractive index matching glycerol was applied to the reverse of the slide, placing a quartz prism on top of this glycerol and holding it in place with a small amount of tape. A drop of immersion oil was also applied to the opposite cover slip surface, before inverting it and placing it on the mechanical stage, so that the cover slip faces downward and the prism upwards.

TIRFM was performed using a customised set up (Rassam *et al.*, 2015), with image stacks recorded on MicroManager. Fluorophores were excited with a 488 nm (GFP) or 561 nm (mCherry) laser, and appropriate band-pass optical filters was used to select GFP (525/50) or mCherry (605/20) fluorescence emission.

2.3.4 Processing of microscopy data

Processing of single microscope images was completed using Fiji (Schindelin *et al.*, 2012), an open source image processing package based on ImageJ. Scale bars were added by inputting the correct properties of the pixel size in μm and using the “scale bar” tool within Fiji, which automatically added a scale bar to the image. Pseudo-colouring of fluorescence images was achieved by the edit LUT feature after selecting the respective colour.

2.3.5 Median filtering of TIRFM data

To reduce the impact of noise, Median Filtering was then applied to these images. To achieve this, scripts produced by members of the Baumann lab were run (scripts not published). Image stacks collected by TIRFM were each saved in a different folder, containing a single “.Tif” file composed of 900 frames. In order to complete Median Filtering it was required that “.Tif” stacks of the same conditions were together. Therefore, a script produced by Rosalyn Leaman (PhD student in Baumann group), “saveas.ijm,” was run in Fiji in order to collect all like stacks in one folder.

Median filtering is completed within MATLAB, once more using scripts encoded by Rosalyn Leaman. 5 scripts were inserted into the MATLAB folder, “extractLeftChannel.ijm,” “extractRightChannel.ijm,” “medianTimeFilter.m,” “medianTimeFilterFolder.m” and “tiffread.m.” The script “medianTimeFilterFolder.m” was run, coordinating the other scripts to extract image stacks from one folder, median filter each stack individually by comparing the pixel array between 3 frames, and save the output into a separate folder, also in “.Tif” format. Median filtering calculates a median image for each time point by using three consecutive video frames (includes frames before and after the designated frame). For this reason, the first and final video frame are not included in the median filtered data, and so in a raw video stack of 900 images, only 898 median filtered images will be produced.

2.3.6. Single particle tracking of TIRFM data

The analysis of TIRFM data was completed using MATLAB R2019a, executing a script package encoded by Alex Scott (PhD student in Baumann group). Following median filtering of the data, the desired folder was selected upon running the script “ParticleTrackerFolder.m” and each .Tif file was analysed for each single fluorescent spot (termed a ‘particle’ here) of

appropriate intensity which would be present continuously for extended periods of time. The trajectory of each particle was then determined from its time-dependent x,y coordinates, and a diffusion coefficient calculated from the mean-squared displacement (MSD) of the particle in time. The MSD gives insight into the diffusiveness of a particle. It describes the movement of a particle from its starting position to a later position at a particular time.

The time averaged mean squared diffusion for the trajectory i , was calculated from the trajectory position of the x-axis and y-axis over time, $x_i(t)$ and $y_i(t)$ respectively. Here, δt denotes the time point and N denotes the overall measurement time.

$$MSD(\delta t) = \frac{1}{N - \delta t} \sum_{n=0}^{N-\delta t} \{ [x_i(n + \delta t) - x_i(n)]^2 + [y_i(n + \delta t) - y_i(n)]^2 \}$$

Linear regression of the MSD data for the first 4 time intervals was done to calculate the diffusion coefficient, D, using the following equation: $MSD(t) = 4D_{2D} t$.

“ParticleTrackerFolder.m” is part of a package of code designed by Alex Scott for tracking of single fluorophores attached to molecules undergoing two-dimensional lateral diffusion (currently not published). Within the code, pixel size in the magnified image was set to $0.096 \mu\text{m pixel}^{-1}$, and the video data was collected with an exposure time of 0.015 s. The maximum allowed number of consecutive missing “dead” frames was set to 5. A particle is defined by the code “createP.m” which uses a two-dimensional Gaussian fit to each fluorescent spot in a video frame to determine the centre of its point spread function (PSF) and therefore its x,y coordinates. The minimum and maximum size of the particle can be defined within “createP.m,” alongside the required signal to noise ratio for particle detection.

Output data was then analysed by further code, “checkAgg.m” to remove trajectories in which high intensity particles were present, representing aggregated fluorescently-labelled molecules. Particles were determined to be high intensity if the measured intensity was more than 1.5 interquartile ranges above the upper quartile, with particles 1.5 interquartile ranges below the lower quartile also removed.

An open source alternative, PaTrack, is also available for single particle tracking of TIRFM data to measure the MSD of fluorophores (Dosset *et al.*, 2016).

2.4. Recombinant protein production and purification.

2.4.1. Separating soluble and insoluble fractions

4mL of each overnight sample is centrifuged at 6,800 g for 2 min, and the supernatant decanted. BugBuster solution (800 μ L) is then added and the pellet is resuspended. The solution is gently mixed at room temperature for 30 min, and then centrifuged (6,800 g, 10 min). The pellet and supernatant are then separated, with the pellet washed in 800 μ L of PBS solution, before being resuspended in 800 μ L of fresh PBS, heating at 60°C to aid this process.

2.4.2. Nickel affinity chromatography

2L LB solution was inoculated with an overnight sample of 20mL cell culture expressing desired plasmid, and was grown at 37 °C, 220 rpm for 4 h. Following sufficient growth, the cell culture was centrifuged at 5000 g, 15 min, 4 °C and supernatant decanted. The pellet was then resuspended in 40 mL KPI wash solution (50 mM KPI, 40 mM imidazole, 200 mM NaCl, 20% v/v glycerol, pH 7.2) and frozen overnight at -80°C. 1 M KPI solution at a pH of 7.8 was produced by mixing 0.145 M potassium dihydrogen orthophosphate and 0.855 M diphosphate hydrogen orthophosphate.

Solution was thawed under warm water. Sample was sonicated with microtip (3 min, 3 s intervals, 7 s pause, 60 w) then centrifuged (20,000 g, 20 min, 4 °C) and supernatant collected.

Nickel affinity chromatography was completed on using a GE healthcare AKTA start pump system, using HisTrap HP 5mL column, the previously discussed KPI wash solution and a KPI elution buffer (50 mM KPI, 200 mM imidazole, 200 mM NaCl, 20% w/v glycerol). Column was washed with KPI wash solution, before loading 35 mL of supernatant, washing with 200 mL KPI wash solution, and eluting with 20 mL KPI elution buffer, collecting the eluent in 1-3 mL aliquots.

2.4.3. Size Exclusion Chromatography (SEC)

For the course of the reaction, 1 L SEC buffer was produced (20 mM Tris, 50 mM NaCl) and filter sterilized. As with Nickel Affinity chromatography, chromatography was completed on a GE healthcare AKTA start pump system, however using a HiPrep™ 16/60 Sephacryl™ S-200 HR column, which is otherwise stored at 4 °C.

The column was equilibrated with SEC buffer, running at a rate of 0.5 mL/min for 2x120 mL column volumes. Once the column was equilibrated, sample loading could be completed.

An aliquot of interest produced by Nickel Affinity Chromatography was obtained and 3.2 mL was inserted into a syringe. The column was flushed once more with SEC buffer, and the sample was loaded into a loading line via the syringe. Once flushing was complete, the sample was run through the column alongside SEC buffer, at a flow rate of 1 mL/min for 1 column volume of 120 mL.

2.4.4. Size Exclusion Chromatography – Multi-Angle Laser Light Scattering (SEC-MALLS)

Protein was prepared by purifying through a combination of nickel affinity chromatography and SEC to remove any background. A 120 μ L aliquot of the desired protein at concentration of 2.4 mg/ml in SEC buffer was then produced, diluting solution in SEC buffer where necessary.

To facilitate elution SEC buffer was used (25 mM Tris, 50 mM NaCl, pH 8.0), filtered to 0.20 μ m, with an additional 0.10 μ m filter present within the flowpath. The experiment was conducted by Andrew Leech of the University of York Technology Facility. Experiments were conducted on a system comprising a Wyatt HELEOS-II multi-angle light scattering detector and a Wyatt rEX refractive index detector linked to a Shimadzu HPLC system. The column was equilibrated with at least 2 column volumes of solvent before use and flow was continued at the working flow rate until baselines for UV (detected at 280 nm), light scattering and refractive index detectors were all stable, this was conducted in the morning of the experiment to maintain flowrate through the column.

Sample injection volume was 100 μ L; Shimadzu LabSolutions software was used to control the HPLC and Astra 7 software for the HELEOS-II and rEX detectors. The Astra data collection was 1 minute shorter than the LC solutions run to maintain synchronisation. Blank buffer injections were used as appropriate to check for carry-over between sample runs. Data were analysed using the Astra 7 software. MWs were estimated using the Zimm fit method with degree 1. A value of 0.182 was used for protein refractive index increment (dn/dc).

The sample was run through a Superdex S200 10/300 GL column, produced by G.E. Healthcare, at a flow rate of 0.5 mL/min for 60 minutes, of which data collection took place for 59. A BSA standard was also run for comparison.

2.4.5. Protein crystallization

A viva spin column was applied to increase concentration of SEC purified protein where appropriate. Two concentrations were selected, a lower concentration between 7-10 mg/ml, and a higher concentration between 15-20 mg/ml. Using the Mosquito LCP, produced by SPT labtech, aliquots of the two concentrations were loaded onto crystallization plates. The plates

selected for this work were the Crystal screen (Hampton research), PEG/Ion (Hampton Research), Index (Hampton Research), CSS 1 (Molecular Dimensions) and CSS 2 (Molecular Dimensions). The plates were then stored in a controlled climate room set to 22°C and observed under microscope for crystallisation.

Crystals that could be recognised were resuspended in a cryoprotectant by Reyme Herman (Technician in Thomas group). Initial trials using 10-20 % v/v glycerol demonstrated a rapid solubilisation of the crystal, therefore necessitating the use of 30 % v/v PEG 400 (polyethylene glycol 400).

X-ray diffraction data were collected at the Diamond Light Source, UK on the i03 beamline on the 24th of June 2020. The data collected were indexed and scaled using the XDS pipeline on xia2. Data reduction was performed using AIMLESS. Molecular replacement was used to obtain initial phase information using the structure of a xylose isomerase from *Thermoanaerobacterium thermosulfurigenes* (Obtained from PDB (Berman *et al.*, 2000) PDB ID: 1A0C) on MOLREP. The constructed model was then refined using REFMAC by Reyme Herman.

2.4.6. Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was prepared in two parts, a stacking gel and resolving gel. The resolving SDS-PAGE gel was prepared using 12% w/v acrylamide, 0.32% w/v bisacrylamide, 0.375 M Tris pH 8.8, 0.1% w/v SDS, 0.1% w/v APS and 0.01% v/v TEMED, and 4 mL was pipetted into a holding mould and left to set under ethanol for 30 min. The stacking SDS-PAGE gel was prepared using 4% w/v acrylamide, 0.1% w/v bisacrylamide, 0.125 M Tris pH 6.8, 0.1% w/v SDS, 0.1% w/v APS and 0.01% v/v TEMED, and was loaded into the remainder of the mould. Running buffer was composed of 2 g/L Tris, 14 g/L glycine and 1 g/L SDS, with pH adjusted to 8.8.

Loading dye is diluted in PBS solution with 3% mercaptoethanol. 10 µL of loading dye is added to 30 µL of either soluble or insoluble cell fractions and heated for 20 min (90 °C for soluble fraction, 60 °C for insoluble fraction). For Coomassie blue, 20 µL of each sample is pipetted into each well; for western blot, 10 µL is pipetted instead. SDS PAGE then follows (200 V, 50 min).

2.4.7. Western blot

The western blot is completed following invitrogen's iBlot® Dry Blotting protocol. Invitrogen iBlot® mini nitrocellulose gel stacks or iBlot® regular PVDF gel stacks were used on the

original iBlot® Gel Transfer Device, running at 20 V for 7-8 minutes in accordance with the protocol.

The produced nitrocellulose/PVDF membrane is then mixed at 4°C overnight in 5% (w/v) skimmed milk blocking agent (40 mL). After which point it is washed for 20 min using 2 rounds of 50 mL PBST, and an additional 40 mL 5% (w/v) skimmed milk solution is added, mixing at room temperature for 30 min. Once complete, 2 washes of 50 mL PBST and 3 rotations of 50 mL PBS wash are completed, each mixing at room temperature for 10 min. Once the membrane is washed, the primary antibody is added at a 1 in 2000 dilution in blocking agent (list of antibodies in Table 2.4). Solution is then mixed for 2 h on an orbital shaker at room temperature.

Following addition of primary antibody, a second washing stage is required: 3 rotations of PBST followed by 2 rotations of PBS. Secondary antibody is then added at a dilution of 1 in 5000 in PBST and mixed at room temperature for 1 h. The solution is finally washed once more with PBST for 1 hour, adding fresh PBST every 10 minutes.

The washed membrane is then treated with HRP substrate. 5-10 mL of Luminata™ Forte HRP substrate (Millipore) is added to the membrane to facilitate chemiluminescence. The sample is then pressed against a light sensitive film within a dark room, with the medical film then processed by the Konica Model SRX-101A, which moves the film through a developer, fixer and finally a water wash tank to develop the image.

2.4.8. Coomassie blue staining

Following gel electrophoresis, the gel is placed in a container and submerged in Coomassie brilliant blue dye, before mixing gently at room temperature. Coomassie brilliant blue dye is prepared by dissolving 2.5 g Remazol Brilliant Blue R (Sigma-Aldrich) in 450 mL methanol (Sigma-Aldrich), followed by addition of 100 mL acetic acid and 450 mL dH₂O.

Gels are stained for 20-60 min, then the dye is removed and a destaining solution (10% (v/v) ethanol, 10% (v/v) acetone) is added, once more submerging the gel. The gel is then mixed gently at room temperature for 8 h, before removing the destaining buffer and observing the gel under white light.

Antibody	Dilution factor	Blocking agent	Source
Mouse monoclonal anti-GFP	1 in 2000	5% (w/v) skimmed milk in PBS	BioRad
Mouse anti-tetraHis	1 in 2000	5% (w/v) skimmed milk in PBS	Qiagen
Mouse monoclonal anti-mCherry	1 in 10,000	3% (w/v) BSA in PBS	Sigma Aldrich
Anti-mouse IgG(Fc specific) HRP conjugate	1 in 5000	PBST	Sigma Aldrich

Table 2.4. List of antibodies used in this work.

2.5. Bioinformatics

2.5.1. Determining Major Facilitator Superfamily protein fusions in nature.

The InterPro database was searched for MFS domains with the identifier 'IPR020846', identifying 3527 architectures. To survey the most commonly found, architectures with more than 100 examples were selected, which also removed any chance of them being false positives from sequencing errors. Furthermore, as the study focussed on bacterial MFS transporters, only architectures with bacterial representatives were included. These architectures would be defined as a single gene containing both an MFS domain region and an up-stream or down-stream secondary domain (Willson *et al.*, 2019).

Each example was then downloaded, giving insight on the organism of origin, alongside a gene/protein identifier suitable for searching for information on the protein in the Uniprot database (Bateman, 2019). Of the architectures discovered, 8 fusion groups could be defined, and architectures were subsequently sorted into one of these fusion groups.

2.5.2. Multiple sequence alignment

Multiple sequence alignment of the protein sequence was completed using one of two online resources, Tcoffee (Notredame, Higgins and Heringa, 2000) and ClustalW (Sievers *et al.*, 2011). The subsequent alignment is then represented using Jalview (Waterhouse *et al.*, 2009) – which features options to visualise properties of the regions of overlap.

Images were coloured using ClustalX annotation, to identify the amino acid profiles for sequences above a defined residue threshold. The ClustalX colour scheme results in red annotation for positively charged (basic) amino acids, magenta for negatively charged (acidic) amino acids, blue for hydrophobic residues, green for polar amino acids and cyan for aromatic amino acids. Some colours are also used for annotating individual amino acids with important properties: pink for cysteines, orange for glycine, and yellow for prolines, Unconserved amino acids below the threshold are not annotated.

2.5.3. Maximum Likelihood Phylogenetic tree

An alignment of the overlapping regions of different genes is produced using ClustalW software and saved as an alignment (.aln) file. The file is subsequently uploaded to the website IQTree (Trifinopoulos *et al.*, 2016) in order to produce a maximum likelihood phylogenetic tree. IQTree is not altered from its default in this work, and is set to automatic detection, producing a .treefile document. To visualize the trees produced, the “.treefile” document is then uploaded to ITOL (Letunic and Bork, 2019, 2021).

Reagent	Purity	Source
Ethanol	Absolute (100%)	VWR Chemicals
Acetone	≥99.8%	Fisher scientific
Methanol	≥99.9%	Sigma-Aldrich
Tryptone	-	Oxoid
Yeast Extract	-	Oxoid
Sodium Chloride	99.5%	Fisher scientific
Agar	Bacteriological grade	Formedium
M9 minimal media salt	5x	Sigma-Aldrich
Thiamine hydrochloride	-	Serva
Glycerol	≥99.5%	Honeywell
Casamino acids	-	Fisher scientific
Magnesium Sulfate Dried	Extra pure	Fisher scientific
Calcium Chloride Dihydrate	≥99%	Fisher scientific
D-(+)-Xylose	≥99%	Sigma-Aldrich
D-(+)-Glucose	≥99.5%	Sigma-Aldrich
L-Arabinose	≥99%	Sigma-Aldrich
Glycine	≥99%	Sigma-Aldrich
Sialic acid	97%	Alfa Aesar
Glucuronic acid	98%	Alfa Aesar
Agarose	Bioreagent	Sigma-Aldrich
Tris	≥99.9%	Invitrogen

Boric acid	≥99.8%	Fisher scientific
EDTA	≥99%	Fisher scientific
SYBR safe stain	-	Invitrogen
Loading dye (DNA loading buffer blue)	5x	Bioline
Di-potassium hydrogen orthophosphate anhydrous	≥99%	Fisher scientific
Potassium dihydrogen orthophosphate	≥99.5%	Fisher scientific
Imidazole	>99.5% (GC)	Sigma-Aldrich
Protogel	30% w/v Acrylamide, 0.9 % w/v Bisacrylamide	National Diagnostics
Sodium Dodecylsulfate (SDS)	-	Melford labs
TEMED	Ultrapure	Biobasic
Ammonium persulfate	98%	Sigma-Aldrich
β-mercaptoethanol	99%	Sigma-Aldrich
Tween-20	≥40% (GC)	Sigma-Aldrich
Phosphate Buffer Saline tablets	-	Sigma-Aldrich
Skimmed Milk	-	Sigma-Aldrich
Bovine Serum Album	Lysophized powder, >99%	Sigma-Aldrich
Remazol Brilliant Blue R	Pure	Sigma-Aldrich
Polyethylene glycol 400	For synthesis	Sigma-Aldrich
Hydrochloric acid	37%	Sigma-Alrich

Sodium hydroxide - pellets	Analytical reagent grade	Fisher Scientific
-----------------------------------	--------------------------	-------------------

Table 2.5. List of reagents used in this work.

Primer	Target	DNA Sequence (5'-3')	Restriction site
VR	DVK_AF	ATT ACC GCC TTT GAG TGA GC	
VF2	DVK_AF	TGC CAC CTG ACG TCT AAG AA	
VCR	pSB1C3	TTC TTA GAC GTC AGG TGG CA	
VCFw	pSB1C3	GAG TAA ACT TGG TCT GAC AGC TC	
Fw Ori	pWKS30	GAG CTG TCA GAC CAA GTT TAC TCC GTA GGA CAG GGT GCC GG	
Rev Ori	pWKS30	CTG TCA GAC CAA GTT TAC TCC GCT GCG TGA CTA CGA TGG T	
VFIXFw4	<i>E. coli</i> MG1655 xylE	TCA CAG GGA AGA CAC ACC TCA GCG TAG CAG TTT GTT GT	Bpil
VxFlipR2	<i>E. coli</i> MG1655 xylE	TCA CAG GGA AGA CAC AAT GAA TAC CCA GTA TAA TTC CAG TTA TAT ATT TTC GAT TAC CT	Bpil
VmXEFw	DVA_EF_XylE	AAA GAA ACC AGT TTC CCG CGA ATA TGA GCT GGA G	
VmXERv	DVA_EF_XylE	GGA AAC TGG TTT CTT TTA ACC AGT TTG CGA TTA TTT TCG	
VxFlfw CF Bsal	DVA_EF_XylE	CTA CTA GTA GGT CTC TAG CGC AGC GTA GCA GTT TGT TGT GTT TTC TT	Bsal
VxFlipR CF Bsal	DVA_EF_XylE	AGA CTA GTG GGT CTC AAA TGA ATA CCC AGT ATA	Bsal
VxexuT CF fw Bsal	DVA_EF_ExuT	TAG TGG GTC TCA AAT GCG TAA AAT TA	Bsal
VxexuT CF Rv Bsal	DVA_EF_ExuT	CTA CTA GTA GGT CTC TAG CGT TAA TGT TGC GGT GCG GG	Bsal
VxNanT CF fw Bsal	DVA_EF_NanT	AGA CTA GTG GGT CTC AAA TGA GTA CTA CA	Bsal

VxNanT_CF rv Bsal	DVA_EF_NanT	CTA CTA GTA GGT CTC TAG CGT TAA CTT TTG GTT TTG ACT AAA TCG TTT TTG GC	Bsal
VEAK2Fw	Assembly PCR	GCG CGG GGA AGA CGT AGG TAT GCT GAA GCT GCA GCG AAG GAA GCT GCA G	Bpil
VEAK2R	Assembly PCR	GCG GTG CAG AAG ACA TAA GCA CAG CCT TCG CTG CAG CTT CCT TCG CTG CAG	Bpil
VFIFw	Assembly PCR Flexible linker	GCG CGG GGA AGA CGT AGG TAT GAA GGT AAA AGC AGC GGT AGT GGT AGC GA	Bpil
VFIRev	Assembly PCR Flexible linker	CGG TGC AGA AGA CAT AAG CAC GGT AGA TTT GCT TTC GCT ACC ACT ACC GCT GCT TTT	Bpil
VProFw	Assembly PCR Polyproline linker	GCG CGG GGA AGA CGT AGG TAT AAG CCA GCT CCT CAA CCG AAG CCA GCT CCT AAG	Bpil
VProR	Assembly PCR polyproline linker	GCG GTG CAG AAG ACA TAA GCA CTG GCT TTG GCT TAG GAG CTG GCT TCG GTT GAG GA	Bpil
VpEAKFw	Assembly PCR LnArai linker	GCT ATA AGA AGA CGT AGG TAT GAA GCT GCA GCA AAA GAA GCT GCA GCA AAA GAA GCT GCA G	Bpil
VpEAKR	Assembly PCR LnArai linker	GTA GAA GAC ATA AGC ACT TTC GCT GCA GCT TCT TTT GCT GCA GCT TCT TTT GCT GCA GC	Bpil
VEAK2Fw	Assembly PCR ShArai linker	GCG CGG GGA AGA CGT AGG TAT GCT GAA GCT GCA GCG AAG GAA GCT GCA G	Bpil

VEAK2R	Assembly PCR ShArai linker	GCG GTG CAG AAG ACA TAA GCA CAG CCT TCG CTG CAG CTT CCT TCG CTG CAG	Bpil
VpWfW	Assembly PCR Waldo linker	GCG CGG AGA AGA CGT AGG TAT GGA TCC GCT GGC TCC GCT GCT GGT TCT G	Bpil
VpWR	Assembly PCR Waldo linker	GCG GTG CAG AAG ACA TAA GCA CGA ATT CGC CAG AAC CAG CAG CGG AGC CAG	Bpil
VpGSFw	Assembly PCR Poly(GS) linker	GCG CGG GGA AGA CGT AGG TAT GGT GGC GGA GGT TCT GGT GGT GGT GGT TCT GGT GGC GGT G	Bpil
VpGSR	Assembly PCR Poly(GS) linker	CGG TGC AGA AGA CAT AAG CAC AGA ACC ACC GCC ACC AGA ACC ACC GCC ACC AGA ACC ACC ACC ACC AGA AC	Bpil
VE040Fw	E0040m_CD	TCA CAG GGA AGA CAC GCT TAT GCG TAA AGG AGA AGA ACT TTT CAC T	Bpil
VE040Rev	E0040m_CD	TCA CAG GGA AGA CAC AGC GTT ATT ATT TGT ATA GTT CAT CCA TGC CAT GTG T	Bpil
VE0040CF Bsal	E0040m_CD	AGA CTA GTG GGT CTC AAA TGA TGC GTA AAG GAG AAG AAC TTT TCA CT	Bsal
VE040Rev Bsal	E0040m_CD	CTA CTA GTA GGT CTC TAG CGT TAT TAT TTG TAT AGT TCA TCC ATG CCA TGT GT	Bsal
VNanT Fw	<i>E. coli</i> MG1655 nanT	TCA CAG GGA AGA CAC AAT GAG TAC TAC AAC CCA GAA T	Bpil
MutaNanT Rv	<i>E. coli</i> MG1655 nanT	GAG AAG ACT CTT TCC AAT CTT CCG CTT	
MutaNanT Fw	<i>E. coli</i> MG1655	GGA AGA CTG GAA AGA GAA ACA CG	

	nanT		
VnanT Rv2	<i>E. coli</i> MG1655 nanT	TCA CAG GGA AGA CAC ACC TAC TTT TGG TTT TGA CTA AAT CGT TTT TG	Bpil
pKD46 Ori Fw	pKD46	GCT CAC TCA AAG GCG GTA ATC ATG GGT ATG GAC AGT TTT CC	
pKD46 Ori Rev	pKD46	GAG CTG TCA GAC CAA GTT TAC TCA AGG ATC TAG GTG AAG ATC CTT TT	
FIXyIBFw	<i>E. coli</i> MG1655 xyIB	TCA CAG GGA AGA CAC ACC TCG CCA TTA ATG GCA GAA GTT GCT G	Bpil
FIXyIBRev#2	<i>E. coli</i> MG1655 xyIB	TCA CAG GGA AGA CAC AAT GTA TAT CGG GAT AGA TCT TGG CAC CTC GG	Bpil
VxFifw CF	DVA_EF_XylE	CTA CTA GTA GGT CTC TAG CGC AGC GTA GCA GTT TGT TGT GTT TTC TT	Bsal
VxFlipR CF	DVA_EF_XylE	AGA CTA GTG GGT CTC AAA TGA ATA CCC AGT ATA	Bsal
VExFw	<i>E. coli</i> MG1655 exuT	TCA CAG GGA AGA CAC AAT GCG TAA AAT TAA AGG GTT ACG TTG GTA TAT GAT CG	Bpil
VExRev2	<i>E. coli</i> MG1655 exuT	TCA CAG GGA AGA CAC ACC TAT GTT GCG GTG CGG GAT CGT	Bpil
pET28 seq For	pEt28aa-BirA	CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CC	
pET28 seq Rev	pEt28aa-BirA	GCT TGT CGA CGG AGC TCG AAT	
MutBirA Fw	pEt28aa-BirA	AGA CGT TCT GAG GAT CCG AAT TCG AGC TCC GTC	

MutBirA Rev	pEt28aa-BirA	GAT CCT CAG AAC GTC TCA GGC TGA ACT CTC CC	
FIXAbpFw	<i>E. coli</i> MG1655 xylA	CTG AAG ACT TCA GGA TCT GGG TTC GTC GCC G	Bpil
FIXAbpRev	<i>E. coli</i> MG1655 xylA	GAG AAG ACA TCC TGA AGT CCT CAG CTG GGC GGC AA	Bpil
FIXylAFw	<i>E. coli</i> MG1655 xylA	TCA CAG GGA AGA CAC ACC TTT TGT CGA ACA GAT AAT GGT TTA CCA GAT TTT CC	Bpil
FIXylARev2	<i>E. coli</i> MG1655 xylA	CAC AGG GAA GAC ACA ATG CAA GCC TAT TTT GAC CAG CTC GAT CGC GT	Bpil
VE040Rev 3 (w/o stop)	E0040m_CD	TCA CAG GGA AGA CAC AGC GTT TGT ATA GTT CAT CCA TGC CAT GTG T	Bpil
Vdva_fg Fw	DVA_EG	GTG GGT CTC ACG CTA TGT CTT CTG CAC CAT ATG CGG TGT G	
Vdva_fg Rv	DVA_EG	GAC ATA GCG TGA GAC CCA CTA GTC TCT AGA AGC GGC CGC	
VAN4_FG Fw	Assembly PCR	GCG CGG GGA AGA CGT CGC TAT GGT AGT GGC GAA ATC GCT GCG CTC GAA CAG GAA ATT GCT GCG CTG GAG AAG GAG AAT GCA GCC TTG	Bpil
VAN4_FG Rev	Assembly PCR	CGG TGC AGA AGA CAT GGC AAC CTA CCC CTG TTC CAG TGC GGC AAT TTC CCA TTC CAA GGC TGC ATT CTC CTT CTC CAG CGC	Bpil
VBN4_FG Fw	Assembly PCR	GCG CGG GGA AGA CGT CGC TAT GGT AGT GGC AAG ATT GCT GCG CTG AAA CAG AAG ATC GCC GCA CTG AAA TAC AAG AAT GCG GCC CTG	Bpil

VBN4_FG Rev	Assembly PCR	CGG TGC AGA AGA CAT GGC AAC CTA ACC CTG TTT CAG TGC CGC GAT CTT CTT TTT CAG GGC CGC ATT CTT GTA TTT CAG TGC	Bpil
VAN4_EF Fw	DVA_FG_AN4	TCA CAG GGA AGA CAC GCT TGA AAT CGC TGC GCT CGA AC	Bpil
VAN4_EF Rev2	DVA_FG_AN4	TCA CAG GGA AGA CAC AGC GCC CCT GTT CCA GTG CGG	Bpil
VBN4_EF Fw	DVA_FG_BN4	TCA CAG GGA AGA CAC GCT TAA GAT TGC TGC GCT GAA ACA G	Bpil
VBN4_EF Rev2	DVA_FG_BN4	TCA CAG GGA AGA CAC AGC GAC CCT GTT TCA GTG CCG C	Bpil
VpLC3 blunt end fw	pLC1	TGC CAT GTC TTC TAC TAG TAG CGG	
VpLC3 blunt end Rv	pLC1	AGA GAC CAC CGC GCG	
VctCoh Fw EF	C. thermocellum Cohesin gBlock	TCA CAG GGA AGA CAC GCT TAT GGC CTC TGA TGG CGT G	Bpil
VctCoh rev EF - no stop2	C. thermocellum Cohesin gBlock	TCA CAG GGA AGA CAC AGC GTT TTG TCG GCG TAG CTG TAT TTG	Bpil
VctCoh rev EF – stop2	C. thermocellum Cohesin gBlock	TCA CAG GGA AGA CAC AGC GTC ATT TTG TCG GCG TAG CTG TAT TTG	Bpil
VctDoc Fw EF	C. thermocellum Dockerin gBlock	TCA CAG GGA AGA CAC GCT TAT GGG CGA TGT AAA CGG TGA	Bpil
VctDoc rev EF - no stop2	C. thermocellum Dockerin gBlock	TCA CAG GGA AGA CAC AGC GGA TTA CTC TTA AAA GGT AGC GTG AAA GAA G	Bpil
VctDoc rev EF – Stop2	C. thermocellum Dockerin gBlock	TCA CAG GGA AGA CAC AGC GTC AGA TTA CTC TTA AAA GGT AGC GTG AAA GAA G	Bpil

VKanpET28 QF Fw2	pEt28a	CGG TAG CAC TTG GGT TGC GCT TTC TAC GGG GTC TGA CGC T	
VKanpET28 QF rev	pEt28a	CGA GTT GAT CGG GCA CGT AAG GGT GGC ACT TTT CGG GGA AA	
VpLC3QF Fw	pLC3	TTT CCC CGA AAA GTG CCA CCT TAC GTG CCC GAT CAA CTC G	
VpLC3QF Rev2	pLC3	GCC GCT CGC GGC CAT CCG AGC GCA GCG AGT CAG	
VpBBR1 rep Fw	pBBR1MCS	CTG ACT CGC TGC GCT CGG ATG GCC GCG AGC GGC	
VpBBR1 rep rv	pBBR1MCS	AGC GTC AGA CCC CGT AGA AAG CGC AAC CCA AGT GCT ACC G	
VxylA_EF Fw	DVA_CD_ FIXylA	TCA CAG GGA AGA CAC GCT TAT GCA AGC CTA TTT TGA CC	Bpil
VxylA_EF Rev	DVA_CD_ FIXylA	TCA CAG GGA AGA CAC AGC GTT TGT CGA ACA GAT AAT GGT TTA	Bpil
VXylA_CF Fw	DVA_CD_ FIXylA	CTA GTG GGT CTC AAA TGC AA	Bsal
VXylA_CF rev Bsal	DVA_CD_ FIXylA	CTA CTA GTA GGT CTC TAG CGT TTG TCG AAC AGA TAA TGG TTT A	Bsal
VCtCoh rev CD	DVA_EF_ctCoh	TCA CAG GGA AGA CAC ACC TTT TTG TCG GCG TAG CTG TAT TTG	Bpil
VctCoh Fw CD	DVA_EF_ctCoh	TCA CAG GGA AGA CAC AAT GGC CTC TGA TGG CGT G	Bpil
VpLD RVseq	pLD1/pLD3	GAG GAA GCC TGC ATA ACG C	
VpolyHis6 FG assemb Fw	Duplex Assembly PCR	CGC TCT GAA CTG TAT AAA CAT GAT GAA CTG CAT CAC CAT CAT CAT CAC TAA GT	

VpolyHis6 FG assemb Rev	Duplex Assembly PCR	GGC AAC TTA GTG ATG ATG ATG GTG ATG CAG TTC ATC ATG TTT ATA CAG TTC AG	
VAN4 CD Fw	DVA_FG_AN4	TCA CAG GGA AGA CAC AAT GGA AAT CGC TGC GCT CGA AC	Bpil
VAN4 CD Rv	DVA_FG_AN4	TCA CAG GGA AGA CAC ACC TCC CCT GTT CCA GTG CGG	Bpil
VBN4 CD Fw	DVA_FG_BN4	TCA CAG GGA AGA CAC AAT GAA GAT TGC TGC GCT GAA ACA G	Bpil
VBN4 CD Rv	DVA_FG_BN4	TCA CAG GGA AGA CAC ACC TAC CCT GTT TCA GTG CCG C	Bpil
VmCherry Bpil Fw EF	pPT100-IM9- mCherry or pAS84-87 (PAmCherry)	TCA CAG GGA AGA CAC GCT TGT GAG CAA GGG CGA GGA G	Bpil
VmCherry Bpil mutcut Fw	pPT100-IM9- mCherry or pAS84-87 (PAmCherry)	GAA GAA GAC CAT GGG CTG GG	Bpil
VmCherry Bpil mutcut Rv	pPT100-IM9- mCherry or pAS84-87 (PAmCherry)	GAG AAG ACA TCC CAT GGT TTT CTT CTG CAT TAC GGG GC	Bpil
VmCherry Bpil Rv EF	pPT100-IM9- mCherry or pAS84-87 (PAmCherry)	TCA CAG GGA AGA CAC AGC GTT ACT TGT ACA GCT CGT CCA TGC	Bpil
VmCherry CD Fw	DVA_EF_ mCherry	TCA CAG GGA AGA CAC AAT GGT GAG CAA GGG CGA GGA GGA T	Bpil

VmCherry CD Rv	DVA_EF_ mCherry	TCA CAG GGA AGA CAC ACC TCT TGT ACA GCT CGT CCA TGC CG	Bpil
VCcelloDoc Fw EF	gBlock for <i>Clostridium cellulolyticum</i> dockerin and cohesin	TCA CAG GGA AGA CAC GCT TAT GAT TGT TTA TGG TGA CTA CAA TAA CGA T	Bpil
VCcelloDoc rev EF - no stop	gBlock for <i>Clostridium cellulolyticum</i> dockerin and cohesin	TCA CAG GGA AGA CAC AGC GCA TTC CCA GAA GGT ATT TCT TTA AAA TAG C	Bpil

Table 2.6. List of primers used in this work.

Plasmid	Description	Resistance	Source
DVA_AB	Level 0 Golden Gate destination vector	Amp	(Iverson <i>et al.</i> , 2016)
DVA_BC	Level 0 Golden Gate destination vector	Amp	(Iverson <i>et al.</i> , 2016)
DVA_CD	Level 0 Golden Gate destination vector	Amp	(Iverson <i>et al.</i> , 2016)
DVA_DE	Level 0 Golden Gate destination vector	Amp	(Iverson <i>et al.</i> , 2016)
DVA_EF	Level 0 Golden Gate destination vector	Amp	(Iverson <i>et al.</i> , 2016)
DVA_EG	Level 0 Golden Gate destination vector	Amp	(Iverson <i>et al.</i> , 2016)
DVA_FG	Level 0 Golden Gate destination vector	Amp	This work
DVK_AF	Level 1 Golden Gate destination vector	Kan	(Iverson <i>et al.</i> , 2016)
pWKS30	Low copy number	Amp	(Rong Fu Wang and Kushner, 1991)
pSB1C3	High copy BioBrick assembly plasmid	Cam	(iGEM, 2015)
pLC1_AF	Low copy number level 1 Golden Gate destination vector	Cam	This work
pKD46	Recombineering plasmid with temperature sensitive origin of replication	Amp	(Datsenko and Wanner, 2000)
pLC2_AF	Temperature sensitive, Low copy number level 1	Cam	This work

	Golden Gate destination vector		
pLC3_AG	Low copy number level 1 Golden Gate destination vector	Cam	This work
pEt28a-BirA	A cloning and expression plasmid. Target genes under control of strong bacteriophage T7 polymerase. Gene expressing BirA embedded in cloning site.	Kan	BirA in pET28a (w400-2) was a gift from Eric Campeau (Addgene plasmid # 26624 ; http://n2t.net/addgene:26624 ; RRID:Addgene_26624) provided by David-Paul Minde.
pBBR1MCS-2	Low copy number plasmid suitable for blue-white screening	Kan	(Kovach <i>et al.</i> , 1995)
pPT100-IM9-mCherry	Plasmid expressing an Im9 gene at the N-terminal of mCherry	Kan	
pAS84-87	pUC18-derived plasmid containing Photoactivatable mCherry	Kan	(Brett, 1989)
pLD1_AF	Low copy number level 1 Golden Gate destination vector	Kan	This work
pLD3_AG	Low copy number level 1 Golden Gate destination vector	Kan	This work
J23100_AB	DVA_AB vector containing iGem	Amp	(Iverson <i>et al.</i> , 2016)

	constitutive promoter BBa_J23100		
R0010_AB	DVA_AB vector containing a pLacI inducible promoter	Amp	(Iverson <i>et al.</i> , 2016)
I13453_AB	DVA_AB vector containing a pBAD araC regulated inducible promotor	Amp	(Iverson <i>et al.</i> , 2016)
B0034m_BC	DVA_BC vector containing a high strength Weiss RBS with a mutation to adjust spacing in MC system	Amp	(Iverson <i>et al.</i> , 2016)
E0040m_CD	DVA_CD containing iGem biobrick GFP Bba_B0040 modified to fix illegal site	Amp	(Iverson <i>et al.</i> , 2016)
DVA_CD_FIXylEm# 5	DVA_CD vector containing <i>E. coli xylE</i> gene with a mutation to remove BsaI binding site	Amp	This work
DVA_CD_ExuT#2	DVA_CD vector containing <i>E. coli exuT</i> gene	Amp	This work
DVA_CD_NanT#2	DVA_CD vector containing <i>E. coli nanT</i> gene with BsaI site removed	Amp	This work
DVA_CD_XylA	DVA_CD vector containing <i>E. coli</i> gene <i>xylA</i>	Amp	This work

DVA_CD_XyIB	DVA_CD vector containing <i>E. coli xyIB</i> gene	Amp	This work
DVA_CD_ctCoh	DVA_CD vector containing <i>Clostridium thermocellum</i> cohesin	Amp	This work
DVA_CD_mCherry	DVA_CD vector containing mCherry	Amp	This work
DVA_DE_Wld	DVA_DE vector containing the waldo linker	Amp	This work
DVA_DE_flex	DVA_DE vector containing the flexible linker	Amp	This work
DVA_DE_pro	DVA_DE vector containing the polyproline linker	Amp	This work
DVA_DE_EAK	DVA_DE vector containing the long Aria linker	Amp	This work
DVA_DE_(EAK)2	DVA_DE vector containing the short Aria linker	Amp	This work
DVA_DE_poly GS	DVA_DE vector containing the poly glycine-serine linker	Amp	This work
DVA_EF_E0040m	DVA_EF vector containing BBa_B0040m	Amp	This work
DVA_EF_GFPm	DVA_EF vector containing BBa_B0040m without a stop codon	Amp	This work

DVA_EF_mCherry	DVA_EF vector containing mCherry	Amp	This work
DVA_EF_PAmCherry	DVA_EF vector containing Photoactivatable mCherry	Amp	This work
DVA_EF_BirA	DVA_EF vector containing BirA	Amp	This work
DVA_EF_XylA	DVA_EF vector containing <i>E. coli xylA</i> gene	Amp	This work
DVA_EF_ctDoc	DVA_EF vector containing <i>Clostridium thermocellum</i> dockerin	Amp	This work
DVA_EF_ctDoc stop	DVA_EF vector containing <i>Clostridium thermocellum</i> dockerin with a stop codon	Amp	This work
DVA_EF_ctCoh	DVA_EF vector containing <i>Clostridium thermocellum</i> cohesin	Amp	This work
DVA_EF_ctCoh stop	DVA_EF vector containing <i>Clostridium thermocellum</i> cohesion with a stop codon	Amp	This work
DVA_FG_His6	DVA_FG containing a hexa histidine tag	Amp	This work
DVA_FG_AN4	DVA_FG vector containing the coiled coil AN4	Amp	This work
DVA_FG_BN4	DVA_FG vector containing the coiled coil BN4	Amp	This work

pLC1_GFP	pLC1 expressing GFP with a J23100 promotor and B0034m RBS	Can	This work
pLC1_FIXylE	pLC1 expressing XylE with a J23100 promotor and B0034m RBS	Can	This work
pLC1_11111	pLC1 expressing XylE-wld-GFP with J23100 promotor	Cam	This work
pLC1_21111	pLC1 expressing XylE-wld-GFP with R0010 promotor	Cam	This work
pLC1_31111	pLC1 expressing XylE-wld-GFP with I13453 promotor	Cam	This work
pLC1_11211	pLC1 expressing ExuT-wld-GFP with J23100 promotor	Cam	This work
pLC1_31211	pLC1 expressing ExuT-wld-GFP with I13453 promotor	Cam	This work
pLC1_11311	pLC1 expressing NanT-wld-GFP with J23100 promotor	Cam	This work
pLC1_31311	pLC1 expressing NanT-wld-GFP with I13453 promotor	Cam	This work
pLC1_11411	pLC1 expressing XylB-wld-GFP with J23100 promotor	Cam	This work

pLC1_11511	pLC1 expressing DjlA-wld-GFP with J23100 promotor	Cam	This work
pLC1_11121	pLC1 expressing XylE-flex-GFP with J23100 promotor	Cam	This work
pLC1_11131	pLC1 expressing XylE-pro-GFP with J23100 promotor	Cam	This work
pLC1_11141	pLC1 expressing XylE-EAK-GFP with J23100 promotor	Cam	This work
pLC1_11151	pLC1 expressing XylE-(EAK)2-GFP with J23100 promotor	Cam	This work
pLC1_11161	pLC1 expressing XylE-polyGS-GFP with J23100 promotor	Cam	This work
pLC1_11112	pLC1 expressing XylE-wld-mCherry with J23100 promotor	Cam	This work
pLC1_11114	pLC1 expressing XylE-wld-BirA with J23100 promotor	Cam	This work
pLC1_31211	pLC1 expressing ExuT-wld-GFP with I13453 promotor	Cam	This work
pLC1_31221	pLC1 expressing ExuT-flex-GFP with I13453 promotor	Cam	This work

pLC1_31241	pLC1 expressing ExuT-EAK-GFP with I13453 promotor	Cam	This work
pLC1_31251	pLC1 expressing ExuT-(EAK)2-GFP with I13453 promotor	Cam	This work
pLC1_31311	pLC1 expressing NanT-wld-GFP with I13453 promotor	Cam	This work
pLC1_31321	pLC1 expressing NanT-flex-GFP with I13453 promotor	Cam	This work
pLC1_31361	pLC1 expressing NanT-polyGS-GFP with I13453 promotor	Cam	This work
pLC1_11116	pLC1 expressing XylE-wld-XylA with J23100 promotor	Cam	This work
pLC3_aXyIAH	pLC3 expressing HexHis tagged XylA with an arabinose promotor	Cam	This work
pLC3_aExuTH	pLC3 expressing HexHis tagged ExuT with an arabinose promotor	Cam	This work
pLC3_aNanTH	pLC3 expressing HexHis tagged NanT with an arabinose promotor	Cam	This work
pLC3_111111	pLC3 expressing XylE-wld-GFP-His ₆ with J23100 promotor	Cam	This work

pLC3_311161	pLC3 expressing XylE-wld-XylA-His ₆ with I13453 promotor	Cam	This work
pLC3_311261	pLC3 expressing XylE-flex-XylA-His ₆ with I13453 promotor	Cam	This work
pLC3_311361	pLC3 expressing XylE-pro-XylA-His ₆ with I13453 promotor	Cam	This work
pLC3_311561	pLC3 expressing XylE-EAK-XylA-His ₆ with I13453 promotor	Cam	This work
pLC3_311661	pLC3 expressing XylE-polyGS-XylA-His ₆ with I13453 promotor	Cam	This work
pLC3_311111	pLC3 expressing XylE-wld-GFP-His ₆ with I13453 promotor	Cam	This work
pLC3_111112	pLC3 expressing XylE-wld-GFP-AN4 with J23100 promotor	Cam	This work
pLC3_111113	pLC3 expressing XylE-wld-GFP-BN4 with J23100 promotor	Cam	This work
pLC3_111114	pLC3 expressing XylE-wld-GFP-His ₁₀ with J23100 promotor	Cam	This work
pLC3_111191	pLC3 expressing XylE-wld- <i>C. thermocellum</i> Doc-His ₆ with J23100 promotor	Cam	This work

pLC3_111291	pLC3 expressing XylE-flex- <i>C. thermocellum</i> Doc-His ₆ with J23100 promotor	Cam	This work
pLC3_111391	pLC3 expressing XylE-pro- <i>C. thermocellum</i> Doc-His ₆ with J23100 promotor	Cam	This work
pLC3_111591	pLC3 expressing XylE-EAK- <i>C. thermocellum</i> Doc-His ₆ with J23100 promotor	Cam	This work
pLC3_111691	pLC3 expressing XylE-polyGS- <i>C. thermocellum</i> Doc-His ₆ with J23100 promotor	Cam	This work
pLC3_11A161	pLC3 expressing GFP-wld-XylA-His ₆ with J23100 promotor	Cam	This work
pLC3_11A191	pLC3 expressing GFP-wld- <i>C. thermocellum</i> Doc-His ₆ with J23100 promotor	Cam	This work
pLD1_11111	pLD1 expressing XylE-wld-GFP with J23100 promotor	Kan	This work
pLD1_11711	pLD1 expressing <i>C. thermocellum</i> Coh-wld-GFP with J23100 promotor	Kan	This work
pLD1_31711	pLD1 expressing <i>C. thermocellum</i> Coh-wld-GFP with I13453 promotor	Kan	This work

pLD1_31112	pLD1 expressing XylE-wld-mCherry with I13453 promotor	Kan	This work
pLD1_31712	pLD1 expressing <i>C. thermocellum</i> Coh-wld-mCherry with I13453 promotor	Kan	This work
pLD1_31C1B	pLD1 expressing mCherry-wld- <i>C. thermocellum</i> Coh with an I13453 promotor	Kan	This work
pLD1_3111B	pLD1 expressing XylE-wld-PAmCherry with I13453 promotor	Kan	This work
pLD1_3171B	pLD1 expressing <i>C. thermocellum</i> Coh-wld-PAmCherry with I13453 promotor	Kan	This work
pLD3_317161	pLD3 expressing <i>C. thermocellum</i> Coh-wld-XylA-His ₆ with I13453 promotor	Kan	This work
pLD3_317261	pLD3 expressing <i>C. thermocellum</i> Coh-flex-XylA-His ₆ with I13453 promotor	Kan	This work
pLD3_317361	pLD3 expressing <i>C. thermocellum</i> Coh-polyPro-XylA-His ₆ with I13453 promotor	Kan	This work
pLD3_317461	pLD3 expressing <i>C. thermocellum</i> Coh-LnArai-	Kan	This work

	XylA-His ₆ with I13453 promotor		
pLD3_317561	pLD3 expressing <i>C. thermocellum</i> Coh-ShArai-XylA-His ₆ with I13453 promotor	Kan	This work
pLD3_317661	pLD3 expressing <i>C. thermocellum</i> Coh-poly(GS)-XylA-His ₆ with I13453 promotor	Kan	This work
pLD3_31B1A1	pLD3 expressing XylA-wld- <i>C. thermocellum</i> Coh-His ₆ with I13453 promotor	Kan	This work
pLD3_31B2A1	pLD3 expressing XylA-flex- <i>C. thermocellum</i> Coh-His ₆ with I13453 promotor	Kan	This work
pLD3_31B3A1	pLD3 expressing XylA-polyPro- <i>C. thermocellum</i> Coh-His ₆ with I13453 promotor	Kan	This work
pLD3_31B5A1	pLD3 expressing XylA-ShArai- <i>C. thermocellum</i> Coh-His ₆ with I13453 promotor	Kan	This work
pLD3_31B6A1	pLD3 expressing XylA-poly(GS)- <i>C. thermocellum</i> Coh-His ₆ with I13453 promotor	Kan	This work
pLD3_31C1A2	pLD3 expressing mCherry-waldo- <i>C.</i>	Kan	This work

	<i>thermocellum</i> Coh-AN4 with I13453 promotor		

Table 2.7. List of plasmids used in this work.

Chapter 3

Designing a Golden Gate assembly protocol to produce a library of membrane protein fusions to soluble proteins.

3. Designing a Golden Gate assembly protocol to produce a library of membrane protein fusions to soluble proteins.

3.1 Production of a Golden Gate expression vector suitable for membrane protein expression.

The over-expression of transporter-encoding genes can be toxic to the *E. coli* cell (Wagner *et al.*, 2006, 2007; Schlegel *et al.*, 2010; Gubellini *et al.*, 2011). Taking this into account, an expression vector for Golden Gate assembly suitable for membrane protein fusions was designed by incorporating features from other vectors routinely used for this purpose (Klesney-Tait *et al.*, 1997; Yoon *et al.*, 2011; Hopkins, Hawkhead and Thomas, 2013; Fischer *et al.*, 2015). An important consideration is the copy number of the plasmid being used. The CIDAR MoClo Golden Gate plasmids were designed for soluble protein expression, therefore have a high copy number (Iverson *et al.*, 2016). To overcome this issue, a new plasmid was designed using Biotool's QuickFusion kit, a one-step, ligation-independent cloning method akin to Gibson Assembly (Gibson *et al.*, 2009).

The plasmid DVK (Fig. 3.1) was selected from the CIDAR MoClo Golden Gate library, containing AF overlaps suitable for inserting up to 5 components (Iverson *et al.*, 2016). The overlaps, alongside an in-frame transcriptional terminator downstream of the plasmid, were cloned using PCR using primers VR and VF2 (see Table 2.6). DVK plasmids contain a kanamycin resistance gene, therefore an alternative antibiotic selection was needed to confirm assembly. As ampicillin resistance is used by Golden Gate holding plasmid DVA, a chloramphenicol resistance gene was identified as an alternative marker (Fig. 3.1). The relevant genes from plasmid pSB1C3 were therefore also amplified by PCR (Table 2.6).

Finally, a low copy number origin of replication was obtained by cloning the relevant region of pWKS30 using primers Fw Ori and Rv Ori (Table 2.6), which would yield six to eight plasmid copies per cell (Fig.3.1) (Wang *et al.*, 1991). The plasmid pWKS30 was selected as the Thomas lab had previous experience in expressing recombinant membrane proteins using this plasmid (Severi *et al.*, 2005) and it had previously been used by Waldo *et al.* as the basis for pWaldo (Waldo *et al.*, 1999), a plasmid suitable for the expression of membrane protein fusions, and therefore there was evidence the low copy number would support membrane protein expression for this work.

For all cloned regions of DNA, the amplicons contained short ~20bp regions up- and down-stream of the amplicon which overlapped with the coding region of another amplicon. Through designing the overlaps, the linear amplicons were positioned to assemble into a new plasmid once QuickFusion enzyme was added (Table 2.6 for details).

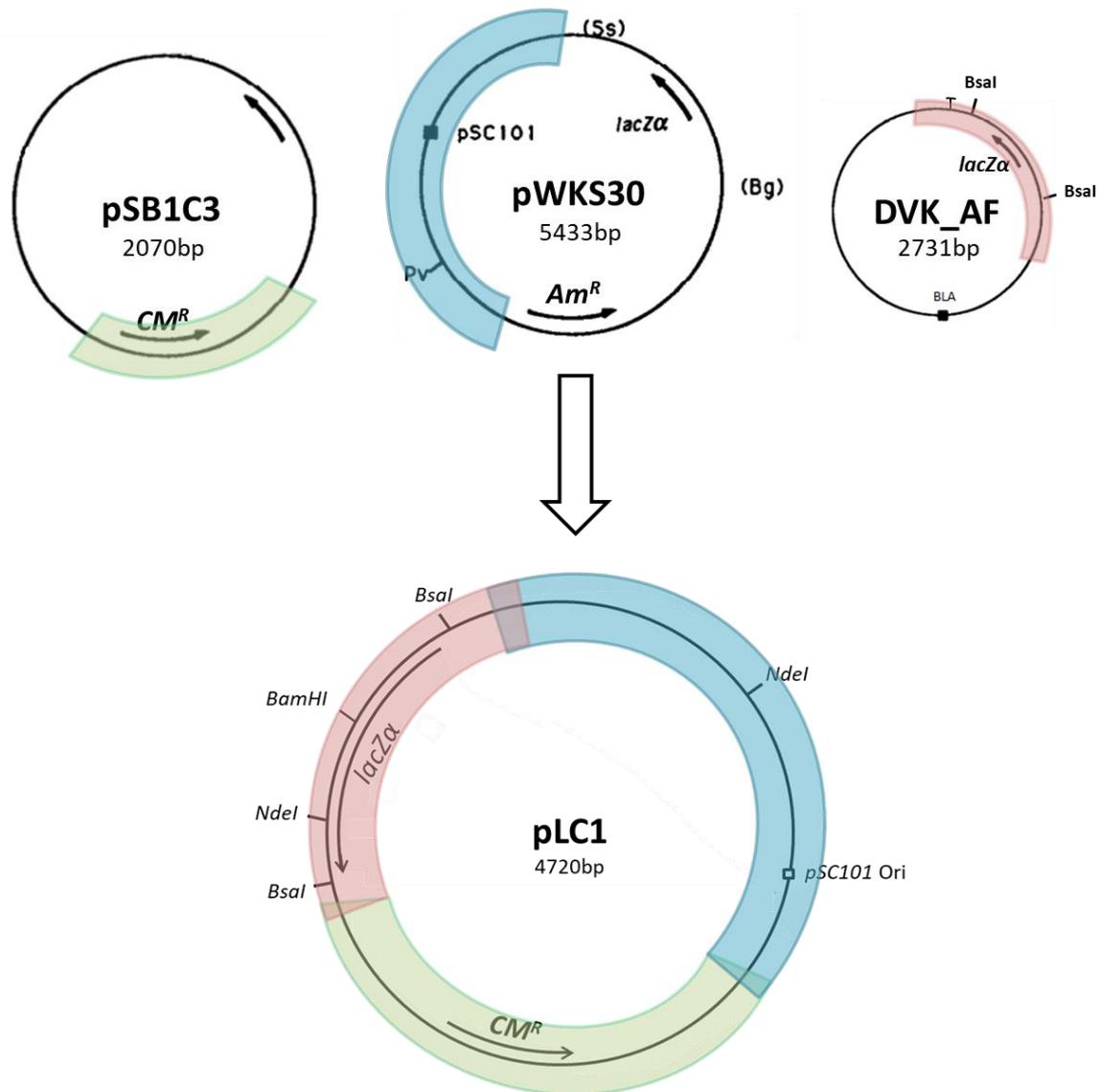


Figure 3.1. The components of pLC1, a very low copy number Golden Gate expression vector. The plasmid pLC1 was produced by QuickFusion cloning of three base components, a low copy number origin of replication (six to eight) was derived from pWKS30, which made the bulk of the new plasmid backbone (3004 bp). The Chloramphenicol resistance genes were derived from pSB1C3 (938 bp). Finally, the Golden Gate cloning region of CIDAR MoClo plasmid DVK_AF (841 bp) was used achieve subsequent insertion of Golden Gate compatible DNA fragments into the pLC1 plasmid. A more detailed plasmid map is found in Appendix A.

The size of each amplicon was confirmed by separating the DNA samples on an agarose gel using electrophoresis, with an appropriate DNA ladder. Three replicates of the origin of replication and backbone of pWKS30 were obtained at the expected 3004 bp size (Fig. 3.2A), and chloramphenicol resistance genes and Golden Gate cloning region at 938 bp and 841 bp respectively (Fig. 3.2B). Following transformation into DH5 α , the cells were grown on chloramphenicol plates containing with X-gal and IPTG to select for blue colonies. Blue colonies indicated expression of the *lacZ α* gene, therefore production of the desired plasmid. Two colonies were obtained, and the plasmids expressed and extracted from overnight cultures.

The produced plasmid, pLC1 (4720 bp), was validated by sequencing with primers VR and VF2, which should recognise either end of the DVK_AF region (Table 2.6). DNA sequencing results confirmed presence of the Golden Gate region in a chloramphenicol resistance plasmid. The size of the plasmid pLC1 was confirmed by restriction endonuclease digestion and gel electrophoresis to observe the expected restriction endonuclease digest pattern. Two enzymes were chosen to validate pLC1, NdeI and BamHI. The pLC1 plasmids should contain two NdeI cut sites, one in the *lacZ α* gene from DVK_AF and a second in the backbone from pWKS30. On the agarose gel, this resulted in two bands at 1848 bp and 2972 bp in length respectively, supporting assembly. On the other hand, there is only a single BamHI, cut site on pLC1, producing a band of 4720 bp in length (Fig. 3.2C).

Upon validating production of pLC1, additional Golden Gate expression vectors suitable for membrane protein expression were produced by using alternative components within the QuickFusion cloning step (Table 3.1). In particular the pLD-family of compatible vectors was produced in order to facilitate co-expression of two proteins simultaneously.

Plasmid pLC2 was produced in order to introduce a temperature sensitive origin of replication derived from pKD46, which could be used to stop expression of proteins at 37°C. The variant pLC3, derived from pLC1, was produced in order to have an A and G overhang so that six DNA fragments could be inserted into the plasmid – this was of particular interest to His tagging proteins.

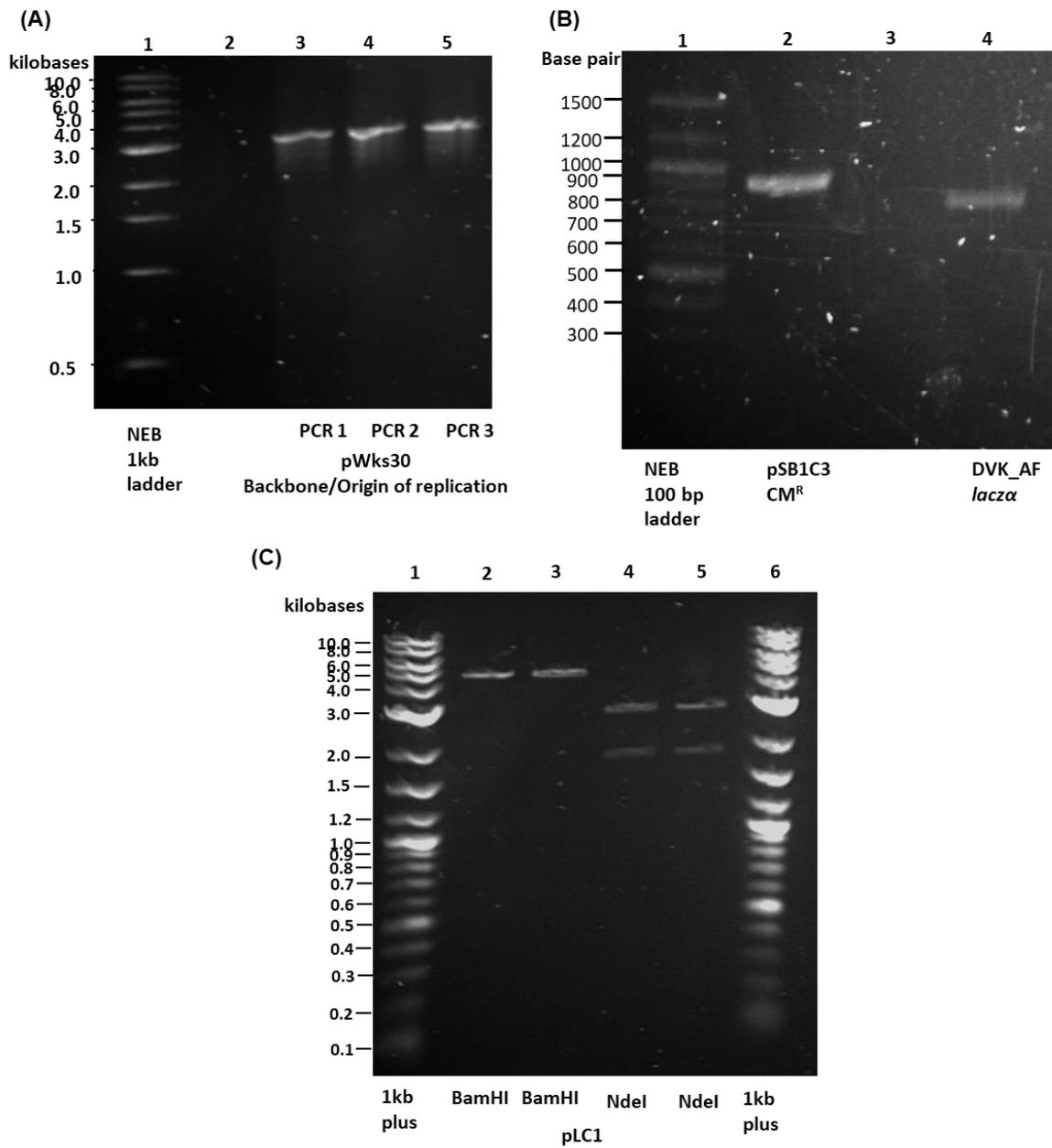


Figure 3.2. Electrophoresis gels of PCR amplicons of desired regions of the plasmids pWKS30, pSB1C3 and DVK_AF. (A) Three candidate PCR amplicons of the origin of replication plasmid pWKS30. (B) PCR amplicons encoding chloramphenicol resistance of PSB1C3 and Golden Gate cloning region of DVK_AF. (C) 0.5% agarose gel of pLC1 cut by either NdeI or BamHI-HF – duplicates displayed side by side. Separated alongside appropriate New England Biolab ladder, DNA was mixed with 6x loading buffer in a 1:5 ratio of buffer to DNA.

Name	Part 1	Part 2	Part 3
pLC1	pWKS30 – low copy number origin of replication	pSB1C3 - Chl ^R	DVK_AF – Golden Gate cloning region
pLC2	pKD46 – temperature sensitive origin of replication	pSB1C3 - Chl ^R	DVK_AF – Golden Gate cloning region
pLC3	pWKS30 - low copy number origin of replication	pSB1C3 - Chl ^R	DVK_AG – Golden Gate cloning region

Table 3.1. Components of the pLC series Golden Gate expression plasmids. The Golden Gate expression vectors, produced by QuickFusion of three amplicon components, are listed alongside which parts were used.

3.2 Testing the function of Golden Gate cloning in novel expression vector pLC1

With expression vector pLC1 produced, it was important to confirm that the plasmid would function correctly in a Golden Gate assembly reaction. In order to validate that pLC1 could be used for Golden Gate assembly to create protein fusions, four components of the CIDAR MoClo kit were inserted into pLC1 and the sequence of the resulting plasmid and expression of the protein was checked. The red fluorescent protein, E1010m was selected as it would allow confirmation of expression through epifluorescence microscopy. In order to express E1010m, a promoter, RBS and terminator were also selected from the CIDAR MoClo kit to create a complete A to F loop.

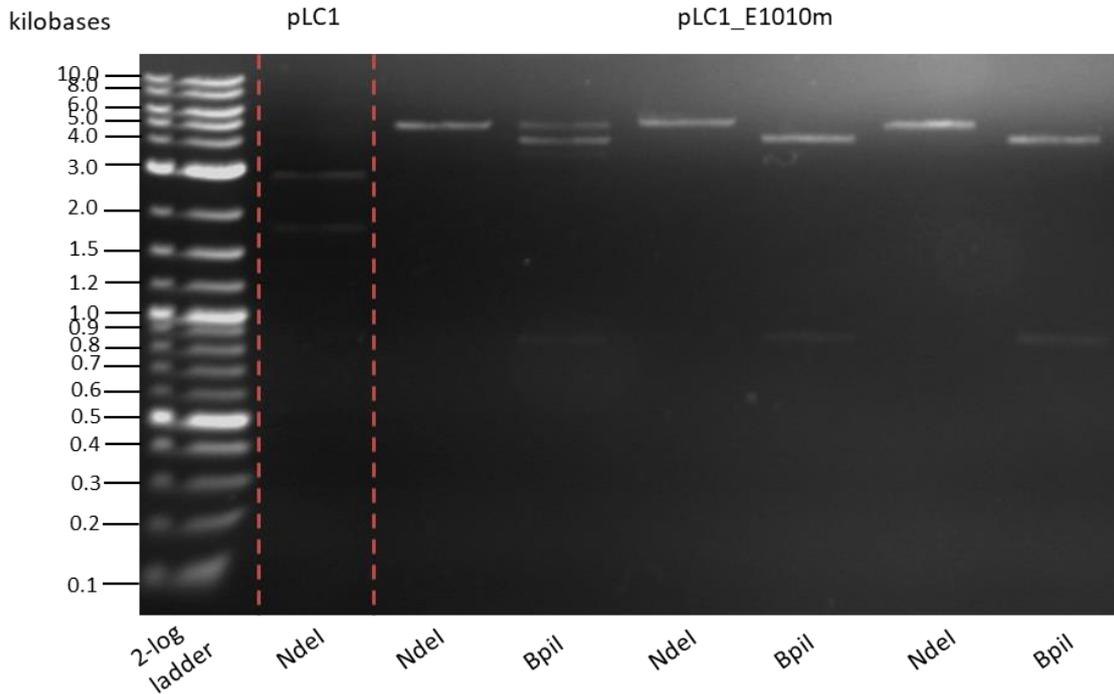
The success of the Golden Gate cloning was confirmed by preparing plasmid from white colonies identified after the assembly process and observing the restriction enzyme digest pattern, with NdeI and BpiI selected due to producing different restriction enzyme digest patterns.

NdeI results in a double cut in pLC1 but only a single cut in pLC1_E1010m, the RFP expressing plasmid. This can be observed in the agarose gel, in which a band at 5102 bp is seen for the pLC1_E1010m plasmid cut by NdeI, compared to the two bands for pLC1, 1848 and 2872 bp respectively (Fig. 3.3). This band length also matches expected size of the plasmid containing the four CIDAR components.

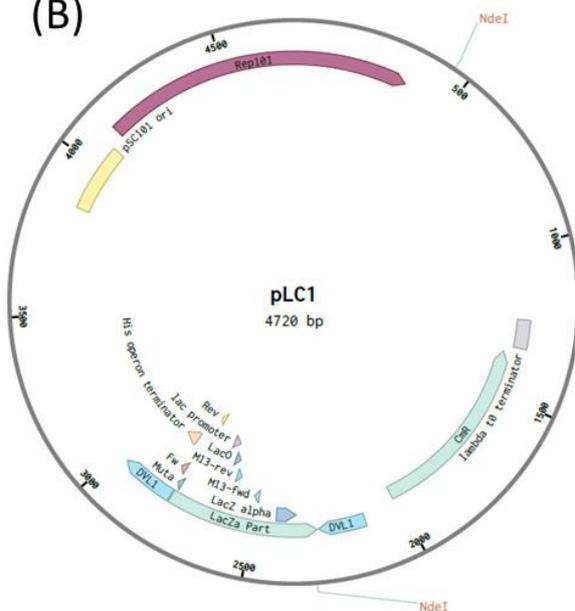
The complete restriction digest of BpiI cut pLC1_E1010m twice, produced two observable bands consistent with the expected sizes of 4220 bp and 882 bp, one restriction enzyme binding site located in the pLC1 plasmid backbone and the second present in the inserted Golden Gate components respectively. In the first well of pLC1_E1010m cut with BpiI it should be noted that a band at 5102 bp can also be seen, suggesting incomplete enzyme digest (Fig. 3.3).

With successful cloning of the pLC1_E1010m plasmid confirmed, the encoded protein was then expressed in DH5 α under control of a constitutive promoter and observed under an epifluorescence microscope. The presence of cells were checked under a confocal microscope at 1000x magnification (Fig. 3.4A and 4C respectively) and compared to epifluorescence imaging. Expression of pLC1 alone did not result in any notable fluorescence when cells were observed under an epifluorescence microscope (Fig. 3.4B), however red fluorescence could be seen when cells were transformed with pLC1_E1010m (Fig. 3.4D). Therefore, the Golden Gate components had been successfully cloned into the plasmid in the correct order for expression, supporting the use of pLC1 as an expression vector for protein fusions going forward.

(A)



(B)



(C)

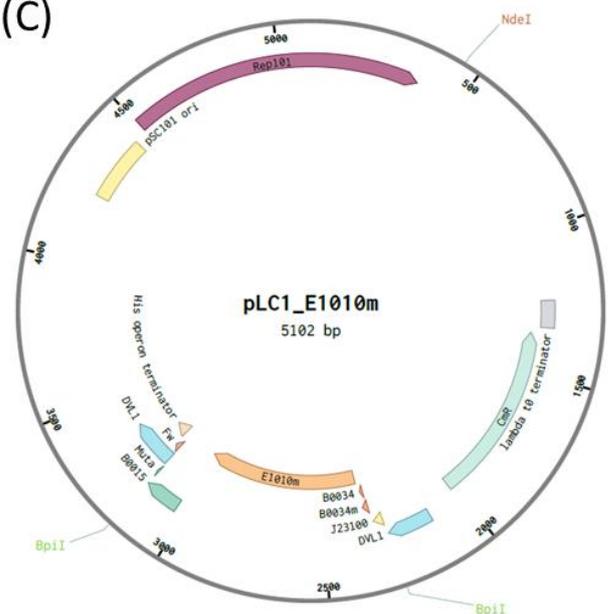


Figure 3.3. Electrophoresis gel showing the enzyme digest pattern of pLC1 and pLC1_E1010m. (A) 0.5% (w/v) agarose gel of pLC1 and pLC1_E1010m cut by either NdeI or BpiI – Triplicate cuts with NdeI and BpiI displayed side by side. DNA was separated alongside an appropriate New England Biolab ladder, DNA was mixed with 6x loading buffer in a 1:5 concentration of buffer to DNA. **(B)** The plasmid map of pLC1 with the NdeI cut sites. **(C)** The plasmid map of pLC1_E1010m with NdeI and BpiI cut sites.

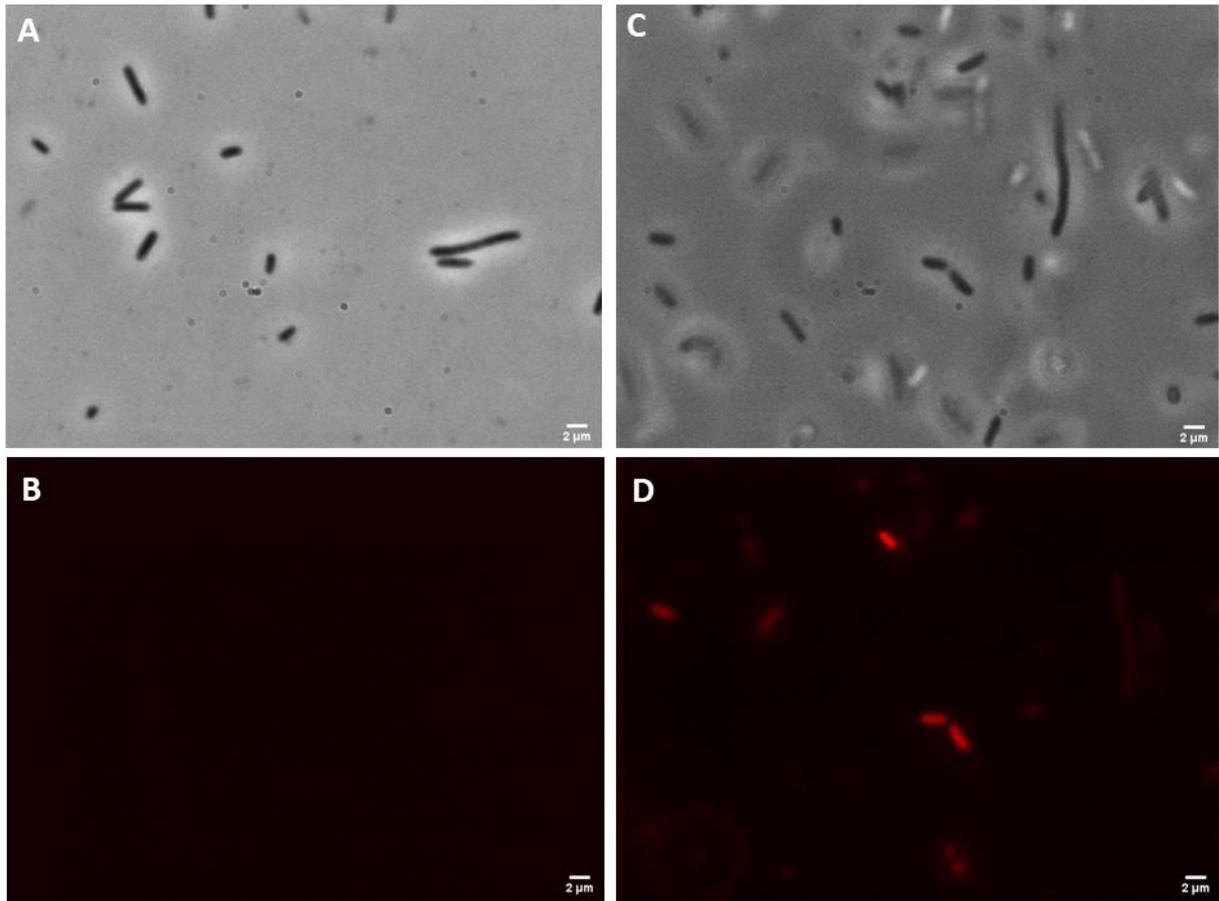


Figure 3.4. Confocal and epifluorescence microscopy images of transformed *E. coli* strain DH5α. Confocal microscopy completed using a Zeiss Axiolioskop 40 microscope, cells analysed under 1000x magnification **(A)** Phase contrast image of DH5α transformed by pLC1 **(B)** epifluorescence image of DH5α transformed by pLC1 at a bandpass of 605/70, Pseudocoloured by image programme Fiji and adjusted for contrast, **(C)** Phase contrast image of DH5α transformed by pLC1_E1010m, **(D)** epifluorescence image of DH5α transformed by pLC1_E1010m, at a bandpass of 605/70, Pseudo-coloured by image programme Fiji and adjusted for contrast.

3.3 Naturally occurring Major Facilitator Superfamily protein fusions demonstrates a need for protein linkers.

To help in our design process for making functional fusions between sugar transporters and the catabolic enzymes used for their breakdown, an analysis of existing examples of membrane transporter fused to other domains was undertaken. This was work undertaken in collaboration with Dr. Benjamin Willson and Ms Lindsey Dalzell (Willson *et al.*, 2019). The scope of the analysis was to study the largest family of secondary transporters, being the major facilitator superfamily (MFS) (Saier *et al.*, 1999) and using the EBI Interpro database to identify and then study examples of naturally occurring MFS fusions. A first observation - that is important to synthesis assembly methods for making fusions - was that almost all natural examples of fusions to MFS proteins were C-terminal (Fig. 3.5). Eight key groups of fusions were identified demonstrating an abundance of fusions at the C-terminal (Fig. 3.5)

Peptide linkers can be vital in stabilising synthetic protein fusions, therefore key structural features of the linkers in natural fusions may give insight into design considerations. Considering the abundance of the MFS fusions, it was of interest whether linker sequences were retained both between and within the groups of fusions identified. Repeating motifs may be important to consider for subsequent linker design.

Two of the MFS fusion groups were analysed in further detail in order to understand the role of the linkers spacing the upstream MFS region to the downstream soluble domain. The MFS-PlsC fusions, designated as group 2 in (Willson *et al.*, 2019), were selected due to the broad range of species expressing the fusion, whereas the MFS-CBS fusions, designated as group 7, were selected as the fusion was predominantly conserved within the Actinobacteria family.

For both types of fusion, examples were collected from all organisms suggested by EBI to contain a fusion, according to genus and species. Of this listing, the most abundant 10 genera were selected and at least one species from each genera was chosen to study in further detail. Following selection of the species of interest, the phyla of the species were also recorded to determine the distribution of the fusions in the domain of bacteria.

The amino acid sequence of the selected MFS fusions were then obtained using Uniprot (Bateman, 2019), which identifies key features of each fusion, with amino acids correlating to the transmembrane helix and subsequent downstream domain being most relevant to this work. The linker sequence within each fusion was defined as the region between the final amino acid of the transmembrane helices recognised by Uniprot, and the first amino acid of the downstream domain. Where downstream domains were not recognised by Uniprot, a different species was selected.

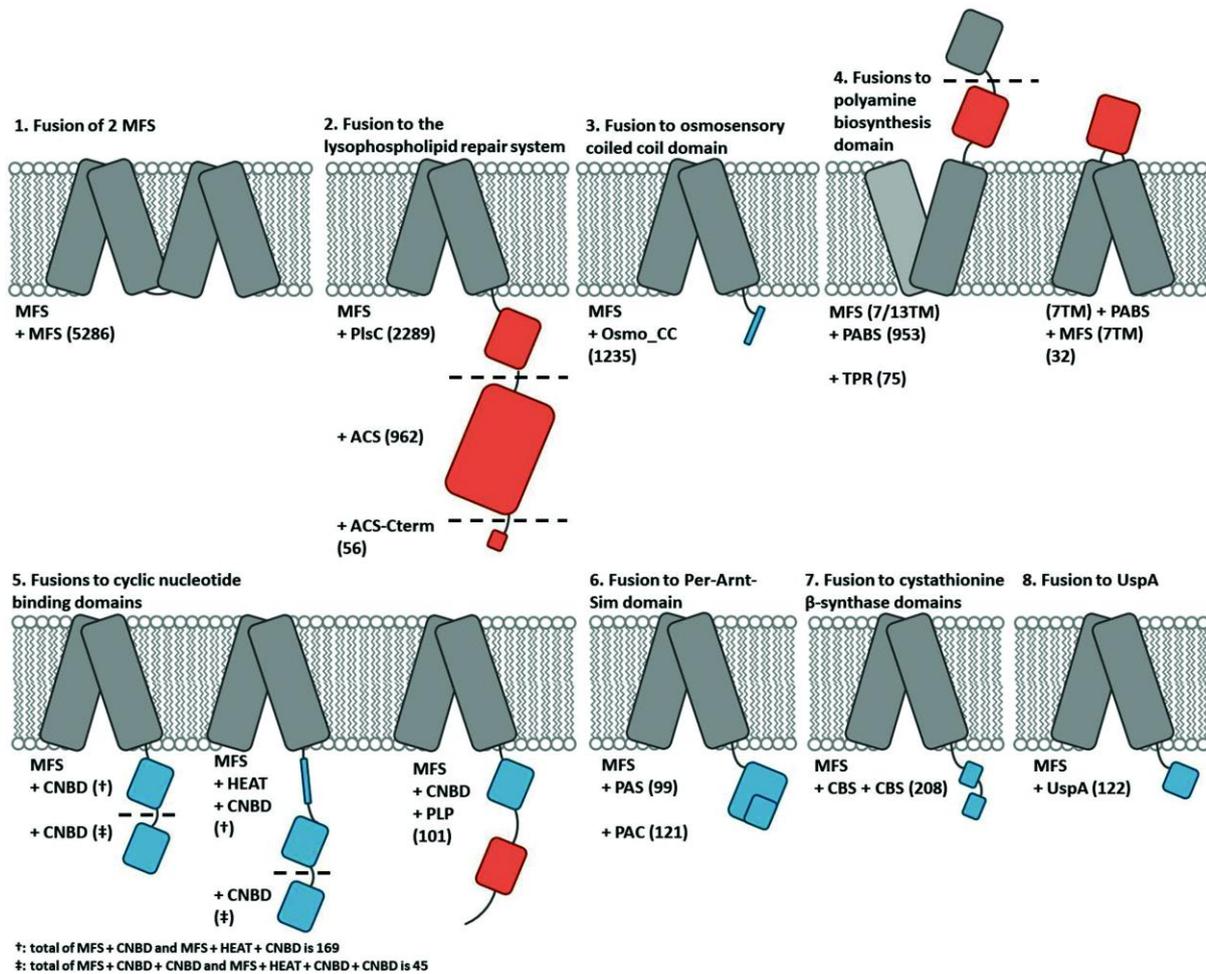


Figure 3.5. Schematic diagram of the eight most common groups of MFS fusion proteins in bacteria. Dashed lines indicate domain boundaries where multiple architectures are observed. Numbers in parentheses indicate the numbers of results in InterPro for each domain architecture. For group 5, numbers of HEAT repeat containing and non-HEAT repeat containing architectures are combined (annotated as † and ‡) as these are not separated in InterPro. Blue highlights represent regulatory domains, and red highlights represent catalytic domains. The results presented were accurate as of 07/03/18. Figure taken from (Willson *et al.*, 2019)

3.3.1 The MFS-PlsC fusions demonstrate significant sequence similarity in the linker region

Of the MFS protein fusions fused to soluble domains, the MFS-PlsC fusion proteins were found to be the most abundant throughout nature, with 3,307 examples of the architecture recorded within Interpro [as of 7th March 2018] (Willson *et al.*, 2019). The fusion appears especially abundant throughout the γ -proteobacteria families – with only the Enterobacteriaceae containing unfused MFS and PlsC domains.

Due to the high abundance of this fusion type it was interesting to determine whether the linkers were conserved. Twenty MFS-PlsC fusions were selected from the domain architectures in order to determine the nature of the linkers throughout bacteria (Table 3.2). The secondary structures were predicted by Jpred (Drozdetskiy *et al.*, 2015) and the amino acid sequence of the group γ -proteobacteria linkers were then aligned using ClustalW (Sievers *et al.*, 2011).

Using ClustalW, it was found that there were regions of high sequence conservation for linkers within both the same and different phyla, with a highly conserved region at either flank, but with a central region where <50% sequence homology was recorded (Fig. 3.6). Many of the MFS-PlsC fusion linkers contained both an acidic (Aspartate or Glutamate) and basic group (Arginine) within close proximity, 3-6 amino acids in distance. These domains have the potential to form salt-bridges within 4Å (Scholtz and Baldwin, 1992; Kumar and Nussinov, 2002), which may have aided to achieve the predicted alpha-helix. Beyond the charged residues, the bulk of the aligned region is hydrophobic, with the exception of the C-terminal, which contains a proline followed by two acidic residues. An overall greater abundance of basic residues compared to acidic residues results in an overall net positive charge for the linkers, which is consistent with the positive inside rule, which describes the propensity of cytoplasmic loops to be positively charged in membrane proteins (Von Heijne, 1986).

For the majority of the linker structures the length ranged between 29 and 35 residues. Two exceptions were initially noted when recording linker structures, the *Mesorhizobium loti* A0A1A5IBS3 linker and the *Cyanobacterium TDX16* linker. The *Mesorhizobium loti* linker was initially suggested to be 61 residues long, however sequence alignment revealed the predicted linker actually contained the final TMH of the MFS domain, therefore was actually much shorter and was corrected. On the other hand, *Cyanobacterium TDX16* appears to have a novel linker in comparison with the others measured, with the conserved arginine retained, yet

Proteobacteria (gamma)	<i>Stenotrophomonas acidaminiphila</i>	A0A0S1B3E0	33	PEFLMRFLSWVMVHALYRLRRTGIEEHVPDEGA --HHHHHHHHHHHHHHHEEEEE-----
Proteobacteria (gamma)	<i>Lysobacter enzymogenes</i>	A0A0S2DC48	32	PEFLMRFLSWVLVRGLYRLRVSGTERIPDEGA --HHHHHHHHHHHHHHHEEEEE-----
Proteobacteria (gamma)	<i>Photobacterium marinum</i>	L8J7U2	35	SQVHEFLLR FVIWVLSHTIYRVSHKDLENIPEQGA ----HHHHHHHHHHHHHHHEEEEE-----
Cyanobacteria	<i>Nostoc sp. RF31Y</i>	A0A252DVF0	29	MRFLSWLLVRTLYRLELHGIEKNVPDEGA --HHHHHHHHHHHHHHHH-----
Cyanobacteria	<i>Cyanobacterium TDX16</i>	A0A254WVA7	42	VRLRRGRFGRAITFVTNFNQFYCRFWARVQRDGICTVPAQGP -----EEEE--HHHHHHHHHH-----EEE-----
Fibrobacteres	<i>Fibrobacteres bacterium CG2_30_45_31</i>	A0A1J5EPB4	35	SKTPQALIRACLRTFFAFRYKLNKNTGIDNIPSEGP ---HHHHHHHHHHHHHHHHHHHH-----
Fibrobacteres	<i>Fibrobacter sp. UWB12</i>	A0A1M6UDT2	31	PQAMLRTLLRFVFSRYRIRVLNVQNIPNEGP -HHHHHHHHHHHHHHHEEEEE-----

Planctomycetes	<i>Planctomyces sp. SH-PL14</i>	A0A142WSB4	35	RLMPEITARFVFWLASHTVYRLQVVGRENIPERGG ---HHHHHHHHHHHHHHHHHEEEEE-----
Chlamydiae	<i>Parachlamydia acanthamoebae</i>	A0A0C1BXH8	30	DYLSRFVAMILSRLHFKIRFNGVENIPKTP -HHHHHHHHHHHHHHHEEEEE-----
Acidobacteria	<i>Acidobacteria bacterium 13_2_20CM_57_7</i>	A0A1Q6X946	32	PDSLLRLILWIATHTLYRLDVEGRENVPARGG ---HHHHHHHHHHHHHHHEEEEE-----
Acidobacteria	<i>Acidobacteria bacterium 13_1_40CM_2_60_7</i>	A0A1Q7RAD0	32	PDSLLRLLLWFATHTLYRIDVQGREYVPVRGG --HHHHHHHHHHHHHHHEEEEE-----
Fusobacteria	<i>Fusobacterium necrophorum subsp. funduliforme B35</i>	A0A017H5R7	29	PRSIQSLLAIFFKVDVKGLETFEKAGKR --HHHHHHHHHHHH-----HHHHH----
Fusobacteria	<i>Fusobacterium equinum</i>	A0A133N6P4	29	PRSIQSLLAIFFKVEVKGLEHFEKAGKR --HHHHHHHHHHHHHH---HHHHH----

Table 3.2. Linker length and amino acid sequence of a random selection of natural MFS fusions within MFS-fusion group 2, belonging to one of the seven phyla with most recognised members. The amino acid sequence of linkers was derived from Uniprot (Bateman, 2019), and can be defined as the sequence between the final amino acid of the MFS domain (final Transmembrane helix unless otherwise specified)

and the first amino acid of the PlsC domain. Beneath each amino acid sequence, the predicted secondary structure is recorded, as predicted by online resource Jpred (Drozdetskiy *et al.*, 2015) (where H represents an α -helix, and E represents an extended secondary structure). The uniprot code is provided for proteins to aid in identification.

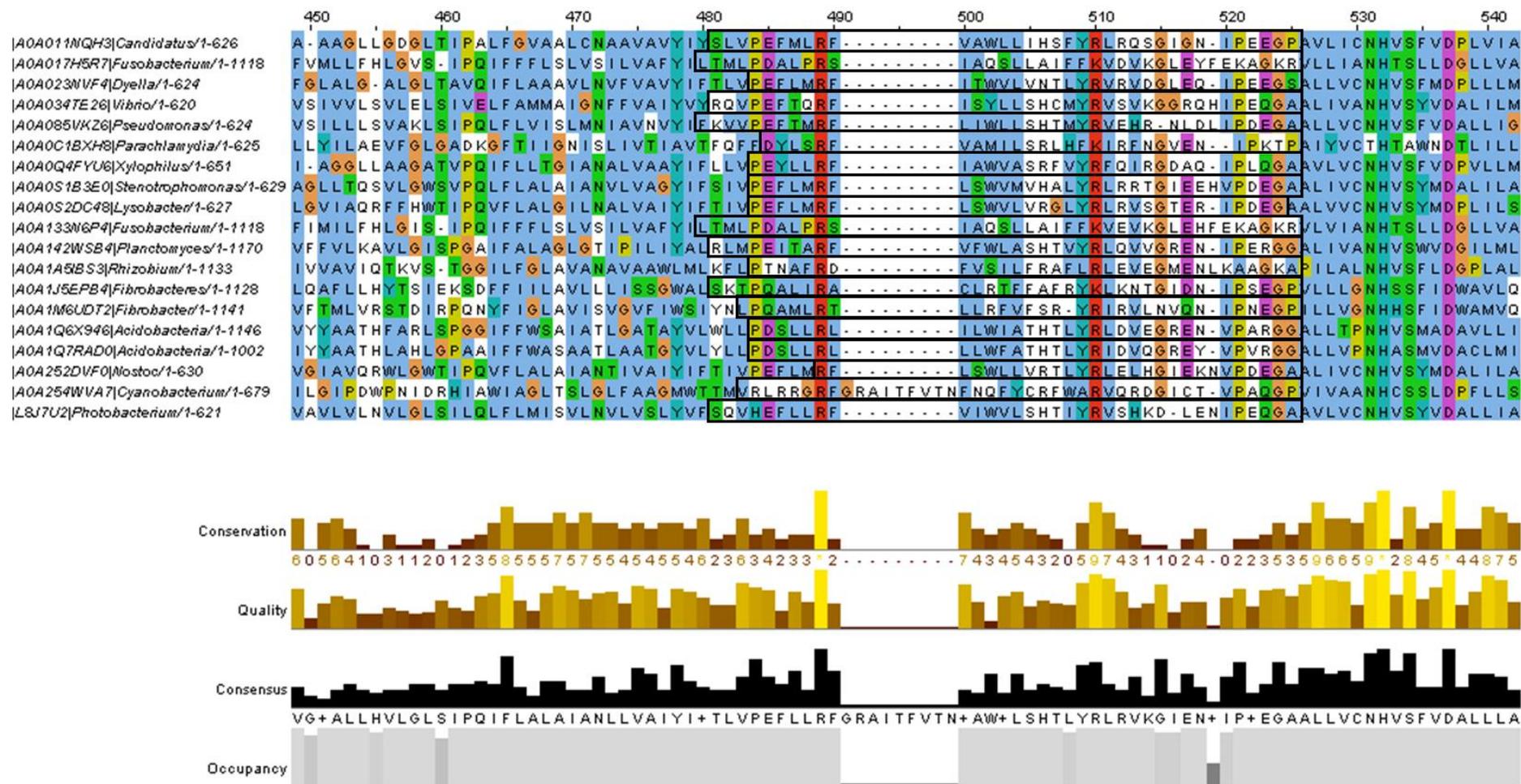


Figure 3.6. Multiple sequence alignment of the MFS-PlsC fusions focused on the C-terminal region of the MFS, whole linker sequence and N-terminal region of PlsC domain. Produced using ClustalW (Sievers *et al.*, 2011) and visualized in Jalview (Waterhouse *et al.*, 2009). Amino acid groups are coloured in accordance with Clustalx annotation in Jalview. Peptide linkers are highlighted.

a larger middle region. BLAST searching of the protein fusion on Uniprot identified only 5 other species containing homologous linkers, all within the planctomycetes phylum.

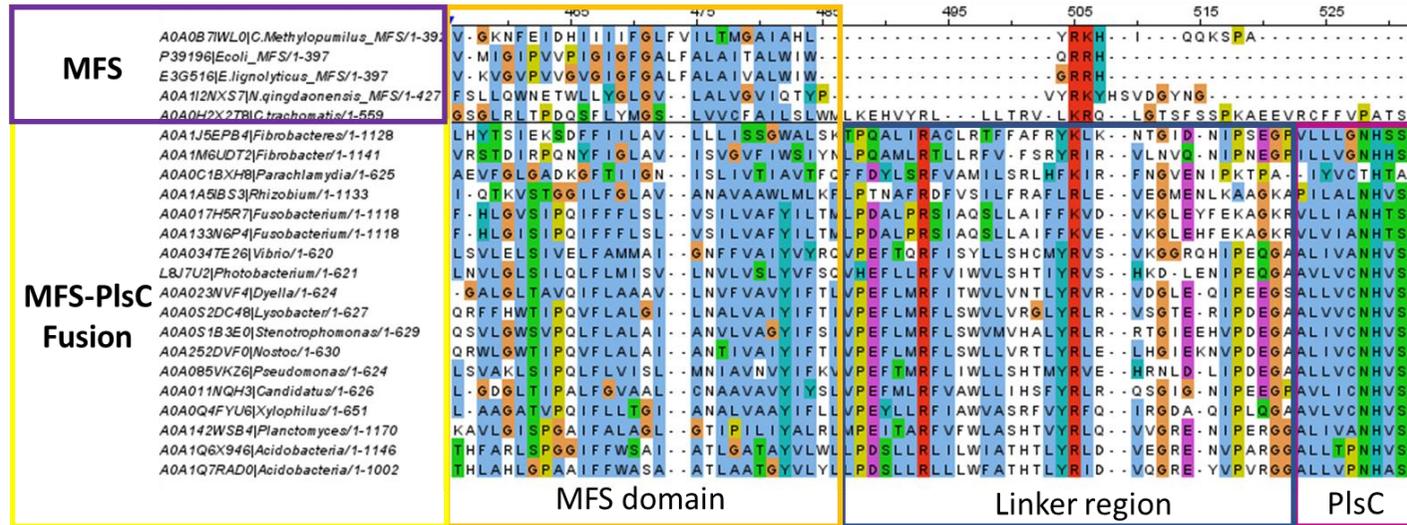
This may suggest either a separate fusion event or an insertion took place for fusions in the planctomycetes phylum, resulting in longer linkers between the two domains. *Cyanobacterium TDX16* would have had to take this fusion up by horizontal gene transfer, as it is not consistent with the other MFS-PlsC fusion seen within the cyanobacterium *Nostoc sp. RF31Y*. Curiously, some planctomycetes species contain the otherwise observed ~32 residue linker, which suggests that the 61-residue linker is not universal even in phylum where it is seen.

Considering that Uniprot failed to recognise a final TMH, causing the predicted *Mesorhizobium loti* linker to be longer than it actually was, it was possible that other linkers were conversely identified as being smaller than they actually were. Regions outside the suggested linker architecture for 29 amino acid linkers often aligned to the 32+ amino acid linkers. Therefore, Uniprot's definition of the transmembrane helix was variable between samples which may have resulted in the variation in length of the ~32 residue linker. The typical linker is therefore 32 amino acids in length, with a net positive charge, featuring a conserved proline at the N-terminal, followed by an acidic amino acid. A few residues downstream, a conserved arginine is seen, followed by about 10 hydrophobic amino acids before a second conserved arginine. At the C terminal a mixture of glycine, proline and glutamic acid is seen, although in a somewhat less conserved order. The Jpred data largely predicts these linkers will be helical towards the N-terminus.

To better understand the context of the linker – what the length of the linker is likely to be and the origin of the residues that make up its bulk, the MFS-PlsC fusion proteins were compared against unfused homologues of the MFS and PlsC domains (Fig. 3.7). Homologues were found by searching for known gene fusions in the Microbes Online database (Alm *et al.*, 2005) and using the phylogenetic tree function to find closely related unfused proteins, or by- using BLAST searching for truncated genes. Homologous MFS proteins showed no alignment at the linker region (Fig. 3.7.A), suggesting the linker begins around 3 amino acids upstream of the conserved N-terminal linker proline. *Cyanobacterium TDX16* was not included in this alignment due to its unique linker architecture.

In contrast, some clear sequence conservation is observed in the linker region between the MFS-PlsC and unfused PlsC-domain containing protein homologues (Fig. 3.7.B). Alignment occurs within 8 amino acids of the MFS-PlsC linker; around a conserved arginine at 475. The high level of conservation suggests that the linker contains structural elements of the unfused PlsC protein, in which case only a small, 8 amino acid region of the linker is novel to the fusion.

(A)



(B)

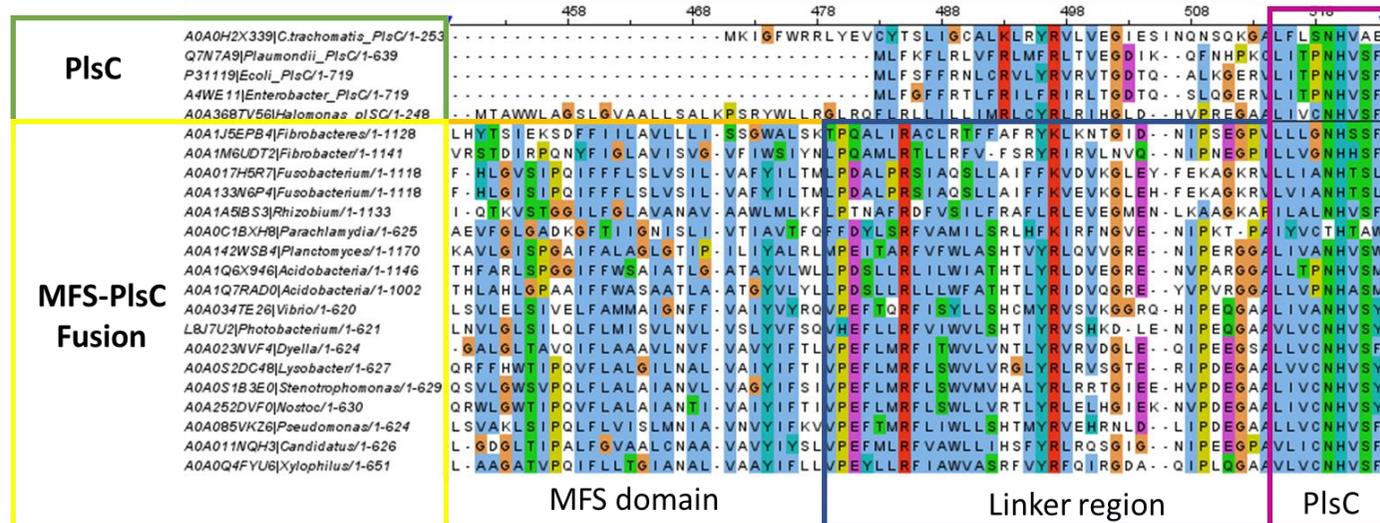


Figure 3.7. Multiple sequence alignment of the C-terminal region of the MFS, linkers and N-terminal region of PlsC domain from group 2 Major facilitator Superfamily fusions. Derived from gamma-proteobacteria, produced using ClustalW (Sievers *et al.*, 2011) visualized in Jalview (Waterhouse *et al.*, 2009). Amino acid groups are coloured in accordance with ClustalX annotation in Jalview. General regions of MFS, PlsC and Linker domains are highlighted on the fusion proteins and overlap with relevant unfused protein. (A) Sequence alignment of the predicted linker region of MFS-PlsC fusion proteins (~480 – 520) with unfused structural homologues of the MFS region along the top five rows: *Escherichia coli* (strain K12) - P39196, *Chlamydia trachomatis* serovar A - A0A0H2X2T8, *Neptunomonas qingdaonensis* - A0A1I2NXS7, *Candidatus Methylopusillus turicensis* - A0A0B7IWL0, and *Enterobacter lignolyticus* (strain SCF1) - E3G516 (B) Sequence alignment of the predicted linker region of MFS-PlsC fusion proteins with unfused homologues of the PlsC domain along the top five rows: *Chlamydia trachomatis* - A0A0H2X339, *Enterobacter sp. (strain 638)* - A4WE11, *Photorhabdus laumondii* - Q7N7A9, *Escherichia coli* - P31119, and *Halmonas sp. TQ8S* - A0A368TV56.

Furthermore, if the bulk of the linker aligns with the PlsC-containing homologues, then it is possible the region defined as the linker is largely composed of structural features of PlsC. A novel proline is present at the C-terminal of the linkers not present within most PlsC-homologues, with only the *Halmonas sp. TQ8S* containing it. It's possible this may have been introduced to offer improved flexibility.

An alternative explanation is that the unfused PlsC homologues are products of gene fission processing, being left with an N-terminal region from the old fusions protein. The MFS-PlsC fusion is abundant within many species, beyond simply a single group or phyla. When searching for a suitable unfused protein an abundance of the fusion throughout gammaproteobacterium phylum was noted compared to the relative infrequency of unfused MFS. The abundance of the MFS-PlsC may suggest that the fusion occurred within a shared ancestral species, and the Enterobacteriaceae, where fusions were less common, instead were an example of protein fission. Therefore, the sequence alignment between the N-terminus of unfused PlsC and the linker of the fusion proteins could alternatively be explained by a fission event, causing linker residues to be retained.

To try and resolve these two opposite 'life stories' of the linker, a phylogenetic tree was produced, focusing on the MFS domains of the fusion proteins and unfused homologues by calculating maximum likelihood (Trifinopoulos *et al.*, 2016) (Fig. 3.8). In order to improve the accuracy of the tree, in addition to the fusions in Table 3.2, additional fusions were obtained through Microbes Online and additional BLAST searching.

The phylogenetic tree is complex yet suggests a variety of events that appear to have occurred in the evolution of these fusions. First, that the *Cyanobacteria TDX16* containing the 42-residue linker were produced in a separate fusion event, with the branch points derived much earlier than all other samples. Conversely, the more common, ~32-residue linker is found to have occurred later, and the fusion may feature an additional fusion to an AMP-dependant Synthetase/Ligase (ACS) domain downstream of the PlsC.

Some unfused MFS proteins, such as B1KN08 of *Shewenella woodyi* are produced at branchpoints following fusion, suggesting domain deletion. The protein A0A112NXS7 on *Neptunomonas qingdaonensis* is an unfused homologue to the MFS-PlsC fusion protein, however also present within *N. qingdaonensis* is A0A112VAB7, in which both the MFS and PlsC are still fused. This suggests a gene duplication event occurred within this species, resulting in two versions of an MFS domain to be present. Other PlsC-containing proteins in *N. qingdaonensis* have a <30% identity with the PlsC of the fusion protein, suggesting that for the unfused MFS the associated PlsC was either not duplicated, or deleted. A duplication-

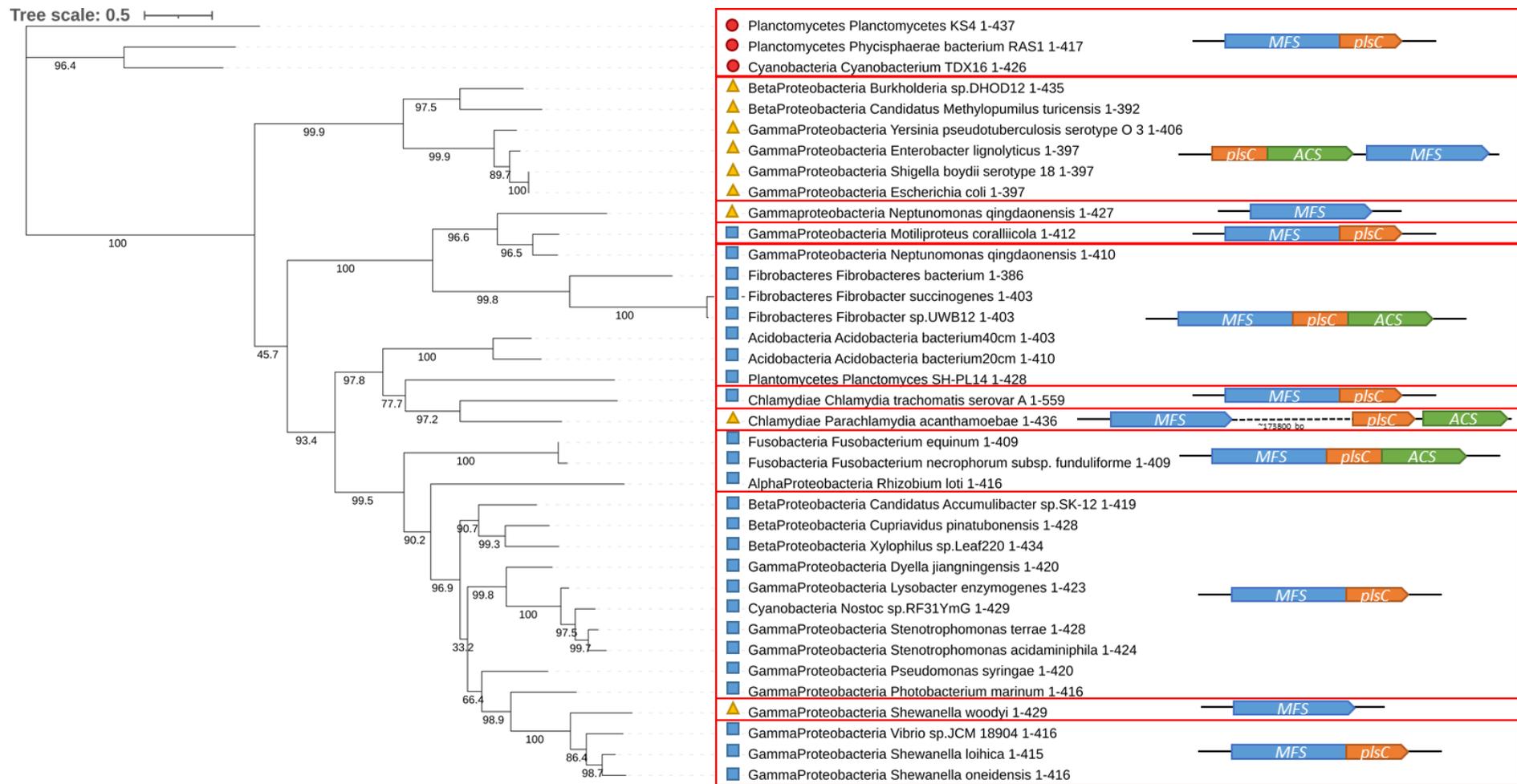


Figure 3.8. Maximum likelihood phylogenetic tree of MFS-PlsC fusions. A Phylogenetic tree demonstrating the average distance of MFS domains of either an MFS-PlsC fusion with a ~32 residue linker (blue square), a fusion with a 42 residue linker (red circle) or an unfused MFS homologues (yellow triangle). Species names and phylum are displayed for each protein, with the exception of proteobacteria, in which class is included instead. Alignment is determined using IQ-TREE (Trifinopoulos *et al.*, 2016) and visualized in iTOL (Letunic and Bork, 2019). Alongside

each protein the domain architecture of the gene, as well as relevant upstream or downstream components are also shown: the MFS domain is represented in blue, the AMP-dependant synthase/ligase (ACS) is represented in green and the PlsC is represented in orange. Where genes are fused they are shown as overlapping with one another. Scale bar represents the branch length, the evolutionary time between two nodes. For bootstrapping 1000 phylogenetic tree reconstructions were completed, with the percentage of replicates in which a branch was produced displayed as a percentage of 100.

The proteins expressed by each organism is as follows: *Cyanobacterium TDX16* - A0A254WVA7, *Phycisphaerae bacterium RAS1* - A0A5C5VS94, *Planctomycetes bacterium KS4* - A0A517YTM9, *Rhizobium loti* - A0A1A5IBS3, *Pseudomonas syringae* - A0A085VKZ6, *Candidatus Accumulibacter sp. SK-12* - A0A011NQH3, *Dyella jiangningensis* - A0A023NVF4, *Xylophilus sp. Leaf220* - A0A0Q4FYU6, *Vibrio sp. JCM 18904* - A0A034TE26, *Stenotrophomonas acidaminiphila* - A0A0S1B3E0, *Lysobacter enzymogenes* - A0A0S2DC48, *Photobacterium marinum* - L8J7U2, *Nostoc sp. RF31YmG* - A0A252DVF0, *Fibrobacteres bacterium CG2_30_45_31* - A0A1J5EPB4, *Fibrobacter sp. UWB12* - A0A1M6UDT2, *Parachlamydia acanthamoebae* - A0A0C1BXH8, *Acidobacteria bacterium 13_2_20CM_57_7* - A0A1Q6X946, *Acidobacteria bacterium 13_1_40CM_2_60_7* - A0A1Q7RAD0, *Fusobacterium necrophorum subsp. funduliforme B35* - A0A017H5R7, *Fusobacterium equinum* - A0A133N6P4, *Chlamydia trachomatis serovar A (strain ATCC VR-571B / DSM 19440 / HAR-13)* - A0A0H2X2T8, *Shewanella woodyi (strain ATCC 51908 / MS32)* - B1KN08, *Shewanella oneidensis (strain MR-1)* - Q8EKA1, *Shewanella loihica (strain ATCC BAA-1088 / PV-4)* - A3Q957, *Enterobacter lignolyticus (strain SCF1)* - E3G516, *Shigella boydii serotype 18 (strain CDC 3083-94 / BS512)* - B2TYQ6, *Yersinia pseudotuberculosis serotype O:3* - B1JQC2, *Candidatus Methylopusillus turicensis* - A0A0B7IWL0, *Burkholderia sp. DHOD12* - A0A4P8IKC2, *Fibrobacter succinogenes (strain ATCC 19169 / S85)* - C9RNC8, *Neptunomonas qingdaonensis* - A0A1I2VAB7 (Fusion protein), *Neptunomonas qingdaonensis* - A0A1I2NXS7 (MFS), *Stenotrophomonas terrae* - A0A0R0D2Z4, *Cupriavidus pinatubonensis (strain JMP 134 / LMG 1197)* - Q472Q7, *Motiliproteus coralliicola* - A0A369WAG5, *Escherichia coli (strain K12)* - P39196.

deletion event could explain why the gene encoding the homologous MFS of *Chlamydia trachomatis* serovar A is present far removed from a *plsC* gene (Fig. 3.8).

The phylogenetic tree suggests that for some unfused homologues of the MFS domain, a gene fission event has occurred to separate or delete a component of the MFS-PlsC fusion. However, for Enterobacteriaceae it is possible we are observing a pre-fused state as the branch point predates the MFS-PlsC fusions with the common linker. Within Enterobacteriaceae such as *E. coli* the PlsC was found fused to an ACS domain, however PlsC domain and the MFS domain are still potentially associated with one another even when unfused. The PlsC-ACS fusion of *E. coli*, *aas*, is found upstream of the MFS-expressing gene, *lpIT*, however, *aas* demonstrates gene overlap with *lpIT* (Fig. 3.9). This out-of-frame overlap suggests that gene translocation or a gene duplication and deletion event would be required to achieve fusion (Durrens, Nikolski and Sherman, 2008), which seems rather unlikely. As the coding sequences for the two genes overlap, it is possible they are translationally coupled – a feature in which genes in close proximity are expressed in a similar manner (Wang, Lercher and Hurst, 2011). There is evidence that gene overlap precedes a gene fusion event (Zakharova *et al.*, 1999; Sakharkar, Sakharkar and Chow, 2006), therefore it is possible that the coupling of these genes encourages subsequent fusion.

Considering the presence of the MFS-PlsC-ACS fusions, which define 30% of MFS-PlsC fusions reported, it is possible that following an early fusion of PlsC-ACS, a second fusion event occurred to transfer the whole gene to the MFS. A proposed model for the evolution of these fusions is presented in Fig. 3.10, however, although this appears to be most parsimonious with the data, there could be alternative patterns of events. It is possible that the MFS-PlsC-ACS fusion occurred from the Enterobacteriaceae architecture, with the ACS lost in many organisms over time due to domain deletion or shuffling, similar to what is seen between MFS and PlsC in *Chlamydia trachomatis* serovar A. Therefore, the linker of the MFS-PlsC fusion can be assumed to be the result of a PlsC-ACS domain being added to the C-terminus of MFS via a gene duplication-deletion event, or gene shuffling.

To further support this conclusion, a larger scale phylogenetic tree was produced. A BLAST search was completed on Uniprot to collect ~500 homologous sequences for each of the MFS proteins: A0A1A5IBS3 (*Rhizobium loti* MFS-PlsC-ACS), A0A034TE26 (*Vibrio* sp. JCM 18904 MFS-PlsC with shorter linker), A0A254WVA7 (*Cyanobacterium TDX16* MFS-PlsC with longer linker) and P39196 (*E. coli* unfused MFS) – redundant sequence were then deleted and absent sequences from the Fig. 3.8 phylogenetic tree were added.



Figure 3.9. Sequence of the *E. coli* genome, focused on the 3- region of the *aas* gene and 5' region of the *lplT* gene. The genetic sequence, obtained from the European nucleotide archive of *Escherichia coli* str. K-12 substr. MG1655. The 3' region of *aas*, the gene containing the PlsC domain, and the 5' region of *LplT*, a lysophospholipid transporter of the MFS family, are both shown to demonstrate the overlapping nature of the two genes.

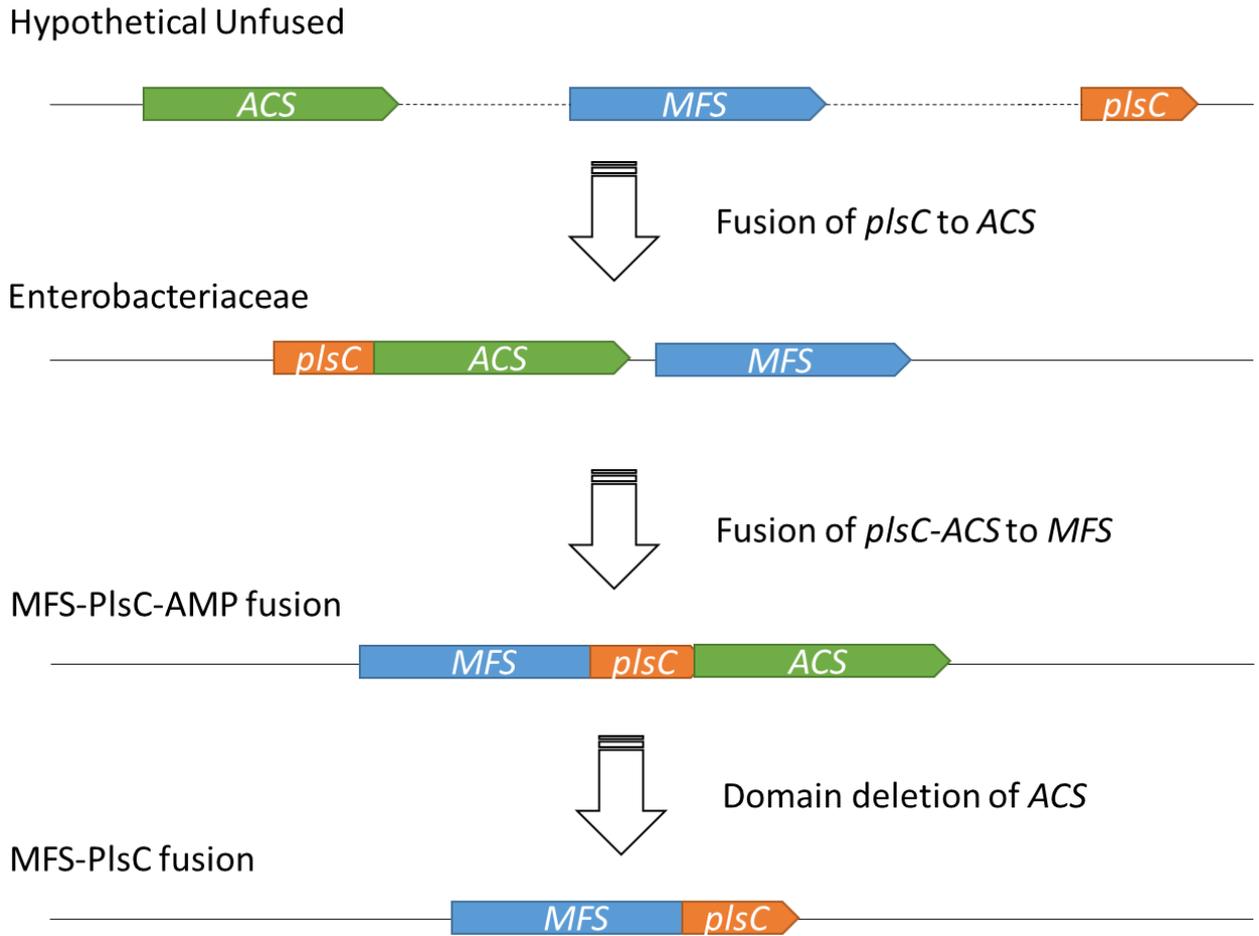


Figure 3.10. Hypothetical evolution pathway of the MFS-PlsC fusion architecture. A representation of a three-step genetic fusion/fission pathway which may explain the architectures and linkers seen in the MFS-PlsC fusion and homologous genes.

The subsequent phylogenetic tree featured 1397 points of data, offering a greater insight into the evolution of the MFS-PlsC fusion (Fig. 3.11). It was found that, as with the smaller tree, the MFS-PlsC-ACS fusion appeared to form first, represented by the purple clades in the phylogenetic tree – subsequently the ACS fusion was lost, resulting in MFS-PlsC, the dominant fusion type. This is consistent with what was seen in Fig. 3.8, further supporting the hypothesis. The red clades contain at least one node where a gene encoding the unfused MFS protein was downstream of genes encoding an PlsC-ACS fusion. These clades are present throughout the branching point prior to the MFS-PlsC-ACS fusion, suggesting the origin hypothesized in Fig. 3.10 could still be argued. Finally, a single green clade, representing 6 proteins, is recognised as an early branching event. The green clade represents the MFS-PlsC fusion, containing the longer 42 amino acid linker, suggesting once more that this MFS-PlsC fusion type is a unique event separate to the fusions containing the other linker.

Overall, assuming that the Enterobacteriaceae PlsC-homologue represents a pre-fused state, the linker region of the MFS-PlsC fusion is composed of an 8 amino acid novel region, followed by structural features of the unfused protein that make up the remaining ~24 residues. Whether the retained amino acids are important for function or stability is not certain – if not then the linker region can be defined as all 32 amino acids, whereas if they are essential, then only the 8 amino acid region would be classified as the linker.

Tree scale: 1

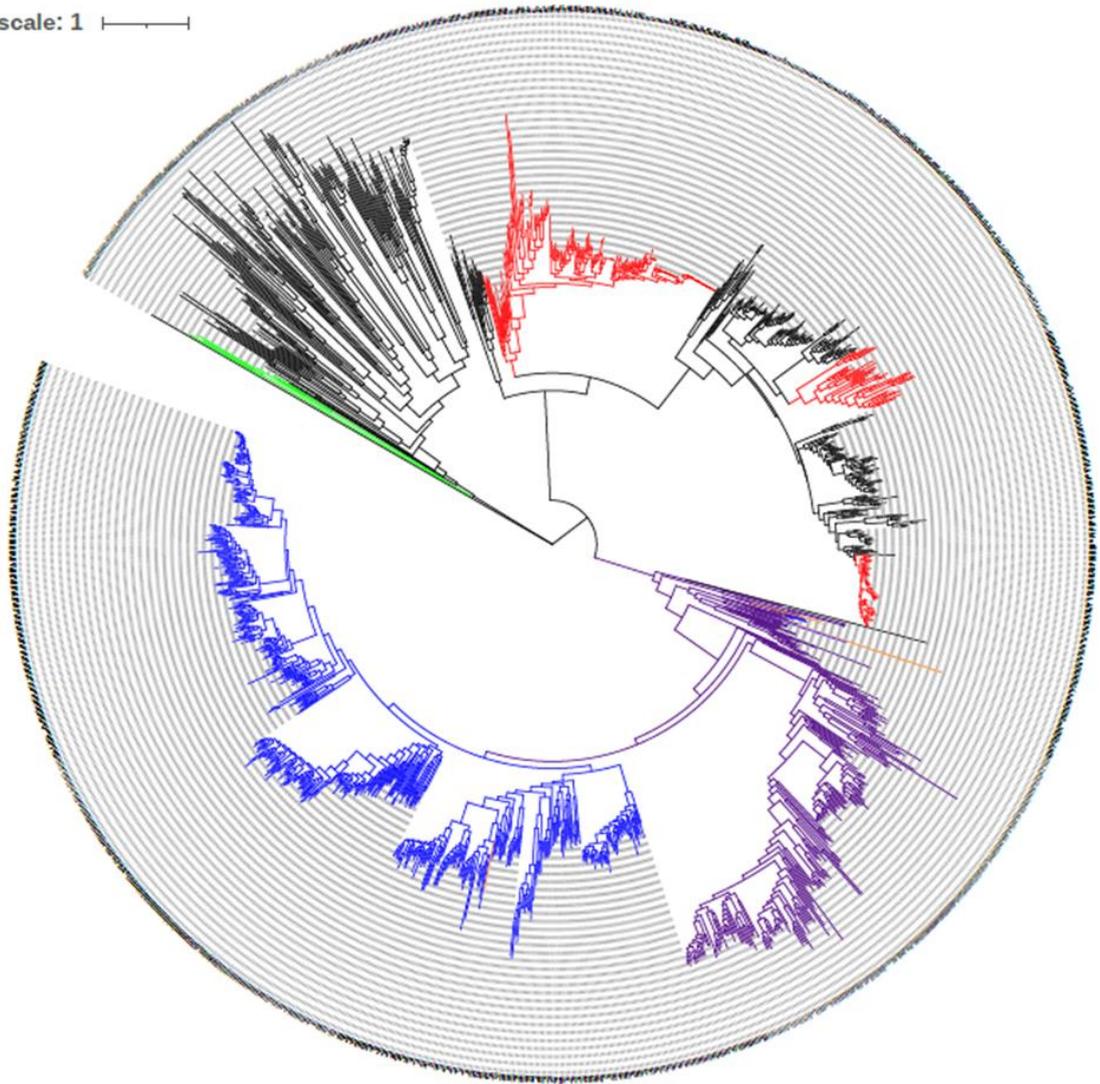


Figure 3.11. Maximum likelihood phylogenetic tree of a broad range of LpIT MFS homologues. A Phylogenetic tree demonstrating the average distance of MFS domains of either an MFS-PlsC fusion with a ~32 residue linker (blue clades), MFS-PlsC-ACS fusions (Purple clades), a fusion with a 42 residue linker (green clades), unfused MFS homologues with an upstream PlsC-ACS fusion (red clades) or unfused MFS derived from a fusion (Orange clades). Alignment is determined using IQ-TREE (Trifinopoulos *et al.*, 2016) and visualized in iTOL (Letunic and Bork, 2019). Scale bar represents branch length, representing the evolutionary time between two nodes.

3.3.2 The MFS-CBS protein fusions demonstrate greater variability in linker architecture in less abundant fusions.

The MFS-CBS fusion is found in 203 incidences, within 190 identified organisms. Of these organisms, 164 belong to the Actinobacteria phylum, with 14 being identified as Firmicutes and just 5 identified as Proteobacteria; other identified species are either Archaea or Eukaryota. In comparison to the MFS-PlsC domains, in which the fusions across different phyla had a relatively conserved linker around 32 residues in length, the MFS-CBS fusions showed greater variability in length of linkers within the same phyla (Table 3.4). Following the same methodology for determining linker length, a range from 15 to 95 amino acid residues in length was recorded. The genus *Bifidobacterium* expressed MFS-CBS fusions with linkers ~56 residues in length, while other actinobacteria and the firmicute species expressed fusions with short 20 amino acid linkers. The proteobacteria were particularly unique, with fusions predicted to contain long, >90 amino acid linkers between the final TMH and the first CBS domain (Table 3.3). Furthermore, the secondary structure predicted by Jpred would range greatly, containing short to large alpha helix regions and the occasional extended secondary structure, however this being inconsistent between different species, with some linkers showing no secondary structure predicted at all. For the shortest linkers a predicted structure could not be produced by Jpred due to a 20 residue requirement (Table 3.4) (Willson *et al.*, 2019).

Using ClustalW to produce a multiple sequence alignment of the selected linkers, an increased variability appears to be present for the MFS-CBS linker (Fig. 3.12) with respect to the MFS-PlsC linkers (Fig. 3.6). The linkers showed surprisingly poor overlap when compared to the MFS-PlsC fusion linkers. The consensus amino acid was generally poor and occupancy varied between linkers, and even similarly sized linkers would show mixed overlap if from different genus. The largest linker, of the MFS-CBS fusion from *Agrobacterium genomosp. 9 str. Hayward 0363*, does show overlap with the MFS domain of other organisms, suggesting that the length is more likely to be around 80 residues, as with the second proteobacteria species selected.

When looking at linkers individually, the linkers from the proteobacteria MFS-CBS fusion are almost identical to one another, as are the linkers expressed in *Bifidobacterium*. Even when removing these organisms, however, there is little improvement in the overall homology. Taking this into account, it appears there is little consensus between linkers for natural MFS-CBS fusions. It is possible the variation in linkers represent poor optimization of the fusion, resulting in reduced uptake without linker mutation. Alternatively, multiple incidents of the fusion may have occurred independently.

Phylum	Species	Protein	Linker length
Actinobacteria	<i>Bifidobacterium breve</i>	C4NXS7	55
Actinobacteria	<i>Bifidobacterium longum</i>	Q8G5T3	56
Actinobacteria	<i>Bifidobacterium pseudocatenulatum</i> IPLA36007	A0A072MU51	56
Actinobacteria	<i>Bifidobacterium bifidum</i> (strain S17)	E3EPJ8	56
Actinobacteria	<i>Corynebacterium</i> sp. JB4	A0A1W1IZT5	27
Actinobacteria	<i>Corynebacterium glutamicum</i>	A0A169S3B6	15
Actinobacteria	<i>Adlercreutzia equolifaciens</i> DSM 19450	S6BXI4	48
Actinobacteria	<i>Alloscardovia macacae</i>	A0A1Y2SXX8	16
Actinobacteria	<i>Gordonibacter urolithinifaciens</i>	A0A1Y4G286	20
Actinobacteria	<i>Eggerthella lenta</i>	C8WJV2	24
Actinobacteria	<i>Parascardovia denticolens</i> DSM 10105	E6K0F3	25
Actinobacteria	<i>Parascardovia denticolens</i> IPLA 20019	I8ULH3	23
Actinobacteria	<i>Dietzia cinnamea</i> P4	E6JAT9	20
Actinobacteria	<i>Alloscardovia omnicolens</i> F0580	U1SHC0	21
Firmicutes	<i>Clostridium kluyveri</i> ATCC 8527	A5N198	19
Firmicutes	<i>Desulfitobacterium hafniense</i>	B8FPR8	16
Firmicutes	<i>Anaerotignum lactatifermentans</i>	A0A1Y3TX30	23
Firmicutes	<i>Lactobacillus fuchuensis</i> DSM 14340	A0A0R1RPE1	20
Firmicutes	<i>Firmicutes bacterium</i> CAG:466	R6PZ71	21
Proteobacteria	<i>Agrobacterium genomosp.</i> 9 str. Hayward 0363	A0A1S7TTP2	95
Proteobacteria	<i>Agrobacterium fabrum</i> str. J-07	A0A1S7P5M4	79

Table 3.3 Linker length of a random selection of natural MFS-CBS(-CBS) fusions. The length of 21 linkers derived from a random selection of fusions from species from one of three phylums. The amino acid sequence of linkers was derived from Uniprot, and can be defined as the sequence between the final amino acid of the MFS domain (final transmembrane helix) and the first amino acid of the CBS domain. The uniprot code is provided for proteins to aid in identification

Species	Amino Acid Sequence and Predicted Secondary Structure
<i>Bifidobacterium breve</i>	KEKAARNAKAYEAQASAEVAAETTEGQPAEHHYAAYVAPAASLFKQAQEQSIGGI -----EE-----E-E---HHHHHHHHH-----
<i>Bifidobacterium longum</i>	KEKAARNAKASEAQASAEVAAETTEGQPAEHHYAGAYVAPAASLFKQAQEQSIGGI -----EE-----HHHHHHHHHH-----
<i>Bifidobacterium pseudocatenulatum</i> <i>IPLA36007</i>	KEKAARNAKAYEAQASAEVAVETTEGQPAEHHYAGAYVAPAASLFKQAREQSICGI -----EEEE-----HHHHHHHHHH-----
<i>Bifidobacterium bifidum</i> (strain S17)	KEKAARNAKAYEAQASAEVAAETTEGQPAEHHYAGAYVAPASSLFKQAQEQSIGGI -----EE-----HHHHHHHHHH-----
<i>Corynebacterium sp.</i> <i>JB4</i>	RAYSARKPQAAQPEAPQVESTGTVAGLM -----E-----
<i>Corynebacterium glutamicum</i>	KTPAASDQELTVQSI
<i>Adlercreutzia equolifaciens</i>	KMPGAAAVTAAERGAEPVDYAAGFDGVTVDGTLGESWDAEHTYTAADV -----HH-----EEE-----

<i>Alloscardovia macacae</i>	TMARPAAEAQDELSAL
<i>Gordonibacter urolithinfaciens</i>	PGETAELDKDNARRTVLESI -----HHHHHH--
<i>Eggerthella lenta</i>	RKRAAVQAVERRAAQPQALSVLASI ---HHHHHHHHHH---HHHHHHHH--
<i>Parascardovia denticolens</i>	ATVSAGAHPADETVEPAAESGIRQL -----HHHH-----
<i>Parascardovia denticolens</i>	ATVSAGAHPADETVEPAAESGIR -----
<i>Dietzia cinnamomea P4</i>	TARPSASETGSRPVDSVESI -----
<i>Alloscardovia omnicolens F0580</i>	ETQNSTADSEDSHATPLIHSL -----HH---
<i>Clostridium kluyveri</i> ATCC 8527	DGKKAVVISENTSFELDIA

<i>Desulfitobacterium hafniense</i>	KTVAAVTEAVYELDIA
<i>Anaerotignum lactatifermentans</i>	RKLRKAASAEAPAKEVHTLAPYM --HHHH-----HHHH----
<i>Lactobacillus fuchuensis</i>	LKLKNQLTSQATPDNIVKTH --HHHHH-----
<i>Firmicutes bacterium CAG:466</i>	RKLKTASAEAPAKEAHTLAPY -----
<i>Agrobacterium genomo</i>	PVAGNSLILLALALAYNNAIGRAYPHGLKLGKTAHGTTDPTPIQKIGFSSNDLDEVLKEYDQVLDIDRDELETILRRTELRSWRRRAL HLDCASV --HHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHH----EE---
<i>Agrobacterium fabrum str. J-07</i>	NNSTGRAYPHGLKPGKAAHGTTDPTPIQKIGFSSTDLDEVLKEYDEVLDIDRDELETILRKTELRSYRRRALHLDCASV -----HHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHH----EE---

Table 3.4 Amino acid sequence of a selection of linkers of natural MFS-CBS(-CBS) fusions. The amino acid sequence of linkers was derived from Uniprot, with the predicted secondary structure is displayed where possible, as predicted by online resource Jpred (Drozdetskiy *et al.*, 2015). Where linker length is less than 20, no Jpred prediction was available. For some predicted secondary structures, a length of less than 5 amino acids are seen, much shorter than can be realistically expected – demonstrating a short coming of prediction software.

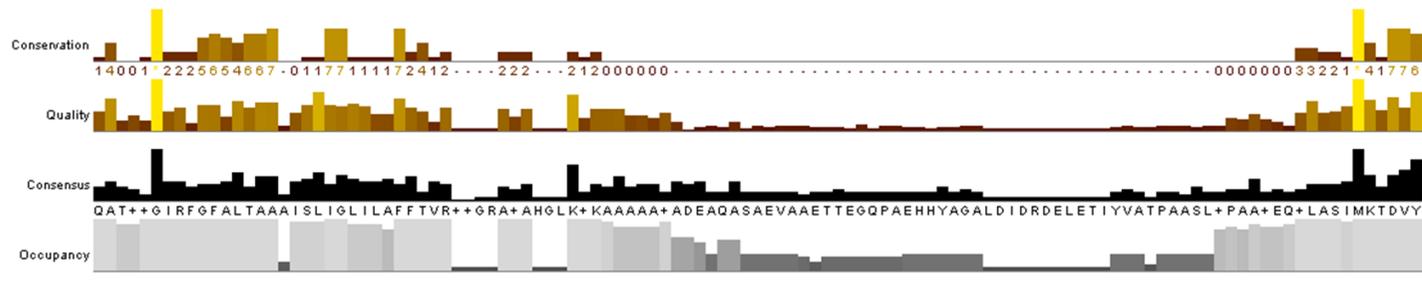
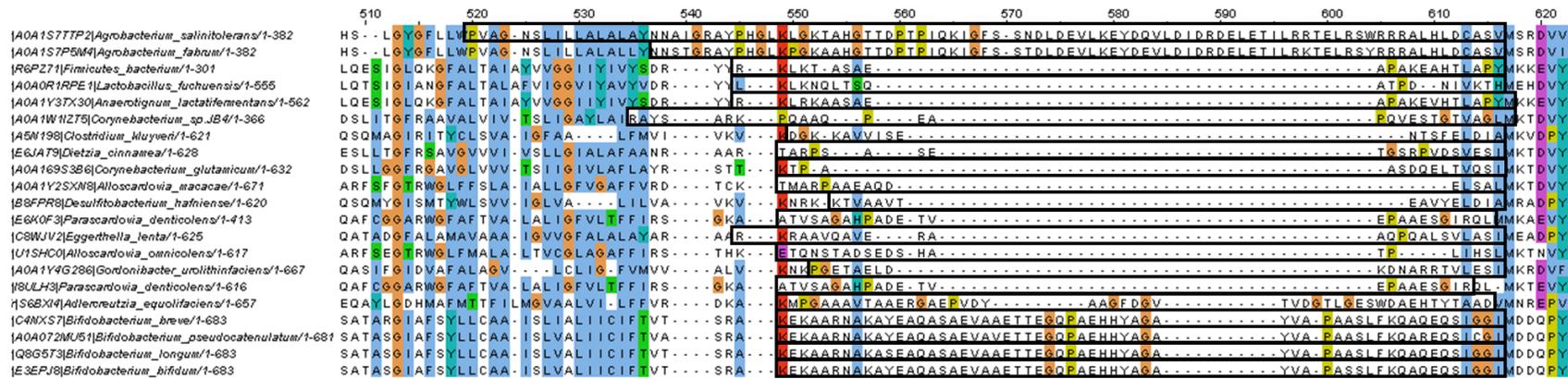
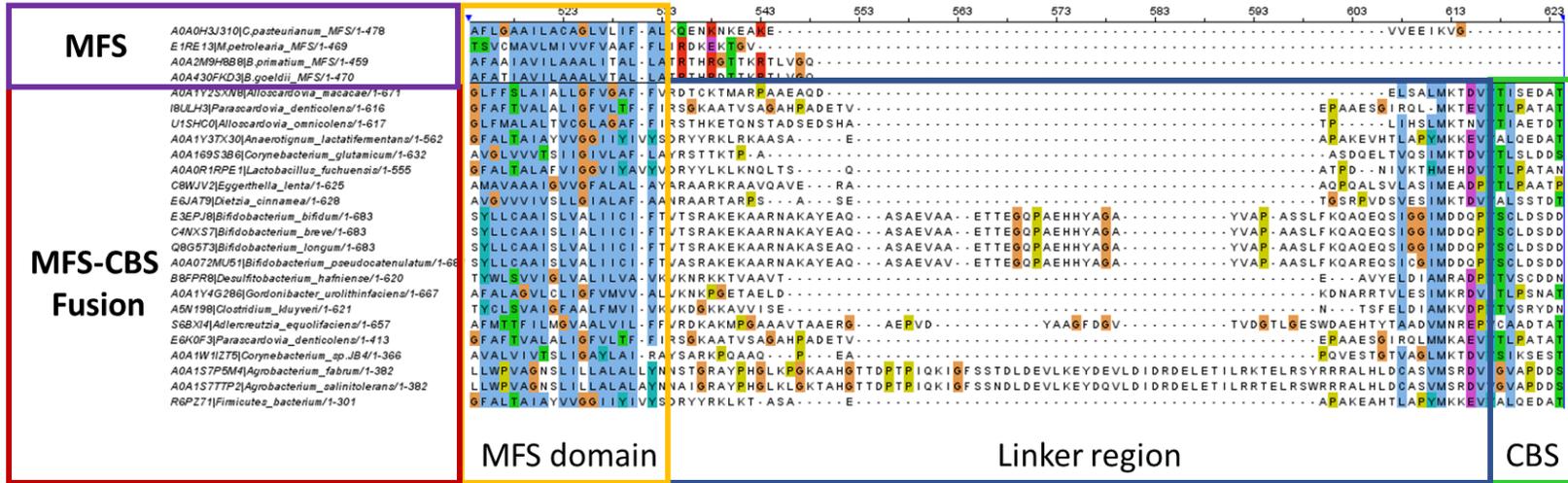


Figure 3.12. Multiple sequence alignment of the MFS-CBS fusions, focusing on the C-terminal region of the MFS, whole linker sequence and N-terminal region of CBS domain. Produced using Clustal Omega (Sievers *et al.*, 2011) and visualized in Jalview. Amino acid groups are coloured in accordance with ClustalX annotation in Jalview.

Looking at the linkers in context with the MFS region and CBS region (Fig 3.13) the linkers largely consist of novel amino acid sequence, with little overlap to the MFS or CBS homologues in this region. Furthermore, charge of the peptide linkers was highly variable, with many being negative. Therefore, unlike the MFS-PlsC fusion, the MFS-CBS fusion appears to be highly novel. The presence of multiple types of linker suggests multiple events in which fusion took place, furthermore, some fusions contain truncated components, such as the proteobacteria species.

Overall, this analysis suggests that the MFS-CBS fusion lacks a defined architecture of fusion. The linkers are not optimised within the fusion, and therefore multiple incidents of fusion or linker modification has occurred rather than uptake of a single stable MFS-CBS fusion. It is hard to know exactly what occurred to result in this variation, if multiple fusions occurred different amounts of upstream or downstream sequences could have been inserted alongside the CBS gene, alternatively if a single fusion occurred as opposed to multiple incidents, then the linker may have been modified through the splicing of amino acid sequences to increase the length of the linker. Altering the length of linkers has been demonstrated to alter the transport activity of an ABC transmembrane domain fused to soluble binding protein (Schuurman-Wolters *et al.*, 2018), therefore the different linkers may result in variations in the fusion stability and activity. It's possible the variation in linker length reflects a lack of optimisation within the fusion.

(A)



(B)

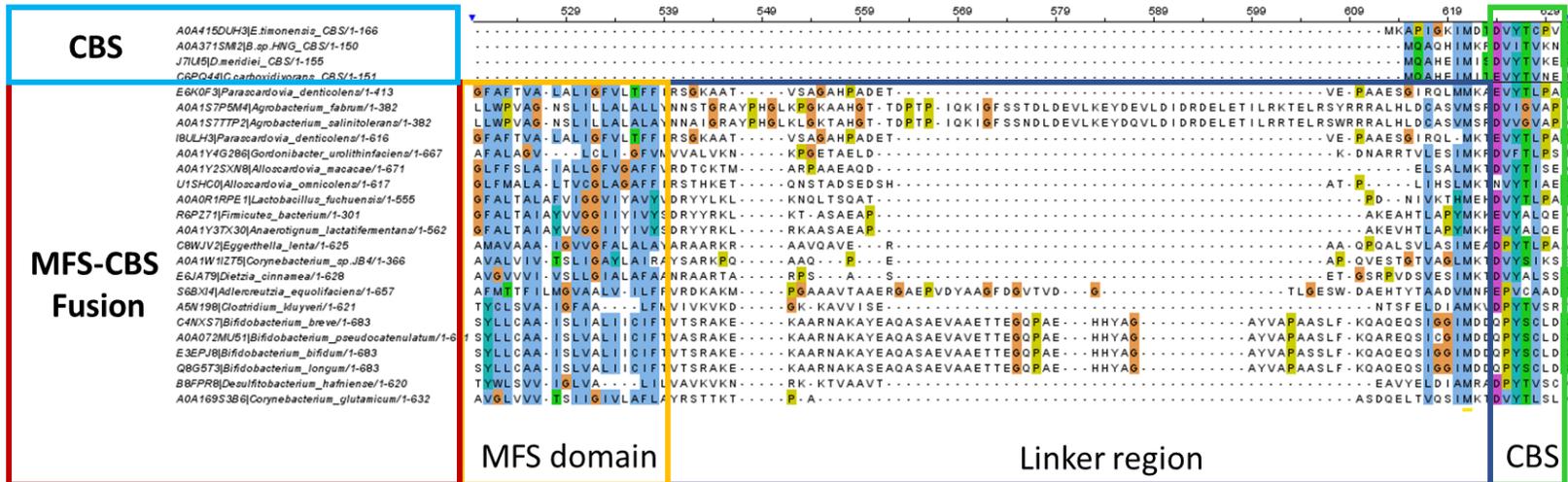


Figure 3.13. Multiple sequence alignment of the C-terminal region of the MFS, linkers and N-terminal region of CBS domain from Major facilitator Superfamily fusions with unfused homologues. Produced using Clustal Omega (Sievers *et al.*, 2011) and visualized in Jalview. Amino acid groups are coloured in accordance with ClustalX annotation in Jalview. General regions of MFS, PlsC and Linker domains are highlighted on the fusion proteins and overlap with relevant unfused protein. (A) With the unfused MFS domain homologues along the top 4 rows - *Methanobacillus_thermautotrophicus* E1RE13_METP4 Drug resistance transporter, *Bifidobacterium_goeldii* A0A430FKD3_9BIFI Multidrug transporter, *Bifidobacterium_primatium* A0A2M9H8B8_9BIFI Multidrug transporter, *Clostridium_pasteurianum* A0A0H3J310_CLOPA Drug resistance transporter (B) With unfused CBS homologues along the top 4 rows - *Bacillus_sp.HNG* A0A371SMI2_9BACI, *Desulfosporosinus_meridiei* J7IU15_DESMD, *Emergencia_timonensis* A0A415DUH3_9FIRM, *Clostridium_carboxidivorans* C6PQ44_9CLOT

3.3.3. Summary of natural linkers

The peptide linkers found to occur in different natural fusions between MFS proteins and downstream soluble domains were of interest for potential application or design ethos going forwards. Two groups were selected, one due to the broadness of fusion across species, the other due to its relative isolation to actinobacter and firmicutes. In some cases, many members of the group use the same peptide linker (MFS-PlsC), whereas other examples show greater variation between different organisms based on group, genus or phyla (MFS-CBS).

The MFS-PlsC domain fusion was abundant within bacteria, and it can be hypothesized that the fusion occurred early within bacteria development. Within the gamma-proteobacteria, which was well represented on microbes online, the relative scarcity of the unfused homologues suggested that fusion was the default assembly within the phylum. The linker was likely to have been composed of elements of the unfused PlsC-containing protein, as overlap was seen between the N-terminal of PlsC homologues and the C-terminal of the MFS-PlsC linker.

Exceptions were seen within MFS-PlsC cyanobacterium TDX16, which had a different linker compared to the other samples tested, however this was localised to just 5 other organisms, whereas all other MFS-PlsC linkers showed high regions of alignment. With over 3000 examples of the fusion, such events were very rare. Comparatively MFS-CBS fusion had more numerous linker variations, which showed homology to like-species, such as the proteobacteria or *Bifidobacterium*, but poor homology between different organisms.

Overall this suggested that abundant fusions were produced in a single instance, and if advantageous, were retained. For the MFS-PlsC there is an abundance of fusions found throughout nature, with similar linkers within the bacteria. This suggests that the fusion is optimal for uptake, whereas the MFS-CBS fusions number only around 200, with variations of linkers likely representing that the fusion is less ideal.

Linkers may therefore give insight into the stability of the fusion – however a more computational approach would be required to make any clear conclusions. Nevertheless, variations between linker lengths and composition suggests multiple linkers should be attempted in order to determine the optimal conditions for protein fusion to MFS sugar transporters.

3.4 Designing peptide linkers suitable for Major Facilitator Superfamily protein fusion

Considering that the linker structure within nature appeared to be random, largely consisting of a few unique amino acids and non-domain regions of the fusion partner, it was therefore of interest to test linkers of different characteristics (Willson *et al.*, 2019). Synthetic linkers allow for known lengths and properties. Examining the literature, two key types of synthetic linkers were relevant to this work – flexible linkers, in which adjacent protein domains are free to move relative to one another, and rigid linkers, in which adjacent peptide linkers are kept at a defined distance from one another due to bulky amino acids or fixed secondary structures (Chen, Zaro and Shen, 2013).

Depending on the type of fusion, a flexible or rigid linker may be preferred (Chen, Zaro and Shen, 2013). A flexible linker can allow the two proteins to be positioned in an advantageous orientation which can reduce steric strain (Li *et al.*, 2016), however rigid linkers may separate two proteins that can have deleterious interactions (Arai *et al.*, 2001).

Six linkers were therefore designed, three rigid and three flexible. These linkers were derived from functional fusions in literature and were designed to be of different sizes and different composition (Table 3.5).

1: Waldo linker

Derived from the Waldo plasmid. This linker has previously been applied to fuse membrane proteins to GFP for correct visualization. Rich in glycine and serine groups to achieve flexibility, the Waldo linker contains minimal hydrophobic groups and few homologous repeats to reduce risk of cleavage (Waldo *et al.*, 1999; Chen, Zaro and Shen, 2013).

2: Single-chain Flexible linker

The presence of glycine and serine groups give this linker flexibility, although it also contains lysine and glutamate, which could be capable of forming salt bridges. As salt bridges increase stability, such groups could risk a reduced flexibility compared to the Waldo linker, however it also aids solubility (Chen *et al.*, 2017). This flexible linker was produced in order to tether the light-chain amino acid sequence to the heavy-chain sequence single-chain within antigen-binding proteins. The linker was designed by a computer-assisted methodology, in which three-dimensional peptide structures derived from the protein data bank were analysed to find a linker which would both facilitate a suitable distance between the two amino acids sequences, while maintaining the correct orientation for function (Bird *et al.*, 1988).

Linker name	Flexibility	Linker length	Linker sequence
Waldo	Flexible	12	GSAGSAAGSGEF
Single-chain flexible	Flexible	14	EGKSSGSGSESKST
Polyproline	Rigid	14	KPAPQPKPAPKPKP
Long Arai	Rigid	20	EAAAKEAAAKEAAAKEAAAK
Short Arai	Rigid	12	AEAAAKEAAAKA
Poly(glycine-serine)	Flexible	20	GGGGSGGGGSGGGGSGGGGS

Table 3.5. Synthetic protein linkers generated for Golden Gate assembly of Membrane Fusion proteins. A listing of each linker used for Golden Gate assembly, with flexibility, linker length and sequence recorded.

3: Proline rich linker

Polyproline chains have been demonstrated by FRET to act as a rigid structure to separate fused proteins (Schuler *et al.*, 2005). The derived formula is $(XP)_n$ in which X may represent any amino acid, although Ala, Lys and Glu are preferred (Ortiz de Montellano, 2005). Within this work, the polyproline linker was produced by selecting a variety of amino acids for X, so to reduce homologous repeats and therefore reduce risk of cleavage, however other references suggest X= alanine is the preferred linker structure (McCormick, Thomas and Heath, 2001).

4: Long 'Arai' linker

The use of glutamic acid and lysine within a peptide, separated by 3-4 amino acids, can result in the production of a stable α -helix via Glu^- — Lys^+ salt bridges (Marqusee and Baldwin, 1987) – Arai *et al.* used the formula $A(\text{EAAAK})_nA$ to produce variable linkers that exploited this idea. Arai used FRET to demonstrate that increasing the number of EAAAK repeats resulted in reduced fluorescence, representing an increased distance between two fluorophores. Flexible linkers, on the other hand, resulted in little difference in FRET measurement when size of the linker was increased (although looking at the SDS-PAGE gel, the mass of the largest linker was similar to the smallest Arai linker measured) (Arai *et al.*, 2001). In this case the modified formula $(\text{EAAAK})_n$ was used to produce the linker, termed “long Arai” due to its length, derived from the work by Lu and Feng on bifunctional fusions of β -glucanase and xylanase (Lu and Feng, 2008).

5: Short 'Arai' linker

A second Arai linker was produced, this time in accordance with the original formula $A(\text{EAAAK})_nA$ (Arai *et al.*, 2001). The Arai linker produced was shorter as an alternative to the length, and derived from Zhao *et al.*'s work on separating interferon and HAS within a fusion, resulting in an increase in anti-viral activity which was 115% greater with respect to a direct fusion, an improvement compared to a flexible linker (39%) and a proline rich linker (68%) (Zhao *et al.*, 2008).

6: Poly Glycine-serine linker

Polypeptide chains rich in Glycine-Serine repeats have been used as flexible linkers in multiple protein fusion experiments (Chen, Zaro and Shen, 2013). In the original pWaldo experiment a glycine-serine rich linker was also used, and the fluorescence reporter showed an unchanged performance within the fusion peptide with respect to the Waldo linker, however the homologous repeats were seen as risking cleavage (Waldo *et al.*, 1999). Furthermore, within

FRET tests, increasing the length of poly glycine serine linkers has little impact on FRET measured between the two separated fluorophores, supporting the flexibility of the linker (Arai *et al.*, 2001).

3.5 Using Golden Gate assembly to produce genetic fusions of XylE to a model GFP protein

Due to the propensity of MFS proteins to form natural C-terminal protein fusions (Willson *et al.*, 2019), it was hypothesized that MFS proteins would be more amenable to forming stable synthetic protein fusions at this end. As the *E. coli* protein XylE, itself a member of the MFS superfamily, was a key focus of this work it was also selected as the modelling protein to confirm that pLC1 can be used for producing membrane protein fusions. Furthermore, the structure of XylE is known, with the C-terminus on the inside of the cell (Sun *et al.*, 2012) making it suitable for substrate channelling with downstream enzymes present in the cytoplasm.

The *xylE* gene was amplified from *E. coli* chromosome using colony PCR, with the amplicon containing Bpil restriction enzyme binding sites cloned at either end, organised to cut a “C” overhang at the 3’ of the amplicon and a “D” overhang at the 5’ region. The amplicon was then cut and ligated into a DVA_CD plasmid using Golden Gate cloning.

The PCR amplification and Golden Gate cloning process was repeated for each of the six linkers (Table 3.5), introducing a DE overhang, and for E0040m, encoding a GFP molecule from the CIDAR MoClo kit, which was cloned to replace the CD overhang with an EF overhang. GFP was selected as the partner for the membrane fusion as it would allow visual confirmation of successful fusion by both fluorescence analysis and western blotting.

Once all holding plasmids were produced, a second stage of Golden Gate assembly could be applied to construct gene fusions in the expression vector pLC1. Five DNA fragments were introduced into pLC1, a constitutive promoter, an RBS, *xylE*, one of the six linkers and GFP. Candidate plasmids for successful Golden Gate assembly could be confirmed using the blue-white screening technique, transforming DH5 α with the reaction mixture and selecting for any white colonies.

The correct insertion of components within the expression vector was confirmed by DNA sequencing of the Golden Gate region. To further validate that the cloning was successful the plasmid size was measured following a single digest with XbaI, finding the mass of the each plasmid to be at the expected band mass of approximately 6500 bp, with the exact size varying a little depending on the linker used (Fig. 3.14).

The plasmids produced and their components are listed in Table 3.6.

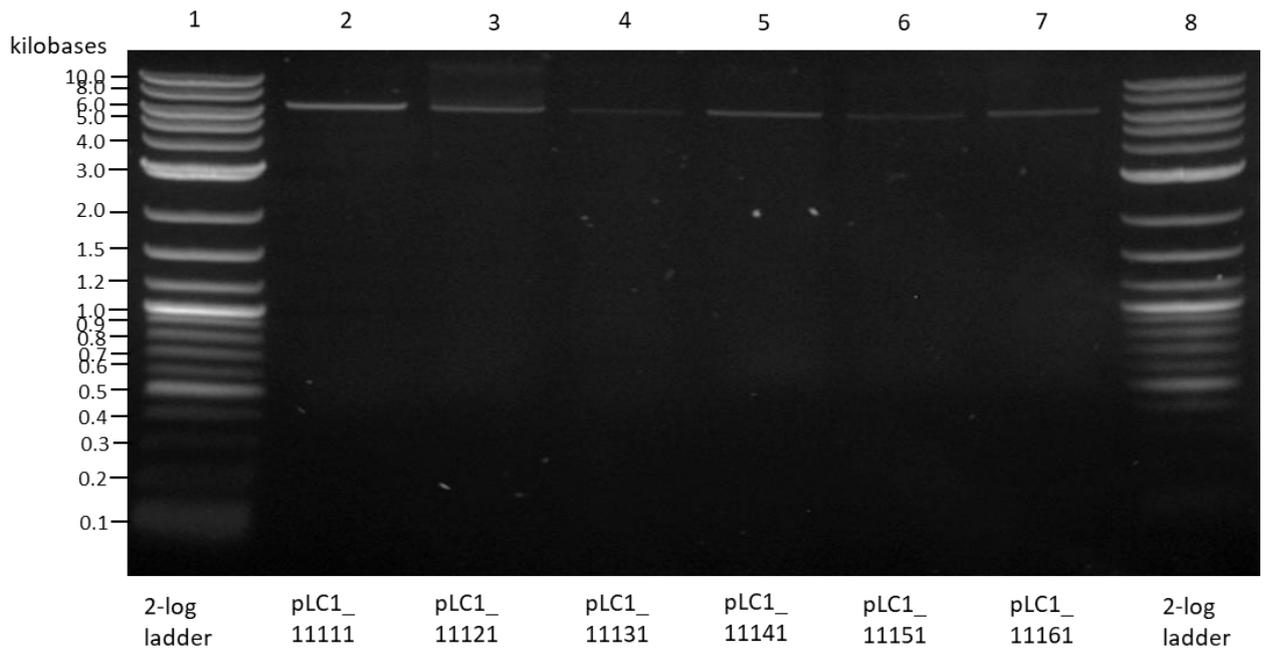


Figure 3.14. Electrophoresis gel of pLC1 fusion constructs following enzyme digest with XbaI. 0.5% (w/v) agarose gel of pLC1 fusion constructs produced by Golden Gate assembly. Plasmids were cut using XbaI and DNA fragments were separated alongside a New England Biolab 2-log DNA ladder. Sample DNA was mixed with 6x loading buffer in a 1:5 concentration of buffer to DNA

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein
pLC1_11111	J23100 (constitutive)	B0034m	XylE	Waldo	E0040m (GFPmut3B)
pLC1_11121	J23100 (constitutive)	B0034m	XylE	Single-chain flexible	E0040m (GFPmut3B)
pLC1_11131	J23100 (constitutive)	B0034m	XylE	Polyproline	E0040m (GFPmut3B)
pLC1_11141	J23100 (constitutive)	B0034m	XylE	Long Arai	E0040m (GFPmut3B)
pLC1_11151	J23100 (constitutive)	B0034m	XylE	Short Arai	E0040m (GFPmut3B)
pLC1_11161	J23100 (constitutive)	B0034m	XylE	Poly(glycine-serine)	E0040m (GFPmut3B)

Table 3.6. The components of XylE-GFP fusion constructs. A table displaying the components making up the six successful XylE-GFP constructs successfully produced by Golden Gate assembly. The plasmid pLC1 can contain up to 5 components.

3.6 Confirming function of the XylE-GFP fusion

Protein fusions could impact on the function of one or both of the fused domains. The fusion may cleave into smaller components, the expression may be unstable or multiple fusion proteins may aggregate. In order to confirm that the production of XylE-GFP genetic constructs resulted in the expression of the predicted protein, 3 key techniques were used in combination – epifluorescence microscopy, western blotting of whole cell lysate and a xylose deficiency growth assay.

3.6.1 Fluorescence analysis of C-terminal XylE-GFP demonstrates successful expression of fusion construct

As the GFP was fused to the C-terminal end of XylE, the expression of this protein would require complete translation of the XylE-GFP construct. Therefore measuring GFP expression using epifluorescence microscopy would support that the fusion has been produced, with the subsequent GFP either present within the fusion or soluble due to peptide linker cleavage.

Four DH5 α cell lines were therefore produced to observe whether fluorophores were produced. The first was transformed with unaltered vector pLC1, acting as a baseline for background fluorescence of DH5 α . On the other hand, transformation with pLC1_E0040m was hypothesized to express native GFPmut3B from the CIDAR MoClo kit as a positive control. For confirming Golden Gate activity, the plasmid pLC1_11411 was used, encoding a soluble xylulose kinase-GFP fusion, XylB-GFP. Finally pLC1_11111 encoded the desired fusion between XylE and GFP.

It was found that cells expressing pLC1 resulted in no notable fluorescence (Fig. 3.15.A and 3.15.B) and expressing pLC1_E0040m resulted in the expected GFP fluorescence (Fig. 3.15.C and 3.15.D), overall confirming the activity of the controls. Comparatively, cells expressing pLC1_11411 demonstrated similar fluorescence to the native GFP (Fig. 3.15.E and 3.15.F) supporting that the soluble XylB-GFP fusion was produced. Expression of pLC1_11111 did result in cells that were less fluorescent than the pLC1_E0040m and pLC1_11411 (Fig. 3.15.G and 3.15.H), which is likely the result of reduced expression or increased aggregation, however the presence of fluorescence supports the assembly of XylE-GFP fusion constructs and subsequent expression. It was important therefore to confirm that the fluorescence corresponded with successful assembly of the protein and not cleavage at the linker.

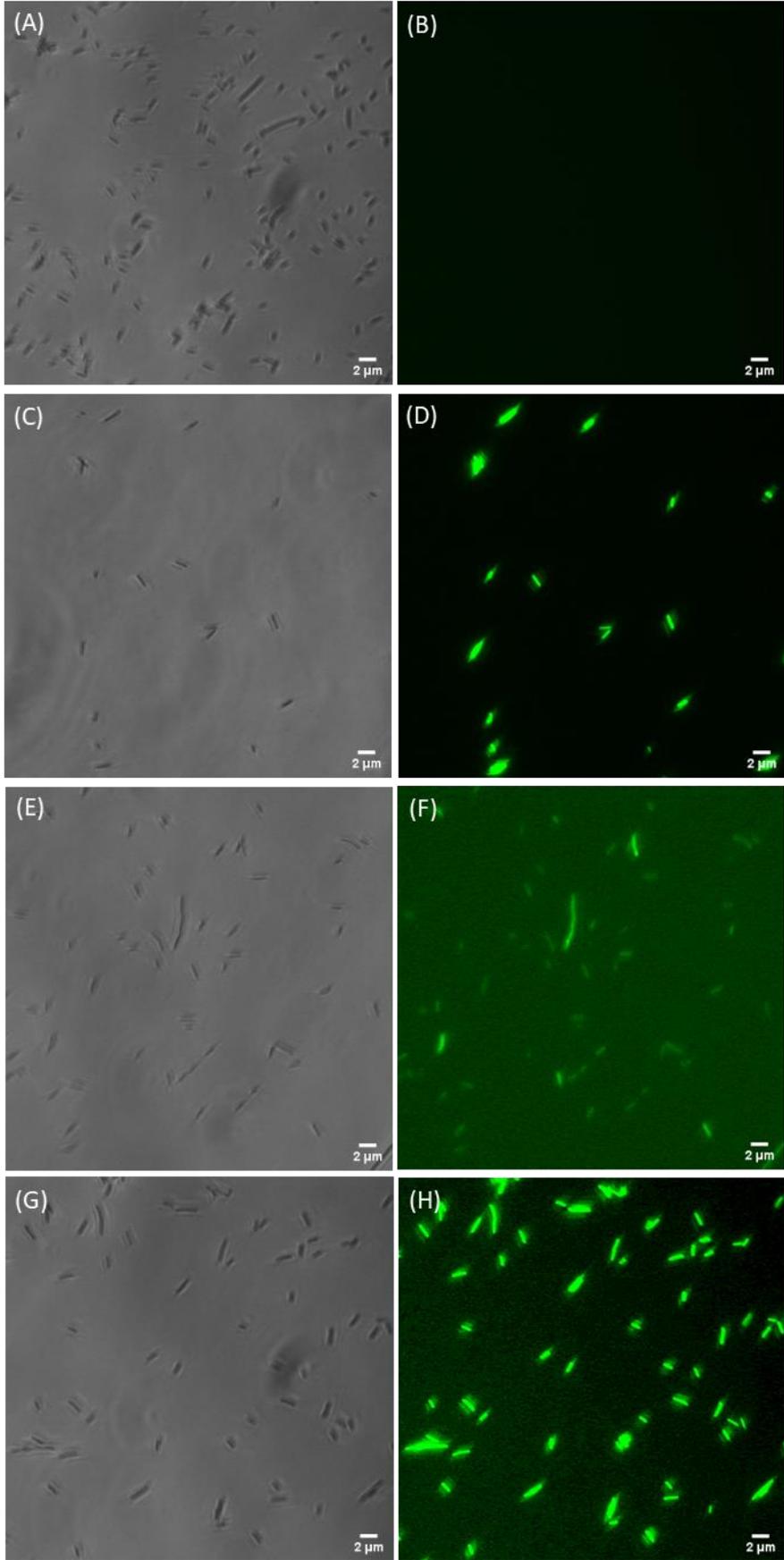


Figure 3.15. Confocal and epifluorescence microscopy of *E. coli* expressing GFP and fusion proteins. Confocal microscopy (left-hand images) and epifluorescence microscopy (right-hand images) of DH5 α expressing: (a and b) pLC1. (c and d) pLC1_E0040m, (e and f) pLC_11411 or (g and h) pLC1_11111. Scale bar represents 2 μ m.

3.6.2 Anti-GFP Western Blot further demonstrates production of the complete XylE-GFP fusion

In order to confirm that the GFP seen under epifluorescence was still fused to XylE, as well as determine whether any linkers resulted in partial or complete cleavage of the fusion due to homologous repeats, the expression of the complete fusion was confirmed using an anti-GFP western blot.

Western blots can confirm the presence of proteins or short amino acid tags of interest using industrial antibodies. The antibody anti-GFP, obtained from Bio-Rad Laboratories LTD, was selected for this work as it would avoid the need to add further His-tags to the proteins. Whole cell lysates were obtained and split into two elements – a soluble fraction derived from the supernatant and an insoluble fraction containing the pellet of the cell lysate. Within the soluble fraction would be cytoplasmic protein and within the insoluble fraction would be membrane bound proteins and any aggregates.

Using anti-GFP, western blots of native GFP and fused GFP were produced following SDS-PAGE derived separation of proteins against a ThermoFisher SuperSignal molecular weight protein ladder. Native GFP could be visualized in the soluble fraction between the 20 and 30 kDa bands of the protein ladder, with the protein mass of GFP being predicted at 26.91 kDa (Fig. 3.16A). The native GFP however was visible in the insoluble fraction in spite of a wash step. On the other hand, expression of native XylE, the negative control, did not result in any bands beyond the background, showing that there was no non-specific binding to the XylE protein.

Conversely, all XylE-GFP fusions produced multiplet bands found exclusively in the insoluble fraction suggesting complete fusion took place (Fig. 3.16.A and 3.16.B). The presence of such multiplet bands appears to be consistent with the type of linker used – doublet for flexible linkers (pLC1_11111, 11121 and 11161), and triplet for rigid linkers (11131, 11141, 11151). Presence of multiplet bands when a membrane protein is fused to GFP, including MFS protein LacY, has been previously observed (Geertsma *et al.*, 2008). Based on the work of Geertsma *et al.*, it is possible that these multiplet bands reflect a degree of misfolding or aggregation. Furthermore, no bands in line with native GFP were observed in either the soluble or insoluble fractions, suggesting that the linkers do not result in cleavage of the fused peptides.

One short-coming of using western blots to confirm membrane proteins is that the membrane protein moves through the SDS-PAGE gel at a different rate to soluble proteins, resulting in the bands being much lower on the gel than expected compared to the mass of the protein (Rath *et al.*, 2009). XylE-GFP fusions would be expected to be around 82 kDa in mass,

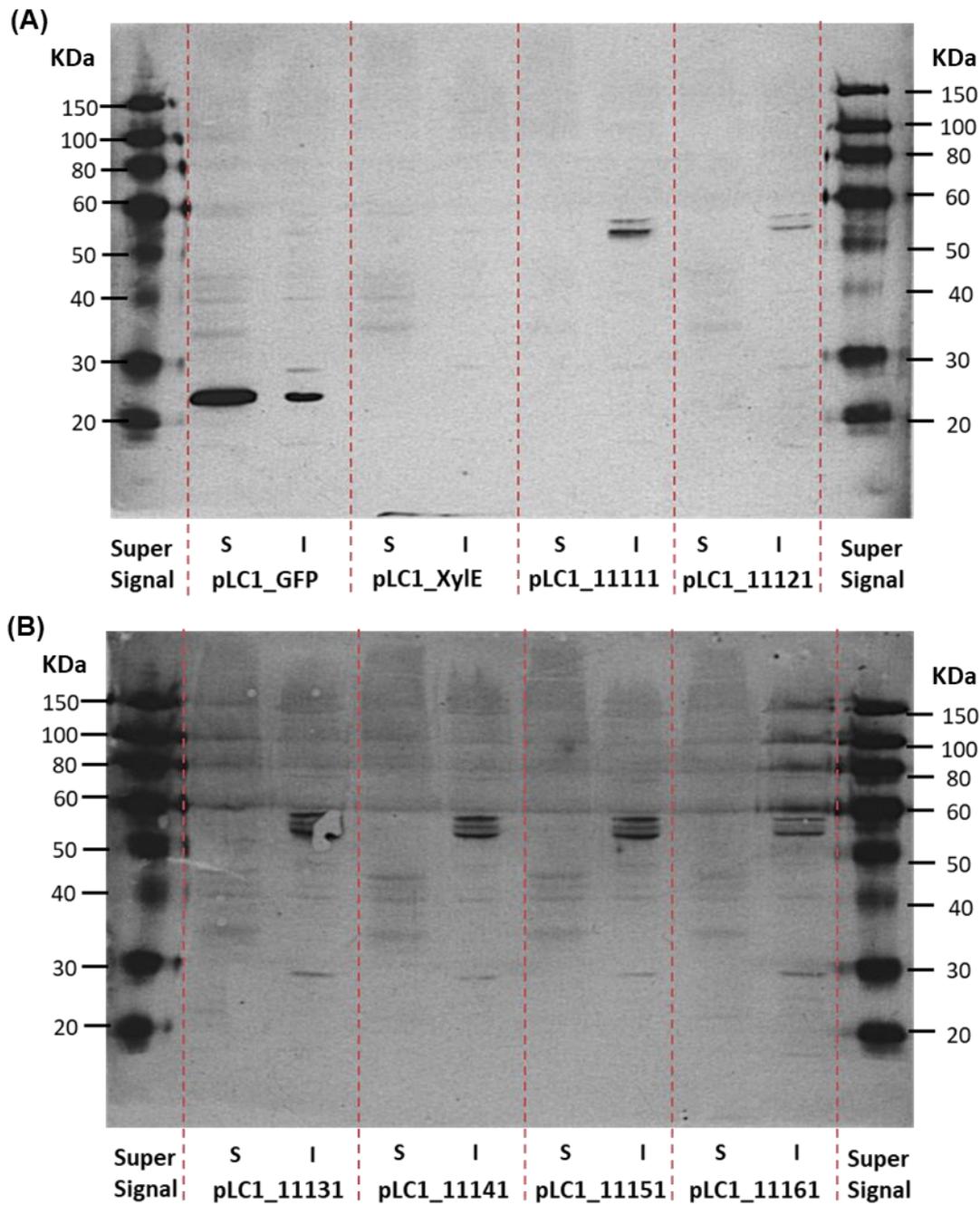


Figure 3.16. Analysis of *E. coli* strain DH5 α cell lysate. Anti-GFP western blot were produced from SDS_PAGE gels of whole cell lysate. DH5 α was transformed with pLC1 plasmids expressing Golden Gate cloned constructs. S represents soluble fraction of cell lysate, I represents the insoluble fraction.

varying somewhat depending on the linkers chosen. Within the western blot, however, they are found at between the 50 and 60 kDa bands on the ladder. While the western blot demonstrates that a fusion has been formed, and is membrane localized as it cannot be detected in the soluble fraction, it does not confirm that the XylE protein has retained the ability to transport xylose following fusion to GFP, necessitating further testing.

3.6.3 Growth assay

The BW25113-derived strain TDXYlose was produced within the Thomas lab in order to study the *E. coli* xylose transporters through growth assays. It contains gene deletions of $\Delta xylE$ $\Delta xylH$ $\Delta araH$ in order to inhibit cell growth on a xylose background solely due to the lack of D-xylose transport. The deletion of *araH* was introduced due to the non-specific transport of xylose by this protein.

As the western blots demonstrated that a membrane bound fusion was being produced, and little cleavage had taken place, this suggested that free XylE was absent. Therefore, the TDXYlose strain was used to test whether the fusion of XylE to GFP would impact the function of the xylose transporter. Furthermore, it allowed the impact of linkers on the fusion functionality to be observed.

Growth assays of the TDXYlose strains on M9+10 mM xylose found that fusing XylE to GFP had minimal impact on the growth of the TDXYlose strain. Cell growth for TDXYlose strains would begin within 8-10 h of incubation for all strains expressing XylE or a XylE fusion to GFP. While this was less than growth of the BW25113 strain, the rate of growth demonstrated that XylE fusions to GFP did not impact XylE function. Furthermore, TDXYlose was confirmed to have reduced growth on xylose with respect to BW25113 as when expressing GFP alone cell growth would not begin until 30 h after incubation (Fig. 3.17).

Within this work the type of linkers used appeared to have little impact on overall growth – while small variants could be seen within the growth of XylE-GFP fusions with different linkers, the variation could be explained with error and repeat experiments demonstrated that there was no retained pattern between linkers used and growth of cells (Fig. 3.17).

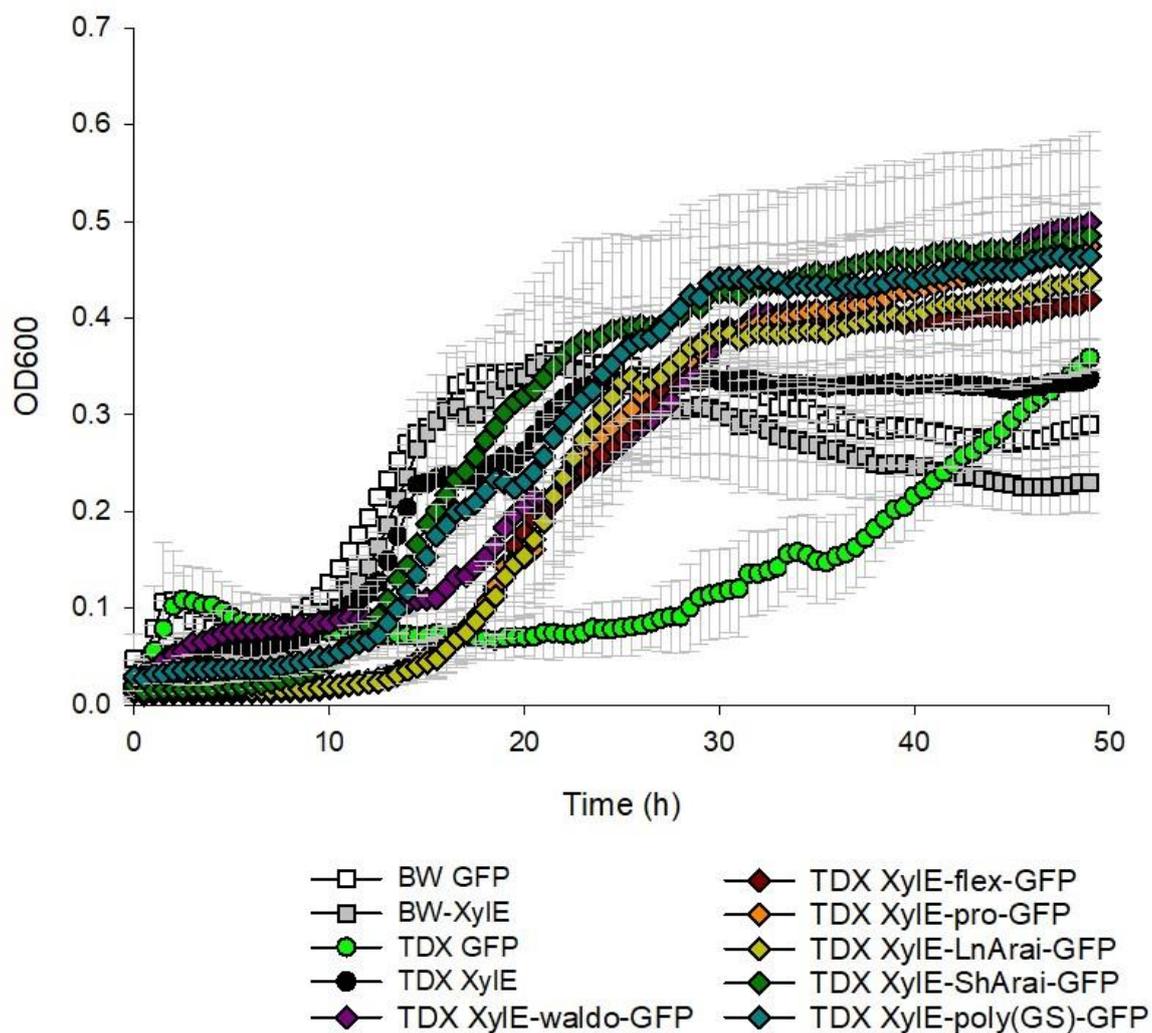


Figure 3.17. Recovery of TDXylose growth on a low concentration xylose growth medium by expressing XylE-GFP fusions. A growth assay of TDXylose on an M9+10mM Xylose background. Fusions of XylE-GFP were expressed and tested against XylE and GFP expressing controls. An additional expression trial in BW25113 was completed as a positive control for growth on xylose medium. N = 5. Error is standard deviation. Cells expressing either XylE or XylE-GFP fusions demonstrate equivalent growth, both of which were greater than the negative control, cells expressing GFP.

3.7 Golden Gate cloning to produce additional MFS-GFP fusions

As Golden Gate cloning was demonstrated to be able to produce a library of XylE-GFP fusion proteins, varying only in the linker, it was of interest to this work to apply the same technique to alternative MFS proteins to assess its general applicability. Two alternative MFS sugar transporters were therefore considered, ExuT and NanT. Both of these transporters were previously studied in the Thomas lab and therefore deletion mutants suitable for assaying in vivo transporter function were available, making them suitable targets for fusion studies.

3.7.1 Producing a fusion of ExuT and GFP

The *E. coli* galacturonate transport protein, ExuT, was selected for fusions with GFP.

Golden Gate cloning was applied as in section 3.5, introducing 5 gene fragments into a pLC1 vector after multiple stages of enzyme digestion and ligation, with the major difference being the use of ExuT as opposed to XylE as the third component. DH5 α was somewhat more recalcitrant to expression of the produced plasmids, often requiring repeat Golden Gate cloning in order to obtain white colonies suitable for plasmid harvesting. Furthermore, an arabinose-inducible promoter was used as opposed to the constitutive promoter used in section 3.5 as success rates were overall higher.

Ultimately 4 ExuT fusions to GFP were produced using Golden Gate cloning, which are listed alongside the components used in Table 3.7. In total two flexible and two rigid linkers were present respectively in successful ExuT-GFP fusions. Taking into account the difficulty in producing these fusions, it was important to determine whether GFP was impacting ExuT in any way.

The Thomas lab previously produced an *E. coli* strain BW25113 Δ exuT, which significantly reduced growth on glucuronic acid compared to wild type. BW25113 Δ exuT was transformed with each of four ExuT-GFP fusions, as well as a positive control, native ExuT, and a negative control, GFP. All ExuT expression vectors were under control of the AraC promoter, which was not induced with arabinose and relied on leaky expression.

For each of the produced fusions, BW25113 Δ exuT growth was recovered within 8 h of inoculation (Fig. 3.18). This rate of growth was comparable to native ExuT expressed in the BW25113 Δ exuT sample. Comparatively, BW25113 was able to grow within 4 h of inoculation, however this may reflect the levels of chromosomal ExuT expressed within *E. coli*. Finally the negative control, BW25113 Δ exuT expressing GFP, did not begin growth until over 40 h, demonstrating the impact of the transporter deletion.

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein
pLC1_31211	I13453 (ara inducible)	B0034m	ExuT	Waldo	E0040m (GFPmut3B)
pLC1_31221	I13453 (ara inducible)	B0034m	ExuT	Single-chain flexible	E0040m (GFPmut3B)
pLC1_31241	I13453 (ara inducible)	B0034m	ExuT	Long Arai	E0040m (GFPmut3B)
pLC1_31251	I13453 (ara inducible)	B0034m	ExuT	Short Arai	E0040m (GFPmut3B)

Table 3.7. The components of ExuT-GFP fusion constructs. A table displaying the components making up the four ExuT-GFP constructs successfully produced by Golden Gate assembly. The plasmid pLC1 can contain up to 5 components.

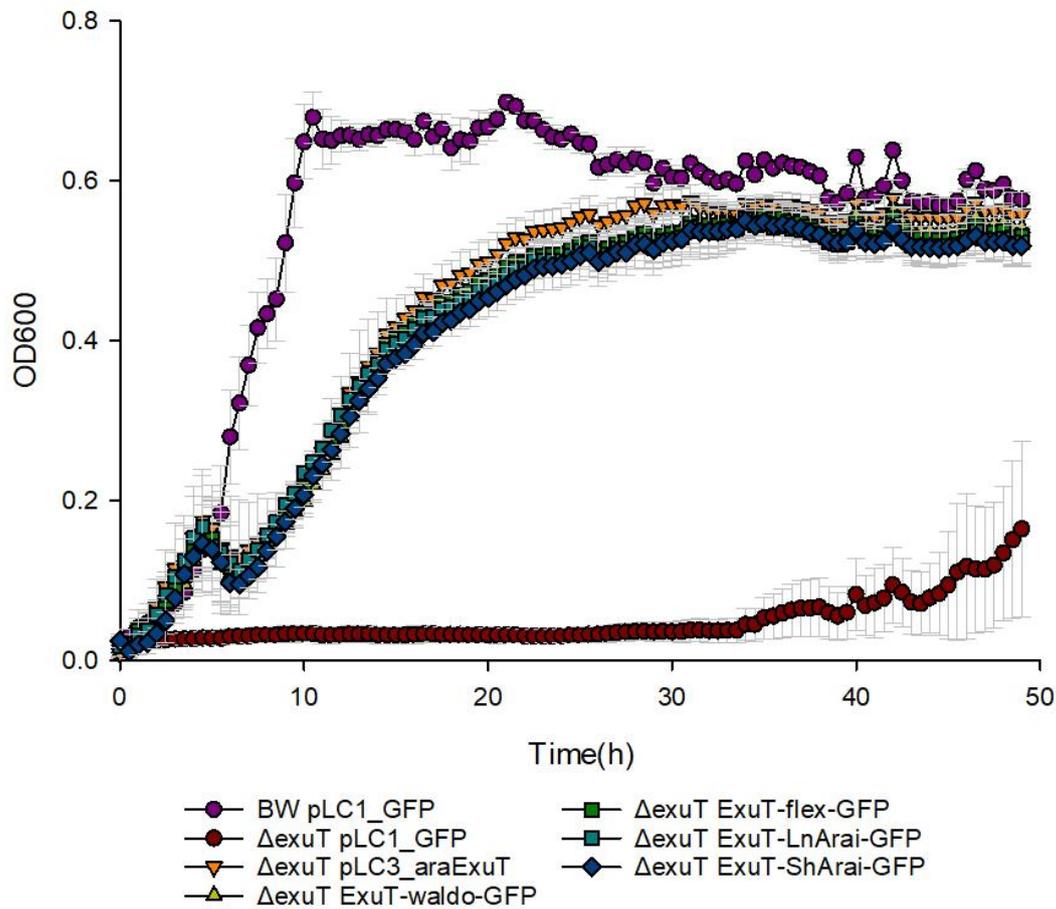


Figure 3.18. Growth of BW25113 wild-type and $\Delta exuT$ strains growth on a low concentration glucuronic acid. A growth assay of BW25113 $\Delta exuT$ on an M9+10mM glucuronic acid background. Fusions of ExuT-GFP were expressed and tested against ExuT and GFP expressing controls. No arabinose induction was required in this work. An additional expression trial in BW25113 was completed as a positive control for growth on glucuronic acid medium. N = 5. Error is standard deviation.

3.7.2 Producing a fusion of NanT to GFP

The third MFS fusion tested to demonstrate the application of Golden Gate assembly for membrane protein fusions was the *E. coli* sialic acid transporter NanT.

Golden Gate cloning was applied to produce the NanT-GFP fusions with one of three linkers. Once more DH5 α was more recalcitrant to expression of all plasmids, resulting in successful fusions containing one of the three flexible linkers but none of the rigid linkers introduced. The resultant fusions are displayed in table 3.8.

As with ExuT, the Thomas lab previously produced an *E. coli* strain BW25113 $\Delta nanT$ strain, which was transformed with each of three NanT-GFP fusions, as well as a positive control, native NanT, and a negative control, GFP. All NanT expression vectors were under control of the AraC promotor, which was not induced with arabinose and relied on leaky expression.

Each sample was grown on an M9+10 mM sialic acid background. The BW25113 $\Delta nanT$ cells transformed with pLC3 alone showed a truncated growth rate, as measured the reduced gradient of the exponential phase, reaching an OD of 0.3 at 20 h with respect to the positive BW25113 control, which reached the same OD of 0.3 at 8 h (Fig. 3.19). Comparatively the BW25113 $\Delta nanT$ sample expressing NanT reached this OD at 10 h, and each fusion reached it between 10 and 12 h.

Furthermore, expression of NanT or the NanT-GFP fusions within BW25113 $\Delta nanT$ resulted in the cell growth reaching stationary phase at an OD of approximately 0.55 between 15 and 20 h. Comparatively, the BW strain reached stationary phase within 12 h, similarly at an OD of approximately 0.55, whereas the BW25113 $\Delta nanT$ expressing pLC3 had yet to reach lag phase or an OD of 0.55. This suggested that NanT activity was retained, however compared to ExuT and XylE, linker properties appeared to have a greater impact on NanT activity.

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein
pLC1_31311	I13453 (ara inducible)	B0034m	NanT	Waldo	E0040m (GFPmut3B)
pLC1_31321	I13453 (ara inducible)	B0034m	NanT	Single-chain flexible	E0040m (GFPmut3B)
pLC1_31361	I13453 (ara inducible)	B0034m	NanT	Poly(Glycine-serine)	E0040m (GFPmut3B)

Table 3.8. The components of NanT-GFP fusion constructs. A table displaying the components making up the three NanT-GFP constructs successfully produced by Golden Gate assembly. The plasmid pLC1 can contain up to 5 components.

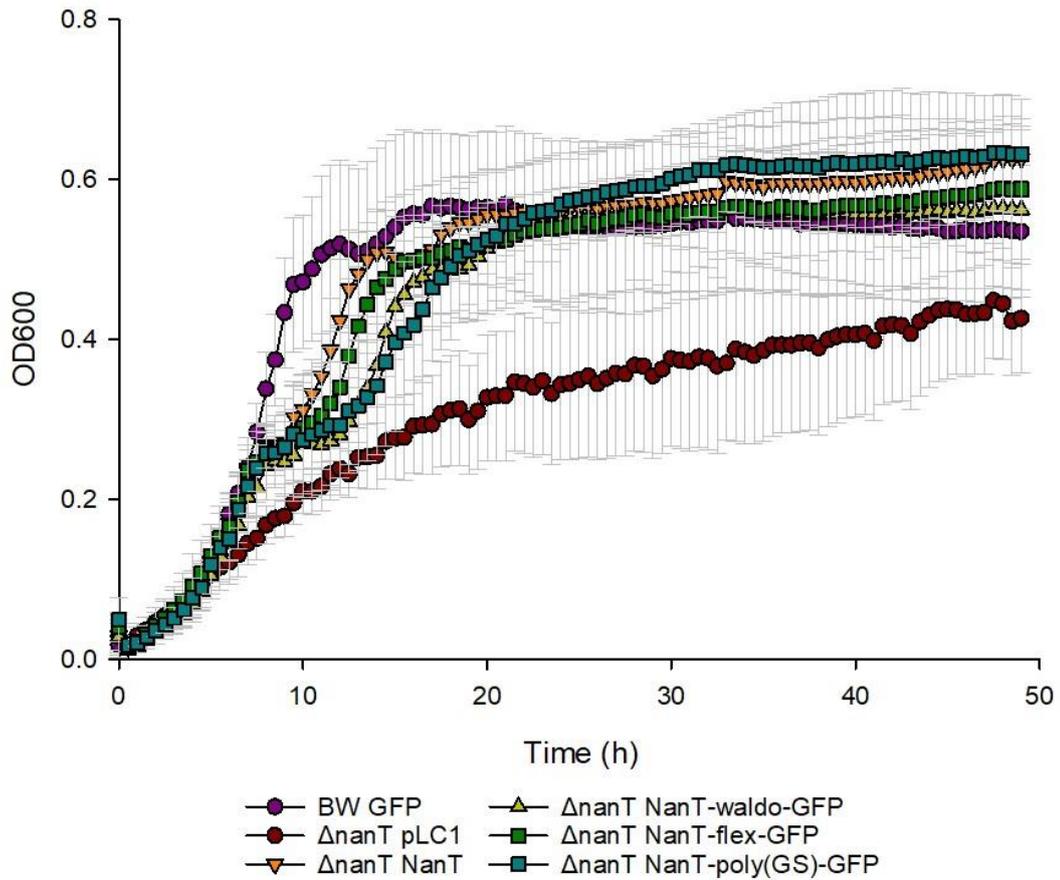


Figure 3.19. Recovery of BW25113 $\Delta nanT$ growth on a low concentration sialic acid background. A growth assay of BW25113 $\Delta nanT$ on an M9+10mM sialic acid background. Fusions of NanT-GFP were expressed and tested against NanT and empty plasmid expressing controls. No arabinose induction was required in this work. An additional expression trial in BW25113 was completed as a positive control for growth on sialic acid medium. N = 5. Error is standard deviation.

3.8 Summary of chapter

The Major Facilitator Superfamily proteins have demonstrated a significant propensity to produce fusions to downstream enzymes within nature. The presence of such fusions can be defined into one of 8 key groups (Fig. 3.5), most commonly to a second MFS or downstream lysophospholipid repair mechanism. Beyond the 8 groups determined within the Thomas lab, additional groups may also be present, although with reduced incidence compared to those identified, which makes determining the presence of such groups through Interpro a greater challenge (Willson *et al.*, 2019).

Within this chapter the stability of MFS proteins when used in synthetic protein fusions were also observed. Three different MFS proteins were studied, all used to transfer different sugar groups. It was found that the fusions were able to successfully recover activity within a relevant deletion strain, and that XylE, the protein to be studied in subsequent chapters, was stably accumulated regardless of the linker used.

From this work it can be concluded that the MFS transporter proteins are a suitable group for studying the impact of direct fusion and scaffolding at the membrane. Two key methods of achieving substrate channelling will therefore be tested using XylE as the target transporter, and applying Golden Gate cloning to assemble new libraries of fusions. The first method will test direct fusion to the downstream enzyme, Xylose isomerase (XylA). The second method will use smaller peptide tags to achieve a synthetic protein scaffold at the membrane.

Chapter 4

The direct fusion of XylE to XylA introduces constraints to enzyme function.

4. The direct fusion of XylE to XylA introduces constraints to enzyme function.

4.1 Validating the oligomeric state of *E. coli* Xylose isomerase.

The utilization of D-xylose by *E. coli* is achieved through an initial isomerisation reaction. Following transport of D-xylose across the inner membrane barrier through action of the Xylose/H⁺ symporter XylE, it is acted upon by the xylose isomerase, XylA. The *E. coli* XylA is a class II xylose isomerase composed of 440 amino acids and contains two divalent metal ions, believed to be Mg²⁺ (Schellenberg *et al.*, 1984; Tucker *et al.*, 1988; Farber *et al.*, 1989; Fatima and Javed, 2020).

Within *E. coli* XylA the metal ions are hypothesized to stabilize the catalytic intermediate, assisting in a 1,2-hydride shift (Farber *et al.*, 1989), which is the highest free energy barrier to xylose isomerisation (Meng, Bagdasarian and Zeikus, 1993; Choudhary, Caratzoulas and Vlachos, 2013). Studies of XylA homologues give more insight into the role of the metal ions, as well as highlighting important side chains within the enzyme. These homologues demonstrate that metal ions are essential in interacting with the substrate. One metal ion undergoes a change in coordination geometry from a tetrahedral to an octahedral geometry to bind the substrate O2 and O4 atoms during the reaction, while the second remains in an octahedral geometry regardless of whether substrate is bound, however moves its physical position to bind the O1 and O2 of the substrate, consistent with metal-mediated 1,2-hydride shifts, (Collyer, Henrick and Blow, 1990; Whitlow *et al.*, 1991; Jenkins *et al.*, 1992; Black, Huang and Cowan, 1994). Two histidine ions within XylA have been recognised as essential for function, one of which has a catalytic role as a hydrogen bond acceptor (Batt, Jamieson and Vandeyar, 1990). Homologues show that this histidine is locked in the correct tautomeric form by interacting with an aspartate ion. In addition a second aspartate ion within the active site acts as a base to attract the C2-OH proton (Meng, Bagdasarian and Zeikus, 1993).

The crystal structure of XylA in *E. coli* has yet to be characterized and as a result of this there is contradicting information available about its oligomeric state. Based upon homology modelling, UniProt, a database of protein sequences and information about biological functions (Bateman, 2019), has suggested XylA has a homotetrameric structure, however, BRENDA, an online enzyme repository system (Chang *et al.*, 2021), suggested a dimeric structure based on the literature. UniProt and BRENDA show a high degree of contradiction against one another on the oligomeric states of XylA in multiple organisms (Table 4.1). BRENDA recognises a diverse mix of oligomers, from monomer to tetramer, comparatively, the same proteins listed in UniProt are predominantly identified as homotetramers, with the notable exception of the homodimeric *Hordeum vulgare* xylose isomerase. Due to the

Species	Brenda Suggested Size	UniProt Suggested Size
<i>Arthrobacter sp.</i>	Tetramer	Tetramer (P12070)
<i>Bacillus coagulans</i>	Tetramer	Tetramer (G2TH70)
	Trimer	
<i>Bacillus sp. (in: Bacteria)</i>	Dimer	Tetramer (P54272)
	Trimer	
<i>Bifidobacterium adolescentis</i>	Tetramer	Tetramer (A1A0H0)
	Trimer	
<i>Escherichia coli</i>	Dimer	Tetramer (P00944)
<i>Geobacillus stearothermophilus</i>	Monomer	Tetramer (P54273)
<i>Hordeum vulgare</i>	Dimer	Dimer (Q40082)
<i>Lactobacillus reuteri</i>	Tetramer	Tetramer (B3XR24)
<i>Lactococcus lactis subsp. lactis</i>	Tetramer	Tetramer (Q9CFG7)
<i>Streptomyces albus</i>	Tetramer	Tetramer (P24299)
<i>Streptomyces corchorusii</i>	Tetramer	Tetramer (Q93RJ9)

<i>Streptomyces diastaticus</i> No. 7	Tetramer	Tetramer (P50910)
<i>Streptomyces olivochromogenes</i>	Dimer	Tetramer (P15587)
<i>Streptomyces rubiginosus</i>	Tetramer	Tetramer (P24300)
<i>Streptomyces</i> sp.	Tetramer	Tetramer (P19149)
<i>Thermoanaerobacterium saccharolyticum</i>	Tetramer	Tetramer (P30435)
<i>Thermotoga neapolitana</i>	Dimer	Tetramer (P45687)
	Tetramer	
	Trimer	

Table 4.1. Oligomeric state of xylose isomerase according to online databases BRENDA and UniProt. The oligomeric state of xylose isomerase proteins recognised by BRENDA were listed according to host organism, alongside the comparative oligomeric state suggested by UniProt. BRENDA identification code for all XylA proteins was EC 5.3.1.5, for uniport, individual IDs are listed below the oligomeric state. If UniProt and BRENDA did not contain information about the same strain, the closest available strain was recorded instead.

importance of the oligomeric state on fusion and scaffold protein assembly (Conrado, Varner and DeLisa, 2008; Lee, DeLoache and Dueber, 2012), understanding whether *E. coli* XylA is a dimer or tetramer is important for understanding whether protein fusion or scaffolding is a suitable technique for metabolic channelling.

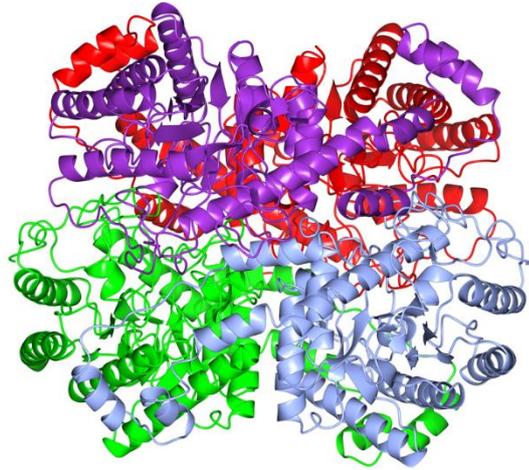
For proteins homologous to *E. coli* XylA, the crystal of the homotetrameric form of the xylose isomerase is the result of an assembly of two dimers, forming a “dimer of dimers” (Fig. 4.1.A) (Rey *et al.*, 1988; Hartley *et al.*, 2000). It is due to the hydrophobic nature of the xylose isomerase monomer that oligomerization occurs. A highly accessible hydrophobic region is present in the monomer, due to extended loop structures increasing surface area, which is otherwise buried with dimers (Fig. 4.1.B). These dimers are then brought together in tetrameric xylose isomerase (Rey *et al.*, 1988).

Examining the literature directly, initial research on *E. coli* XylA supports that XylA functions as a dimer. High performance size exclusion chromatography completed by Tucker *et al.* produced a purified protein with a molecular weight equivalent to a dimeric structure (Tucker *et al.*, 1988). A subsequent mutation study supported this conclusion, as XylA and a series of active site mutants were present as dimers according to HPLC and circular dichroism measurements (Batt, Jamieson and Vandeyar, 1990). More recent literature however has been prone to refer to *E. coli* XylA as a homotetramer (Fatima and Javed, 2020), or fail to identify its oligomeric state with respect to tetrameric xylose isomerase from different species (Epting *et al.*, 2005); however, this may reflect misrepresentation on databases such as UniProt, leading into confusion as to the true oligomeric state. Statements of the tetrameric state of XylA do not feature any attached references (Fatima and Javed, 2020).

Of particular interest is the effect of expressing xylose isomerase of different species in *E. coli*. Both the tetrameric xylose isomerase of *Thermotoga neapolitana* and *Thermoanaerobacterium thermosulfurigenes* are expressed as both a tetramer and a dimer in *E. coli*, with *T. neapolitana*-derived xylose isomerase expressed at a 20:1 ratio of dimer to tetramer (Hess *et al.*, 1998). This could be a result of the thermophilic nature of the bacteria, however, and is not true of all tetrameric xylose isomerase as *Bacillus licheniformis* xylose isomerase maintains its tetrameric confirmation when heterologously expressed (Vieille *et al.*, 2001).

Of the two methods selected to improve substrate channelling through the D-xylose utilization pathway, direct fusion is the focus of this chapter. Direct protein fusion requires that each monomer of an oligomeric protein is fused to another protein or monomer, which can result in a complex fusion. Larger multimeric metabolic enzymes have shown inefficient folding when fused to one another, resulting in insoluble aggregates or reduced activity

(A)



(B)

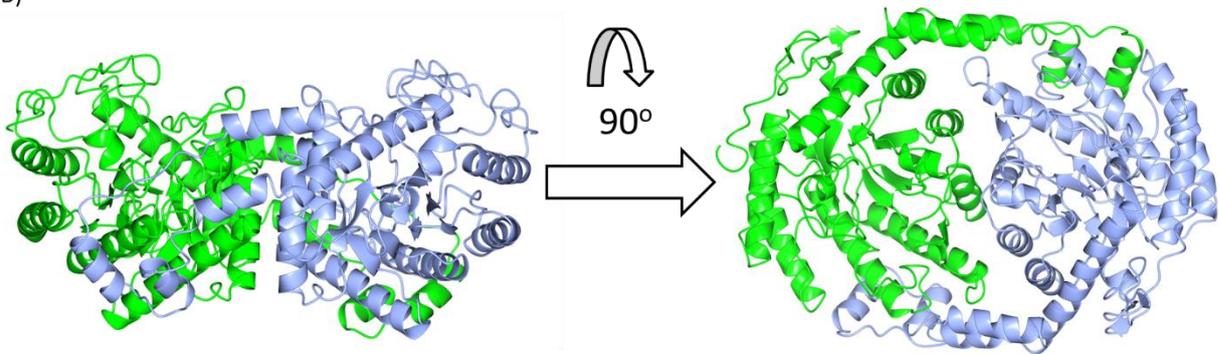


Figure 4.1. The crystal structure of *Thermoanaerobacterium thermosulfurigenes* xylose isomerase. The crystal structure of a xylose isomerase protein representative of tetrameric assemblies (A) the complete structure of the *Thermoanaerobacterium thermosulfurigenes* xylose isomerase (TatXI), PDB: 1A0C – with a tetrameric oligomerization formed by the assembly of a dimer of dimers. (B) The interface of two TatXI monomers forming a dimer. The N-terminal of either monomer forms an extended loop which interacts with the second monomer.

(Conrado, Varner and DeLisa, 2008). In the context of this work, the oligomeric XylA is to be fused to the C-terminus of the monomeric, membrane bound XylE. In order to understand the impact this fusion may have on both XylE and XylA functionality, it is important to understand the oligomeric state of *E. coli* XylA.

Two methods were used in order to determine whether *E. coli* XylA exists as a dimer or a tetramer: Size Exclusion Chromatography – Multiple Angle Light Laser Scattering (SEC-MALLS) of the soluble protein alongside analysis of the structure of the protein.

4.1.1. Size Exclusion Chromatography – Multiple Angle Light Laser Scattering (SEC-MALLS) of *E. coli* Xylose Isomerase.

SEC-MALLS was completed alongside Andrew Leech of the University of York Technology Facility. SEC-MALLS takes advantage of the differing refractive index of solvent and solute, resulting in an excess scattering of light compared to the use of solvent alone, allowing determination of the molar mass. It was used to determine the structure of *E. coli* XylA as the combination of light scattering and SEC allows for determination of the molar mass of differing oligomeric states.

The *E. coli* XylA was expressed in a pLC3 vector, with a 6xHis tag added to the C-terminus. The protein was first purified using nickel affinity chromatography, followed by size exclusion chromatography.

Over the course of the SEC-MALLS separation, completed at a flow rate of 0.5 mL/min, 3 well-resolved peaks were identified (Fig. 4.2). The first peak was identified as material eluting around the void volume (15 min/ 7.5 mL). Both Light Scattering (LS) and Quasi-Elastic Light Scattering (QELS) showed a significant peak, suggesting a bulky structure, however there is a much smaller, yet noticeable response in the UV and Refractive Index (RI) signals which indicates this is a non-negligible weight fraction of the material and is therefore likely to represent large aggregate of *E. coli* XylA.

As aggregate was accounted for, the other two resolved peaks represent oligomeric states of *E. coli* XylA: a major peak at ~24.5 min/ 12.25 mL and a minor peak just after 30 min/ 15 mL. Some additional peaks were recorded in the RI trace at the bed volume, the volume of the resin inside the column, however these were determined to be buffer artefacts.

For analysis the peak regions were set as follows: (1) centre of the major peak, 12.05 to 12.45 mL; (2) the whole of the major peak, 11.2 to 13.5 mL; (3) the high MW material at the void volume, 6.55 to 8.75 mL; (4) the late minor peak 14.45 to 15.95 mL; and (5) a catch-all peak to quantify the total material, 6.5 to 16 mL.

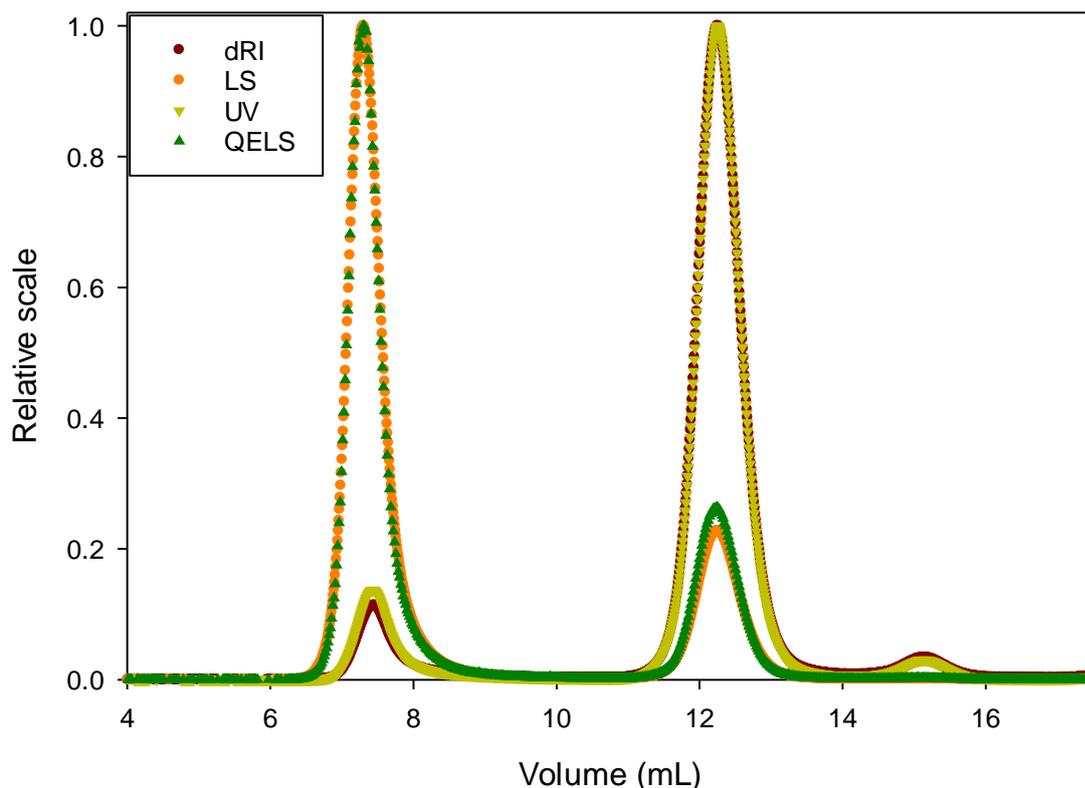


Figure 4.2. The elution profile of *E. coli* xylose isomerase during SEC-MALLS demonstrates 3 key peaks. The SEC-MALLS elution profile demonstrating 3 peaks across the elution volume. The elution profile shows UV (280 nm) in yellow, refractive index (RI) in red, light scattering (LS) in orange and Quasi-Elastic Light Scattering (QELS) in green. The RI signal is proportional to weight concentration of protein and is analogous to a UV absorption trace. The LS signal is proportional to concentration x molecular mass. A major peak is found from 11.2 to 13.5 mL makes up the major construct, with high abundance shown by UV and RI; a peak at the void volume, 6.55 to 8.75 mL demonstrates a low amount of high molecular weight aggregate (large LS and QELS peaks suggest high molecular weight); and a late minor peak 14.45 to 15.95 mL demonstrates a small amount of a protein of low molecular weight. The flow rate was 0.5 mL/min. Original image was produced with time on the x-axis, and used to analyse the data, therefore is shown in Appendix E.

The molecular weight of the proteins contained within each elution peak were obtained by looking at the narrow central region of the peaks. Looking at peak region 1, an estimated molecular weight of 207.4 kDa was determined (Fig. 4.3). The molecular weight of the His-tagged XylA monomer was predicted to be 51.85 kDa, which means the dimeric state for *E. coli* XylA would be 103.7 kDa and a tetrameric state would be 207.4 kDa. Curiously there is no peak consistent with the mass of 103.7 kDa, suggesting no dimerization has taken place.

The minor peak, captured by band 4, was determined to have a molecular weight of 67.9 kDa. This mass was much higher than anticipated but is believed to reflect the presence of a tail from the major peak overlapping the minor peak, obscuring the result somewhat and therefore causing the molecular weight to be recorded higher than in actuality. Overall, this minor band is expected to represent the monomer.

Mass of the three bands were determined to be 8.8 mg for the band at void volume representing aggregate, 87.5 mg for the tetrameric band and 3.3 mg for the monomeric band. Based on the mass of protein associated, the oligomeric states were therefore purified at a ratio of 14:2 of the tetramer to monomer, and 450:1 of tetramer to aggregate (assuming aggregate mass of 9 MDa as predicted by SEC-MALLS).

The total mass of each band together was determined to reach a total of 101.9 mg. This is much lower than the expected 240 mg that was loaded, which may reflect poor dilution or components of the protein solution becoming stuck within the column. It is possible that proteins becoming stuck within the column reflects the presence of additional aggregates that could not be eluted under the conditions chosen. As there was some delay between purification of protein using SEC and completing the SEC-MALLS trial, it is possible increased aggregation took place during that time.

In conclusion, this SEC-MALLS data suggests that *E. coli* XylA exists as a tetramer as opposed to a dimer. This was an unexpected result, as early literature suggested a dimer was the more likely oligomeric state through experimental data (Tucker *et al.*, 1988; Batt, Jamieson and Vandeyar, 1990), however a peak consistent with a dimeric state was not even present. To further validate this conclusion, more information on the structure of XylA was required.

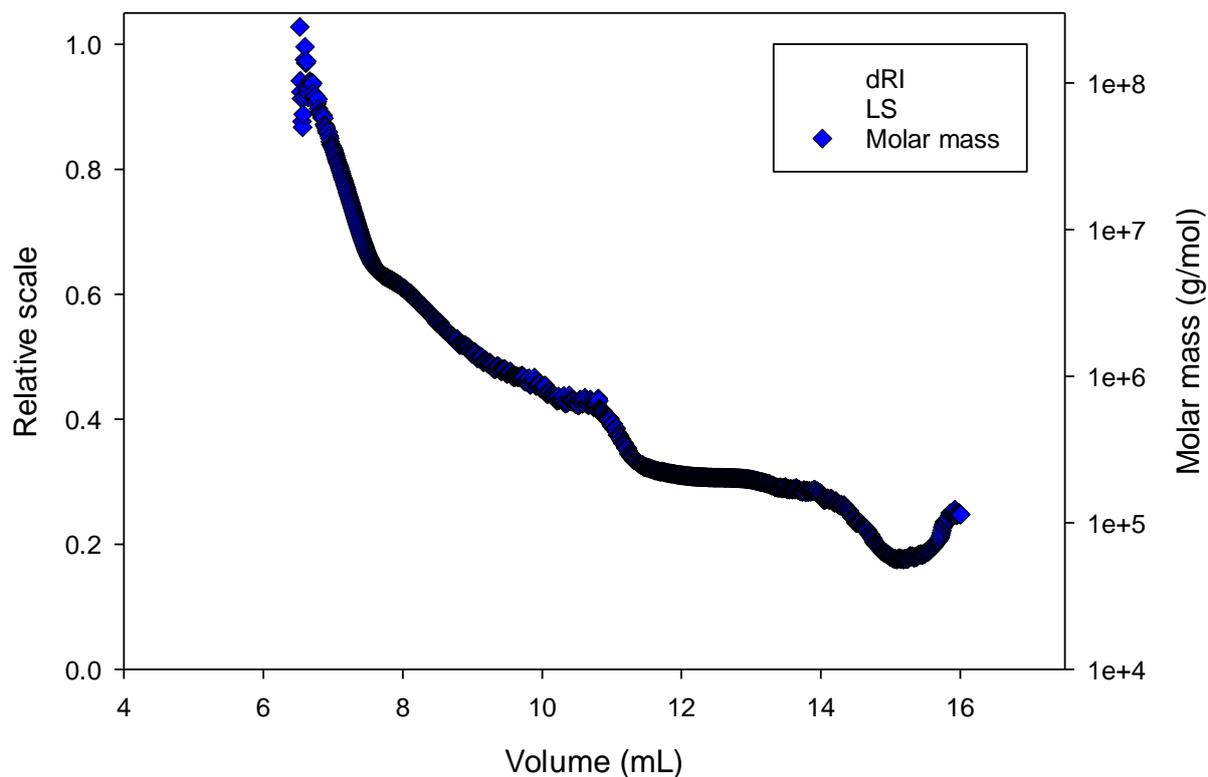


Figure 4.3. The molar mass of protein eluted over the duration of the SEC-MALLS procedure. The elution profiles of differential refractive index (RI) signal as a red line. The RI signal is proportional to weight concentration of protein. The Rayleigh ratio (light scattering signal, LS) is shown as an orange line; in this case the signal is proportional to concentration x molecular mass. The scales are normalised to the largest peak. The blue line on the plot is the calculated estimate of MW at each volume (elution) point and its value is indicated on the right-hand axis. For regions of low eluate concentration, the estimates become indeterminate and often appear very scattered. Aggregates and particulate material eluting at the column void volume may have very large MW but only comprise a very small weight fraction of the sample. The estimated MW plot on a logarithmic axis indicates that the material at the void volume is of very high MW. Flow rate of 0.5 mL/min

4.1.2. Crystallization of *E. coli* xylose isomerase

Following size exclusion chromatography of XylA, the concentration of the purified protein was determined to be suitable for running a crystal screen. A viva tube spin column was used to concentrate the protein to 15 mg/mL and 7.5 mg/mL, acting as the high and low concentration respectively.

A Mosquito robot was used to aliquot small volumes of the protein solutions into 96 wells of one of five different crystal screen plates, which were then stored overnight at 22°C. It was subsequently determined that H1 in Hampton Research's "Crystal screen" plate had immediately crystallized. Using 10% (v/v) glycerol as a cryoprotectant, a crystal was selected with an appropriate loop – It is important to note that during this procedure it was determined that glycerol was a poor cryoprotectant for XylA, as it resulted in resolubilization of the XylA crystals – 30% (v/v) PEG 400 was found to be a superior cryoprotectant. The conditions of XylA crystallisation within H1 was 0.1 M HEPES (pH 7.5), 10% w/v polyethylene glycol 8000, 8% v/v Ethylene glycol.

4.1.2.1. The oligomeric state of *E. coli* XylA

The recombinant *E. coli* XylA crystal from Crystal Screen H1 was analysed by X-ray diffraction at Diamond Light Source (data and refinement tables in Appendix F). The crystal structure was solved at a resolution of 2.13 Å by Reyme Herman through molecular replacement using the xylose isomerase of *Thermoanaerobacterium thermosulfurigenes* (TatXI, PDB: 1A0C), which had a 48% protein sequence identity. While the core of both the XylA and TatXI was similar between the proteins, the surface showed less structural similarity and therefore had to be rebuilt manually. The diffraction pattern showed XylA as a tetramer (Fig. 4.4.A) produced as a dimer of dimers, which was supported by the unit cell, showing that at 50% solubility four monomers were present with 65% probability. XylA is therefore similar to homologous xylose isomerase such as that from *Actinoplanes missouriensis* and *Streptomyces rubiginosus* (Rey *et al.*, 1988; Hartley *et al.*, 2000) and consistent with the previous SEC-MALLS result.

The structure appears to contain a core of similar to that of a $(\beta/\alpha)_8$ -barrel (TIM barrel) structure, composed of 8 parallel α -helices and 9 parallel β -strands (α 1- β 4, α 2- β 5, α 3- β 6/7, α 4- β 8, α 5- β 10, α 6- β 12, α 8- β 13, α 9- β 14). This TIM barrel-like structure differed from a true TIM barrel due to an interrupted third β sheet, as both β 6 and β 7 exist within the expected region (Fig. 4.4.B). This is consistent with previous protein crystals of homologous xylose isomerase, such as TatXI and *Bacteroides thetaiotaomicron* xylose isomerase (PDB: 4XKM).

At either terminus, additional secondary structures could be seen. Three small β -strands can be identified at an N-terminus extension from Ala1 to Arg37 (β 1- β 3). At the C-terminus

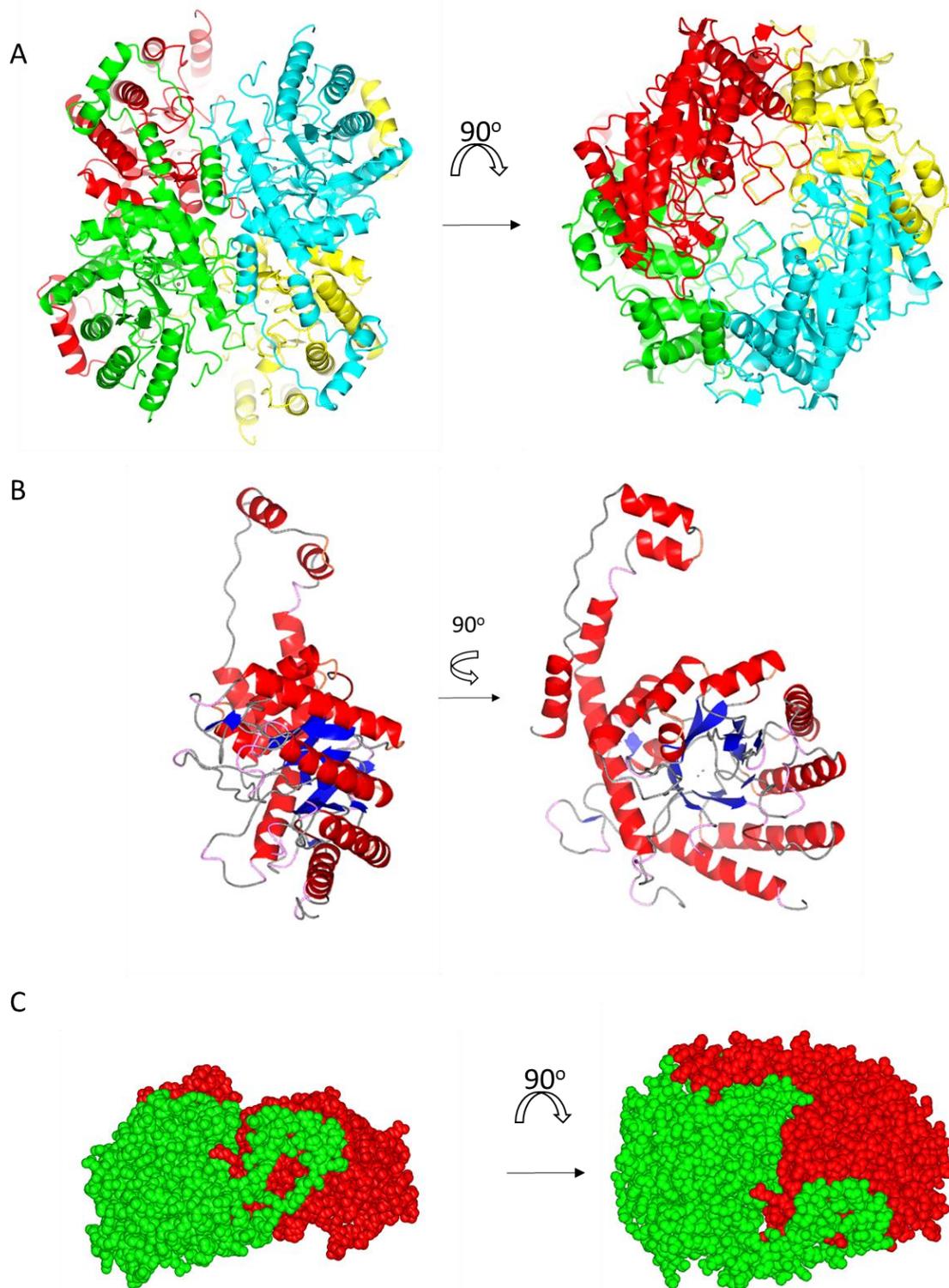


Figure 4.4. The solved structure of *E. coli* XylA suggests a tetrameric state. (A) A model of the tetrameric conformation of *E. coli* XylA, each monomer is identified by a different colour. Data collection and refinement table of the crystal in Appendix F. **(B)** A model of the XylA monomer. XylA has been labelled by colour according to secondary structure, with α -helices labelled in red and β -sheets in blue. **(C)** Space filling model of the dimer interface of two XylA monomers. The C-terminal extended structure overlaps with the second monomer.

extension from Gly377 to Asp439 four α -helices are present (α 11- α 14). The N-terminus extension is much smaller than the C-terminus extension and appears close to the active site. Furthermore, it appears that the C-terminus extension is responsible for interacting with a second monomer for dimerization (Fig. 4.4.C). Overall, this structure is consistent with the crystal structure of TatXI, with the key variation being the relative position of the amino acids.

A unique element of the crystal structure was the region between the Ala52 and Gly73, which was unable to be analysed due to low electron density between these residues (Fig. 4.5.A). The low electron density suggests a flexible region within the crystal structure between β 4 and α 2, possibly a flexible loop. This low electron density “loop” is located at the region where D-xylose enters the active site. Considering the flexibility of the structure, we hypothesize it could function as a cap to the active site to control substrate and product flux or could be responsible for interaction with another protein or the membrane.

The presence of a flexible loop within xylose isomerase has precedence within literature. *Paenibacillus* sp. xylose isomerase (PbXI) is a class II xylose isomerase homologous to XylA (Fig. 4.5.B). The crystal structure of the PbXI contains a loop region between β 1 and α 2 (residues 48–70) that was of low electron density and believed to be a flexible region. This flexible loop region was determined to be essential for protein activity through site-directed mutagenesis studies, and was hypothesized to function as a gatekeeper to control substrate entry and product release (Park *et al.*, 2019).

Park *et al.* identified two important Glycine residues were recognised in the PbXI β 1- α 2 loop, Gly44 and Gly67 (Park *et al.*, 2019). Multiple sequence alignment of structural xylose isomerase homologues derived from PDB (Berman *et al.*, 2000), showed high homology between class II xylose isomerase (Fig. 4.6), however only PbXI contains these glycine (Fig. 4.6.B). Furthermore, other than XylA, the crystal structures of the class 2 homologues do not show the flexible loop.

This suggests that the XylA flexible loop has been produced through alternative flexible amino acid combinations when compared to the G44/G67 amino acids within PbXI. Alternatively, it is important to consider that the flexible loop could represent a region that has been cleaved separate to the central structure. The XylA was stored for 4 weeks following initial extraction for SEC-MALLS study, therefore it is possible the protein degraded in this region.

4.1.2.2. The active site of *E. coli* XylA

Two metal ions were found within the active site of the protein. The protein sequence database Uniprot has identified 7 potential amino acids that are responsible for metal binding: E232,

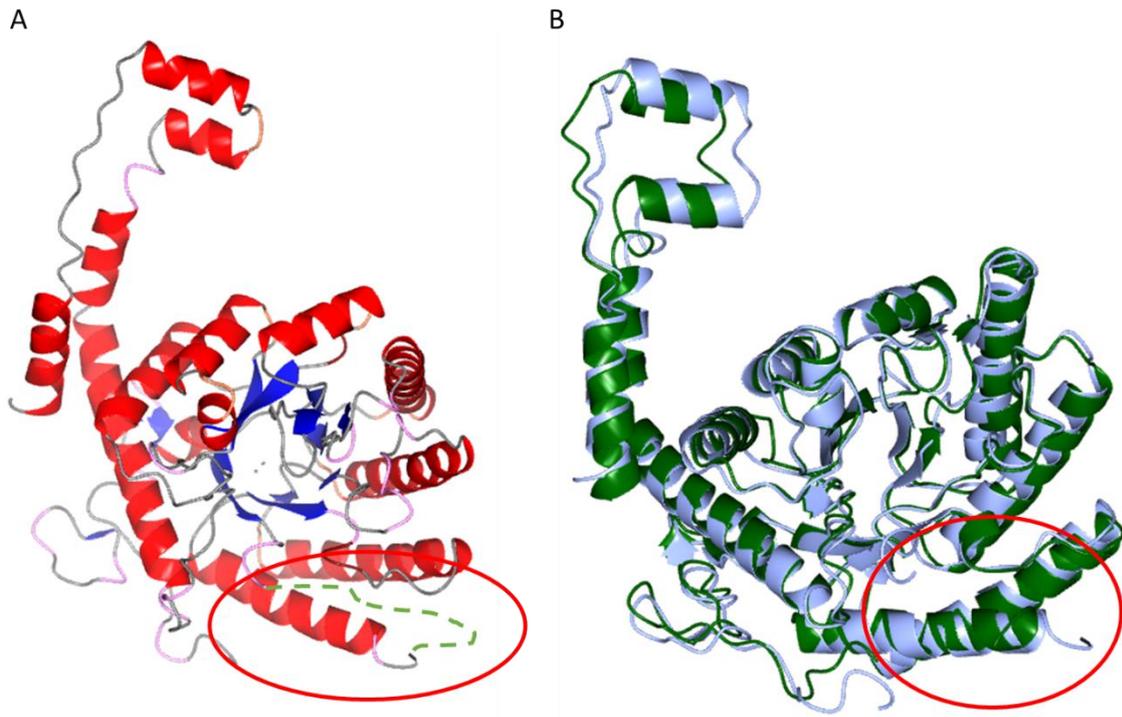
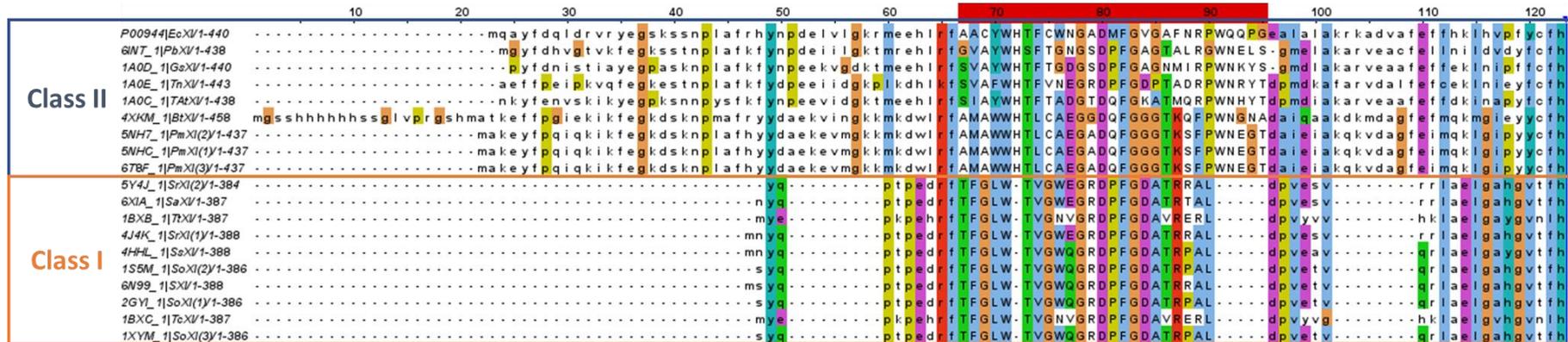


Figure 4.5. The missing loop of the *E. coli* XylA monomer (A) The model of *E. coli* XylA from Figure 4.4.B with the missing region represented by a dashed line – further highlighted by the red circle. **(B)** Overlap of *E. coli* XylA (grey) and PbXI (green) to demonstrate the missing region, highlighted by the red circle. Protein overlap in the image was produced by CCP4MG.

(A)



(B)

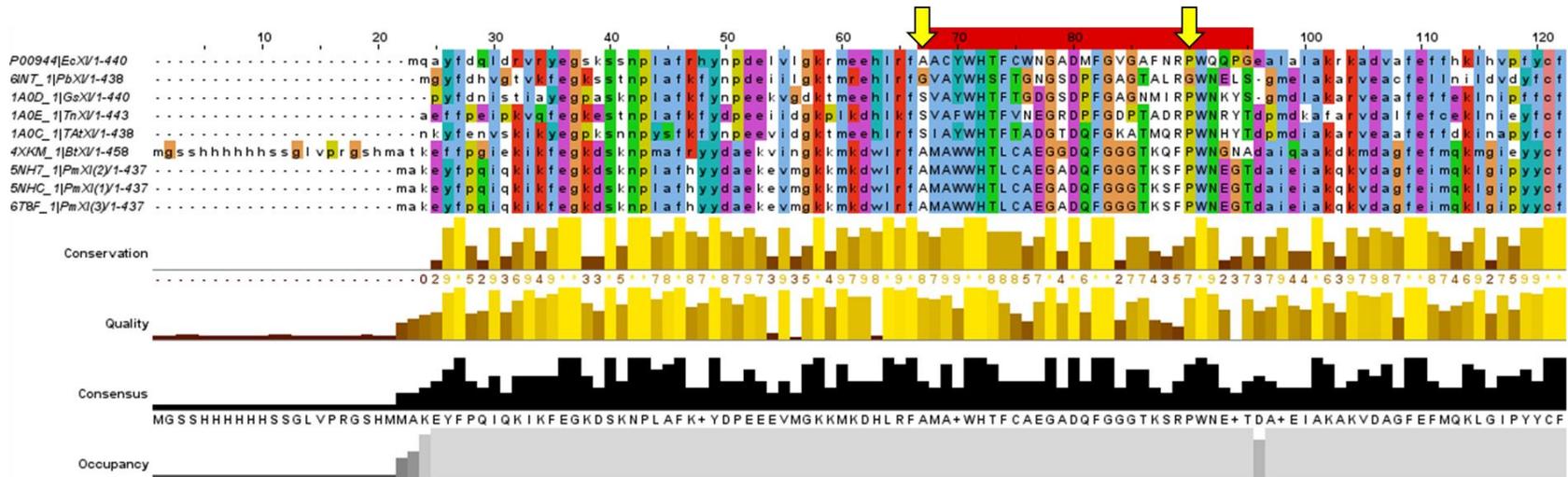


Figure 4.6. *E. coli* Xylose Isomerase shows strong homology with other class II Xylose Isomerase structures deposited on the PDB database, but does not share the two glycine domains predicted to be responsible for the PbXI flexible loop. Multiple alignment sequence of the xylose isomerase that are deposited on the PDB database (Berman *et al.*, 2000). **(A)** alignment of both Class I and Class II xylose isomerase. **(B)** alignment of Class II xylose isomerase alone. Amino acid sequences in lower case with the exception of the missing loop of *E. coli*, which is also highlighted in red, and the overlapping amino acids of the xylose isomerase of other organisms. The two glycine amino acids identified as responsible for the flexible loop in PbXI have been identified by the yellow arrows in (B). Protein structures were identified by their PDB number alongside an abbreviation of their species followed by XI (Xylose Isomerase). **Class II XI:** EcXI – *Escherichia coli* XI (XylA), TAtXI - *Thermoanaerobacterium thermosulfurigenes*, GsXI - *Geobacillus stearothermophilus*, TnXI - *Thermotoga neapolitana*, PbXI - *Paenibacillus* sp. R4, BtXI - *Bacteroides thetaiotaomicron*, PmXI - *Piromyces* sp. E2. **Class I XI:** TtXI - *Thermus thermophilus*, TcXI - *Thermus caldophilus*, SoXI - *Streptomyces olivochromogenes*, SXI - *Streptomyces* sp. F-1, SrXI - *Streptomyces rubiginosus*, SsXI - *Streptomyces* sp. SK, SaXI - *Streptomyces albus*. Where multiple protein structures were present, this was represented by adding a number to the identification - i.e. PmXI(1)

E268, H271, D296, D307, D309 and D339. The amino acid H271 was previously identified by Batt *et al.* as a potential metal binding protein, supporting this assumption (Batt, Jamieson and Vandeyar, 1990). Within the paper, Batt *et al.* also suggested that H101 was part of the active site, which Uniprot identified alongside D104, which was obtained through homology modelling.

Fig. 4.7.A highlights these side chains associated with metal binding or the active site within the XylA crystal. The amino acids are found on short loops outside the β -barrel structure; however, all side chains face towards one end of the inner barrel. Xylose is therefore bound at the opening of the β -barrel and catalysed by both the metal ion and the activity of H101 and A104. It is possible that the catalytic activity proposed for TatXI by Meng *et al.* could apply here also, in which His-101, Asp-104 and Asp-339 form a catalytic triad. Within the proposed triad His-101 becomes a hydrogen-bond acceptor after being locked into a tautomeric form by Asp-104, stabilising the substrate and transition state. Asp-339 could then act as a base, attacking the C2-OH to facilitate hydride shift and ring opening (Fig. 4.7.B) (Meng, Bagdasarian and Zeikus, 1993).

Forming the correct oligomeric state of an enzyme can be vital to ensure its activity (Parsonage *et al.*, 2005; Gogoi and Kanaujia, 2019). The location of the active site is an important consideration for the tetrameric state, as interaction with the other monomers could be required to facilitate a stable fold in the active site (Wente and Schachman, 1987), to produce an otherwise unavailable substrate binding sites (Parsonage *et al.*, 2005) or may impact allosteric regulation (Frieden, 2019). Looking at the oligomeric state, XylA exists as a tetramer composed of two dimers. Each monomer's active site is present facing inwards towards the active site of a different dimer in the dimer-dimer interface (Fig. 4.7.C). While there is some distance between each active site, suggesting they do not directly interact, it is possible the tetrameric structure shields the active site, perhaps to control access to the active site and increase selectivity. The missing density creates a point of access in the crystal, and perhaps suggests the flexible loop, if it does exist, plays a role in further controlling substrate access. Finally, the other end of the β -barrel faces outward and could therefore be another avenue for the substrate to enter the active site, or the product to leave.

4.1.2.3. The metal bound by *E. coli* XylA

To determine the metal ion present in the active site of XylA, X-Ray Fluorescence (XRF) of the XylA crystal was completed, which suggested that Mn^{2+} was present (Fig. 4.8). The XRF spectrum of manganese exists as two peaks: a large K_{α} peak at 5.89 keV and a smaller K_{β} peak at 6.49 keV (Schötzig, 2000; Lioliou and Barnett, 2018). Comparatively, the XRF spectra of the XylA crystal produced multiple peaks associated with iron and nickel, which can exist

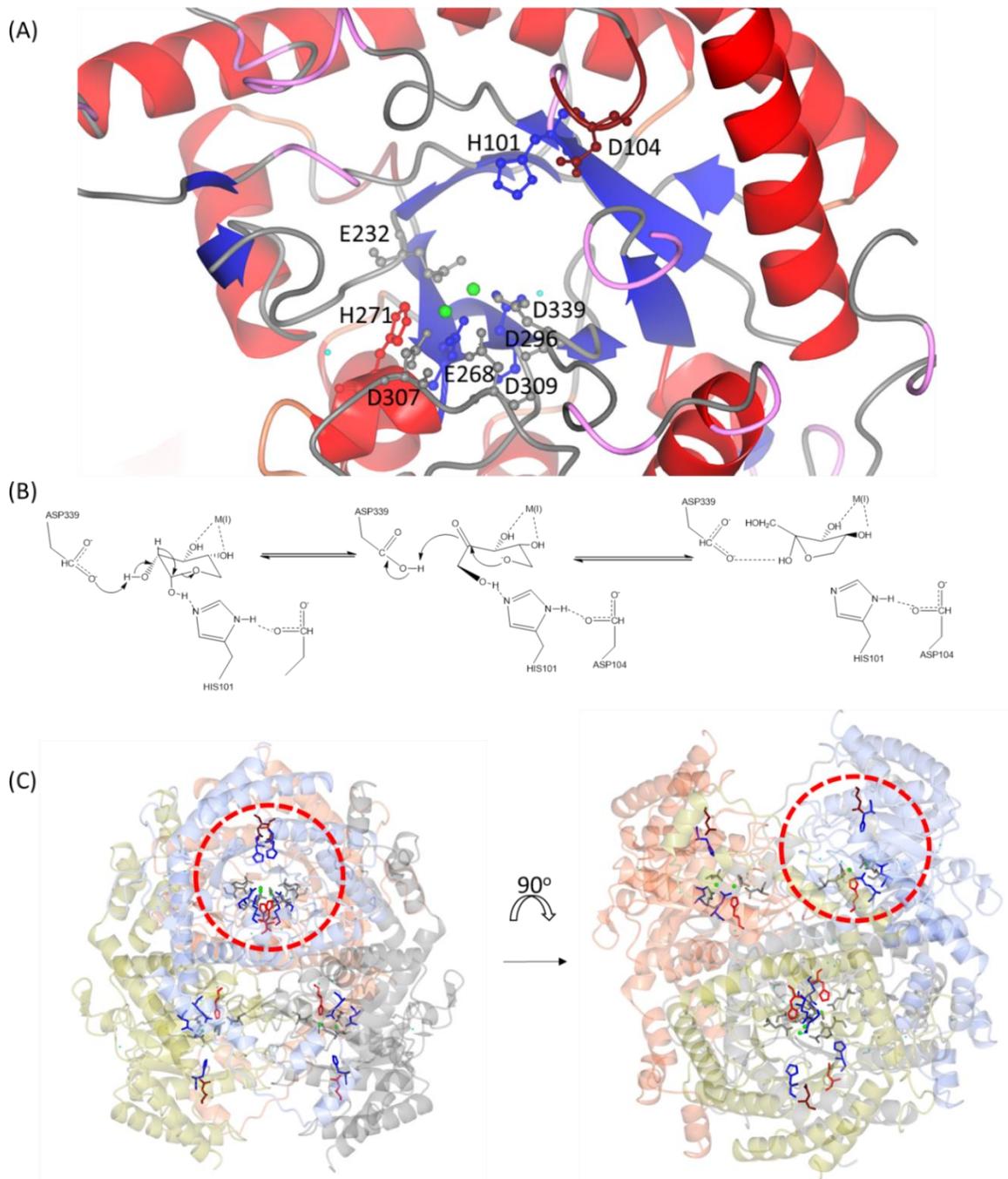


Figure 4.7. The active site of *E. coli* XylA sits within the opening of the TIM-barrel like structure – with active sites of each monomer appearing to act independent of each other (A) The active site of the XylA monomer – amino acids predicted by Uniprot to be responsible for the active site or metal binding have been selected. The metal ions are represented as green spheres and water molecules as cyan spheres. (B) Figure adapted from Meng *et al.* demonstrating the proposed catalytic mechanism for D-xylose isomerisation by TatXI (Meng, Bagdasarian and Zeikus, 1993) (C) The location of the active site in the XylA tetramer. The active site is highlighted by the red circle and faces inwards, parallel to the active site of the monomer at the dimer-dimer formation.

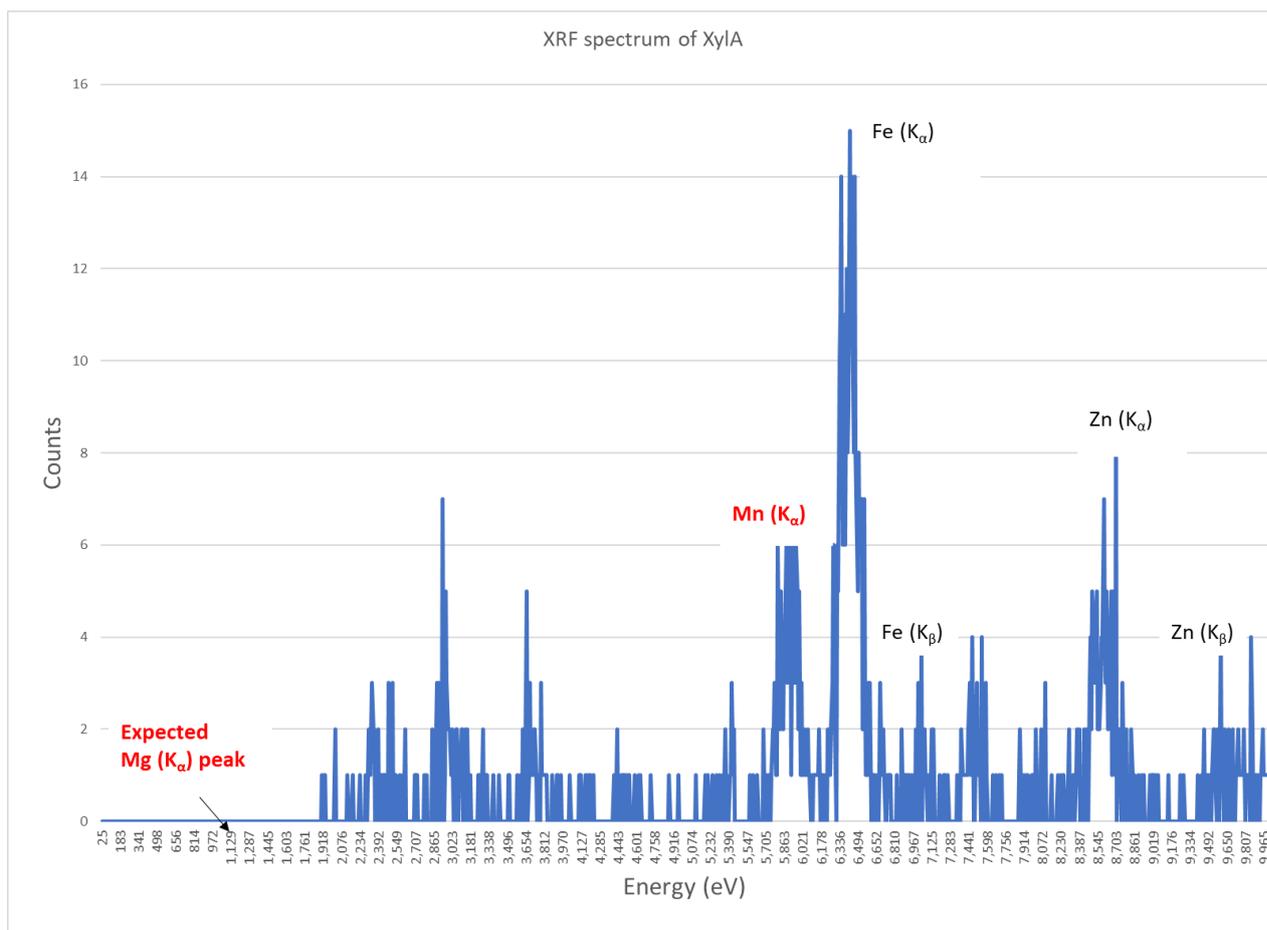


Figure 4.8. The X-Ray Fluorescence Spectrum of *E. coli* Xylose Isomerase. XRF spectra of *E. coli* XylA, data produced by Diamond Light Source. Key peaks have been identified derived from known spectra (Uo, Wada and Sugiyama, 2015). Pure Mn: K α - 5.89, K β - 6.49, Pure Zn: K α - 8.64, K β - 9.57, Pure Fe: K α - 6.4, K β - 7.06, Pure Mg: K α - 1.25.

as inhibitors of XylA. Of interest to this work, a clear peak at 5.863 keV was present, which is consistent with Mn ($K_{\alpha 1}$). It is important to note that the expected Mn (K_{β}) peak has been obscured by Fe (K_{α}) at 6.494 keV. On the other hand, the XRF data produces no peaks consistent with Mg^{2+} , which would be expected around 1.2 keV.

XylA was expressed in *E. coli* strain BW25113, grown overnight in LB broth, which contains a greater concentration of Mg^{2+} ($90 \pm 3 \mu M$) compared to Mn^{2+} ($324 \pm 32 nM$) (Nies and Herzberg, 2013) and therefore the media is unlikely to have caused an increased Mn^{2+} uptake within the cell. This suggests there is preferential binding of Mn^{2+} for XylA.

4.1.2.4. Concluding remarks

The structure of *E. coli* XylA was found to be tetrameric, supporting the evidence from SEC-MALLS data. The structure of XylA has high similarity to other tetrameric xylose isomerases, such as TatXI, containing a TIM-barrel like structure and forming a dimer of dimers. A unique element to XylA is that at the edge of the interface between the two dimers, a region of missing density can be seen which may represent a flexible loop. Beyond this, there are no obvious interactions at the dimer-dimer interface. This contrasts with the monomer, which features a C-terminus extension that increases surface area and facilitates dimerization. The dimer itself does feature such obvious extended structures, and the interaction instead appears to be a complete alignment of one edge of each dimer to the other.

Looking into crystals structures available on the PDB for reference, only one incident of a dimeric xylose isomerase crystal is recorded, the glucose isomerase of *Streptomyces rubiginosus* (PDB: 101H), which did not differ significantly from the structure of the homologous tetramers – and the active form of the enzyme is known to be tetrameric (Waltman *et al.*, 2014), suggesting this crystal was produced under unfavourable conditions. Formation of the tetramer may have a role in controlling access to the active site, as the tetramer both shields the active sites of all monomers from the cytoplasm and orientates these active sites to face the active site of another monomer.

Overall, we argue that *E. coli* XylA is a tetramer, contradicting early literature, but otherwise supporting predicted structures produced by Uniprot through homology modelling. The tetrameric state of XylA does raise questions about the feasibility of fusion to XylE, however. For example, will fusion to a membrane protein negatively impact the formation of this tetramer, and if so, what are the consequences of an altered oligomerisation? In addition, could the fusion of a large protein to the N-terminus of XylA limit access to the active site? It will therefore be important to keep the structure of XylA in mind going forward when designing protein fusions and other methods of substrate channelling.

4.2. Producing fusions between *E. coli* Xylose/H⁺ transporter and *E. coli* Xylose isomerase.

In the previous chapter Golden Gate assembly was used in order to produce a series of membrane protein fusions to soluble reporter protein GFP. It was determined that Golden Gate assembly was an effective method to interchange components for direct protein fusions. Therefore, Golden Gate assembly was applied in order to produce a series of direct fusions between *E. coli* XylE and *E. coli* XylA, using the previously established peptide linkers.

Five of the six desired XylE-XylA fusions were produced in this work (Table 4.2). The exception was the long Arai linker, which was found to be difficult to recover viable transformations within *E. coli* strain DH5 α . In all initial assemblies attempted, no white colonies could be observed and often few blue colonies were present also, suggesting successful digestion of the pLC3 backbone had taken place but that the complete target plasmid could not be assembled, either due to difficulties digesting or ligating all components. After some troubleshooting, fresh Bsal enzyme, Bsal-HF V.2., was obtained and shown to aid with production of fusions containing Long Arai linkers, however at this time the Long Arai linkers began to show a habit of truncation, losing a EAAAK repeat when inserted – changing length from 20 amino acids to 15. For this reason, Long Arai linkers were no longer used in future fusion experiments.

In order to validate the expression of a complete fusion between XylE and XylA, an anti-tetraHis western blot was performed (Fig. 4.9-4.12). Western Blots are produced using samples prepared following cell lysis through the lysozyme containing solution, BugBuster. Two fractions can be produced following BugBuster induced lysis, a soluble fraction containing cytoplasmic soluble protein, and an insoluble fraction containing large aggregates, membranes and membrane proteins. 3 types of protein were expressed in this work. Either a positive control of GFP fused to a His-tagged Dockerin, the His-tagged XylA protein or one of the candidate XylE-XylA fusions.

The first western blot contained soluble and insoluble fractions on two gels. The first gel, showing the soluble fraction, did not produce any bands, including the positive controls, and was therefore not shown here. Conversely, the second gel, showing the insoluble fraction, detected the GFP-Dockerin control (37.01 kDa) alongside two key types of band (Fig. 4.9). Cells expressing XylA produced a single band near the 40 kDa range, which is 10 kDa lower than the expected mass of 51.85 kDa. The mass of the XylA monomer purified by nickel affinity chromatography and size exclusion chromatography was determined by SEC-MALLS to be of the determined size, if not larger than the expected mass (Figures 4.2 and 4.3). This may therefore be a fault of the ladder used, Super Signal.

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein	Tag
pLC3_311161	I13453 (ara inducible)	B0034m	XylE	Waldo	XylA	6xHis
pLC3_311261	I13453 (ara inducible)	B0034m	XylE	Single-chain flexible	XylA	6xHis
pLC3_311361	I13453 (ara inducible)	B0034m	XylE	Polyproline	XylA	6xHis
pLC3_311561	I13453 (ara inducible)	B0034m	XylE	Short Arai	XylA	6xHis
pLC3_311661	I13453 (ara inducible)	B0034m	XylE	Poly(glycine-serine)	XylA	6xHis

Table 4.2. The components of XylE-XylA fusion constructs. A table displaying the components making up the five successful XylE-XylA constructs successfully produced by Golden Gate assembly. The plasmid pLC3 can contain up to 6 components.

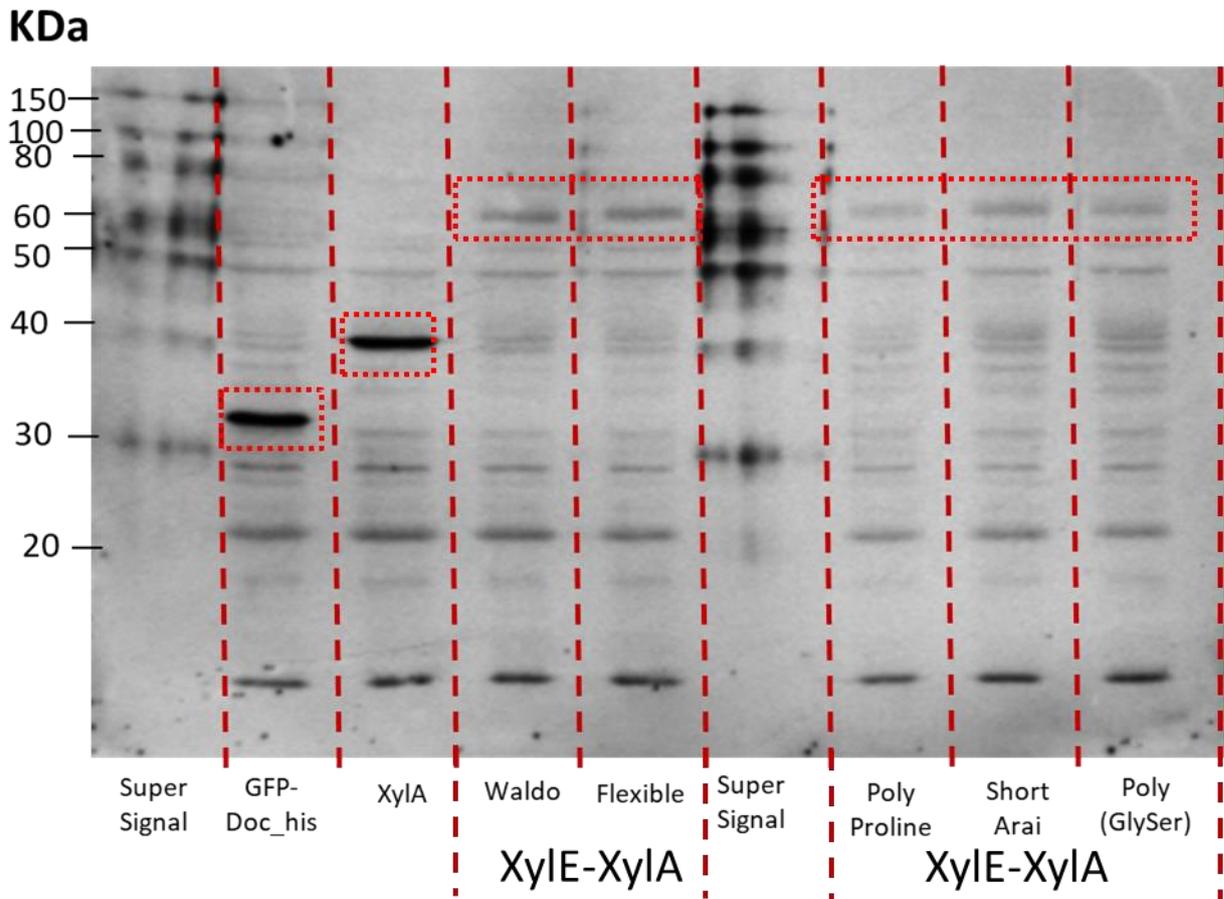


Figure 4.9. In absence of induction a XylE-XylA fusion is produced by transformed TDXYlose cells, however appears at lower levels than soluble proteins. An anti-tetra His western blot of insoluble fraction of uninduced XylE-XylA fusion expressed by TDXYlose, alongside XylA control and soluble control. The western blot was produced from SDS_PAGE gels of cell lysate following standard protocol.

Compared to the cells expressing the positive control or XylA, faint bands can be seen in all lanes expressing the XylE-XylA fusions, with the bands correlating to a weight greater than 60 kDa (Fig. 4.9). The mass of these bands were much lower than the expected mass of ~105 kDa. As with the XylE-GFP fusions, this could be explained by the movement XylE through the SDS-PAGE, as the folded membrane protein will not reach the same charge as soluble proteins, in addition to Super Signal being a poor ladder.

Looking at the uninduced insoluble fraction, which are representative of the conditions of upcoming D-xylose growth assays, cells expressing native XylA appears to produce their protein at a greater expression than those expressing the XylE-XylA fusions, represented by a much clearer band (Fig. 4.9). A Total Cell Lysate was also prepared without induction with arabinose and further demonstrated strong expression of XylA and weak expression of other components (Fig. 4.10). The Total Cell Lysate was prepared by lysis with BugBuster for 1.5 hours, followed by mixing the solution and taking an aliquot for loading onto an SDS-PAGE gel. The remaining cell suspension was then stored overnight at 4°C.

As opposed to Super Signal, the Precision Plus™ All Blue Protein Standard ladder was used (Well 1 Fig. 4.9). This showed the XylA expressed within well 3 at a clear 50 kDa region, once more suggesting that the Super Signal is at fault when suggesting expressed protein was 40 kDa in mass. A faint amount of XylA is also seen in wells 4-8, however no clear bands appear at higher molecular weights.

To determine whether there was expression of the XylE-XylA fusion, the following morning the Total Cell Lysate was removed from 4°C and separated into soluble and insoluble fractions. It is possible in this time further lysis occurred within the cell suspension, although enzyme activity would have been low due to the storage temperature.

A western blot was then produced under identical conditions (Fig. 4.11), however there was no "All Blue" ladder available at the time, so Super Signal ladder was used once more. As a result, the confusion relating to the bands can be seen. Within well 3 and 4 of both gels, XylA showed significant expression in both the soluble and insoluble fraction. A band at high molecular weight is present only within the insoluble fraction of well 4 in Fig. 4.10.A and 4.10.B, suggesting some aggregation which could have formed overnight or could have only been present at a low concentration and therefore not detectable within the total cell.

It appears that the XylE-XylA fusions were all expressed around the expected >60 kDa region, however high expression levels of native XylA can be seen in each fusion well. Curiously there appears to be an inconsistent level of expression between fusions, with the clearest band present for the flexible linker containing fusion. The faintest band, however, appeared for the

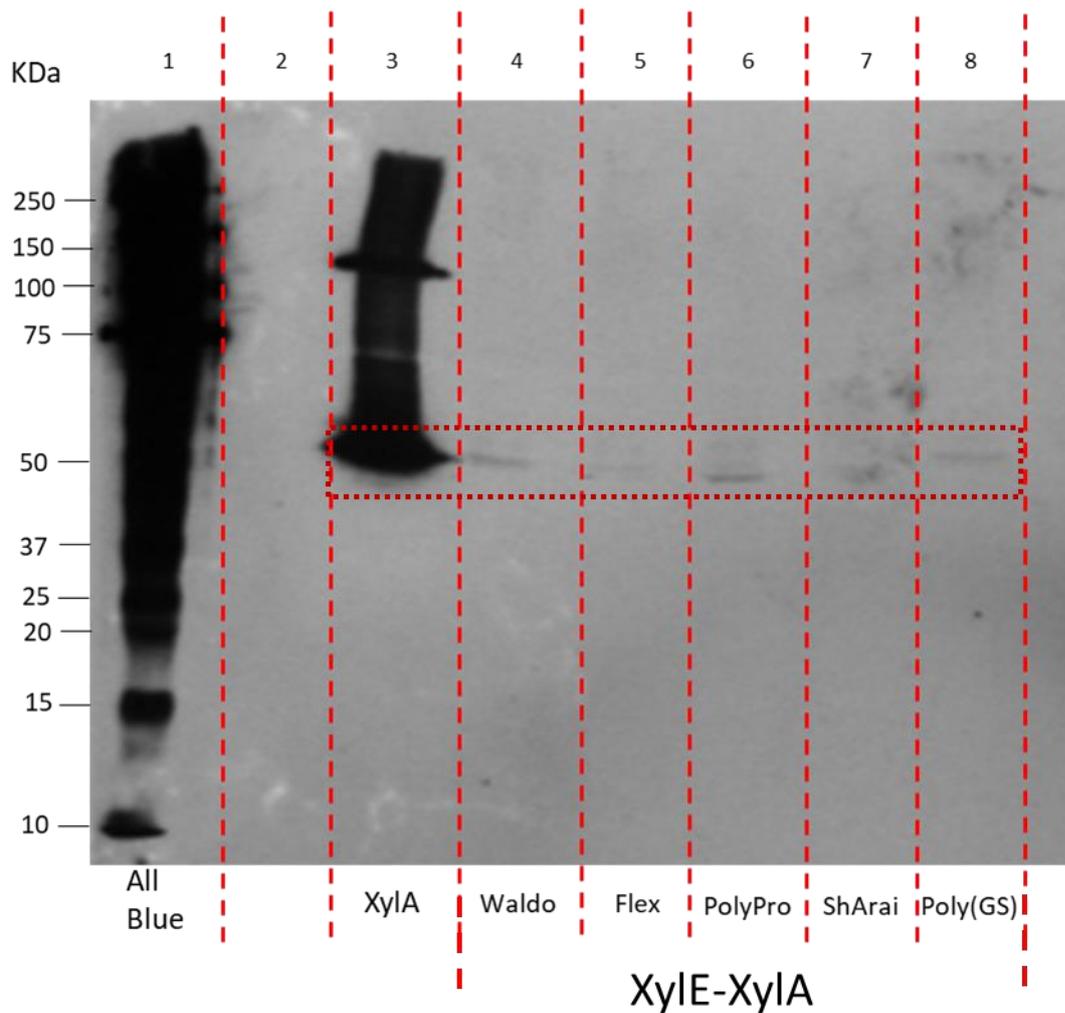


Figure 4.10. Under induction conditions only XylA is detected in the total cell lysate, with a different ladder confirming correct protein mass. A western blot identifying either Xylose Isomerase, XylA, or the Xylose transporter-Xylose Isomerase fusion, XylE-XylA, expressed from a PLD series plasmid within BW25113 strain $\Delta xylA$ without induction. Each fusion is identified by its connecting linker. Primary antibody was Mouse anti-TetraHis added in 3% (w/v) BSA, mixed overnight, secondary antibody was anti-Mouse added in 5% (w/v) skimmed milk in PBST. Forte HRP substrate was used to trigger chemiluminescence. A key feature of this western blot was the use of the All Blue ladder. The All Blue protein ladder demonstrates that the size of XylA is ~52 KDa, the predicted mass. This contrasts with the mass suggested when using the Super Signal ladder, ~40 KDa. Accounting for the SEC-MALLS data, where protein mass was also measured as 52 KDa per monomer, this suggests that the Super Signal ladder is inaccurate.

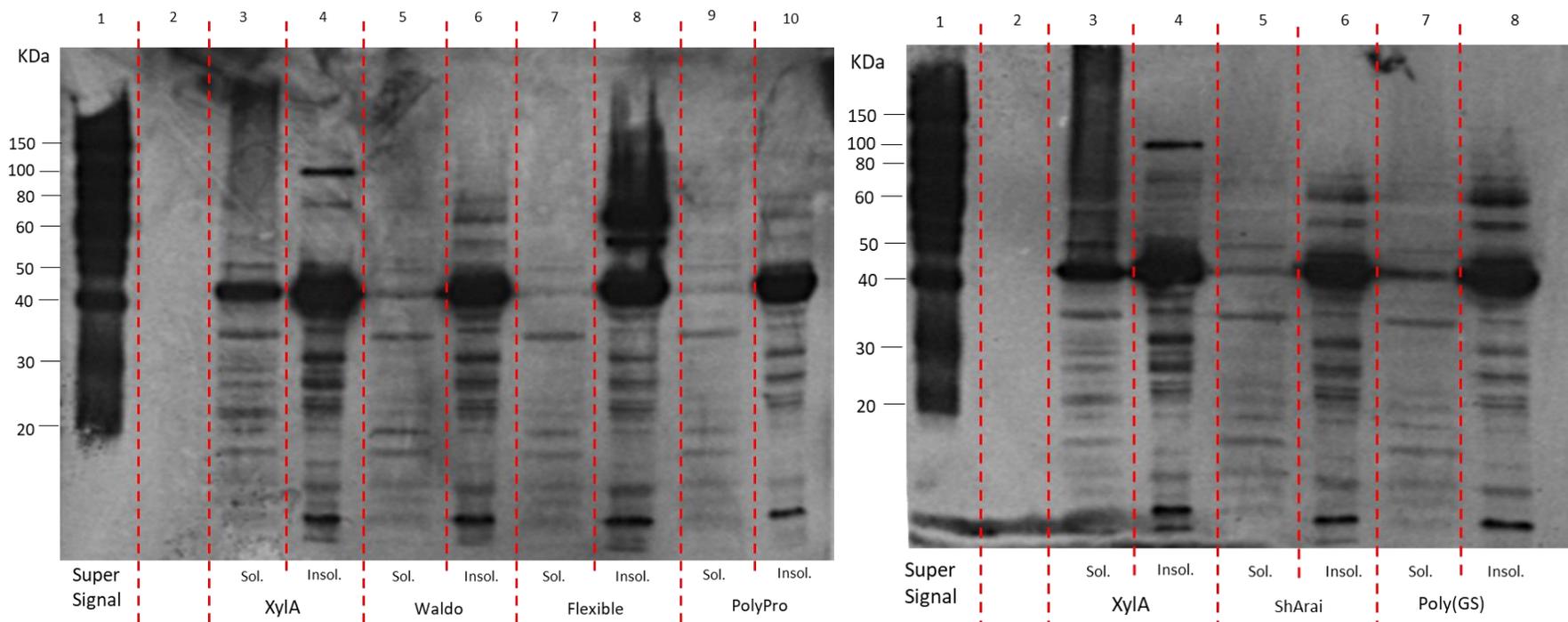


Figure 4.11. Variable amounts of the XylE-XylA fusion are detected by western blot depending on the fusion linker, with all samples showing high degrees of linker cleavage. A Western blot of the soluble and insoluble fractions of the Total Cell Lysate following overnight lysis at 4°C. Xylose Isomerase, XylA, or the Xylose transporter-Xylose Isomerase fusion, XylE-XylA, expressed from a PLD series plasmid within BW25113 strain $\Delta xyIA$ without induction. Derived from same lysis reaction as Fig. 4.5. however stored overnight at 4°C for continued lysis. Each fusion is identified by its connecting linker. Primary antibody was Mouse anti-TetraHis added in 3% (w/v) BSA, mixed overnight, secondary antibody was anti-Mouse added in 5% (w/v) skimmed milk in PBST. Forte HRP substrate was used to trigger chemiluminescence.

PolyProline linker containing fusion. Fusions containing either the Waldo, Short Arai or Poly(Glycine-Serine) linkers showed similar levels of expression.

The high levels of unfused XylA within the wells of Fig. 4.11 contrasts with what is seen in Fig. 4.9. This may suggest that the XylE-XylA fusion is highly unstable and prone to lysis if stored. Furthermore, it appears that the levels of native XylA expressed are greater than the XylE-XylA fusion and residual XylA in the fusion wells. This could be explained either by the XylE-XylA fusion being less available to bind by antibodies, causing fainter bands within the western blot, or that the XylE-XylA fusion being more strenuous for the cell to produce, therefore not being seen at equivalent levels to the expressed native XylA under uninduced conditions.

Overall, the western blots covered in Fig. 4.9-4.11 suggest that in absence of an inducer, the membrane protein fusions appear to express at a weaker level than the soluble XylA. While in practise an arabinose inducer would not be applied for growth assays, a separate western blot was completed with arabinose induction to determine whether increased expression would result in higher concentration of the stable fusion (Fig. 4.12). In this case, cells expressing XylA did not produce a band in the insoluble fraction, however bands within the soluble fraction had also failed to be detected (as seen by the positive control), therefore western blot can only be used to compare the relative expression levels and any degradation of the different XylE-XylA fusions.

Within Fig. 4.12, each XylE-XylA fusion was clearly detectable. Variance is seen between the expression levels of the XylE-XylA fusion, as seen within Fig. 4.11. In this case, however, the waldo linker showed weakest relative accumulation, whereas the polyproline and short Arai linker showed the greatest. Within multiple wells evidence of unfused XylA was present as two faint bands are seen under the expected XylE-XylA band (One at near 50 KDa and a second near 40 KDa). It is interesting to note that the polyproline linker appeared to have had the largest amounts of additional bands representing native XylA protein. This may suggest that although the fusion is well expressed, cleavage is still taking place within the linker region and producing unfused XylA. In addition, both Flexible and Short Arai linker-containing fusions showed clearly detectable bands. Within the waldo and poly(Glycine-Serine) linkers these bands may also be present, but are highly faint suggesting these were the two most stable fusions.

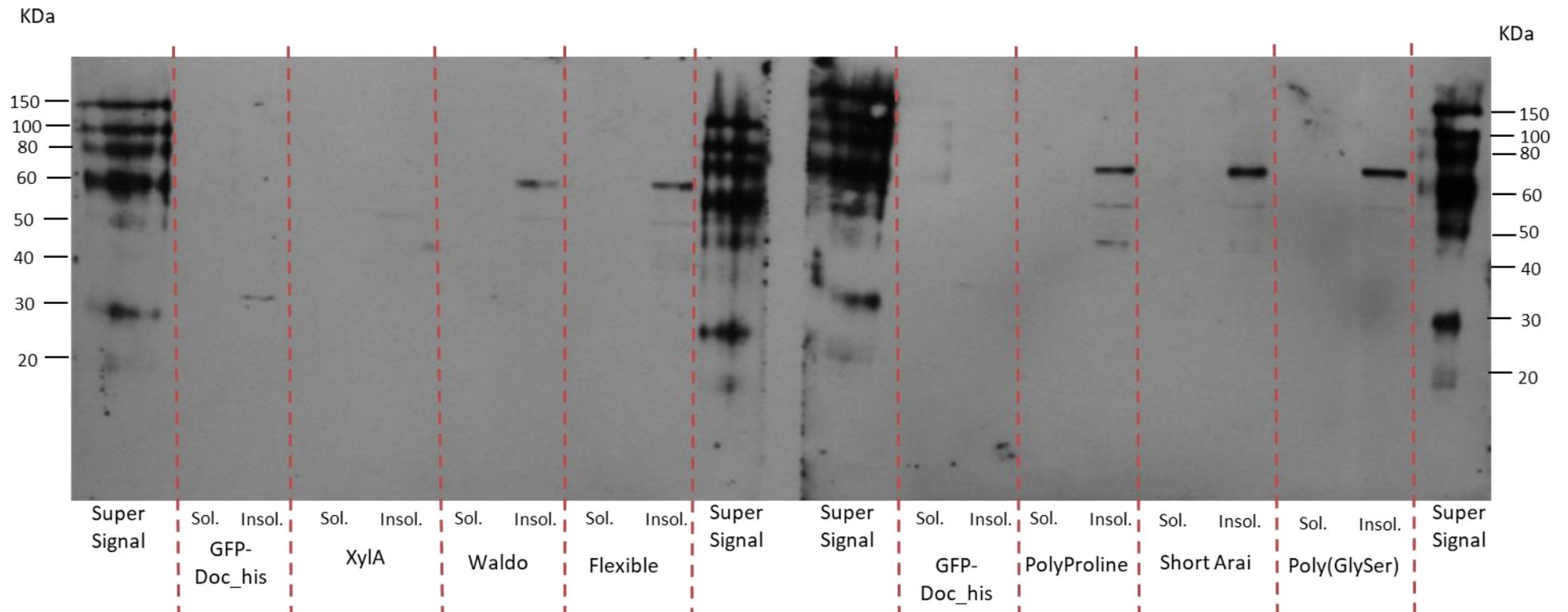


Figure 4.12. The polyproline, short Arai and poly(GlySer) linkers show strong detection by a western blot, however the rigid linkers appear more prone to linker cleavage. An anti-TetraHis western blot of the soluble and insoluble fractions of the XylE-XylA fusions following arabinose induction. The western blot was produced from SDS_PAGE gels of whole cell lysate. BW25113 $\Delta xylA$ was transformed with pLC3 plasmids expressing Golden Gate cloned constructs. “Sol.” Represents the soluble fraction of cell lysate, “Insol.” represents the insoluble fraction. Soluble fractions did not appear to correctly transfer in the western blot, either due to loading issues or an issue during blotting.

4.3. Fusion of XylE to XylA results in a decreased lag in cell growth over wild type XylE

Following the production of XylE-XylA fusions, it was important to understand the impact of the fusion on the activity of the two components. Using the XylE-GFP fusions as a model in the previous chapter suggested that XylE function was permissive to direct fusion, however, XylA has a molecular weight of over 50 kDa, much greater than GFP. Furthermore, this work has demonstrated XylA forms a tetrameric conformation, which creates a much bulkier fusion partner as well as potentially introducing conformational constraints to the fusion. It is possible that the size of the partner or any conformational constraints brought about as a result of the XylA tetramer may inhibit XylE activity or create issue for insertion of XylE into the membrane.

In order to determine whether xylose transport was deleteriously impacted by the direct fusion, a xylose deficiency assay was applied. The BW25113 strain TDXYlose was used because the strain demonstrates a phenotype of reduced growth on M9+10 mM xylose compared to wild type BW25113, which can be recovered by transforming with a plasmid expressing XylE, as demonstrated by the XylE-GFP fusions.

TDXYlose was transformed with each of the five XylE-XylA fusions. In addition, three controls were prepared: BW25113 expressing GFP as a positive control of wild type growth conditions, TDXYlose expressing GFP as a negative control of TDXYlose growth and BW25113 expressing XylE to determine whether growth can be enhanced by excess XylE expression. Arabinose induction was not used in this experiment as initial attempts to overexpress the fused proteins resulted in a reduced cell growth compared to the uninduced sample.

The expression of all the tested XylE-XylA fusions by TDXYlose showed that the fusions were able to facilitate partial recovery of growth on M9+10 mM Xylose when compared to the negative control (Fig. 4.13). All fusions ended lag phase within 14 hours, compared to 35 hours seen in the TDXYlose strain expressing GFP, and reached a final OD₆₀₀ of 0.35 within 22 hours. This OD₆₀₀ was the same as the final OD of the BW25113 strain expressing GFP. In addition, this lag phase was shorter than the TDXYlose cells expressing unfused XylE, 17 hours.

The lag phase of all these fusions is higher than the BW25113 cell line expressing GFP (8 hours), suggesting that in absence of induction, levels are too low to fully recover cell function. However, all fusions, irrespective of the linker used, show a reduced lag phase of about 12-14 hours compared to TDXYlose expressing native XylE (~18 hours).

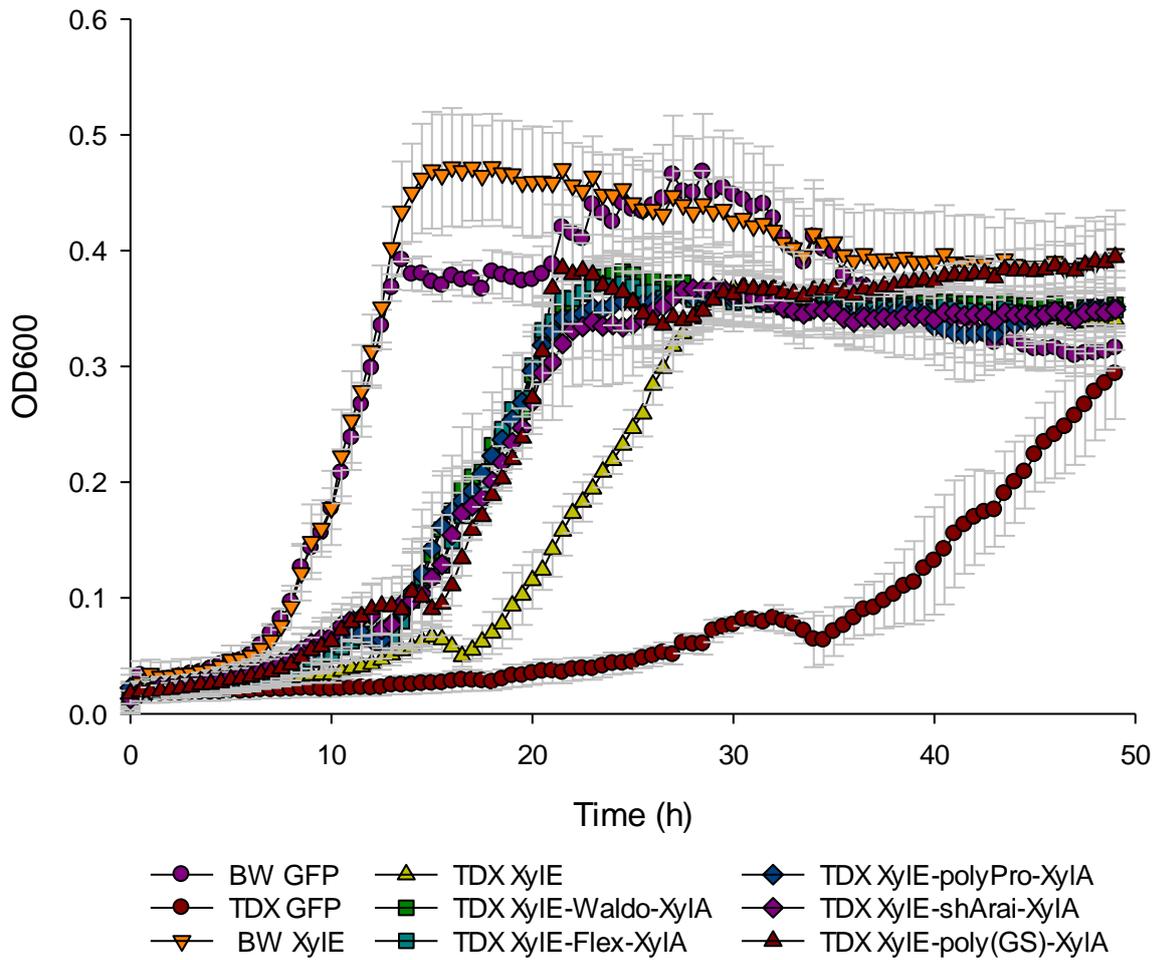


Figure 4.13. Recovery of TDXylose growth on a low concentration xylose growth medium by expressing XyIE-XyIA fusions. Growth assay of TDXylose on an M9+10mM Xylose background without arabinose induction. Fusions of XyIE-XyIA were expressed from a pLC series vector and tested against XyIE and GFP expressing controls. Two additional expression trial in BW25113 was completed as a positive control for growth on xylose medium. N = 5. Error is standard deviation.

Growth rate for TDXYlose expressing any of the XylE-XylA fusion was equal, or even greater, BW25113 expressing GFP (Fig. 4.14). The slowest growth rate of the fusions was found to be using the poly(Glycine-Serine) linker at 0.351 .h^{-1} , with the highest for the flexible linker at 0.425 .h^{-1} . Comparatively the BW25113 strain expressing GFP had a growth rate of 0.392 .h^{-1} . The difference recorded between these two samples was within error, which was confirmed by measuring the p value of the upper and lower rates to the positive control, gaining a value of 0.9405 and 0.1140 respectively, meaning $p > 0.05$.

As TDXYlose lacks direct xylose transporters, it was interesting to determine whether the recovered growth rate seen using the XylE-XylA fusion was greater than the non-specific transport of xylose using other transporters. The growth rate of TDXYlose expressing GFP was found to be 0.244 .h^{-1} , much lower than seen for any XylE-XylA the with p values below 0.05 for the waldo, flexible, polyproline and poly(GS) linkers. Only the short Arai linker showed a value greater than 0.05.

Overall, this work demonstrates that the XylE-XylA fusions facilitate TDXYlose growth comparable to that of the positive BW25113 control. Furthermore, cells grew at a greater rate compared to the negative control expressing GFP, demonstrating that the XylE is active within the fusion. This conclusion is further supported by the significantly reduced lag phase compared to the negative control. TDXYlose cells demonstrate a lag phase almost equivalent to, but still less than, that of BW25113. As BW25113 contains not only XylE, but also the XylFGH and AraC transporters, which are transporters associated with xylose transport, it would be expected to outperform cells expressing XylE alone, as seen when compared to the TDXYlose cells expressing native XylE. Comparatively, the fusions outperformed the unfused native XylE when expressed in TDXYlose, which may suggest that the fusion to XylA is either stabilizing XylE or the higher levels of XylA were increasing catabolism of D-xylose, thereby increasing cell growth.

Going forward it is important to consider the impact of fusion to XylE on the activity of the XylA protein.

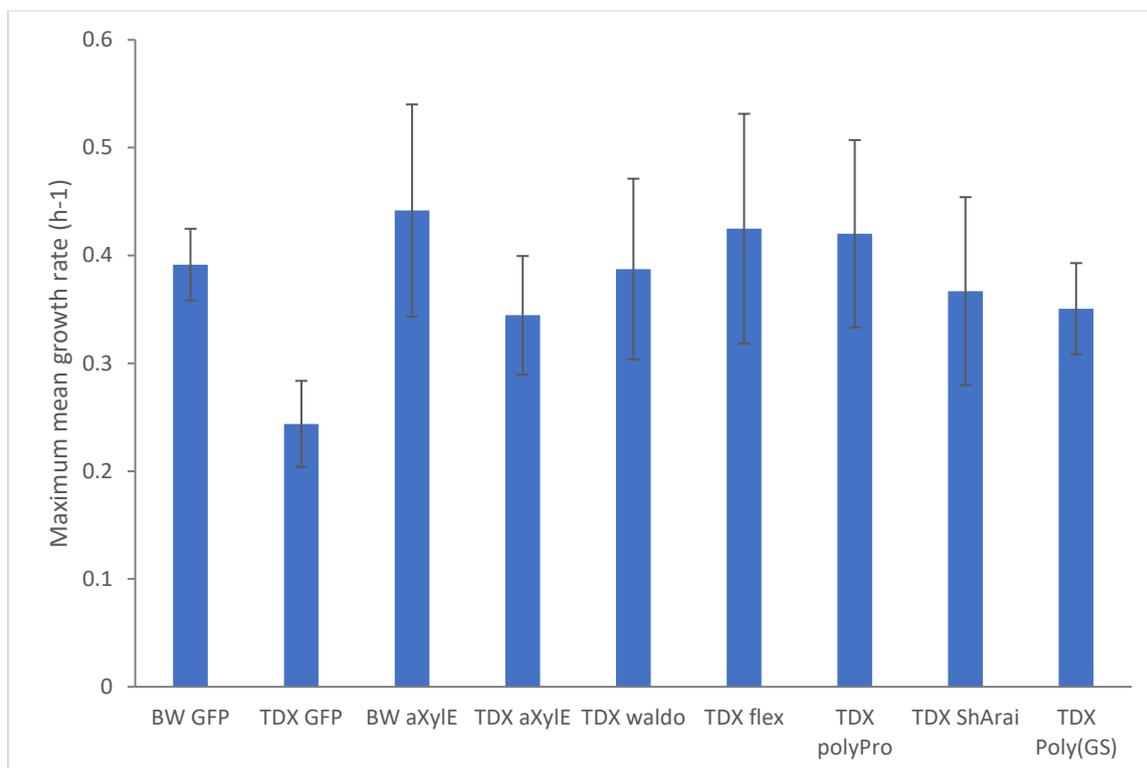


Figure 4.14. Maximum growth rate of TDXylose growth on a low concentration xylose growth medium by expressing Xyle-Xyla fusions. Maximum growth rate of assay in Figure 4.13. Growth rate is calculated as the rate of the slope over 5 time points multiplied by the absolute value of the Pearson correlation over the same five time points, calculated throughout the reaction for each condition. The maximum is this obtained for each repeat and the mean maximum growth rate calculated. N=5. Error is standard deviation.

4.4. Xylose isomerase activity is negatively impacted by its fusion to XylE.

As the activity of the xylose transporter XylE did not appear to be negatively impacted by the fusion to XylA, it was important to determine whether the fused xylose isomerase was also able to retain function equivalent to native XylA. A BW25113 strain containing a XylA deletion, $\Delta xyIA::Kan$, was obtained from the KO collection (Baba *et al.*, 2006). This deletion mutant lacks the ability to grow on xylose as it is unable to convert xylose to xylulose. In order to recover growth of this line on xylose, $\Delta xyIA::Kan$ must be transformed with a plasmid expressing *xyIA*.

To determine the activity of xylose isomerase within the XylE-XylA fusions, $\Delta xyIA::Kan$ was transformed with one of each. BW25113 expressing an empty plasmid was used as a positive control for growth. Comparatively, $\Delta xyIA::Kan$ expressing an empty plasmid was used as a negative control to determine whether any growth took place in absence of XylA. Finally, a second positive control, $\Delta xyIA::Kan$ expressing XylA on a pLC3 plasmid was also produced, to determine the maximum recovery of growth on xylose possible by expressing native XylA (Fig. 4.15).

In contrast to TDXYlose, which demonstrated an extended lag phase but the ability to grow on xylose over time, the $\Delta xyIA::Kan$ expressing empty pLC3 failed to grow, with a final optical density at 600 nm (OD_{600}) of 0.04, reinforcing the essential need of *xyIA* for growth on D-xylose. Both positive controls on the other hand demonstrated growth on xylose. BW25113 expressing empty pLC3 and $\Delta xyIA::Kan$ expressing native XylA showed a similar lag phase of around 8 hours, however the positive control of $\Delta xyIA::Kan$ expressing XylA has a reduced final OD_{600} of 0.45 compared to the native BW25113 strain expressing empty plasmid pLC3, 0.58 (Fig. 4.15). This means that native XylA has the ability to recover growth of $\Delta xyIA::Kan$, however not to the level of the native BW25113 cell, with a maximum growth rate of 0.277 h^{-1} and 0.417 h^{-1} respectively (Fig. 4.16). This likely reflects the reduced concentration of XylA expressed due to both the low copy number plasmid and the lack of induction. This demonstrates that there is the potential to improve xylose utilization within the deletion mutant through substrate channelling.

In spite of the potential to improve xylose utilization and therefore cell growth, the $\Delta xyIA::Kan$ cells expressing different XylE-XylA fusions showed an extended lag phase of between 14-30 hours and significant reduction in the final OD_{600} when compared to the control conditions (Fig. 4.15). Of the fusions tested, the fusion with the best performance was the XylE-XylA fusion containing the polyproline linker, with a lag phase of 14 hours and a final OD_{600} of

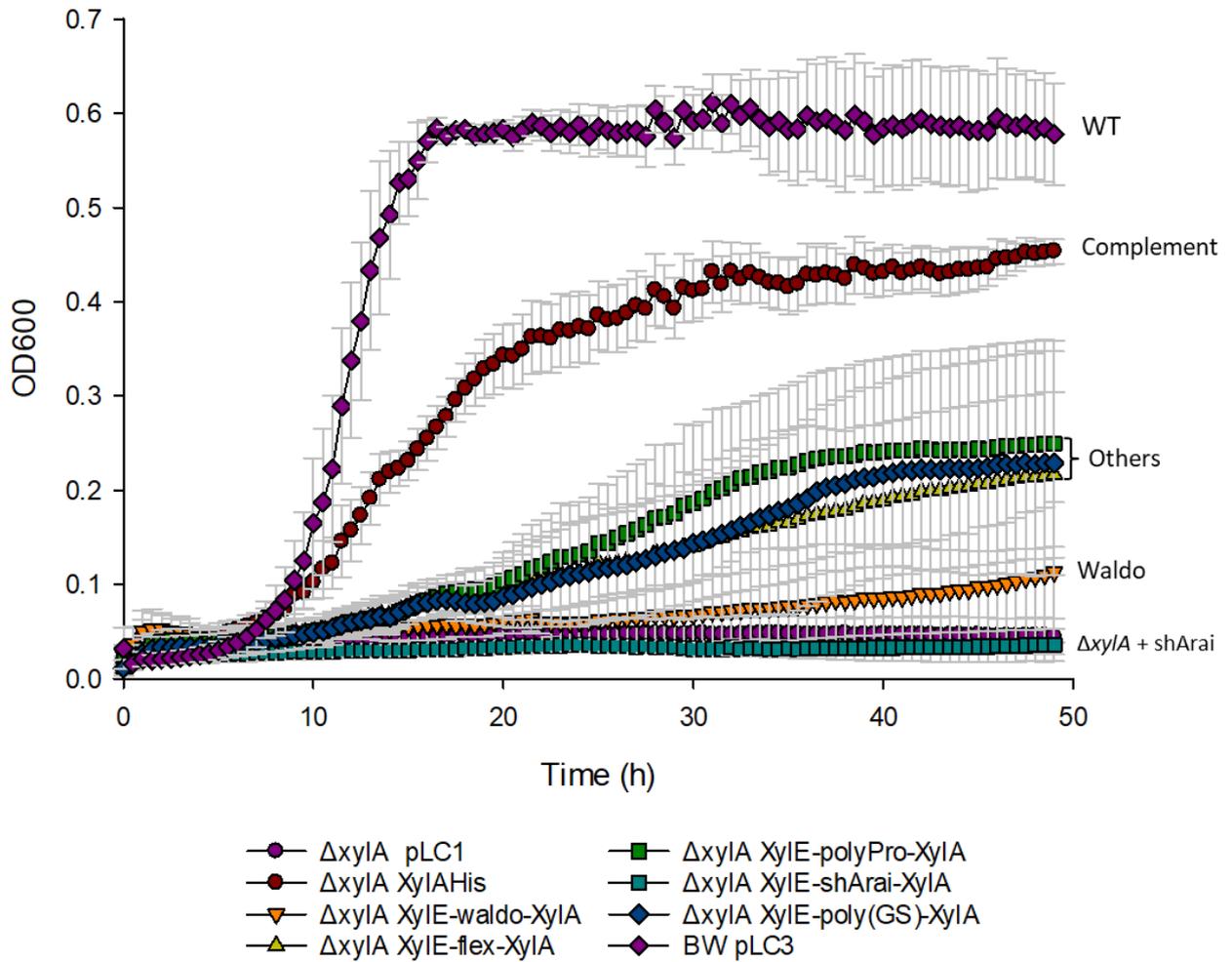


Figure 4.15. Recovery of BW25113 $\Delta xyIA$ growth on a low concentration xylose growth medium by expressing XylE-XylA fusions. Growth assay of BW25113 $\Delta xyIA$ on an M9+10mM Xylose background. Fusions of XylE-XylA were expressed and tested against XylA and GFP expressing controls. An additional expression trial in BW25113 was completed as a positive control for growth on xylose medium. N = 5. Error is standard deviation.

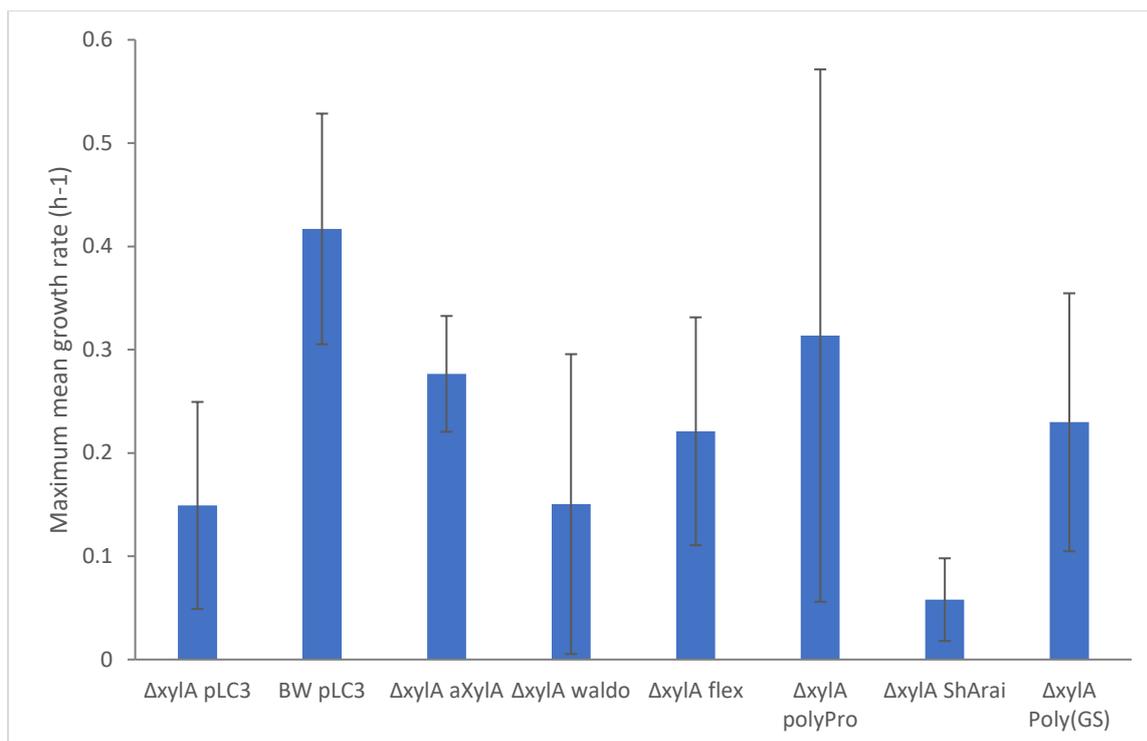


Figure 4.16. Maximum growth rate of BW25113 strain $\Delta xylA::Kan$ growth on a low concentration xylose growth medium expressing XylE-XylA fusions. Maximum growth rate of assay in Figure 4.15. Growth rate is calculated as the rate of the slope over 5 time points multiplied by the absolute value of the Pearson correlation over the same five time points, calculated throughout the reaction for each condition. The maximum is this obtained for each repeat and the mean maximum growth rate calculated. N=5. Error is standard deviation.

0.25. This fusion also demonstrated a higher maximum growth rate than the native XylA, at 0.313 h^{-1} , however the result had high error.

Two additional linkers performed at a similar level: the poly(Glycine-Serine) and Flexible linkers. Both linkers performed similarly, ending lag phase at 14 hours alongside the polyproline linker-containing fusion, however with slower growth. The growth curve of both these fusions was nearly identical for the next 20 hours, before a slight variation in growth rate could be observed - however both curves ultimately plateaued towards a final OD_{600} of approximately 0.23. The final OD was within error of the polyproline linker (Fig. 4.15), therefore both linkers could be argued to be equivalent of the polyproline containing fusion. Similarly, they demonstrated a growth rate of 0.221 h^{-1} and 0.230 h^{-1} for the flexible and poly(Glycine-Serine) linkers respectively, which was within error of the polyproline linker and were not statistically different to the native XylA control.

Cells expressing the fusion containing the waldo linker performed better than the negative control however had an extended, nearly 30 hour lag phase, with a final OD_{600} of 0.12. The worst performing fusion was that containing the short Arai linker, with which the $\Delta xyIA::\text{Kan}$ cells expressing the fusion failed to grow, with a final OD_{600} of 0.03. Both samples had low maximum growth rates that were not statistically significant compared to the negative control, 0.149 h^{-1} . For the Waldo linker containing fusion, the growth rate of demonstrated the high error associated with the condition, with some repeats showing no growth and others high, resulting in a rate of 0.151 h^{-1} .

By expressing the XylE-XylA fusions within $\Delta xyIA::\text{Kan}$, alongside the relevant controls, this work demonstrated that the fusions showed a significant loss in activity. While not a direct measure of enzyme activity, this *in vivo* assay depends upon conversion of xylose to xylulose to facilitate cell growth, a procedure that can only occur using XylA. The assay demonstrated a large difference between growth of cells when XylA fused or unfused.

It is important to note that linkers appear to have a major role in the function of the XylE-XylA fusion, with different linkers changing how much xylose isomerisation can be achieved. The largest contrast in linker effectiveness is between the short Arai linker and the polyproline linker. The short Arai linker did not facilitate growth at any point within 48 hours, whereas the polyproline linker reached an OD_{600} of 0.25 under identical conditions. This therefore means that depending on the linker chosen, the fusion may be unable to convert xylose to xylulose.

The property of the linkers could explain the variation. While both the polyproline linker and the short Arai linker are defined as “rigid”, rigidity was produced through different mechanisms, the presence of repeated proline groups and a salt bridge respectively. While variable in length

between one another, all other linkers were flexible. This highlights the importance of varying linkers when producing novel fusions to find the optimal condition.

It is also important to consider the western blots of the XylE-XylA fusions produced (Fig. 4.9-4.12). Each western blot demonstrated that there was variability between expression of the different proteins – this could therefore explain the low performance of the fusion containing the waldo linker, as it showed low levels of expression in all western blots compared to the flexible linker. Comparatively, the polyproline linker containing fusion showed notable variability between western blots, however consistently underwent noticeable cleavage of the linker, as seen by the presence of a shorter band representing unfused XylA. The high activity of the $\Delta xylA::Kan$ expressing the XylE-polyproline-XylA fusion could be representative of the presence of cleaved XylA. Conversely, the activity of the short Arai linker, which showed similar levels of expression to both the poly(Glycine-serine) and polyproline linkers, performed equivalent to the negative control, which likely suggests that expression levels is not a direct, or consistent, explanation to low growth rates.

As all products demonstrated some cleavage under uninduced conditions in Fig. 4.11, it is likely that all compounds were capable of producing free XylA. While there was some inconsistency between experiments, as discussed in section 4.2, the most stable linkers appeared to be the waldo linker and the poly(Glycine-Serine) – it is possible the poor performance of the XylE-Waldo-XylA fusion reflected the strength of the fusion, and therefore the reduced concentration of free XylA present in the cells compared to fusions with other, more readily cleaved linkers. This does not explain the situation with the XylE-shArai-XylA fusion, however, which may suggest that a combination of high cleavage and a favourable linker for uncleaved fusions is necessary for successful recovery of growth.

Finally, the concentration of the enzymes may also explain this effect. Western blots showed high relative concentration of XylA expressed in the positive compliment control compared to the samples expressing one of the fused proteins, which had lower levels of both the XylE-XylA fusions and free XylA. The reduced concentration of both of these compounds could play a key factor in the reduced levels of recovery of $\Delta xylA::Kan$ expressing XylE-XylA fusions.

4.5. Summary of Chapter

The direct fusion of xylose isomerase, XylA, to the C-terminus of the xylose transporter, XylE, was hypothesized to have a negative impact on one of the two components. This was due to the size of both components and therefore the potential for steric hinderance. The XylA, if tetrameric, could prevent fused XylE from entering the membrane, ensuring misfolded XylE aggregated within the cell, which could introduce sterical constraints for the proteins and reduce XylE concentration in the plasma membrane (Fig. 4.17.A). Alternatively, the insertion of XylE into the membrane may force XylA to take a dimeric conformation and reduce overall activity (Fig. 4.17.B). If linker length and flexibility were favourable, however, it is possible an orientation could be reached for production of a XylE-XylA fusion with all components undisturbed (Fig. 4.17.C).

For XylE a propensity of MFS proteins to produce C-terminus fusions in nature, plus early tests with GFP, suggested that the transporter would be stable within a fusion. Conversely, little was known about the nature of XylA.

Within this chapter the oligomeric state of *E. coli* XylA was determined to be tetrameric through a combination of SEC-MALLS analysis and the crystal structure, which conflicted with initial assumptions of a dimeric structure based off previous literature. The *E. coli* xylose isomerase contained a TIM-barrel like structure with the N-terminus located near the active site, whereas the C-terminus was located towards the oligomerization region.

Fusion of the C-terminus of XylE to the N-terminus of XylA did not negatively impact the activity of XylE, supporting earlier data with GFP fusions. A xylose deficiency growth assay demonstrated similar, if not greater, activity of the XylE-XylA fusions compared to native XylE when expressed by the TDXylose strain.

Conversely, a loss of xylose isomerase activity could be seen for the XylE-XylA fusion - observed as a reduced growth on xylose and prolonged lag phase when compared to native XylA. The loss of activity in fusion could be alleviated somewhat by using different linkers, with the fusion containing a polyproline linker performing most alike the native XylA.

This work suggests that fusion of a large protein to the N-terminus of XylA results in a loss of enzyme activity. It is possible that XylE blocks the xylose isomerase active site, thereby preventing access of the sugar to the enzyme. Alternatively, steric hinderance may prevent oligomerization of the XylA tetramer. Regardless of the explanation, it appears that direct fusion will not facilitate an improved substrate channelling within the xylose utilization pathway. This therefore necessitates an alternative method of substrate channelling must be applied.

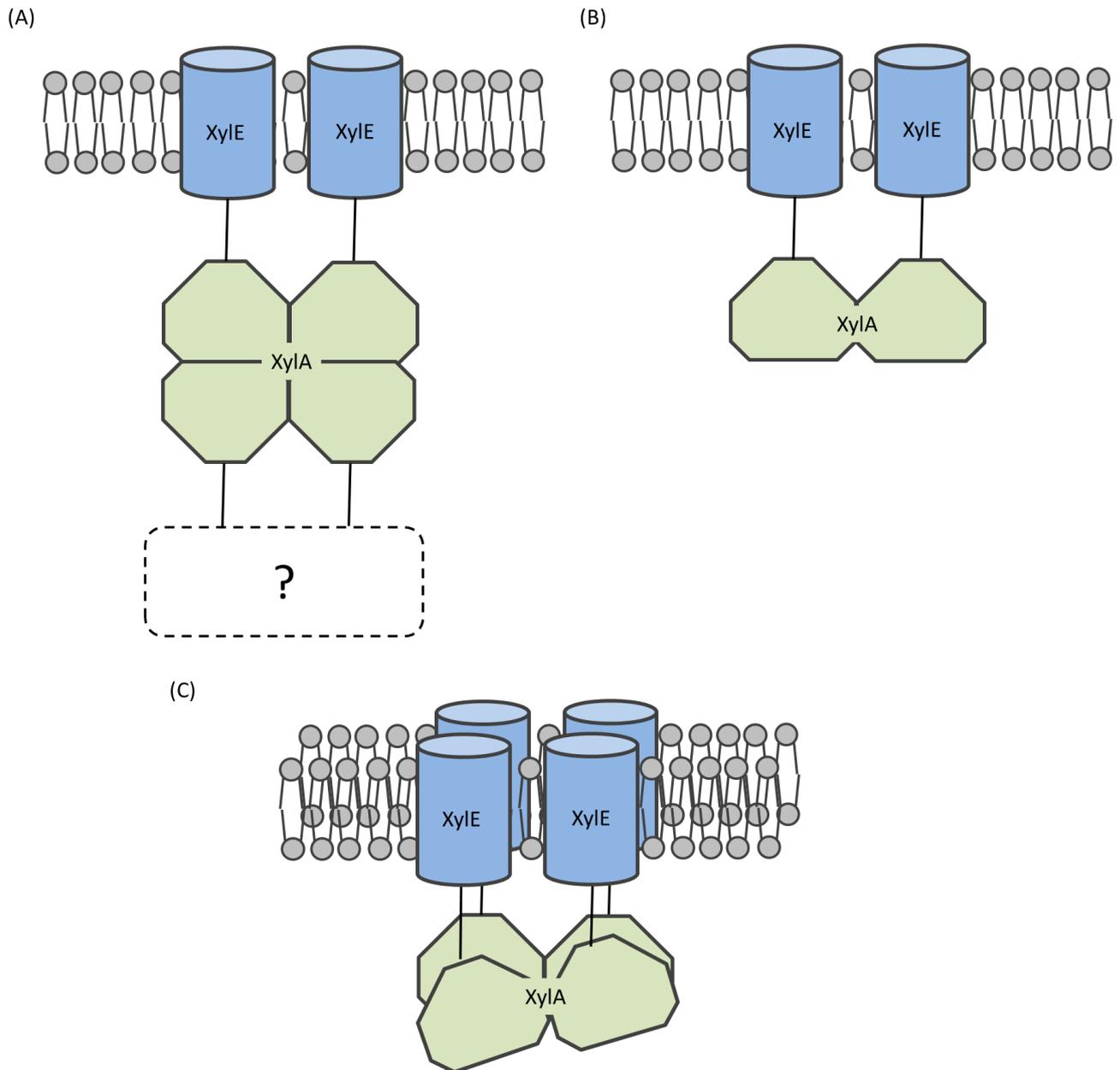


Figure 4.17. A cartoon representation of potential XylE-XylA fusion conformations (A) Fusion of XylE to XylA produces a large aggregate as attached XylE is unable to enter the membrane due to steric hinderance from the fused XylA or linkers are cleaved to facilitate unfused XylE to enter the membrane. **(B)** The steric hinderance from XylE prevents complete oligomerisation of XylA, the fusion assembling with less functional XylA dimers as opposed to XylA tetramers. **(C)** Linker length and flexibility facilitate a favourable orientation in which both all XylE can be inserted into the membrane and the XylA tetramer can be formed.

Chapter 5.

Assessment of the use of Scaffold components to overcome the enzyme activity issues associated with direct fusion.

5. Assessment of the use of Scaffold components to overcome the enzyme activity issues associated with direct fusion.

5.1. Designing a scaffold system for localization of xylose isomerase at the XylE transporter.

Increasing the metabolism of D-xylose is a major focus when lignocellulosic feedstock is fermented to produce biofuels and other biomaterials (Aristidou and Penttilä, 2000; Feng *et al.*, 2018; Cunha *et al.*, 2019). To increase metabolism of xylose in the model organism *E. coli*, this work aims to enhance substrate channelling through the initial enzymes of the xylose utilization process. We hypothesized that xylose isomerase XylA would need to be localized to the xylose transporter XylE to achieve enhanced substrate channelling. Direct fusion of XylE to XylA was the first method attempted to achieve such localization, however it showed a series of shortcomings - when compared to native XylA, there was a notable reduction in the expression levels of the XylE-XylA direct fusion, which was further compounded by growth assays showing that the activity of the enzymes in the complex was also reduced. While there was little impact on the activity of XylE, conversely, the fused XylA demonstrated that the fusion had negatively affected activity.

To enhance substrate channelling through the xylose utilization system it was therefore important to apply an alternative method: protein scaffolding. Protein scaffolding uses small peptide tags which form non-covalent interactions with one another to achieve strong binding (Dueber *et al.*, 2009; Siu *et al.*, 2015). The use of scaffolds necessitates the co-expression of at least two gene fusions, each containing the protein of interest fused to one of two protein tags (Tippmann *et al.*, 2016). This allows the components to be expressed individually, which can be useful for when direct fusion hampers expression (Fierobe *et al.*, 2001). The two tags then interact with one another to bring two bound target proteins together.

In order to express the components of the scaffold system simultaneously, a second series of level 1 Golden Gate plasmids were required, with a different antibiotic resistance marker to the pLC series chloramphenicol selection marker. This would allow the second plasmid family, from here termed the pLD series, to express a protein cooperatively alongside a pLC series plasmid. The pLD series were produced using the same protocol as pLC series, however some consideration was needed to ensure that the components did not interfere. In order to achieve co-expression a different low copy number origin of replication was used, derived from pBBR1-MCS2 (Kovach *et al.*, 1995) alongside a different antibiotic selection cassette, selecting the kanamycin-resistance coding region from pET28 (Table 5.1, Fig. 5.1).

Name	Part 1	Part 2	Part 3
pLD1	pBBR1MCS2 - low copy number origin of replication	pET28 – Kan ^R	pLC1_AF – Golden Gate cloning region
pLD3	pBBR1MCS2 - low copy number origin of replication	pET28 – Kan ^R	pLC3_AG – Golden Gate cloning region

Table 5.1. Components of the pLD series Golden Gate expression plasmids. The Golden Gate expression vectors, produced by QuickFusion of three amplicon components, are listed alongside which parts were used. A pLD2 plasmid was not produced, as there was no pressing need to develop a derivative of the temperature sensitive pLC2.

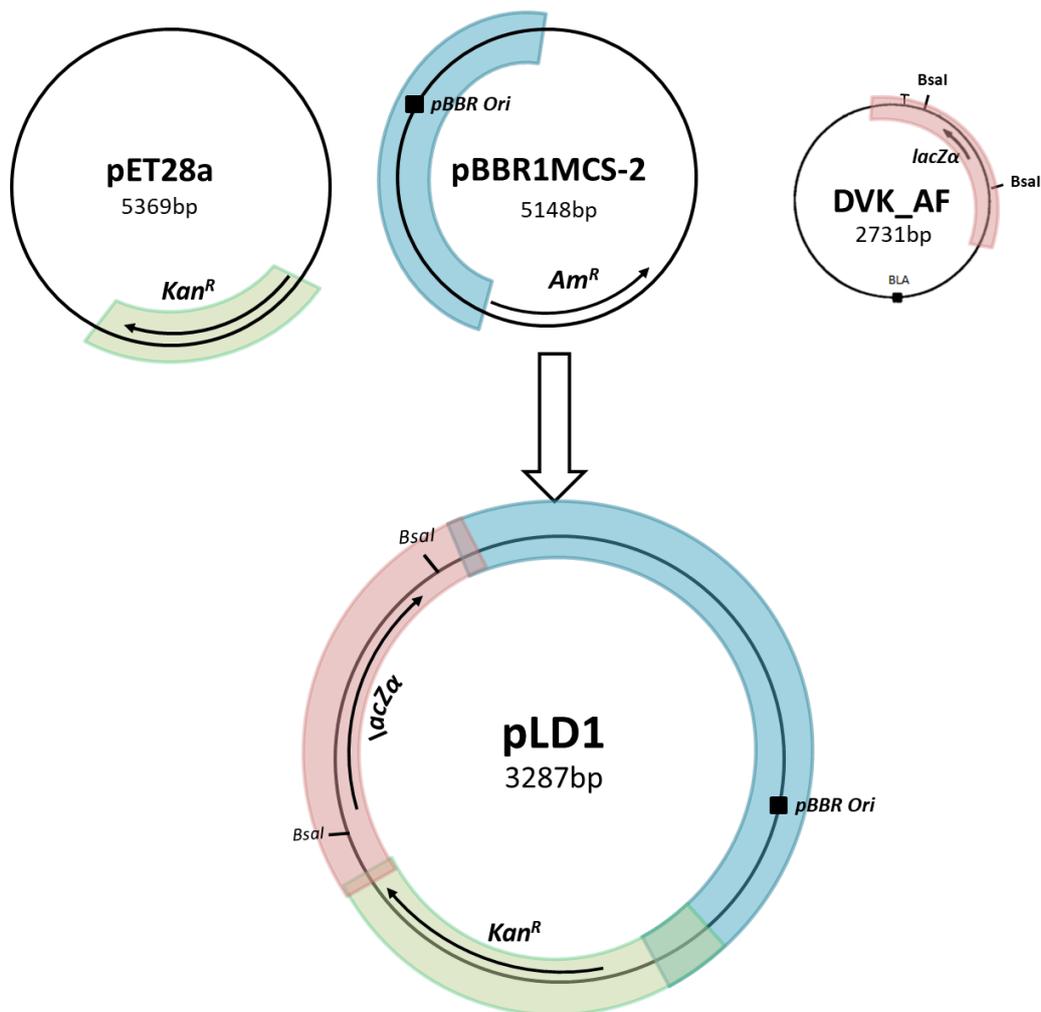


Figure 5.1. The components of pLD1, a low copy number Golden Gate expression vector compatible with pLC1. The plasmid pLD1 was produced by QuickFusion cloning of three base components, a low copy number origin of replication (15 to 20) was derived from pBBR1MCS-2, which made the bulk of the new plasmid backbone (1458 bp). The Kanamycin resistance genes were derived from pET28a (1011 bp). Finally, the Golden Gate cloning region of CIDAR MoClo plasmid DVK_AF (841 bp) was used achieve subsequent insertion of Golden Gate compatible DNA fragments into the pLD1 plasmid. A more detailed plasmid map is found in Appendix A.

Following the production of the pLD plasmid, it was used in combination with pLC to fuse scaffold components to the two target proteins and express them simultaneously in the same cell. Scaffold components considered for this work included the coiled coils AN4 and BN4 (Thomas *et al.*, 2013), complementary Cohesin-Dockerin assemblies (Lin, Zhu and Wheeldon, 2017) and the Spytag/Spycatcher system (Pröschel *et al.*, 2015; Zhang, Quin and Schmidt-Dannert, 2018). The Spytag/Spycatcher system was ultimately discarded as it would form covalent bonds post-expression and would therefore was not a true enzyme scaffold, but more of a delayed fusion (Schoene *et al.*, 2014).

Ultimately, the cohesin-dockerin system was selected as the type 1 cohesion-dockerin system of *Clostridium thermocellum* demonstrated a low dissociation constant, $K_D < 1 \times 10^{-9}$ M (Jindou, Soda, *et al.*, 2004; Stahl *et al.*, 2012) suggesting a strong non-covalent interaction that would mimic the localization properties of direct fusion. While the AN4-BN4 coiled coil also demonstrated a $K_D < 1 \times 10^{-10}$ M (Thomas *et al.*, 2013), both of these values simply reflected instrument limit, however the cohesin-dockerin system has a history of its use in synthetic scaffolding to increase metabolic channelling (Tsai *et al.*, 2009; You and Zhang, 2014; Siu *et al.*, 2015). If time had been permitting, the coiled-coils were considered to test whether different scaffolds alter substrate channelling with respect to one another.

To achieve colocalization at the membrane between soluble XylA and membrane transporter XylE, Golden Gate assembly was therefore applied to fuse the *Clostridium thermocellum* dockerin, hereafter referred to as ctDoc, to the C-terminus of XylE. This assembly was completed within the plasmid pLC3. A constitutive promoter, J23100, was used in this assembly to ensure continuous expression of the XylE-ctDoc fusion. Five successful fusions were produced, each with a different linker (Table 5.2).

For expression of the XylA, it was decided that Golden Gate assembly would construct a plasmid containing an inducible promoter, thereby allowing for direct control over expression levels. The pLD3 plasmid was used as the backbone for assembly and the cohesin from *Clostridium thermocellum*, hereafter referred to as ctCoh, was fused to both the C- and N-terminal of XylA. Ultimately eleven fusions between ctCoh and XylA were produced using pLD3. The fusion ctCoh-InArai-XylA, containing a functional long Arai linker, was only successfully produced in N-terminal fusion to XylA, as this linker was truncated when attempting fusion in the opposite orientation (Table 5.3).

When considering the which terminal to fuse the tag, the use of the scaffold offered more flexibility compared to direct fusion with a membrane protein. The crystal structure of XylA (Fig. 4.3.) showed that both the C- and N-terminal of XylA exist as protrusions from the central TIM-barrel like structure. The N-terminal, however, is much closer to the active site and

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein	Tag
pLC3_111191	J23100 (constitutive)	B0034m	XylE	Waldo	ctDoc (<i>Clostridium thermocellum</i> dockerin)	6xHis
pLC3_111291	J23100 (constitutive)	B0034m	XylE	Single-chain flexible	ctDoc (<i>Clostridium thermocellum</i> dockerin)	6xHis
pLC3_111391	J23100 (constitutive)	B0034m	XylE	Polyproline	ctDoc (<i>Clostridium thermocellum</i> dockerin)	6xHis
pLC3_111591	J23100 (constitutive)	B0034m	XylE	Short Arai	ctDoc (<i>Clostridium thermocellum</i> dockerin)	6xHis
pLC3_111691	J23100 (constitutive)	B0034m	XylE	Poly(glycine-serine)	ctDoc (<i>Clostridium thermocellum</i> dockerin)	6xHis

Table 5.2. The components of XylE-ctDoc fusion constructs. A table displaying the components making up the five successful XylE-ctDoc constructs successfully produced by Golden Gate assembly. The plasmid pLC3 can contain up to 6 components

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein	Tag
pLD3_317161	I13453 (ara inducible)	B0034m	XylA	Waldo	ctCoh (<i>Clostridium thermocellum</i> cohesin)	6xHis
pLD3_317261	I13453	B0034m	XylA	Single-chain flexible	ctCoh	6xHis
pLD3_317361	I13453	B0034m	XylA	Polyproline	ctCoh	6xHis
pLD3_317461	I13453	B0034m	XylA	Long Arai	ctCoh	6xHis
pLD3_317561	I13453	B0034m	XylA	Short Arai	ctCoh	6xHis
pLD3_317661	I13453	B0034m	XylA	Poly(glycine-serine)	ctCoh	6xHis
pLD3_31B1A1	I13453	B0034m	ctCoh	Waldo	XylA	6xHis
pLD3_31B2A1	I13453	B0034m	ctCoh	Single-chain flexible	XylA	6xHis
pLD3_31B3A1	I13453	B0034m	ctCoh	Polyproline	XylA	6xHis
pLD3_31B5A1	I13453	B0034m	ctCoh	Short Arai	XylA	6xHis
pLD3_31B6A1	I13453	B0034m	ctCoh	Poly(glycine-serine)	XylA	6xHis

Table 5.3. The components of XylA-ctCoc and ctCoh-XylA fusion constructs. A table displaying the components making up the eleven fusion constructs between XylA and ctCoh successfully produced by Golden Gate assembly. The plasmid pLD3 can contain up to 6 components

therefore fusion may have a negative impact on the activity of the isomerase. For direct fusion to XylE, we were hesitant to fuse XylA to the XylE N-terminus as we could find no examples of a natural N-terminal MFS fusion, which may suggest that such a fusion could impact transporter function or insertion of the protein into the membrane. On the other hand, fusion to either terminal of the cohesin domain is unlikely to impact activity of the tag, as ctCoh is found in nature within the Cellulosomal-scaffolding protein A, a series of cohesin peptides fused together to form a large complex (Carvalho *et al.*, 2003).

The expected design of the XylA-ctCoh fusions would therefore mean each XylA monomer was tagged with ctCoh (Fig. 5.2). Unlike the direct fusion to XylE, however, these proteins are both small and are not localized to the membrane directly, therefore increasing the likelihood of forming the correct oligomeric state. It is possible that all 4 components could interact with a XylE-ctDoc fusion, however an advantage of this scaffold system would be that if steric constraints would limit XylA assembly, the interaction would be unlikely to form, protecting the component structure.

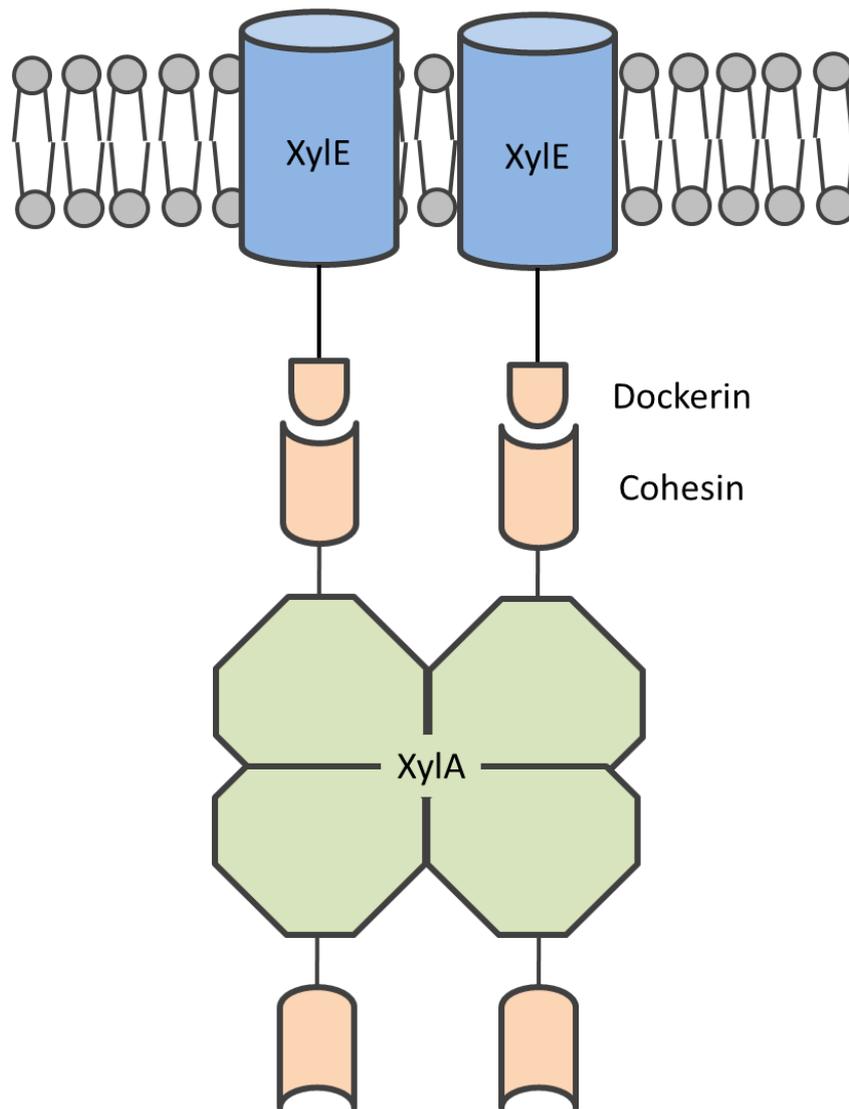


Figure 5.2. Cartoon representing the planned scaffold organisation. A cartoon representing the interaction between transporter XylE and xylose isomerase XylA through the association of scaffoldin peptides dockerin and cohesin, derived from *C. thermocellum*. Each XylA monomer is fused to one cohesin peptide, which can interact with the dockerin fused to XylE. Two cohesin-dockerin interactions are shown here - whether all cohesin peptides will bind a dockerin and if that is sterically possible is unknown. Through this system XylA can form its correct oligomeric state.

5.2. Curing the Kanamycin resistance cassette of Keio collection BW25113 $\Delta xyIA::kan$

Before the pLD series plasmid could be applied, a new strain would need to be produced. This was because a kanamycin resistance marker was selected for the pLD series plasmids; however, the Keio collection removes genes of interest by swapping the target gene out with a kanamycin resistance cassette. Therefore, $\Delta xyIA::kan$ contained a resistance marker which reduced the need for pLD plasmid uptake.

In order to reintroduce selection pressure, the kanamycin resistance cassette needed to be removed. This was achieved by action of the plasmid pCP20, which encodes a flippase enzyme that will remove the kanamycin cassette from the chromosome, excising it in a circular form which is unable to replicate and is subsequently lost by the cell (Merlin, McAteer and Masters, 2002). The $\Delta xyIA::kan$ cells were therefore transformed with pCP20 and grown at 30°C, the cells were then grown overnight at 43°C in a cell culture, thereby inactivating pCP20. The culture was then streaked on an LB agar plate and grown at 37°C to produce candidate $\Delta xyIA$ colonies.

In order to confirm that candidate colonies were kanamycin resistant and no longer contained active pCP20, 9 individual colonies were selected and were compared via patching. A streak of each colony was applied to one of three plates: LB agar (to confirm growth), LB agar + 50 µg/mL Kanamycin (facilitating growth only if the cassette was still present) and LB agar + 100 µg/mL ampicillin (facilitating growth only if pCP20 was still active). Two colonies were able to grow only on LB agar, showing no potential growth on LB + 50 µg/mL Kanamycin nor LB + 100 µg/mL ampicillin. These two candidate colonies, $\Delta xyIA1$ and $\Delta xyIA7$ were selected for further tested.

To validate that the new cell lines were indeed not kanamycin resistant and had retained the mutation, BW25113, $\Delta xyIA::kan$, $\Delta xyIA1$ and $\Delta xyIA7$ were streaked on three different plates (Fig. 5.3). The first plate was the positive control for cell growth, an M9 minimal media + 10 mM glucose plate, which could support growth of all four species (Fig. 5.3.A). The second plate was M9 + 10 mM xylose, which while facilitating growth of BW25113, was not able to support growth of any of the $\Delta xyIA$ strains, as no xylose isomerase was present. The M9 + 10 mM xylose plate therefore confirmed that $\Delta xyIA1$ and $\Delta xyIA7$ were derived from $\Delta xyIA::kan$ (Fig. 5.3.B).

The final plate was important for confirming cassette deletion, an M9 + 10 mM glucose + 50 µg/mL kanamycin plate. It was found that the $\Delta xyIA::kan$ strain produced a lawn of growth on

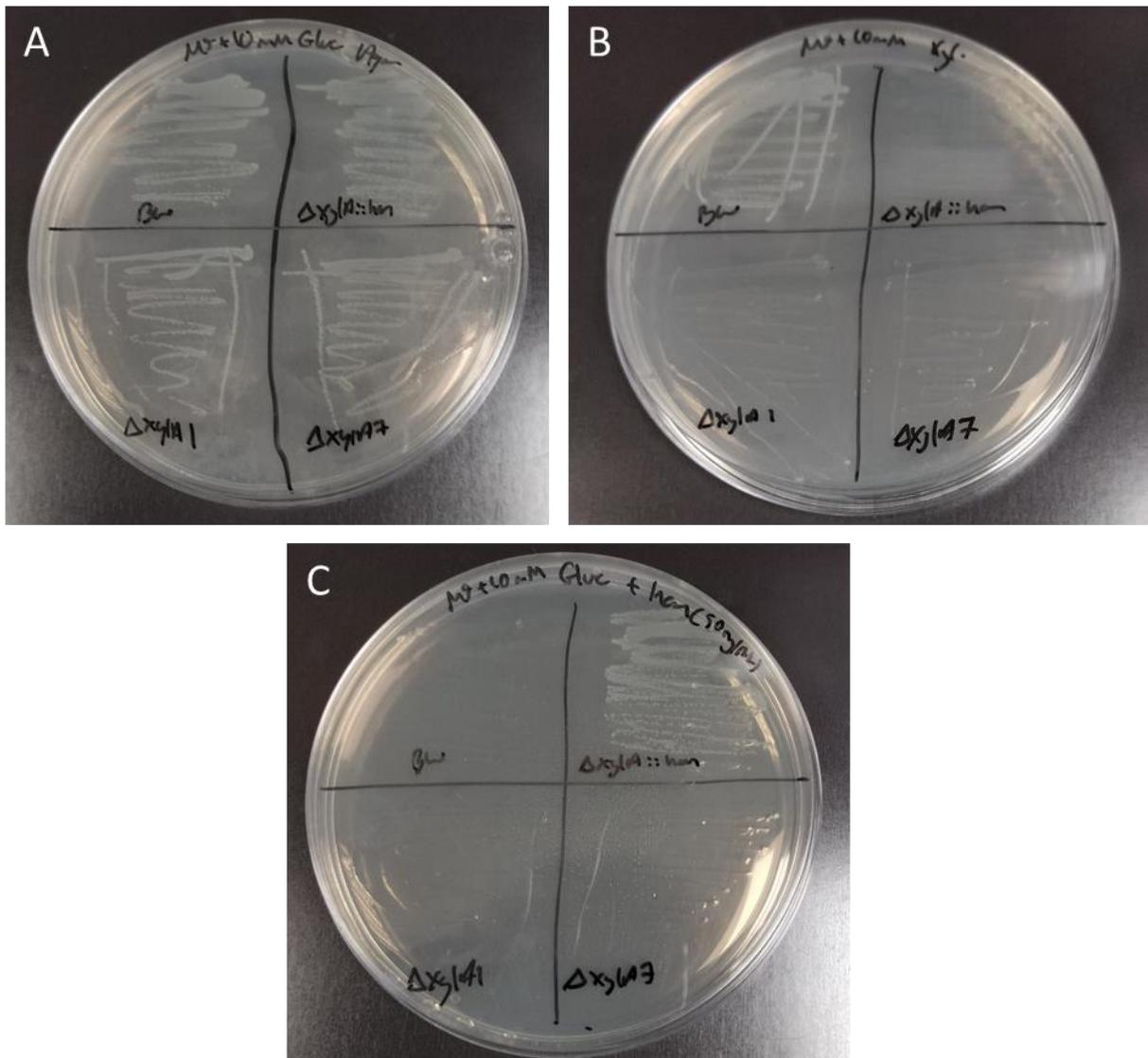


Figure 5.3. Spread plates of candidate *xyIA* deletion mutants $\Delta xyIA1$ and $\Delta xyIA7$. *Escherichia coli* BW25113 strain and its mutants Keio collection $\Delta xyIA::kan$ and cured $\Delta xyIA1$ and $\Delta xyIA7$ streaked on agar plates containing (A) M9 minimal media + 10 mM glucose (B) M9 minimal media + 10 mM xylose (C) M9 + 10 mM glucose + 50 μ g/mL Kanamycin.

the kanamycin-containing plate, as expected due to the kanamycin resistance cassette. Comparatively, the absence of lawn growth for neither $\Delta xyIA1$ nor $\Delta xyIA7$ suggested that the cassette had been removed (Fig. 5.3.C). However, it was important to verify cassette deletion through another method.

Colony PCR was therefore used to confirm the deletion of the kanamycin resistance cassette. The *xyIA* gene is 1323 bp in length. Comparatively, the kanamycin resistance cassette, which includes the FLP binding sites, is around 1256 bp. PCR primers were encoded ~250 bp upstream and downstream of the *xyIA* region of BW25113 in order to confirm removal of both the *xyIA* gene and the kanamycin resistance cassette in 6 colonies (identified through the letters a-f) of $\Delta xyIA1$ and $\Delta xyIA7$.

The BW25113 and $\Delta xyIA::kan$ controls were amplified by the PCR primers to produce an amplicon of around 1800 bp in length, representing the presence of *xyIA* and the kanamycin resistance cassette respectively. In comparison, each colony of both $\Delta xyIA1$ and $\Delta xyIA7$ showed an amplicon ~650 bp in length, suggesting the removal of the genes via the action of the pCP20 encoded flippase (Fig. 5.4).

To further validate an inability to grow in the presence of kanamycin, each colony, as well as the two controls, were measured in triplicate on a plate reader. First, each colony was used to induce 3x200 μ L wells of LB within a 96 well plate. The samples were grown at 37°C, 200 rpm for 24 hours to confirm colony growth (Fig. 5.5.A).

A second 96 well plate was then prepared, with wells containing 200 μ L LB + 30 μ g/mL kanamycin. Following preparation of this second plate, the first 96 well plate was removed from the plate reader and a 9 μ L aliquot of each well from was then transferred to the equivalent position in the LB + 30 μ g/mL kanamycin 96 well plate. The samples were then incubated at 37°C, 200 rpm for 72 hours to determine if any colonies were capable of consistent growth (Fig. 5.5.B).

As expected, the $\Delta xyIA::kan$ colonies were all able to grow on LB + 30 μ g/mL kanamycin, confirming that conditions had been suitable for growth. On the other hand, all candidate colonies showed reduced, if any, growth under the weak kanamycin selection. Under no condition did the candidate colonies demonstrate uniform growth, however for some candidates, such as $\Delta xyIA1$ -b, one of the three repeats that was able to grow. The growth would occur at any point after the 20 hour lag period, which may suggest that trace amounts of uncured $\Delta xyIA::kan$ was present and was selected for over multiple replications. On the other hand, some growth was seen even within the BW25113 negative control, however, and may suggest some small cross contamination or that the kanamycin concentration was too

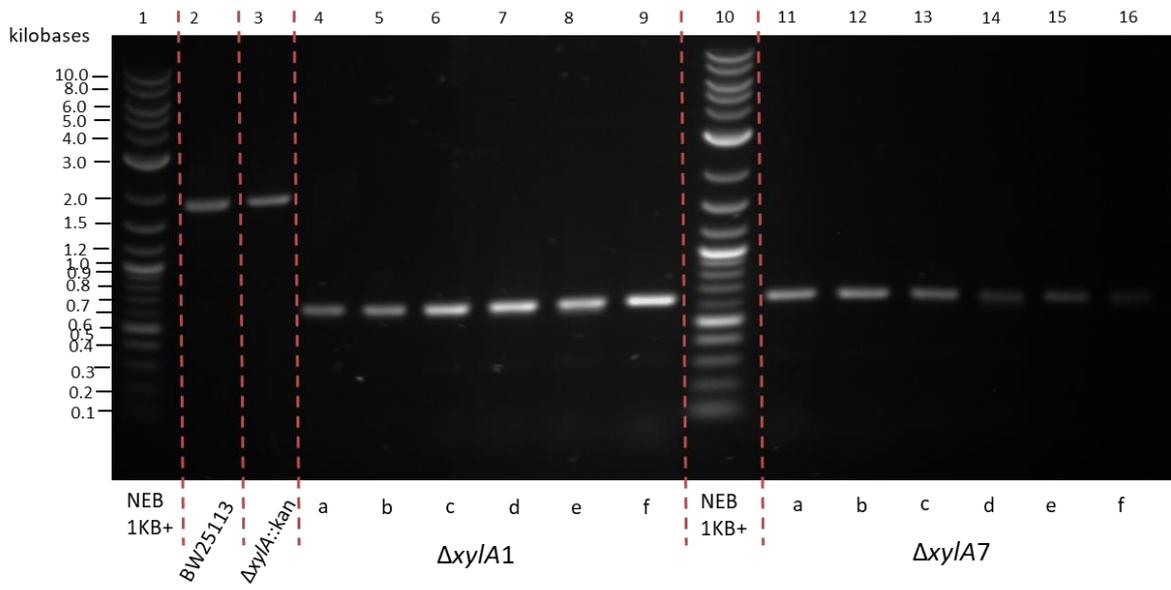


Figure 5.4. Electrophoresis gels of colony PCR of candidate cured Keio collection strains $\Delta xy/A1$ and $\Delta xy/A7$. Seven colonies of either candidate strains $\Delta xy/A1$ and $\Delta xy/A7$ were amplified by colony PCR and compared to the wild type BW25113 and uncured $\Delta xy/A::kan$ strains.

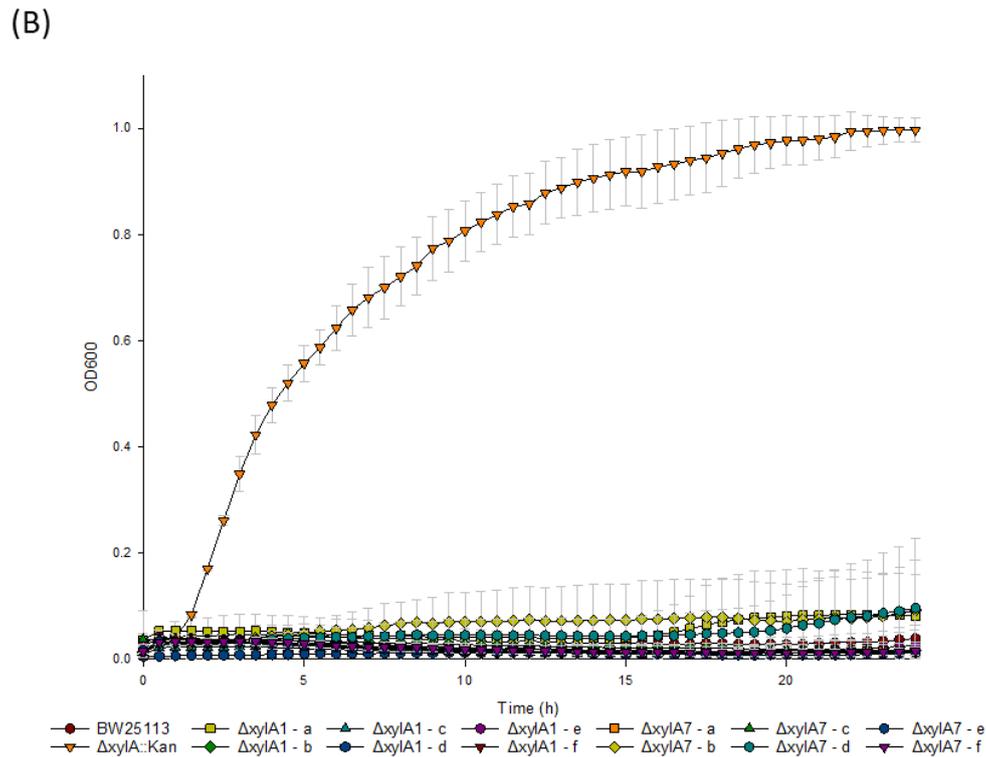
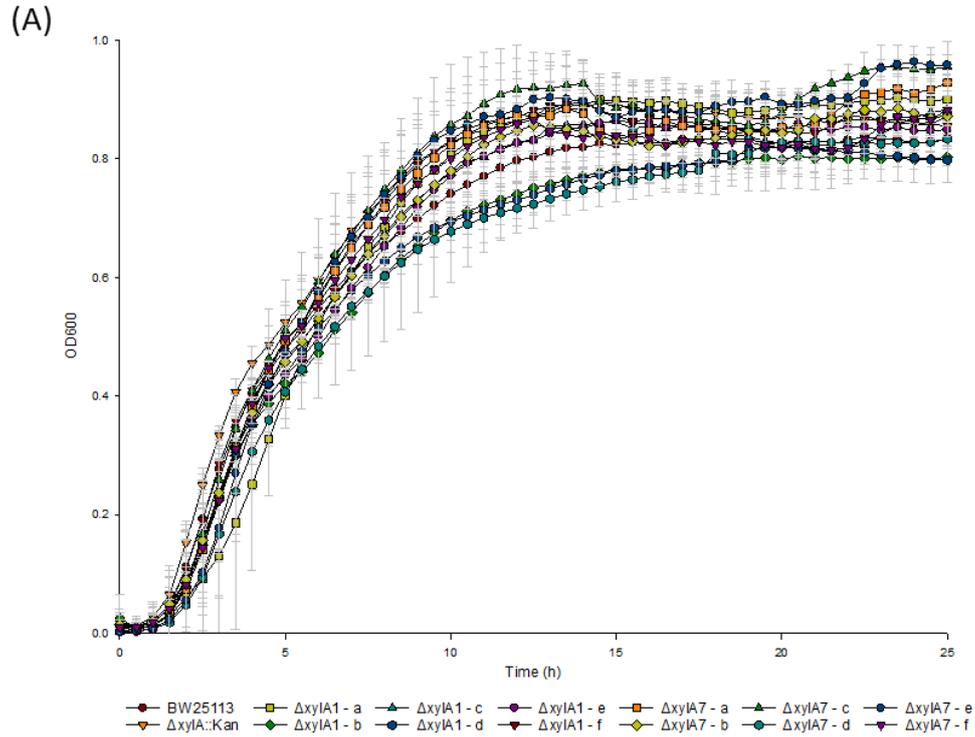


Figure 5.5. Growth of candidate $\Delta xyIA$ cured strains on LB growth media over 24 hours. (A) Growth assay of BW25113, $\Delta xyIA::kan$ and the candidate colonies of $\Delta xyIA1$ and $\Delta xyIA7$ on an LB solution with no antibiotic selection. (B) Growth assay of BW25113, $\Delta xyIA::kan$ and the candidate colonies of $\Delta xyIA1$ and $\Delta xyIA7$ on an LB solution with 30 $\mu\text{g}/\text{mL}$ kanamycin antibiotic selection. N = 3, error is standard deviation.

low and was selected to determine if any degree of kanamycin resistance existed within any of the strains.

Nevertheless, the colonies $\Delta xy/A1$ -d and $\Delta xy/A7$ -c were shown to feature no growth in the presence of kanamycin, even a lower concentration (Fig. 5.5.B). As both PCR and the growth profiles of these strains had demonstrated no kanamycin resistance cassette should be present whatsoever, both candidates could therefore be used for subsequent transformation with pLD series plasmids. From this point forwards, $\Delta xy/A1$ -d was used and is referred to as $\Delta xy/A$, reflecting the deletion of the kanamycin cassette.

5.3. Anti-tetraHis western blots demonstrate expression of fused XylE-ctDoc and XylA-ctCoh/ctCoh-XylA fusions.

To facilitate scaffold formation, the three gene fusions encoding XylE-ctDoc, XylA-ctCoh and ctCoh-XylA were produced through Golden Gate assembly (Table 5.2 and 5.3). Each component was then confirmed through DNA sequencing.

With the plasmids validated and a $\Delta xylA$ knock out successfully generated, the next stage was to confirm that the XylE-ctDoc and XylA-ctCoh fusions were expressed at suitable levels and to determine the stability of the fusion. This was determined, as with the XylE-GFP and XylE-XylA fusions, using western blots. The stability of the different fusions was of particular interest as the XylE-XylA fusions had demonstrated a reduced expression level compared to native XylA and a propensity towards linker cleavage.

This work therefore aimed to determine whether fusion to the scaffold proteins was a more stable alternative compared to direct fusion – which would further validate the use of scaffolds as an alternative to direct fusion.

5.3.1. The XylE-ctDoc shows comparable, if not greater, expression levels to XylE-XylA

The first fusions tested were that of XylE and ctDoc (Fig. 5.6). To prepare the western blot, TDXYlose cells expressing the fusions of interest were produced alongside two positive controls: a TDXYlose cell line expressing the XylE-waldo-XylA fusion and a XylA protein expressed by $\Delta xylA$. The use of mutants was to better represent the conditions of growth assays and to reduce the potential of aggregation with native transporters. The direct fusion was a positive control for the membrane fraction and the XylA for overall expression. Cells were grown overnight and then treated with BugBuster, mixing for 1 hour at room temperature then leaving overnight at 4°C. The supernatant, containing cytoplasmic protein was separated and the pellet, containing membranes, was resuspended in PBS. The products were then separated by SDS-PAGE and blotted onto a PVDF membrane.

Expression of the XylA control demonstrates the functionality of the anti-His antibody. The XylE-XylA fusion was expressed to compare to the five XylE-ctDoc fusions, and showed high cleavage producing a band at ~50 kDa (Fig. 5.6.A well 2 and 3), consistent with XylA. The minor band present, representing the complete fusion, was the XylE-XylA fusion at around 100 kDa, with an expected molecular weight of 106.95 kDa (Fig. 5.6A well 4 and 5).

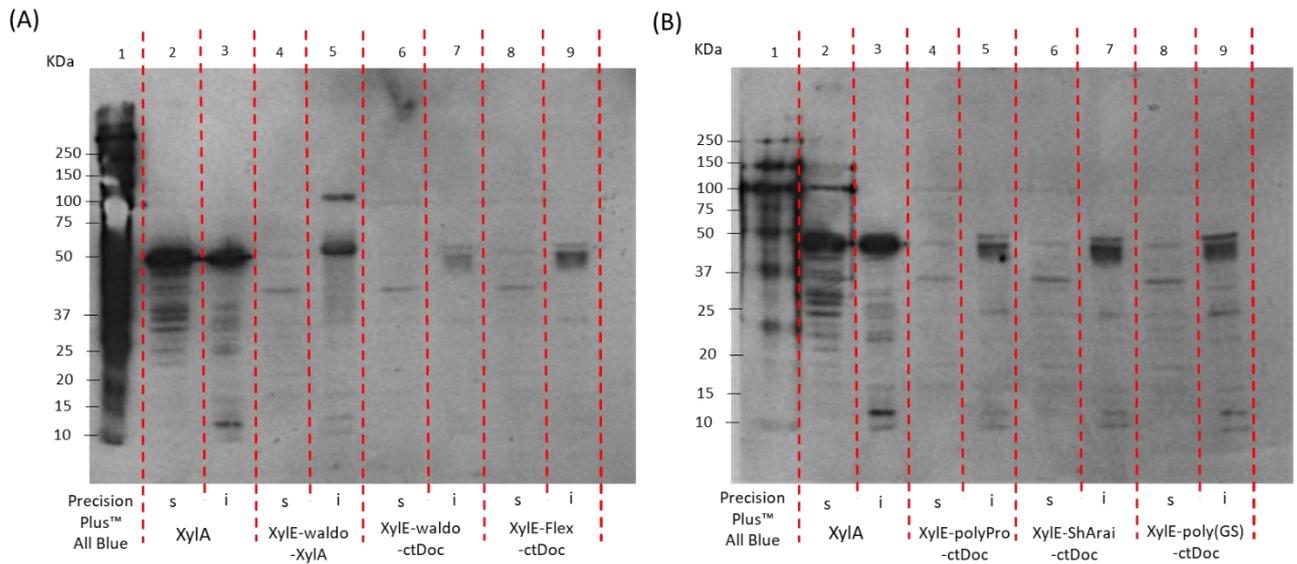


Figure 5.6. Anti-tetraHis western blot of the XylE-ctDoc fusions. Primary antibody was mouse antiHis in 3% (w/v) BSA, secondary antibody was Anti-mouse in 5% skimmed milk (w/v) in PBST. Blocked using a combination of 10% (w/v) skimmed milk and 5% (w/v) BSA. Components separated out into insoluble and soluble fractions, represented as “i” and “s” respectively. As it was not possible to fit all samples on a single gel, two blots are present, A and B, containing different XylE-ctDoc fusions alongside XylA as a positive control for antibody binding.

Each XylE-ctDoc fusion had an expected molecular weight of 63.71 kDa, however was expressed within the insoluble fraction as a series of bands, the clearest existing at >50 kDa, followed by a series of multiplets at mass <50 kDa. This variance may suggest partial digestion of the protein. While the size is smaller than anticipated, such variation is consistent with an understanding of how membrane proteins move through SDS-PAGE (Rath *et al.*, 2009).

The expression levels appear to be greater than that of the complete XylE-XylA fusion, with each XylE-ctDoc band appearing at similar intensity to the single 100 kDa band in well 5 of Fig 5.6.A. The exception to this rule is that of the fusion containing the waldo linker, which accumulated to lower levels (Fig. 5.6.A well 7). As background bands within well 7 are however lower, this could simply reflect a reduced overall protein concentration. As opposed to the XylE-XylA fusions there is no clear secondary band which would represent the cleaved Dockerin. This suggests that the fusion between XylE and ctDoc was more stable than that of XylE and XylA, however it must be acknowledged that the peptide, with a mass of 8.6 kDa, could have moved through the length of the SDS-PAGE gel during electrophoresis and been lost.

5.3.2. High expression levels of the XylA-ctCoh and ctCoh-XylA fusions.

Following confirmation of the XylE-ctDoc fusions, antiHis western blots were then applied to the XylA-ctCoh and ctCoh-XylA fusions to confirm that they too were successfully assembled and accumulate in the cell. A series of western blots were therefore produced for total cell lysate of $\Delta xyIA$ strains expressing one of the protein fusions or a native XylA control. In order to produce a suitable level of protein production, arabinose induction was used on each sample during overnight growth.

Compared to membrane fusion of XylE-XylA, high levels of the soluble protein fusions could be easily recognised for both XylA-ctCoh and ctCoh-XylA (Fig. 5.7+5.8) and were clearly larger than the 51.85 kDa XylA. By comparing to the All-Blue protein ladder, it is possible to see the band for the fusions is slightly smaller than 75 kDa, which is consistent with the predicted molecular weight of 71.26 kDa (Fig. 5.8). This likely reflects the increased stability of the fusion, as ctCoh is a small cytoplasmic protein without the same constraints as the membrane protein XylE.

For most of the fusions of ctCoh to the N-terminal of XylA, there is no clear production of unfused XylA (Fig. 5.7.B wells 2-7, Fig. 5.8.B wells 4-10). There is one key exception to this however, as the ctCoh-polyPro-XylA fusion, which contains the polyproline linker, produced two bands: the expected ctCoh-XylA fusion at 72 kDa, alongside a notable band at around 50 kDa, likely representing cleaved XylA (Fig. 5.8B, well 6). Furthermore, in well 4 on Fig. 5.7.B the band for the fusion protein is fainter than the rest, suggesting lower levels of stability for the fusion. Overall, this suggests that the polyproline linker is prone to cleavage, which appears to be a consistent pattern based on previous fusions containing the linker. The long Arai linker may have also demonstrated some fusion cleavage, as a band of similar size to XylA is present in Fig. 5.8.B, however this linker has proven to be difficult for use in Golden Gate assembly, and therefore has largely been excised from fusions by this point

Comparatively, the XylA-ctCoh fusion demonstrated slight differences in protein stability when compared to the ctCoh-XylA fusion (Fig. 5.8.A). While high levels of expression were seen for bands representing the C-terminal fusion of XylA to ctCoh, it is important to note that in many cases, other bands were also present in the wells expressing the fusions. Wells expressing the XylA-ctCoh fusions often produced a clear low molecular weight band which can be seen in all western blots, likely reflecting the presence of cleaved cohesin, at a molecular weight of 17.93 kDa. Unlike the XylE-XylA fusion or the ctCoh-XylA, it is the cohesin that is His-tagged in the C-terminal XylA-ctCoh fusion, which will mean the small protein is detected instead of

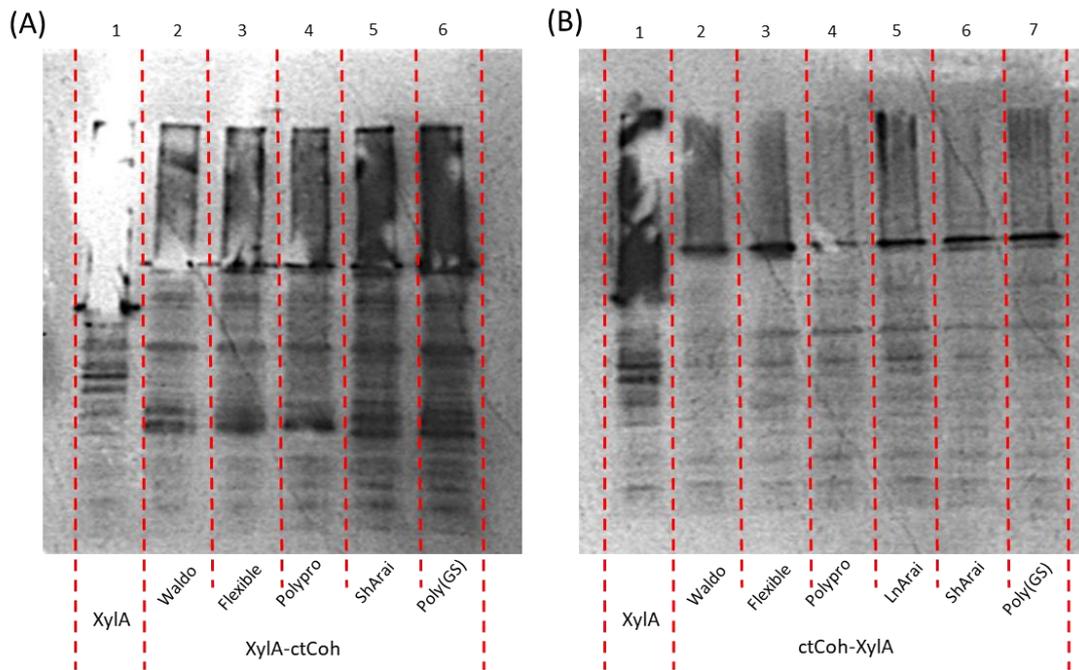


Figure 5.7. Anti-tetraHis western blot of the XylA-ctCoh and ctCoh-XylA fusions. Fusions are identified according to the linker contained, primary antibody was mouse antiHis in 3% (w/v) BSA, secondary antibody was Anti-mouse in 5% (w/v) skimmed milk in PBST. Blocked using a combination of 10% (w/v) skimmed milk and 5% (w/v) BSA. **(A)** XylA-ctCoh fusions (C-terminal fusion to XylA) **(B)** ctCoh-XylA fusion (N-terminal fusion to XylA).

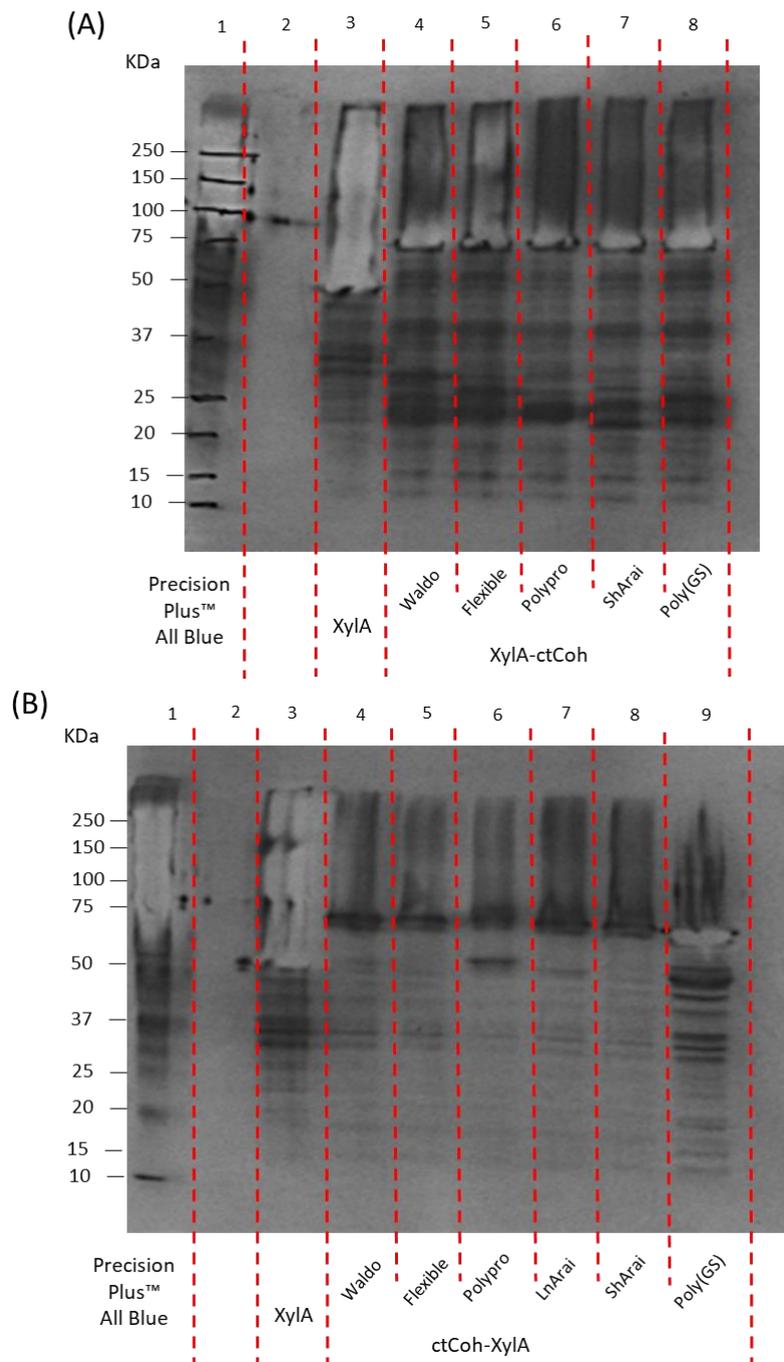


Figure 5.8. Anti-tetraHis western blot of the XylA-ctCoh and ctCoh-XylA fusions with All Blue ladder. Fusions are identified according to the linker contained, primary antibody was mouse antiHis in 3% (w/v) BSA, secondary antibody was Anti-mouse in 5% (w/v) skimmed milk in PBST. Blocked using a combination of 10% (w/v) skimmed milk and 5% (w/v) BSA. Expression was still high in this image, to the point that bands of the highest expression level were present as white blotches. Ladder was Precision Plus Protein™ All Blue Prestained Protein Standards, provided by Bio Rad. Due to poor clarity of the individual bands, marks were drawn on in A by overlaying the blotting film with PVDF membrane, in which the bands

were clearly visible, and marking each one. **(A)** XylA-ctCoh fusions (C-terminal fusion to XylA)
(B) ctCoh-XylA fusion (N-terminal fusion to XylA).

the XylA following cleavage. This suggests that a noticeable level of fusion cleavage has occurred, although much lower than the overall levels of fused protein present. Nevertheless, this work suggests that the XylA-ctCoh protein is likely to be less stable than the ctCoh-XylA fusion.

An additional total cell lysate western blot had been produced using the super signal ladder. As this ladder has been shown to contradict both the All Blue ladder and known protein standards, it necessitated repeating the experiment in Fig. 5.8. However, this result did demonstrate greater protein cleavage and demonstrated that, as with XylE-XylA and ctCoh-XylA fusions, the least stable fusion is that containing the polyproline linker (Fig. 5.9). Within this total cell lysate sample, significant amounts of unfused ctCoh can be seen present in well 5, containing the XylA-polypro-ctCoh fusion, with a reduced overall expression of the band representing the fusion itself.

In conclusion, the XylE-ctDoc, XylA-ctCoh and ctCoh-XylA were all successfully expressed and showed high stability. While linker cleavage took place, the products of this cleavage (XylA and ctCoh) showed expression levels lower than that of the assembled fusion, suggesting that there was some stability present. The most stable of all the fusions appears to be the ctCoh-XylA fusion, which both demonstrated high expression and rare cleavage, apart from the polyproline linker.

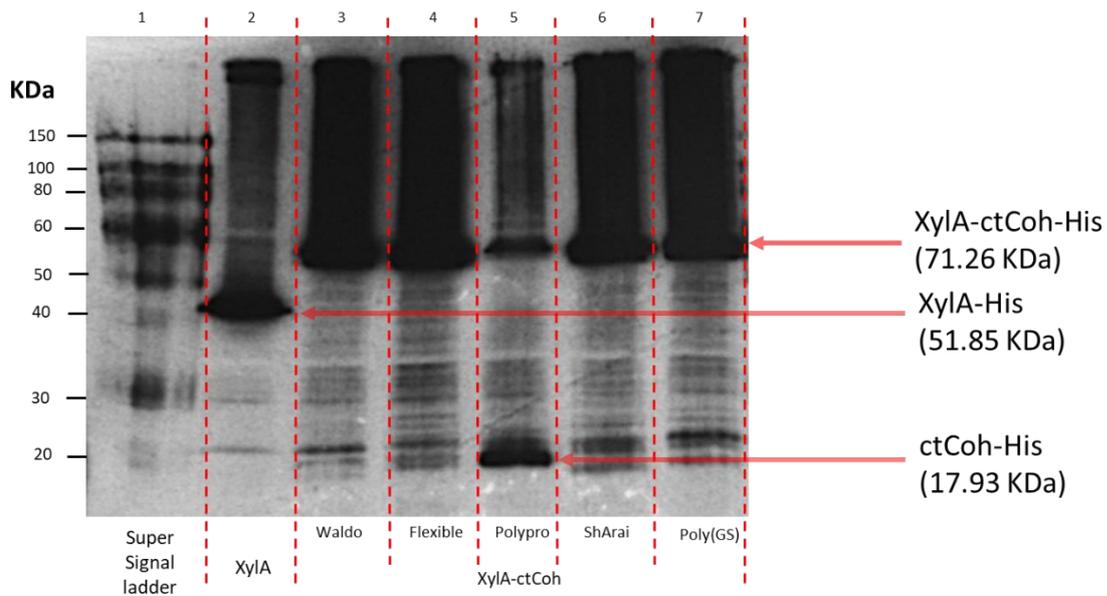


Figure 5.9. Anti-tetraHis western blot of XylA-ctCoh fusion with Super Signal ladder. Fusions are identified according to the linker contained, primary antibody was mouse antiHis in 3% (w/v) BSA, secondary antibody was Anti-mouse in 5% (w/v) skimmed milk in PBST. Blocked using a combination of 10% (w/v) skimmed milk and 5% (w/v) BSA. Ladder in the image was SuperSignal™ Molecular Weight Protein Ladder, provided by Thermofisher, however it did not appear to match expected masses of any protein larger than 30 kDa.

5.4. Determining the functionality of the scaffolds via a pull-down assay

5.4.1 Designing an *ex vivo* pull-down assay to confirm scaffold formation

With fusions successfully produced, it was important to validate that the cohesin and dockerin were able to interact. Considering the strong interaction between ctCoh and ctDoc (Jindou, Kajino, *et al.*, 2004), a pulldown assay was developed to ensure scaffold formation occurred following codon optimisation of the scaffold tags. The principle of this pulldown assay relied on the strong cohesin-dockerin interaction to prevent wash-stage elution of non-tagged proteins in a nickel affinity column. Two protein fusions would therefore need to be produced using two separate plasmids, to facilitate coexpression containing either the dockerin or cohesin respectively, however with the only one fusion His-tagged so to facilitate binding to the nickel column. This *ex vivo* system would also give insight into scaffold formation within the cell, as the interaction is likely to first form during cell growth.

An early design of the pulldown assay simply used a His-tagged ctDoc, however at 8.60 kDa this proved too small for accurate analysis. Following nickel affinity chromatography, we were unable to detect the dockerin in any of the collection columns using an anti-His western blot, possibly because the dockerin was too small for the SDS-PAGE Gel and had moved through the length of the gel during electrophoresis. Therefore, it was decided to fuse a larger protein to ctDoc to aid in detection. A novel fusion was therefore produced for this work, adding GFP to the His-tagged ctDoc to increase its overall mass. An mCherry-ctCoh fusion was designed to be the untagged protein – allowing for an anti-mCherry western blot to determine the amount of product present. In total, 4 unique fusions were designed for this experiment (Table 5.4).

The pull-down assay is developed around the presence of a His tag on only one of the two co-expressed proteins, as the nickel will only directly interact with the GFP-dockerin fusion out of all the proteins present in the cell (Fig. 5.10). In cells expressing both the GFP-ctDoc-His and ctCoh-mCherry fusions, the GFP-ctDoc-His would bind to the column specifically while the strong interaction between ctCoh and ctDoc would hold the mCherry within the column and reduce the amount captured in the wash column. Comparatively, in absence of the GFP-ctDoc-His, the ctCoh-mCherry would be unable to remain in the column and be promptly washed out.

To validate the nickel affinity column would not bind the negative control, the untagged ctCoh-mCherry fusion was run through the system in absence of another protein. 2x1L of cells induced with 0.01% (w/v) arabinose were grown for 6 hours, then spun down and resuspended in 35 mL wash solution, before being sonicated to produce a protein solution. 35 mL of the

Plasmid identification number	Promotor	RBS	Protein 1	Linker	Downstream peptide	Tag
pLC3_11A191	J23100 (constitutive)	B0034m	GFP	Waldo	ctDoc (Clostridium thermocellum dockerin)	6xHis
pLC3_11A161	J23100 (constitutive)	B0034m	GFP	Waldo	XylA	6xHis
pLD1_31712	I13453 (Arabinose induced)	B0034m	ctCoh (Clostridium thermocellum cohesin)	Waldo	mCherry	N/A
pLD3_31C1A2	I13453 (Arabinose induced)	B0034m	mCherry	Waldo	ctCoh (Clostridium thermocellum cohesin)	AN4

Table 5.4. The components of the novel fusions produced for the pull-down assay. A table displaying the components making up the four novel fusion constructs between produced for the pull-down assay. The plasmids used were pLC3, pLD1 and pLD3. The plasmid pLD1 can contain up to 5 components. The plasmids pLC3 and pLD3 can contain up to 6 components.

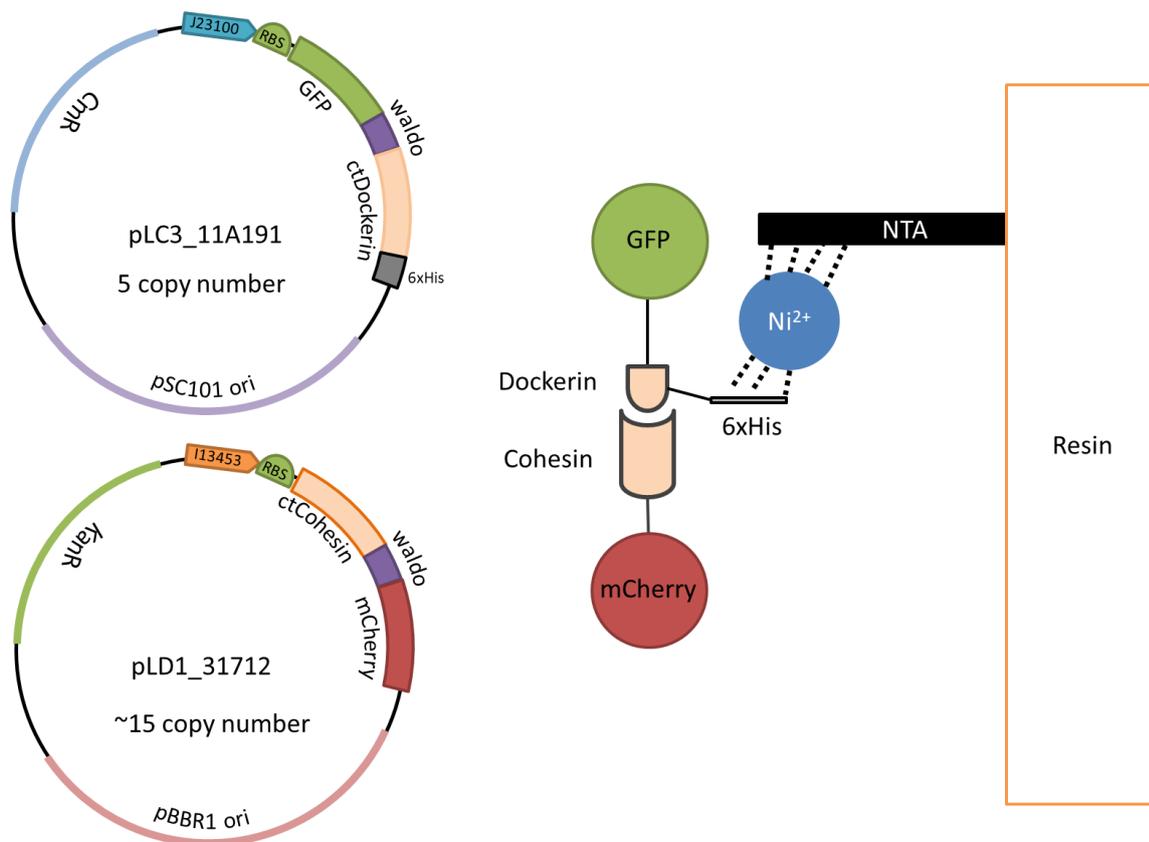


Figure 5.10. Cartoon representing the pull-down assay conditions. A cartoon representing the interaction between ctCoh and ctDoc within the Nickel affinity column. GFP-ctDoc-His is expressed by the pLC series plasmid and interacts with the NTA bound nickel. The tag is placed on the ctDoc to replicate the conditions of the Xyle-ctDoc fusion, which is also His-tagged at the ctDoc domain. The ctCoh-mCherry is expressed by the pLD series plasmid, and the ctCoh forms a strong, non-covalent, interaction with the ctDoc. This allows the ctCoh-mCherry to be retained in the column until imidazole is added in the elution stage.

protein solution was loaded onto the column and eluted into 1 mL aliquots, alongside an approximately 1 mL aliquot of flowthrough and column wash. The resultant eluate solution was found to be colourless, suggesting that little protein was obtained. The eluted sample, alongside the wash and flowthrough solution was diluted 10 fold and loaded onto an SDS-PAGE and western blot (Fig. 5.11).

Multiple faint bands were present within the Coomassie blue-stained SDS-PAGE gel for fractions 9 and 10 (Fig. 5.11.A). This reflects that some proteins present were non-specifically bound to the nickel column, either due to the high levels of expression or intrinsic His-tags. Comparatively both the flowthrough, which represents sample solution that has moved through the column and decanted into waste, and wash, which contains wash solution that flushed the column to remove unbound protein, have high levels of protein present.

An anti-tetraHis western blot followed, which showed none of the bands present in the wash solution, flowthrough nor the collected fractions were His-tagged (Fig. 5.11.B). An anti-mCherry western blot was then applied (Fig. 5.11.C). This demonstrated that some ctCoh-mCherry was detected within the eluate fractions 9 and 10, at a mass of 45.10 kDa. Nevertheless, this level is trace compared to the much greater amount of ctCoh-mCherry detected in both the flowthrough and wash samples, suggesting that the fusion was non-specifically bound.

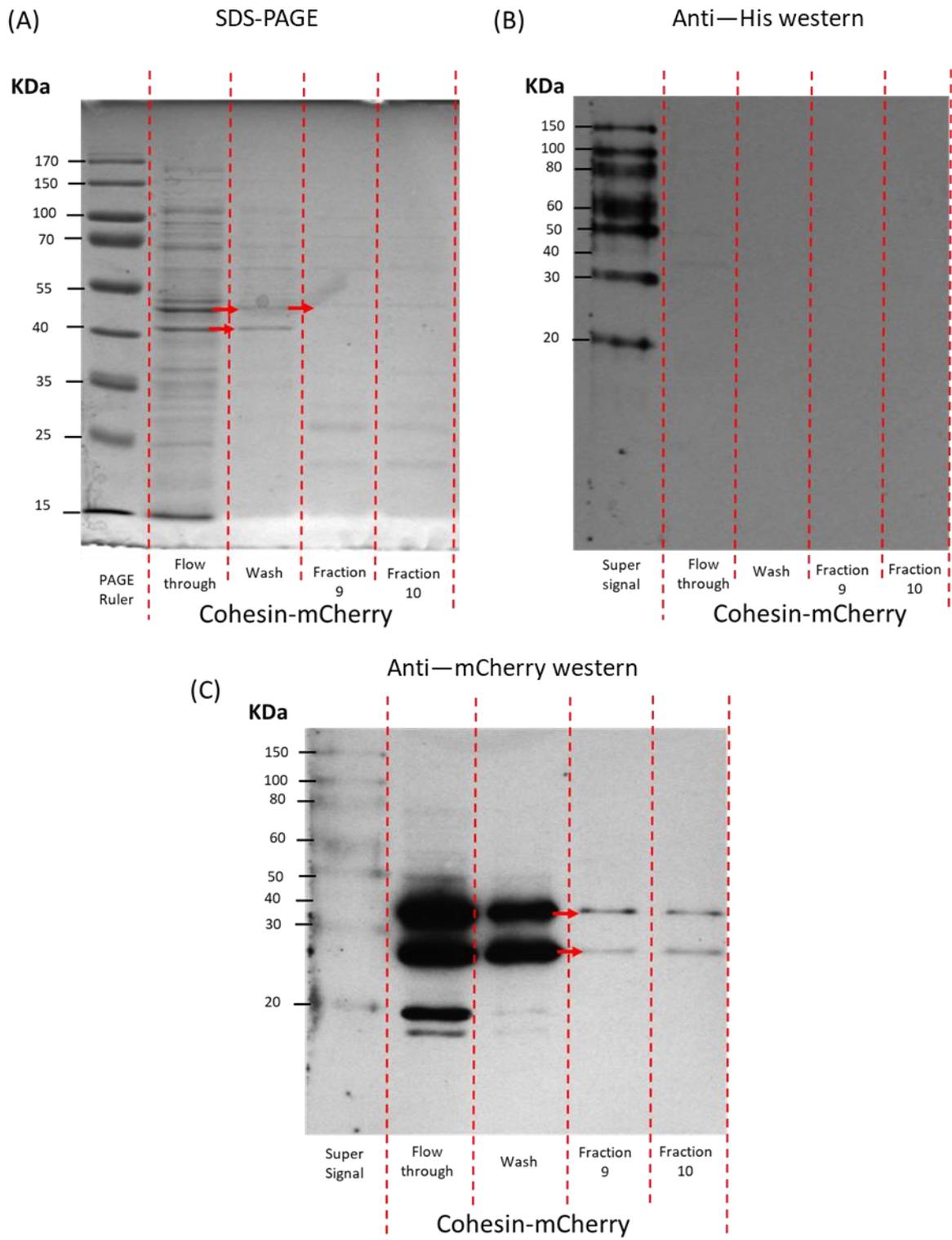


Figure 5.11. SDS-PAGE and Western Blot of cells expressing mCherry-ctCoh following nickel affinity chromatography. Cells transformed with pLC3 (empty vector) + pLD1_31712 (encoding mCherry-waldo-ctCoh) were sonicated and the supernatant run through a nickel affinity column. Regions of eluate were then obtained and run through an SDS-PAGE gel alongside an aliquot the flowthrough and wash. Samples were diluted 10-fold due to high

concentration of the mCherry in flowthrough overpowering results. **(A)** SDS-PAGE gel – ladder was PAGE Ruler **(B)** Anti-tetraHis western blot – ladder was SuperSignal™ Molecular Weight Protein Ladder **(C)** Anti-mCherry western blot - was SuperSignal™ Molecular Weight Protein Ladder.

5.4.2. The cohesin-dockerin scaffold confirmed to assemble in the pull-down assay

With an understanding of the base levels of mCherry bound by the nickel affinity column, the experiment was repeated alongside a sample condition of cells co-expressing both ctCoh-mCherry and GFP-ctDoc-His. As opposed to the first experiment, eluate was collected in 3 mL aliquots in collection tubes, with a photo taken of eluate fractions predicted to contain the protein (fractions 3 and 4) of both the sample and control conditions, alongside the flowthrough of both (Fig. 5.12.A). Looking at the fractions collected from the GFP-ctDoc-His/ctCoh-mCherry coexpression strain, it is clear to see that mCherry is present in the eluate due to the red tint of the fractions. Conversely, the columns for cells expressing ctCoh-mCherry alone did not display any colour in the eluate solution. One important thing to note, however, is that the flowthrough of the coexpression conditions has redder tinge compared to the flowthrough of cells expressing ctCoh-mCherry alone. This may suggest there was a greater concentration of ctCoh-mCherry expressed within the coexpression cells, despite both samples being equally induced.

The SDS PAGE of the ctCoh-mCherry sample expressed without GFP-ctDoc-His lacks clear bands that would correlate with the ctCoh-mCherry in the eluate samples (Fig. 5.12.B). Conversely, the cells coexpressing both ctCoh-mCherry and GFP-ctDoc-His appear to produce two bands suggesting high protein accumulation, identified by the red arrows in Fig. 5.12.C, column 3. These bands were found to be at a ~35 kDa and ~45 kDa, expected to represent the GFP-ctDoc-His (37.01 kDa) and ctCoh-mCherry (45.1 kDa) respectively. A previous pulldown assay of this condition further supported this conclusion by confirming identity with an anti-tetraHis and anti-mCherry western blot respectively (Appendix G).

Looking at the anti-mCherry western blots of both purification procedures, neither fraction 3 nor 4 of the cells expressing ctCoh-mCherry alone contain bands that could represent either the fusion or cleaved mCherry, which can be seen in both the wash and flowthrough conditions (Fig. 5.13.A). Conversely, we see high expression levels of the ctCoh-mCherry fusion in both of the eluate fractions of the coexpression sample, with a second smaller band expected to be cleaved mCherry also in fraction 3 (Fig. 5.13.B). Therefore, in comparison to cells expressing the ctCoh-mCherry fusion alone, this work demonstrates that cells coexpressing both ctCoh-mCherry and GFP-ctDoc-His can facilitate binding of the ctCoh-mcherry to the nickel affinity column, resulting in ctCoh-mCherry being collected in the eluate fractions at significantly higher levels than the control. This means that the scaffold components are able to strongly interact with one another to achieve a non-covalent complex.

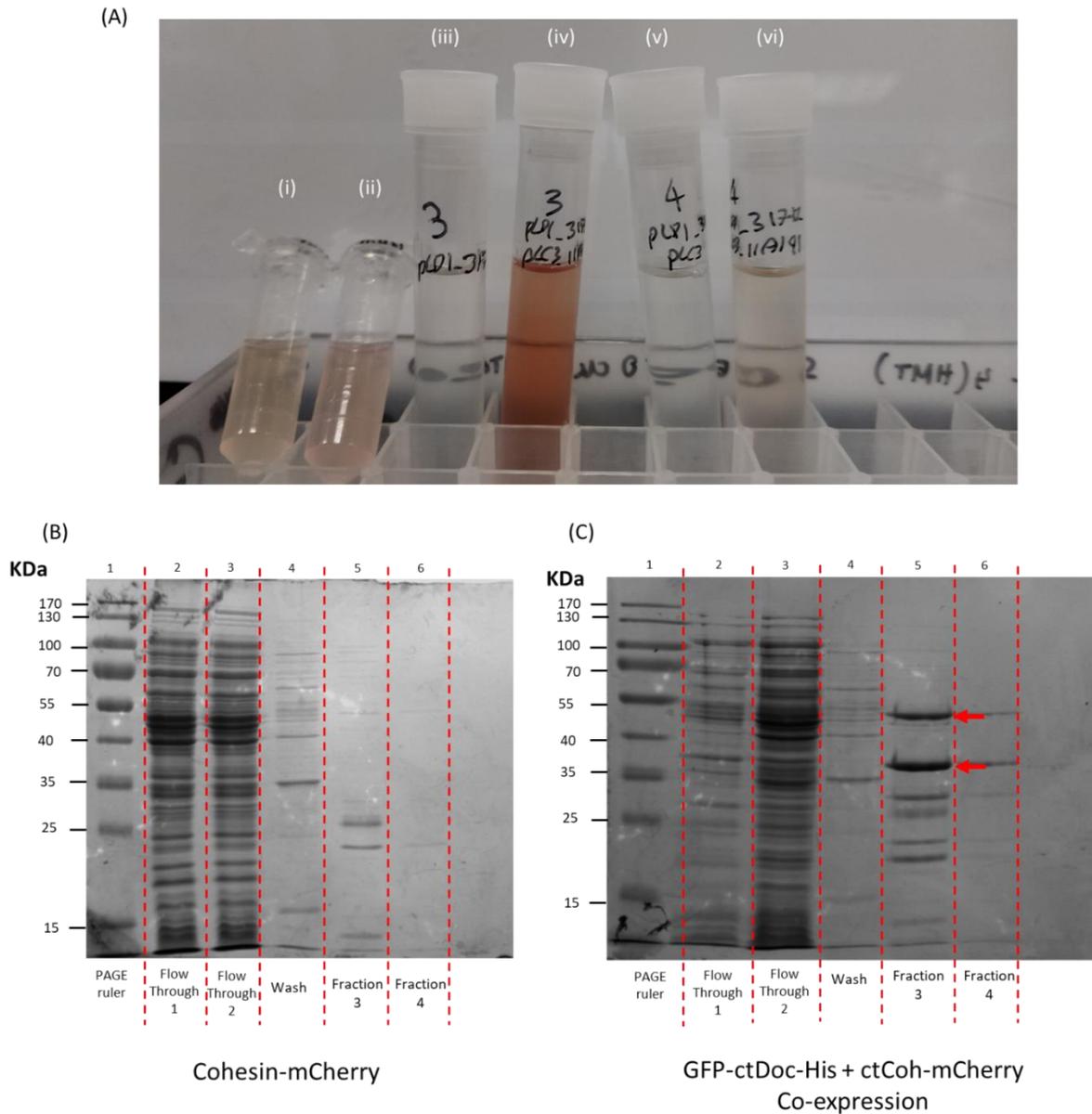


Figure 5.12. Products of Nickel affinity chromatography of the ctCoh-mCherry expressed either in isolation or coexpressed with GFP-ctDoc-His. Cells transformed with either pLC3 (empty vector) + pLD1_31712 (encoding ctCoh-waldo-mCherry), or pLC3_11A191 (GFP-ctDoc-His) + pLD1_31712 were sonicated and the supernatant run through a nickel affinity column. Regions of eluate were then obtained and run through an SDS-PAGE gel alongside aliquots of the flowthrough and wash. (A) An image of the columns obtained from nickel affinity chromatography (i) ctCoh-mCherry flowthrough 2 (ii) coexpression flowthrough 2 (iii) ctCoh-mCherry fraction 3 (iv) coexpression fraction 3 (v) ctCoh-mCherry fraction 4 (vi) coexpression fraction 4. (B) SDS PAGE of ctCoh-mCherry following nickel affinity chromatography, Samples were diluted 10-fold due to high concentration of the mCherry in flowthrough overpowering results (C) SDS PAGE of co-expression of ctCoh-

mCherry and GFP-ctDoc-His following Nickel affinity chromatography, Samples were diluted 10-fold due to high concentration of the mCherry in flowthrough overpowering results.

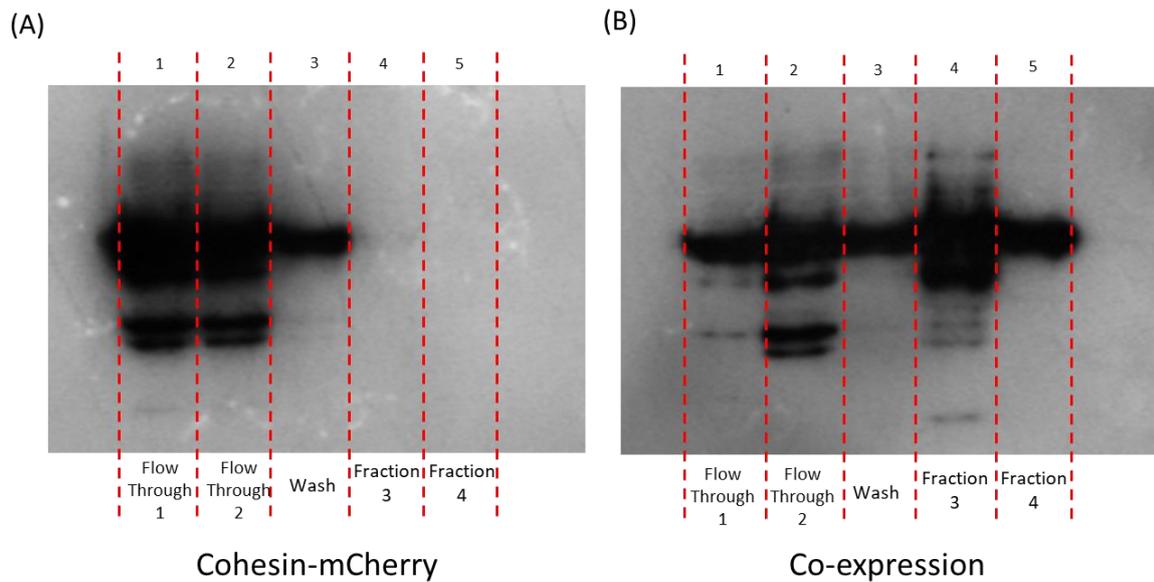


Figure 5.13. Anti-mCherry western blot of supernatant purified by nickel affinity chromatography. Cells transformed with either pLC3 (empty plasmid) + pLD1_31712 (ctCoh-waldo-mCherry), or pLC3_11A191 (GFP-ctDoc-His) + pLD1_31712 were sonicated and the supernatant run through a nickel affinity column. Regions of eluate were then obtained and run through an SDS-PAGE gel alongside aliquots the flowthrough and wash. Samples were diluted 10-fold in this experiment. The gels were then blotted onto a PVDF membrane and mCherry was detected using an anti-mCherry antibody. **(A)** The expression ctCoh-mCherry fusion **(B)** The coexpression of GFP-ctDoc-His and ctCoh-mCherry.

To validate that the interaction between ctCoh-mCherry and GFP-ctDoc-His is responsible for accumulation of ctCoh-mCherry in the eluate, and that this accumulation was not simply due to overexpression of two proteins, a control construct was produced: GFP-XylA-His. This construct lacks the dockerin responsible for binding ctCoh, and therefore while still able to interact with the nickel column, it should not bind the ctCoh-mCherry fusion directly.

In addition, it was of interest to determine whether the ctCoh would be able to bind to the ctDoc if it was bound to the opposite terminal of mCherry. This would represent the XylA-ctCoh fusions and was therefore an important consideration for scaffold assembly. The Golden Gate assembly components that had been produced for this work did not include a ctCoh gene with a stop codon, and due to time constraints, the construct produced, pLD3_31C1A2, featured a coiled coil AN4 at the 6th insert, which was used due to its small size and stop codon. This facilitated testing in the nickel-affinity column without concern that it would interact with the resin, as the AN4 was not His tagged.

BW25113 cells were transformed with either pLC3_11A161+pLD1_31712 (expressing GFP-XylA-His and ctCoh-mCherry respectively) or pLC3_11A191+pLD3_31712A (expressing GFP-ctDoc-His and mCherry-ctCoh-AN4 respectively) prepared as previously discussed, before being run through the Nickel-affinity column. During the GFP-XylA-His + ctCoh-mCherry coexpression, the AKTA suffered a calibration issue that resulted in a reduced volume of solution being passed through the column at all stages, meaning the column was loaded with 25 mL of sample and washed with a reduced volume of wash solution. Furthermore, the eluate collected per collection tube was between 1.2-1.5 mL as opposed to the expected 3 mL aliquot per tube. To compensate for this issue, every two aliquots were combined for this condition.

Both coexpressions produced eluate that was initially green after nickel affinity chromatography, however for the coexpression of GFP-ctDoc-His/mCherry-ctCoh-AN4 this colour gradually turned red over the course of the day as the mCherry matured (Fig. 5.14.A). This meant that the GFP-XylA-His/ctCoh-mCherry coexpression retained a green colour over 5 days, whereas the GFP-ctDoc-His/mCherry-ctCoh-AN4 coexpression turned fully red within 24 hours.

A resultant SDS-PAGE gel of the GFP-XylA-His + ctCoh-mCherry coexpression was produced (Fig. 5.14.B). To ensure that the levels of mCherry were not understated the samples were not diluted, in contrast to the previous control. Within this SDS-PAGE two key bands could be seen within the eluate columns, one near the approximately 70 kDa band, likely representing the complete GFP-XylA-His fusion (80.25 kDa), and a second around 50 kDa and may have represented XylA when cleaved from the fusion (51.85 kDa) or alternatively may reflect the

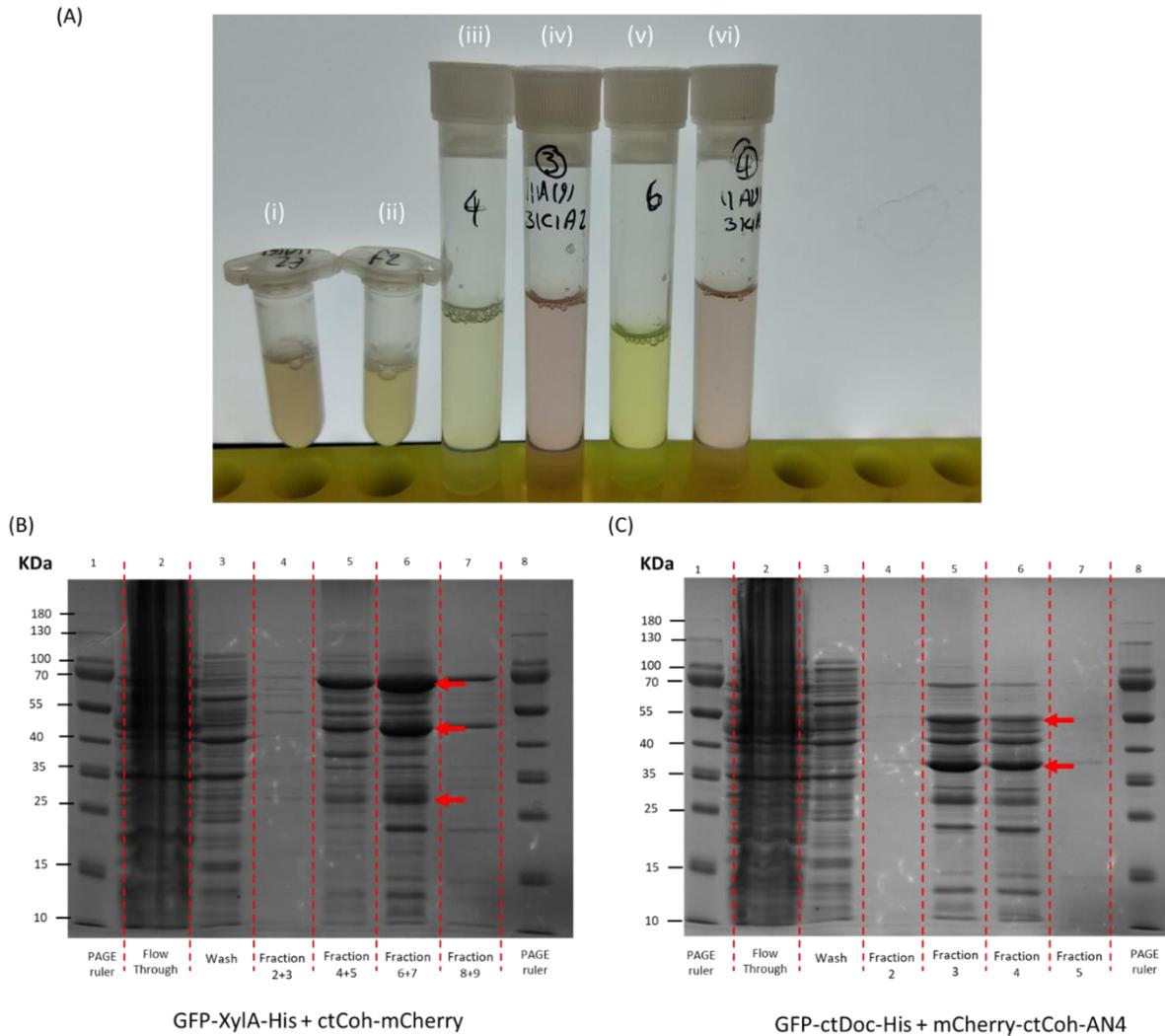


Figure 5.14. Products of Nickel affinity chromatography of cohesin-mCherry fusions expressed either in presence of GFP-XylA-His or GFP-ctDoc-His. Cells transformed with either pLC3_11A161 (GFP-XylA-His) + pLD1_31712 (ctCoh-waldo-mCherry), or pLC3_11A191 (GFP-ctDoc-His) + pLD1_31C1A2 (mCherry-waldo-ctCoh-AN4) were sonicated and the supernatant run through a nickel affinity column. Regions of eluate were then obtained and run through an SDS-PAGE gel alongside aliquots of the flowthrough and wash. The GFP-XylA-His/ctCoh-mCherry coexpression was subjected to chromatography 4 days before the GFP-ctDoc-His/mCherry-ctCoh-AN4 condition. (A) An image of the columns obtained from nickel affinity chromatography (i) GFP-XylA-His + ctCoh-mCherry coexpression flowthrough (ii) GFP-ctDoc-His + mCherry-ctCoh coexpression flowthrough (iii) GFP-XylA-His + ctCoh-mCherry fraction 4+5 (iv) GFP-ctDoc-His + mCherry-ctCoh fraction 3 (v) GFP-XylA-His + ctCoh-mCherry fraction 6+7 (vi) GFP-ctDoc-His + mCherry-ctCoh fraction 4. (B) SDS PAGE of co-expression of GFP-XylA-His + ctCoh-mCherry following nickel affinity chromatography – red arrows point to bands expected to represent GFP-XylA-His (largest band) and the degradation products XylA-His (second band) and GFP (smallest band) (C)

SDS PAGE of co-expression of mCherry-ctCoh-AN4 and GFP-ctDoc-His following Nickel affinity chromatography. Red arrows point to bands expected to represent mCherry-ctCoh-AN4 (larger band) and GFP-ctDoc-His (smaller band).

presence of ctCoh-mCherry. If the ~50 kDa band represented cleaved XylA, then a faint band around the region of 25 kDa would likely represent GFP (26.91 kDa).

The SDS-PAGE of the GFP-ctDoc-His/ mCherry-ctCoh-AN4 coexpression was also produced – to facilitate comparison to the control, this too was not diluted (Fig. 5.14.C). Once more two key bands could be seen within the eluate fractions, one around 50-55 kDa and a second greater than 35 kDa. It is likely the smaller band represents the target GFP-ctDoc-His molecule (37.01 kDa), meaning the larger band is expected to be mCherry-ctCoh-AN4 (48.89 kDa).

To determine whether any cohesin-mCherry fusions are present in the eluate, both samples underwent western blotting against an anti-mCherry primary antibody (Fig. 5.15). The coexpression of GFP-XylA-His/ctCoh-mCherry did demonstrate that some mCherry was present within the eluate (Fig. 5.15.A), as two bands, likely representing ctCoh-mCherry and cleaved mCherry at 45.1 kDa and 26.59 kDa respectively, were present in collection columns 4+5 and 6+7. In both cases however, the intensity of the bands were lower than the those of the wash column, which suggests that the mCherry present was non-specifically bound and simply not washed fully out during the chromatography reaction.

In contrast, the GFP-ctDoc-His/ mCherry-ctCoh-AN4 produced substantial signal for mCherry in fractions 3 and 4, much greater than that of the wash solution, suggesting that the mCherry-ctCoh-AN4 fusion had remained bound to the His-tagged dockerin (Fig. 5.15.B). Within these columns the individual bands present could not be made out, however a band around 50 kDa can be seen in both column 2 and 5. Curiously, a smaller band representing the cleaved mCherry cannot be seen in either of these images.

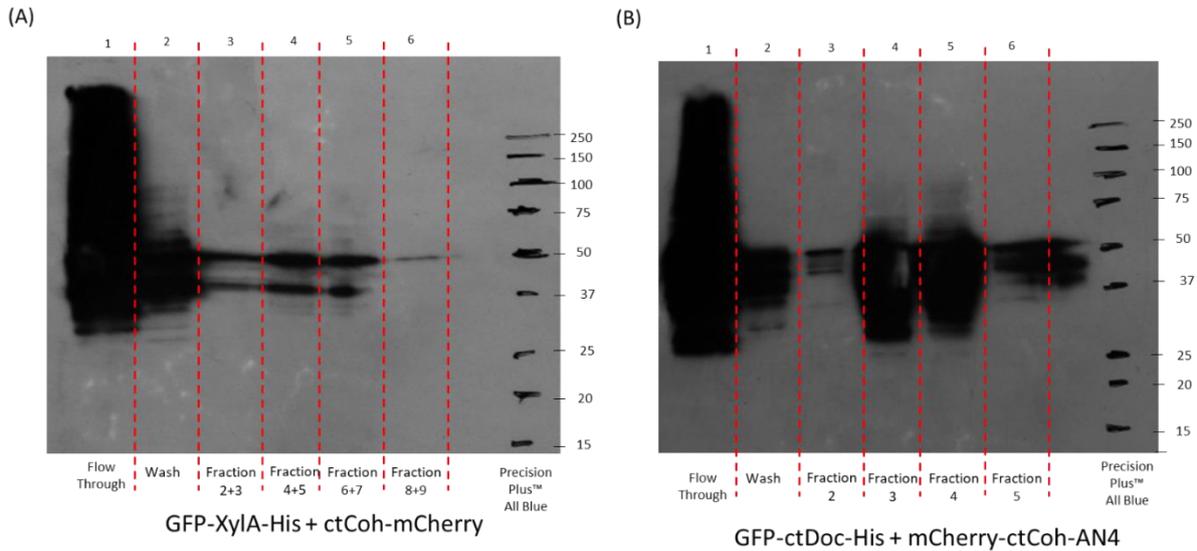


Figure 5.15. Anti-mCherry western blot of supernatant purified by nickel affinity chromatography. Cells transformed with either pLC3_11A161 (GFP-XylA-His) + pLD1_31712 (encoding ctCoh-waldo-mCherry), or pLC3_11A191 (GFP-ctDoc-His) + pLD1_31C1A2 (encoding mCherry-waldo-ctCoh-AN4) were sonicated and the supernatant run through a nickel affinity column. Regions of eluate were then obtained and run through an SDS-PAGE gel alongside aliquots the flowthrough and wash. Samples were not diluted, in contrast to previous western blots. The gels were then blotted onto a PVDF membrane and mCherry was detected using an anti-mCherry antibody. Ladder did not auto-fluoresce and therefore bands have been drawn on by comparing visible regions on the PVDF membrane to the photosensitive film. **(A)** The coexpression of GFP-XylA-His + ctCoh-mCherry fusion **(B)** The coexpression of GFP-ctDoc-His and mCherry-ctCoh-AN4.

5.4.3. Concluding remarks

Overall, these data demonstrate that there is a strong interaction between the ctCoh and ctDoc peptide tags, which can facilitate binding in the Nickel Affinity column and prevent the non-His tagged proteins from being washed out of the column. In the absence of ctDoc, it is possible for some non-specific binding of the ctCoh to the column, as seen by the presence of attached mCherry in western blots – however, these levels are low, and the resultant eluate does not demonstrate the red colour seen when His tagged ctDoc is present.

In addition, within this work we demonstrate that changing the terminal of the ctCoh does not appear to prevent scaffold formation. Both the C-terminal and N-terminal fusion of Cohesin to mCherry resulted in a high enough amount of mCherry being retained in the column that the eluates of both conditions took a red colour. While the terminal may impact the exact stability of the scaffold and propensity to form, in purely qualitative terms it is important to know that the terminus does not matter with this assay.

There is a clear loss of mCherry within the flowthrough, however. This is an important consideration, as while the His-tagged Dockerin is under the control of a constitutive promoter in a very low copy number plasmid, pLC (copy number ~5), the pLD plasmid is both arabinose induced and a slightly higher copy number (~15-20). This suggests that there is some imbalance associated between the two conditions, particularly when the ctCoh-mCherry is induced. This is an important consideration going forwards, as it is possible that the product of pLD series plasmids may swamp the products of pLC series plasmids.

Overall, it appears the cohesin and dockerin form a strong interaction capable of withstanding chromatography conditions. These interactions are likely to have formed during cell growth, although it is also possible, they were formed *in vitro* or broken and reformed during assay conditions. The orientation of the proteins used are representative of the XylE and XylA fusions generated in Table 5.2 and 5.3, therefore suggesting that scaffolds will also form at the membrane.

5.5. The use of growth assays to determine the impact of scaffolds on D-xylose utilization

After demonstrating that the codon optimised cohesin and dockerin components interacted through the pull-down assay, the scaffold system was introduced to the xylose utilization pathway to enhance cell growth. Within this work growth assays have been used to demonstrate the negative impact of direct fusion between XylE and XylA. Measuring the recovery of growth in a deletion mutant is a strong method for determining whether the fusion can achieve an equivalent function as the native proteins, and was therefore applied to the scaffold components to determine whether substrate channelling could be seen.

By using deletion mutants $\Delta xylA$ and TDXYlose it is possible to measure any improvements in growth rate facilitated by increased xylose metabolism. Increased metabolism of xylose is likely to result in an increased rate of growth, thereby demonstrating substrate channelling if a measure of growth, such as lag phase, growth rate or final OD is greater than the untagged controls.

The XylE-ctDoc fusions were encoded on a pLC3 plasmid with a constitutive promoter expressing the genes (Table 5.2). Conversely, the XylA-ctCoh fusions were encoded on a pLD3 plasmid with an arabinose inducible promoter activating gene expression (Table 5.3). No arabinose induction took place in this experiment, both to ensure that the ctCoh-XylA and XylA-ctCoh proteins were not overly expressed, swamping the XylE-ctDoc and masking scaffold formation, and to facilitate comparison with the XylE-XylA fusions.

Prior to measuring the two scaffold tagged proteins expressed together, it was important to also consider the activity of the individual components. The transporter XylE has been shown to be highly stable to C-terminal fusion, unaffected by fusion to bulky XylA or GFP. Considering the high stability of XylE to fusion as well as the small size of ctDoc, the XylE-ctDoc fusions were not tested in absence of a co-expressor. On the other hand, the XylE-XylA fusion demonstrated a loss of xylose isomerase activity and it was therefore important to measure if a similar impact occurred when XylA was fused to cohesin.

Based on the structural information obtained about XylA (Fig. 4.4-4.8) it can be hypothesized there are two potential issues with fusion at either terminal, regardless of the size of the fused domain. The N-terminal exists as an extension near the active site, and therefore fusion to this terminal may create a physical barrier against the substrate, preventing it from entering the active site and directly impacting function. The C-terminal extension is far removed from the active site, however this extension is associated with the monomer-monomer interface, and therefore fusion may prevent correct oligomerization, indirectly impacting isomerisation.

5.5.1. Fusion of cohesin to the N-terminal of XylA results in a loss of activity compared to C-terminal fusion.

Within this work the cohesin of *C. thermocellum* was fused to *E. coli* XylA at either N-terminal or the C-terminal through Golden Gate assembly into pLD3. Following the production of ctCoh-XylA (N-terminal fusion) and XylA-ctCoh (C-terminal fusion), it was important to determine whether fusion of cohesin impacted the performance of XylA. As ctCoh is a smaller protein when compared to XylE, it was hypothesized that the negative impact of N-terminal fusion seen in the XylE-XylA fusions would not be seen for ctCoh.

To understand the impact of fusion on XylA activity, $\Delta xylA$ strains were transformed with pLD3 vectors expressing either the fusion proteins or wild type XylA. Expressing wild type XylA on pLD3 results in partial recovery of $\Delta xylA$ growth on M9+10 mM xylose when compared to the negative control of $\Delta xylA$ transformed with empty vector (Fig. 5.16.A). This growth, however, was less than that of the wild type BW25113 cells. Cells expressing native XylA ended lag phase at 15-16 hours, twice that of BW25113, however quickly reach similar rates of growth, 0.334 h^{-1} to 0.312 h^{-1} respectively (Fig. 5.16.B) and a final OD of 0.4.

For all the ctCoh-XylA fusions, the lag phase appeared to end at 25 hours, an addition of 10 hours compared to native XylA, and the cells varied in growth rate. Cells expressing the ctCoh-XylA fusion containing either the short Arai linker or the poly(Gly-Ser) linker were able to reach an OD of 0.4 or greater and demonstrated a similar maximum rate of growth to the native XylA (0.283 h^{-1} for the short Arai linker, 0.284 h^{-1} for the Poly(GS) linker, $p > 0.1$ for both compared to unfused XylA). Curiously, these linkers are dissimilar in nature, with the poly(GS) linker being a 20 amino acid, flexible linker, while the short Arai linker was a 10 amino acid rigid linker.

Of the other fusions, those containing the long Arai linker and the Waldo linker both demonstrated similar levels of growth, reaching an OD of around 0.28, whereas the Flexible linker demonstrated the lowest OD of approximately 0.17, although demonstrated a high initial growth rate, reaching a maximum of 0.261 h^{-1} . Curiously, during initial growth, the worst performing fusion was that containing the polyproline linker – which in western blots had demonstrated a propensity towards cleavage (Fig. 5.9.B). An assumption could therefore be made that the fusion itself was poor in terms of enzyme activity, however over time cleavage would have resulted in a higher amount of free XylA, resulting in the small improvement in growth seen after 40 hours. Taking into account the late stage increase in growth rate, the ctCoh-polyproline-XylA fusion resulted in cells with a final OD of 0.2, although the maximum growth rate was still the lowest at 0.147 h^{-1} .

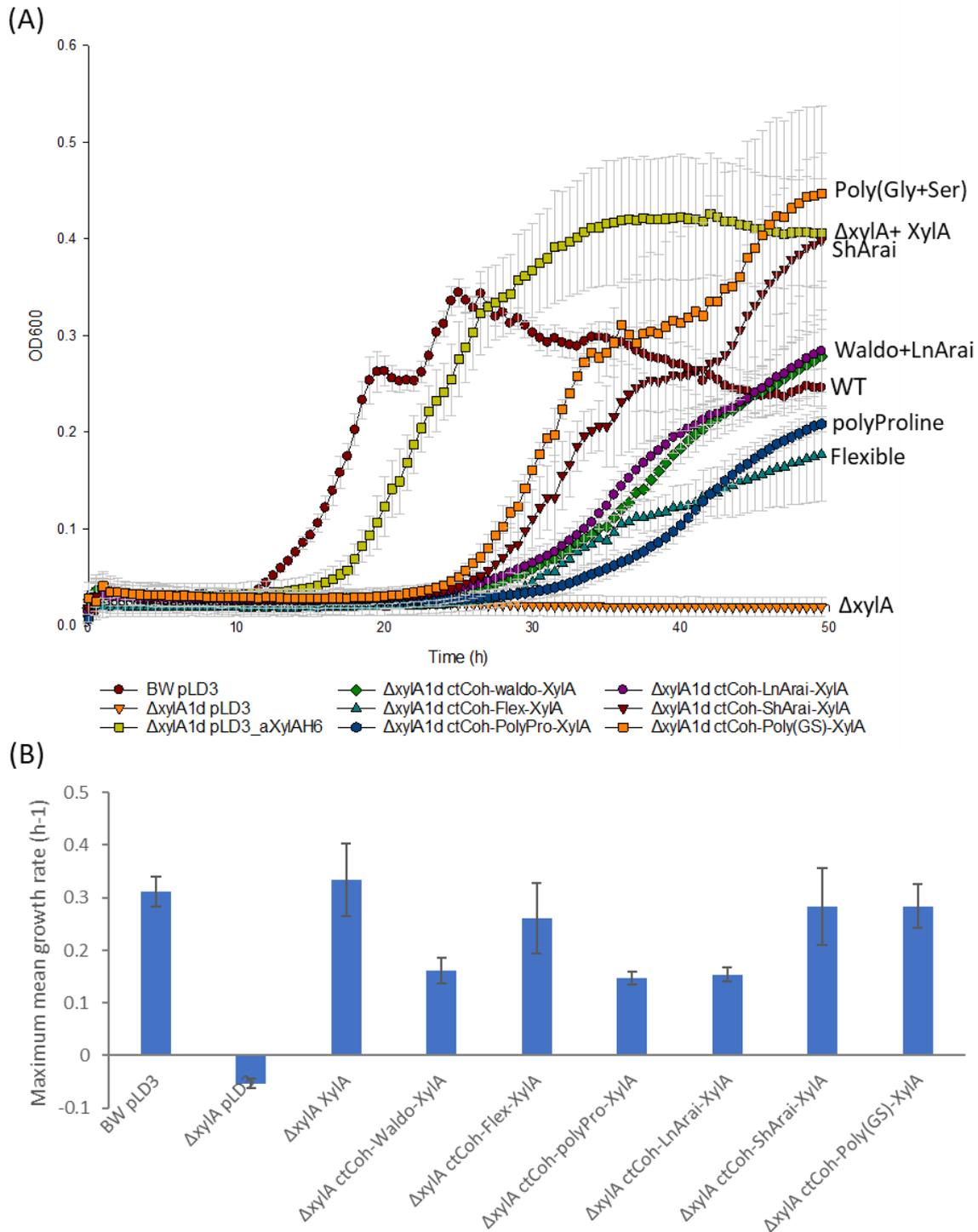


Figure 5.16. Fusion of cohesin to the XylA N-terminal reduces xylose isomerase activity

(A) Growth assay of $\Delta xyIA$ cells transformed with ctCoh-XylA fusions on a xylose deficiency background. Deletion mutant $\Delta xyIA$ was transformed with pLD3 plasmids expressing either native XylA or an N-terminal fusion of XylA to ctCoh. The cells were grown for 50 hours on a M9+10 mM Xylose + 60 μ g/mL kanamycin growth media. A BW25113 strain expressing pLD3 alone was present as a control. N=5. Error is standard deviation **(B)** Mean maximum growth rate of $\Delta xyIA$ cells transformed with ctCoh-XylA fusions. N=5. Error is standard deviation.

Ultimately, this work disproved the hypothesis that smaller peptide tags would not impact XylA when fused to the N-terminal. It appears that even ctCoh, a 17.93 kDa protein, has a negative effect on XylA activity when bound to the N-terminal. As the N-terminal is close to the active site, this may reflect steric hinderance within protein fusion that limited substrate access.

Therefore, to determine whether it was possible for proteins tags to have any non-negative impact to the activity of XylA, the C-terminal fusions were examined. The C-terminal of XylA is believed to be the contact point between the XylA monomers, increasing the surface area of the monomer to facilitate dimerization (Henrick, Collyer and Blow, 1989; Park *et al.*, 2019). Therefore, while the cohesin would be far removed from the active site when compared to the N-terminal, it is possible that C-terminal fusions could impact oligomerization.

Compared to the ctCoh-XylA fusions, the C-terminal XylA-ctCoh fusions demonstrate a reduced lag phase, comparable to the 15-16 hours of native XylA (Fig. 5.17.A), alongside a similar growth rate and final OD. Fusions containing either the Waldo or Flexible linker performed equivalent to the native XylA, with a growth rate of 0.191 h^{-1} or 0.211 h^{-1} respectively compared to XylA growth rate of 0.201 h^{-1} (Fig. 5.17.B). The two samples also demonstrated a p value > 0.1 . Conversely, both the short Arai linker and the polyproline linker performed worse, with a lower initial growth rate, however the XylA-ShArai-ctCoh fusion did have an equivalent maximum growth rate of 0.223 h^{-1} .

The best performing $\Delta xyIA$ strain, however, was that expressing the XylA-poly(GS)-ctCoh fusion, which grew with a growth rate of 0.346 h^{-1} , greater than the native XylA with a p value of 0.007, suggesting statistical significance. Furthermore, this strain, alongside the strain expressing the short Arai linker, achieved a greater final OD than the strain expressing native XylA. This is a curious result as there is no reason to expect the tagging of XylA to offer any improvement vs the native strain. It is possible this simply reflects the variability of growth assays, however an alternative explanation is that the tagging in some way stabilises the XylA protein.

Ultimately, this data suggests that fusion of ctCoh to the C-terminal XylA results in a stable product that behaves equivalent to native XylA. Conversely, the N-terminal fusion appears to reduce XylA activity, which in combination with structural information may suggest there is reduced access to the active site at this terminal. The successful fusion at the C-terminal is important as it means that any beneficial localization effect would not need to overcome the potentially negative impact of tagging XylA.

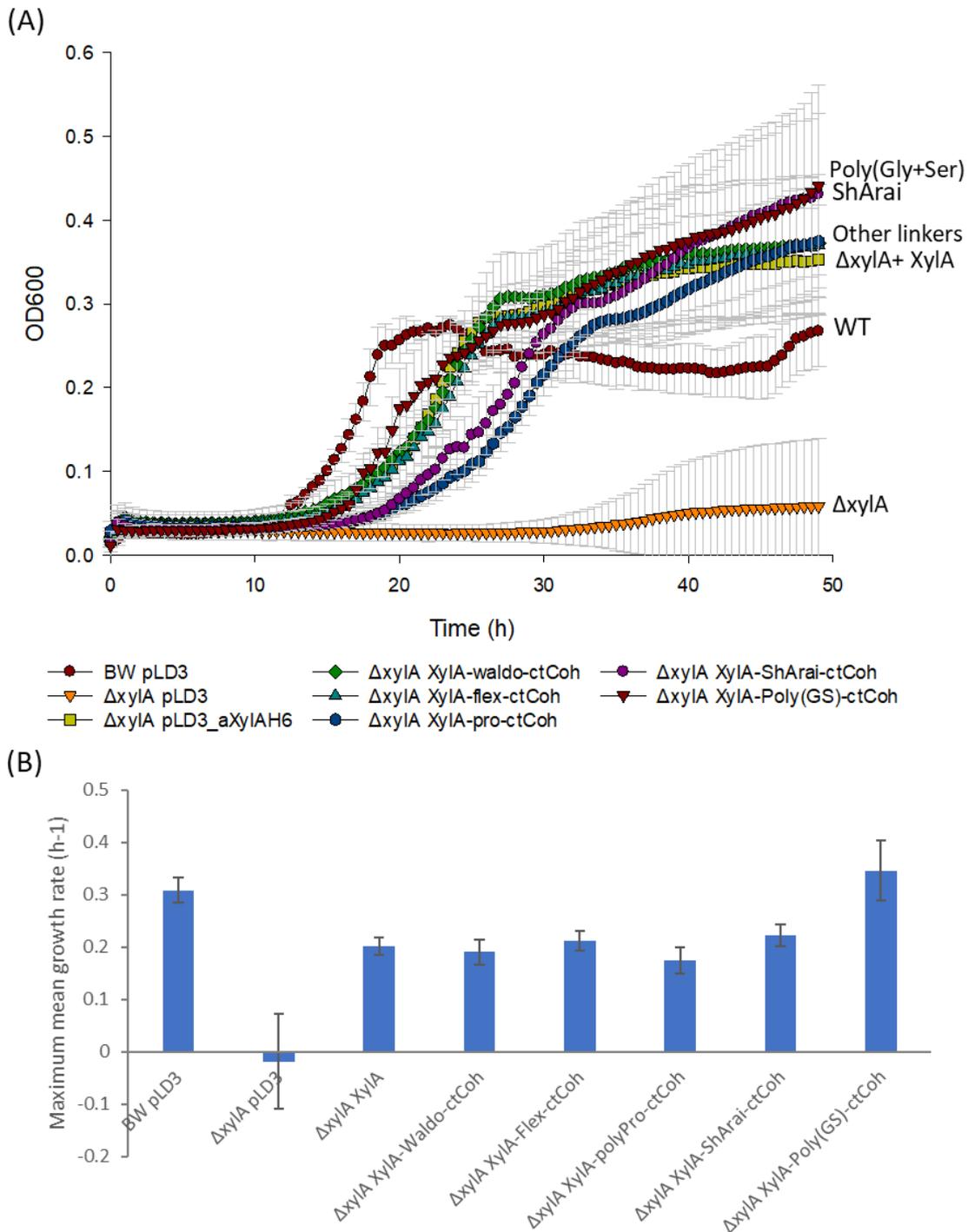


Figure 5.17. Fusion of cohesin to the C-terminal of XylA does not impact activity. (A) Growth assay of $\Delta xyIA$ cells transformed with XylA-ctCoh fusions on a xylose deficiency background. Deletion mutant $\Delta xyIA$ was transformed with pLD3 plasmids expressing either native XylA or a C-terminal fusion of XylA to ctCoh. The cells were grown for 49 hours on a M9+10 mM Xylose + 60 μ g/mL kanamycin growth media. A BW25113 strain expressing pLD3 alone was present as a control. N=5. Error is standard deviation. **(B)** Mean maximum growth rate of $\Delta xyIA$ cells transformed with XylA-ctCoh fusions. N=5. Error is standard deviation.

5.5.2. The transport of xylose across the membrane can be partially enhanced by coexpression of XylE-ctDoc and XylA-ctCoh.

The *E. coli* strain TDXYlose can be used to measure the impact of protein fusions on the transport of D-xylose. Both the primary xylose transporters XylE and XylF are deleted, alongside a component of arabinose transporter AraFGH, which is also associated with xylose transport (Desai and Rao, 2010). Reintroduction of the XylE can return growth to levels comparable to that of BW25113, however this recovery may be affected by protein fusion and scaffolding. Within this work, the use of TDXYlose allowed for measurement of XylE activity and to determine the impact of scaffold components on recovery of growth.

TDXYlose cells were transformed with both pLC3 plasmids encoding XylE-ctDoc fusions and pLD3 plasmids encoding either XylA-ctCoh or ctCoh-XylA fusions. Expression of both XylE and XylA within the TDXYlose strain was hypothesized to increase the overall growth rate compared to XylE expression alone. This is because the conversion of xylose to xylulose is the rate limiting step in the Xylose utilization pathway. Taking this into account, we hypothesize that increasing local concentration of XylA further through the action of scaffolds could further increase cell growth.

The first experiment aimed to contextualize the impact of coexpression on the growth of TDXYlose. To achieve this, cells were transformed with either XylE, XylA or XylE/XylA coexpressions (Fig. 5.18). Cells expressing either XylE or the XylE/XylA coexpression began growth within 8 hours of inoculation, with a much higher growth rate compared to that of cells either containing empty plasmid or expressing XylA alone. Furthermore, both XylE and the coexpression show a similar pattern of growth to one another in the first 24 hours, before reaching a maximum OD of ~0.32 around 25 h, with the cells expressing XylE gradually lowering in OD for the remainder of time measured. Therefore, there was not a significant difference between the TDXYlose cells coexpressing XylE/XylA compared to cells expressing XylE alone when accounting for lag phase, growth rate or maximum OD.

This suggests that while co-expression does not cause problems for cell growth, it does not benefit cell growth either, which is a surprise as expressing XylA alone results in an improved overall growth in the initial 24 hours compared to the negative TDXYlose control containing empty plasmid. While this did not result in a higher growth rate during the late-stage growth, consistent with TDXYlose, there is a consistent difference between OD for the negative control and cells expressing XylA before that. Improved growth of TDXYlose expressing XylA alone suggests that there is an increased catabolic efficiency of the limited amount D-xylose present

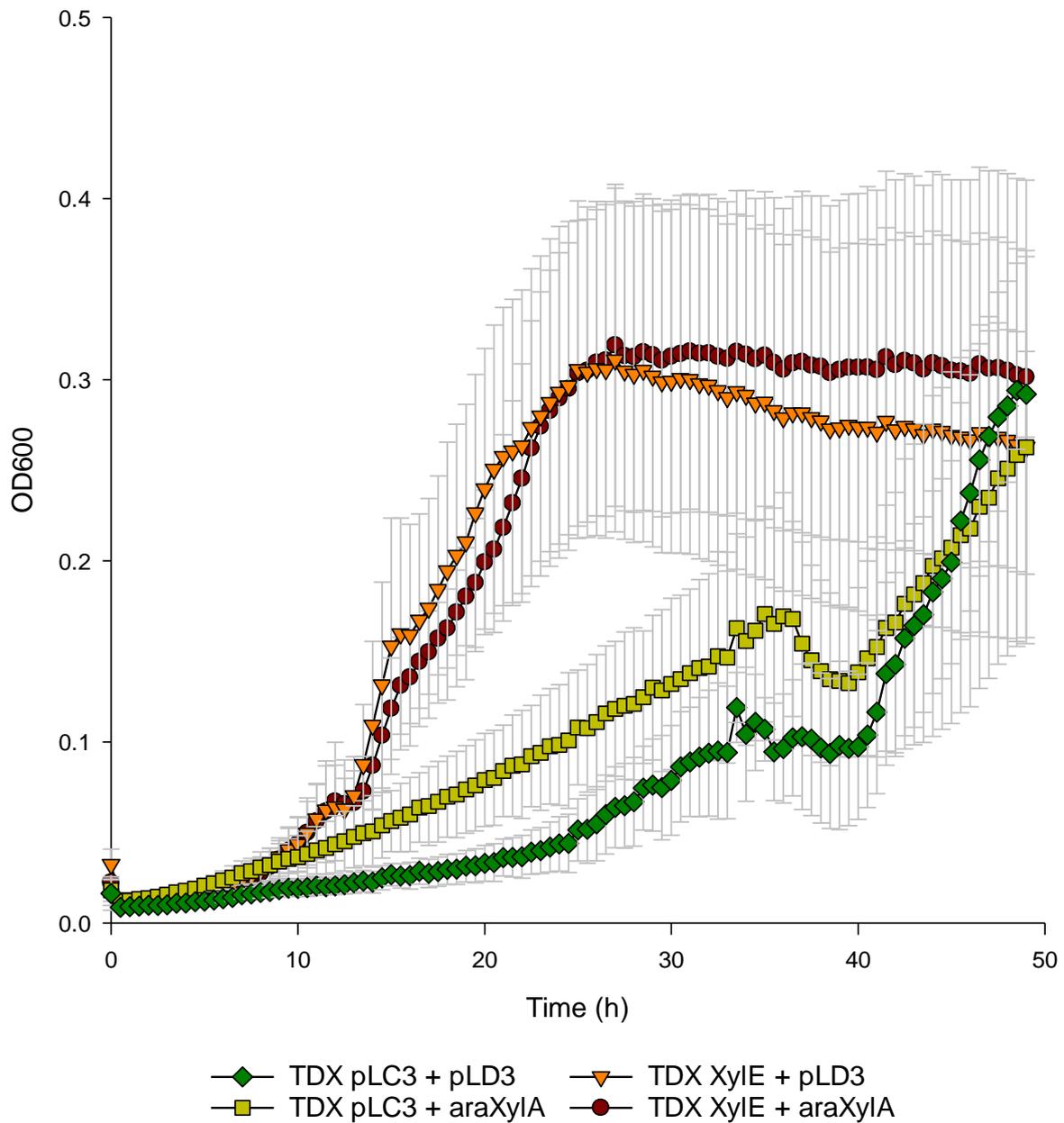


Figure 5.18. The impact of coexpression on a growth assay of TDXylose cells on a low Xylose background. Deletion mutant TDXylose was cotransformed with pLC3 plasmids expressing unfused XylE and pLD3 plasmids expressing unfused XylA. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 $\mu\text{g}/\text{mL}$ kanamycin + 34 $\mu\text{g}/\text{mL}$ chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. N=5. Error is standard deviation.

in the TDXYlose cell, or that higher catabolism promotes increased D-xylose diffusion into the cell. As the coexpression of both XylE and XylA simultaneously did not show a similar improvement, this may suggest that increasing xylose catabolism does not guarantee an improvement in growth.

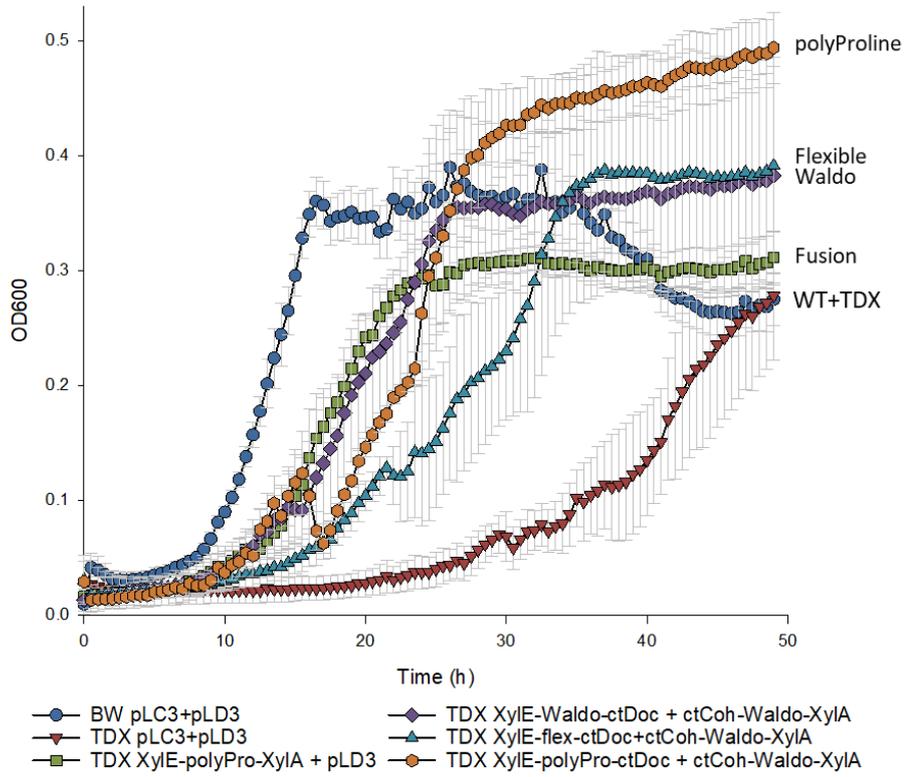
Based on modelling studies with GFP and the use of XylE-XylA fusions, evidence suggests that the impact of C-terminal fusion to XylE is not deleterious to transport activity. To determine whether the individual components could have had an impact, control cells were produced expressing either XylE-ctDoc/XylA or XylE/ctCoh-XylA. These were then compared to three XylE-ctDoc/ctCoh-XylA coexpressions as well as the XylE-polyproline-XylA fusion (Fig. 5.19.A+B). In addition, within this work maximum growth rate was measured only of the first 24 hours (Fig. 5.19.C), in order to determine the activity of the enzymes, while ignoring the late stage growth seen in TDXYlose strains.

It was found that addition of protein tags in the form of cohesins or dockerins did not appear to significantly impact the activity of the proteins. Native XylE expressed alongside ctCoh-waldo-XylA showed a similar lag phase to that of XylE-waldo-ctDoc expressed alongside native XylA, around 8 hours (Fig. 5.19.B). The key difference in growth behaviour between the two was a dip in growth around 15 hours into the assay, followed by recovery into a second exponential growth phase, a pattern seen for several cell lines. The use of ctCoh-XylA meant that xylose isomerase activity was lower than the native XylA, however was selected to determine whether scaffolding could improve cell growth – as changes in lower activity can be easier to see in the TDXYlose cell line.

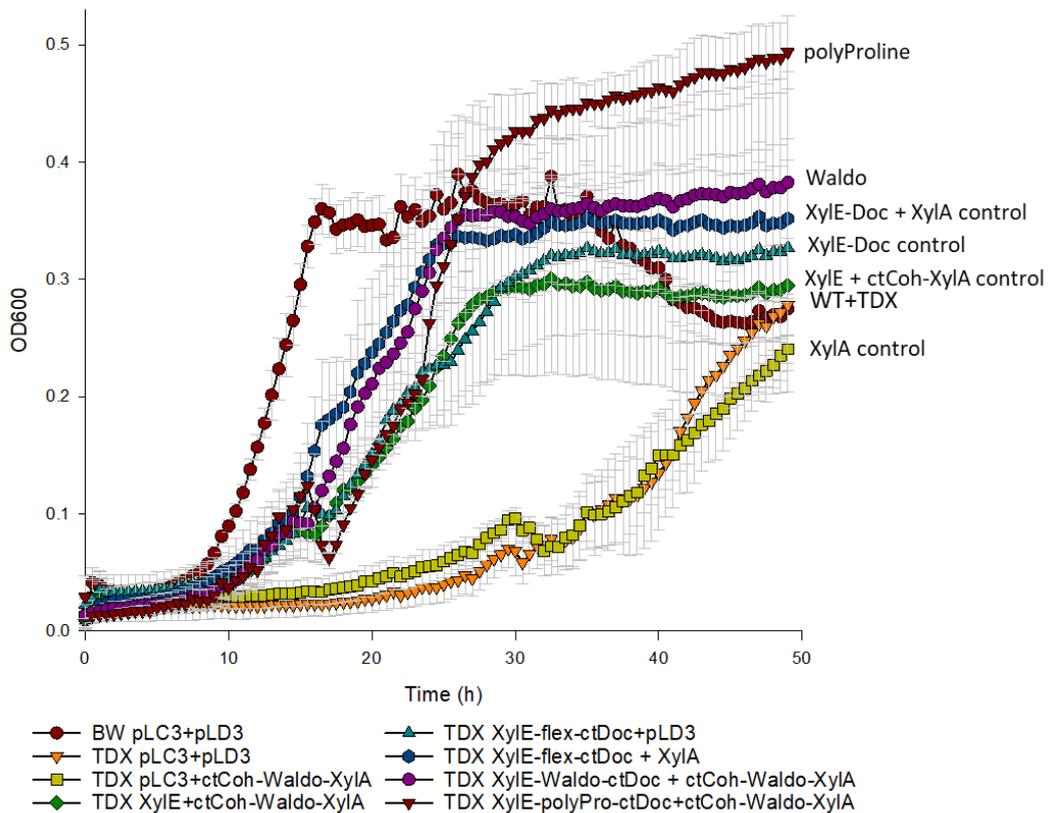
Looking into maximum growth rate, the XylE/ctCoh-XylA coexpression maximum growth rate of 0.238 h^{-1} was within error of cells expressing XylE-ctDoc alone at 0.257 h^{-1} (Fig. 5.19.C), which is consistent with the XylE/XylA coexpressions seen in Fig. 5.18. While the cells expressing XylE-ctDoc/XylA performed better than either sample, at 0.316 h^{-1} , the associated error meant the value compared to XylE/ctCoh-XylA was not found to be statistically significant. However, partially due to the high error, where statistical significance can be seen against the XylE/ctCoh-XylA coexpressions, it cannot be seen for the XylE-ctDoc/XylA coexpression, which may suggest there is some improvement in growth when native XylA is used. Ultimately, again, the key difference between these strains appears to be the dip 15 hours into cell growth (Fig. 5.19.B).

For the conditions expected to form scaffolds measured, the coexpression XylE-polypro-ctDoc/XylA-ctCoh was found to perform the best overall (Fig. 5.19.A). The coexpression had a lag phase of around 9 hours, grew to an OD of 0.39 and had a maximum growth rate of 0.357 h^{-1} , with much lower error compared to the XylE-ctDoc/XylA control. The key variation

(A)



(B)



(C)

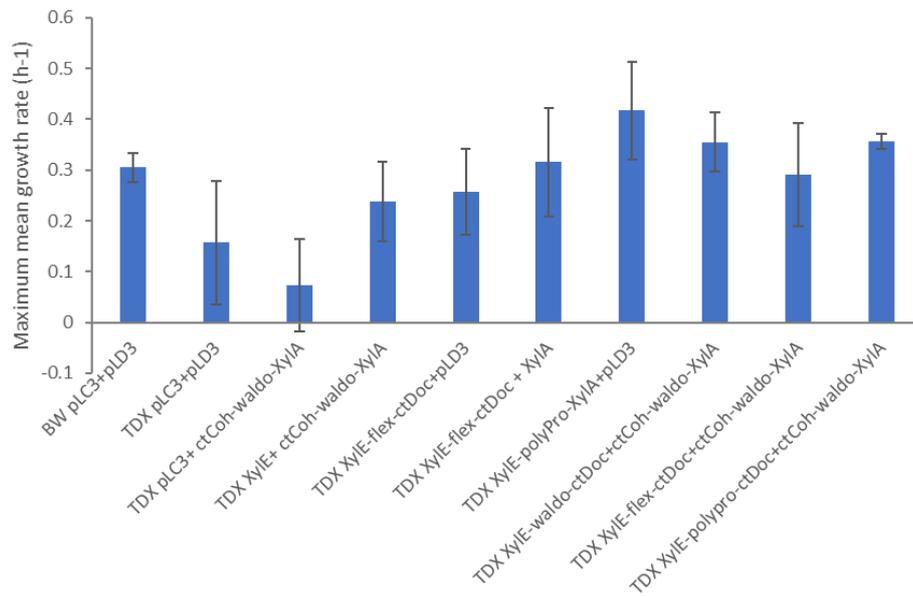


Figure 5.19. The impact of linkers in XylE-ctDoc on a growth assay of TDXylose cells on a low Xylose background when cotransformed alongside XylA-waldo-ctCoh fusions. Growth assay of TDXylose cells transformed with XylE-ctDoc/XylA-ctCoh fusions on a xylose deficiency background. N=5. Error = standard deviation **(A)** The three tested XylE-ctDoc + ctCoh-XylA scaffolds alongside a XylE-XylA fusion. Wild type control (BW25113) and TDXylose control (TDX expressing empty vector) also present **(B)** Growth rate of cells expressing XylE-ctDoc + ctCoh-XylA with Waldo and polyproline linkers alongside controls. Deletion mutant TDXylose was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native XylA or an C-terminal XylA-ctCoh fusion. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 µg/mL kanamycin + 34 µg/mL chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. N=5. Full figure presented in Appendix D. **(C)** Maximum growth rate for initial 24 hours of TDXylose growth. N=5. Error is standard deviation.

between the positive controls and the XylE-polypro-ctDoc coexpression was a small period of reduced growth around 15 hours that was the most significant of all conditions, followed by a phase of steep exponential growth between 17 and 25 hours.

The final OD of the XylE-polypro-ctDoc/ctCoh-XylA coexpression was 0.494, higher than any other sample by a significant margin (Fig. 5.19.C). This was a curious result as it may suggest that despite the unexpected decrease in optical density at 15 hours, this linker could have improved catalytic efficiency. It is important to note, however, that the polyproline linker has displayed a tendency towards linker cleavage when compared to the other linkers, as seen when looking at the XylE-XylA (Fig. 4.12) and XylA-ctCoh fusions (Fig. 5.9). While cleaved ctDoc could not be found in western blots of the XylE-ctDoc fusions (Fig. 5.6), it is difficult to declare with confidence that the improved final OD is a product of successful scaffolding. Furthermore, final OD can be prone to high variance between experiment, which means it is difficult to make conclusions based off of the result (See appendix D, pg 344 for another growth assay in which polyproline results in a lower final OD).

The ctCoh-XylA coexpression containing the XylE-waldo-ctDoc fusion produced a less pronounced pattern closer to the controls (Fig. 5.19.B). The maximum growth rate was found to be 0.355 h^{-1} , statistically significant compared to XylE/ctCoh-XylA ($p = 0.012$) but not compared to XylE-ctDoc/XylA ($p = 0.267$). This may suggest that some scaffolding occurred to increase xylose isomerisation to levels more comparable to the native XylA. Finally, the XylE-flex-ctDoc/ctCoh-XylA coexpression performed far worse than any sample expressing XylE. The coexpression demonstrated an extended lag phase of 12 hours, followed a slow acceleration in growth. The maximum growth rate was 0.291 h^{-1} , but high error made drawing conclusions a challenge. The final OD of 0.391 was consistent with full growth of cells by 48 hours.

From this growth assay, a few key points can be noted. Firstly, the best performing TDXYlose control expressed XylE-ctDoc/XylA and showed a notable difference in growth compared to the XylE/ctCoh-XylA condition - therefore the growth of TDXYlose can be enhanced by the increased isomerisation performance. This reinforces the conclusion that increased D-xylose catabolism can also enhance D-xylose transport into the cell. Furthermore, there is evidence that strains expressing both scaffold components outperform the XylE/ctCoh-XylA control – suggesting that some substrate channelling may have taken place. Secondly, by altering the linker used in the XylE-ctDoc fusion from Flexible to Waldo or polyproline, cell growth is enhanced, further suggesting a scaffold complex has formed, which can be altered by changing linker. Thirdly, for most outputs of cell growth, including lag phase, growth rate and rate at which stationary phase is reached, cells expressing scaffolds or fusions did not

outperform the XylE-ctDoc/XylA control, with one exception, the XylE-polyproline-ctDoc/ctCoh-XylA condition. This condition reached a final OD significantly greater than the control, which may suggest a role of linkers in late-stage catalytic efficiency – however final OD can be variable between experiments. Finally, it is important to note that the XylE-XylA fusion performed similarly to the XylE-ctDoc/XylA coexpression, suggesting the increased TDxylose growth seen for fusions in chapter 4 was the result of increased expression of XylA and not substrate channelling.

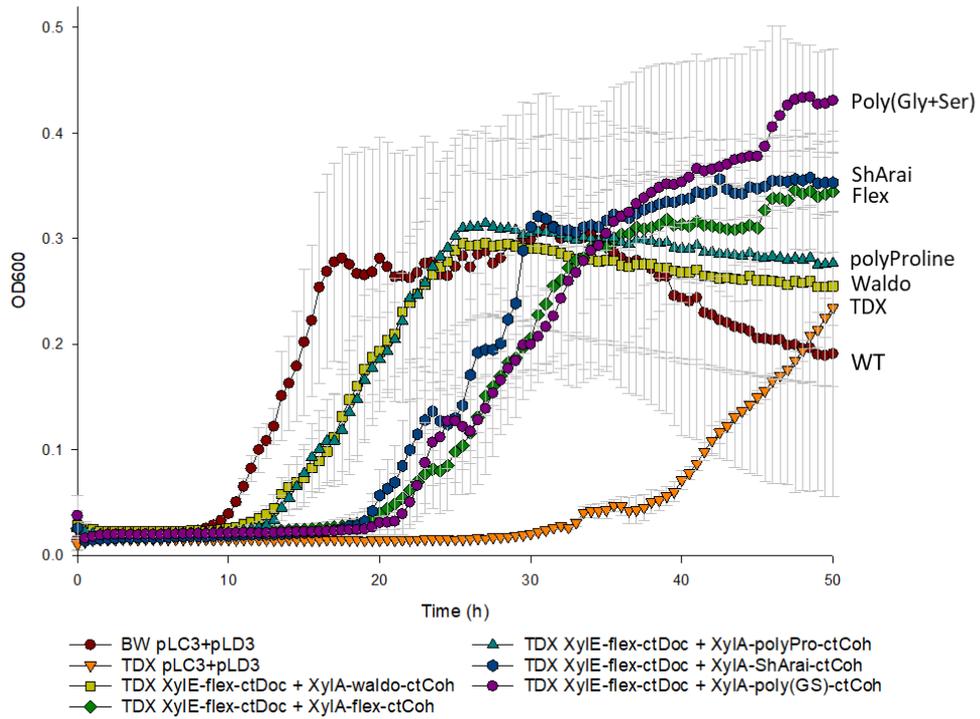
To further determine whether scaffold formation was taking place, and whether it was possible to enhance cell growth through scaffolding, XylA-ctCoh fusions were coexpressed with XylE-ctDoc. As previously established, the XylA-ctCoh fusions perform equivalently to unmodified XylA, therefore, if scaffolding is taking place, it is possible that the growth of TDxylose on the xylose minimal media background can be improved through scaffolding compared to native enzymes.

Due to time constraints this experiment was prepared prior to obtaining the results in Fig. 5.19, and the fusion XylE-Flex-ctDoc had been coexpressed with each XylA-ctCoh construct. This selection was due to the strong performance of the XylE-flex-XylA fusion in previous growth assays (Fig. 4.13-4.16). Unfortunately, the XylE-flex-ctDoc appeared to be the worst performing fusion for the scaffold condition, and it was not possible to rapidly produce new coexpression samples with the waldo or polyproline linker for this experiment. Nevertheless, while the flexible linker demonstrated poor growth rate in the previous coexpression test, it was hoped that it would still be suitable for testing the impact of expressing the different XylA-ctCoh fusions to assist in obtaining a rate equivalent to, or greater than, that of XylE/XylA.

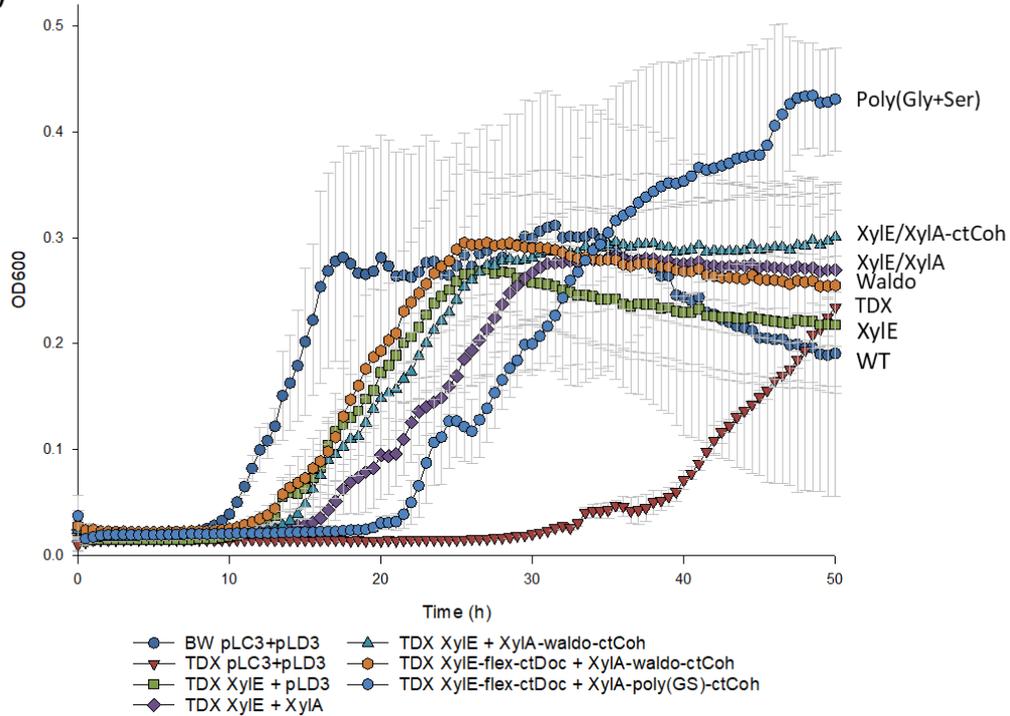
It was found that the best performing strains were those expressing XylA-waldo-ctCoh and XylA-polypro-ctCoh alongside XylE-ctDoc (Fig. 5.20.A). Both strains exited lag phase between 10 and 12 hours, the same time as the XylE control and XylE-ctDoc/XylA control, with the XylE-ctDoc/XylA-waldo-ctCoh sample consistently growing at faster rate than either of the control conditions (Fig. 5.20.B). The maximum rate of growth for the XylA-waldo-ctCoh and XylA-polypro-ctCoh was found to be 0.366 h^{-1} and 0.460 h^{-1} respectively (Fig. 5.20.C), the later comparable to XylE/XylA-waldo-ctCoh at 0.456 h^{-1} , however the difference was not found to be statistically significant for either ($p > 0.05$). This suggests that XylE-ctDoc/XylA-waldo-ctCoh sits within error of the control conditions. Ultimately, these samples demonstrated a small improvement in overall growth compared to the control, however there was no key area where a significant difference can be seen.

The other coexpressions were less effective, exiting lag phase between 19-21 hours (Fig.5.20.A). The fusion containing the poly(Gly-Ser) linker was of interest due to its high

(A)



(B)



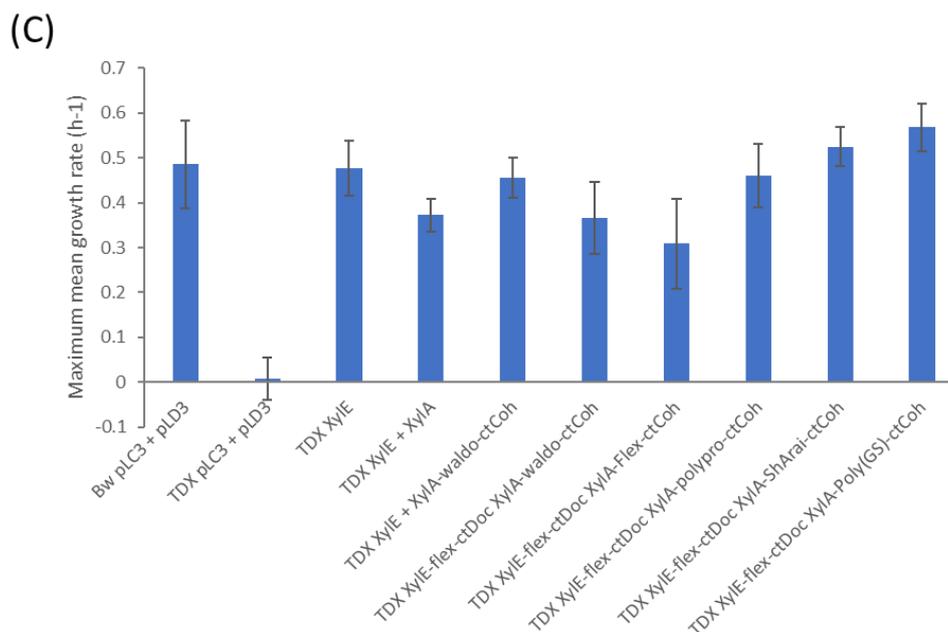


Figure 5.20. The impact of linkers in XylA-ctCoh on recovery of growth of TDXylose cells on a low Xylose background when cotransformed with both XylE-flex-ctDoc fusions. Growth assay of TDXylose cells transformed with XylE-ctDoc/XylA-ctCoh fusions on a xylose deficiency background. N=5. Error is standard deviation. **(A)** The tested XylE-ctDoc + XylA-ctCoh scaffolds containing five different linkers. Wild type control (BW25113) and TDXylose control (TDX expressing empty vector) also present **(B)** Growth rate of cells expressing XylE-ctDoc + XylA-ctCoh with Waldo and poly(Gly-Ser) linkers alongside controls. Deletion mutant TDXylose was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native XylA or an C-terminal XylA-ctCoh fusion. Deletion mutant TDXylose was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native XylA or a C-terminal XylA-ctCoh fusion. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 µg/mL kanamycin + 34 µg/mL chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. Full figure present in Appendix D. N=5. **(C)** Maximum growth rate for initial 24 hours of TDXylose growth. N=5. Error is standard deviation.

maximum growth rate of 0.568 h^{-1} , which reflected the first 5 hours of growth after lag phase and was found to be statistically significant compared to the XylE/XylA-ctCoh control ($p=0.00420$). The high cell-growth slowed after this period of 5 hours however and was followed by a reduced growth rate for the remainder of the experiment. Curiously, this sample reached the highest final OD of 0.427, much higher than the next greatest OD, measured at 0.353 for the XylE-ctDoc/XylA-ShArai-ctCoh coexpression.

Comparatively, the XylE-ctDoc/XylA-flex-ctCoh coexpression showed both a long lag phase and a low growth rate of 0.319 h^{-1} which was found to be statistically significant compared to the 0.456 h^{-1} of the XylE/XylA-ctCoh control. The final OD however was ultimately comparable to the XylE-ctDoc/XylA-ShArai-ctCoh coexpression. This suggests that the most successful cells in terms of overall growth were those that took the longest to overcome the lag phase.

Finally, it is important to also note that in this experiment the XylE/XylA coexpression performed much worse than the cells expressing XylE alone, with a lag phase of 15 hours and a maximum growth rate of 0.371 h^{-1} . This suggests high variability within the TDXylose cells when coexpressing two proteins.

This work demonstrated that the variations in the XylA-ctCoh fusions can also alter the growth of TDXylose, even when coexpressed with a consistent XylE-flex-ctDoc fusion. However, this data also demonstrated that there was only limited improvement to be seen in cell growth when expressing the scaffold components in a TDXylose background. This is most evident when looking at lag phase for each condition, which is the most significant difference between BW25113 and TDXylose cells expressing XylE alone or the XylE/XylA coexpression. Regardless of whether a C-terminal or N-terminal tag was used with XylA, cells did not demonstrate a reduced lag phase nor greater initial growth rate than at least one of the XylE/XylA controls. The potential for improvement can be seen by the BW25113 controls expressed alongside these TDXylose cells, as growth would often be within the first few hours and at a rate greater than most samples.

It is important to acknowledge that in some cases, a greater optical density can be seen for the strains coexpressing scaffold components when compared to the controls, however this value can be highly variable and it is difficult to compare this to BW25113 positive control, as this control often shows an immediate reduction in OD upon reaching stationary phase. In addition, it is concerning that both increased lag phase and reduced growth rate that can be seen for some of the tested scaffold conditions, which suggests that certain linker combinations can have a negative impact on enzyme or transporter function in spite of strong performance during individual expression.

Ultimately it is difficult to conclude the impact of linkers and whether scaffold formation has taken place, as there is a significant variability between experiments which demonstrates the challenge of using TDXYlose. This raises questions as to whether any of the improved growth, such as that seen in the XylE-ctDoc/ctCoh-XylA coexpressions (Fig. 5.19), provided evidence of scaffold formation, or were perhaps just further examples of inconsistent growth of TDXYlose strains. It was therefore reasoned that a better method of measuring the growth of cells on xylose would be through coexpression in a $\Delta xyIA$ cell line.

5.5.3. Coexpression of XylE-ctDoc and XylA-ctCoh fusions in $\Delta xyIA$ results in an inverse of the effects seen for expression of XylA-ctCoh fusions alone.

As the impact of coexpressing scaffold components on TDXYlose appeared largely negligible, it was therefore important to consider the impact of the scaffolding system on $\Delta xyIA$ activity. Coexpression of the XylE-ctDoc and XylA-ctCoh fusions were performed within $\Delta xyIA$ cells and grown on a M9+xylose minimal medium. The XylE-waldo-ctDoc fusion was selected to maintain consistency throughout the tests.

The impact of linkers on the activity of XylA-ctCoh fusions expressed in isolation was measured by a growth assay, producing a pattern of descending activity: poly(Gly-Ser) > Waldo = Flex > short Arai > polyproline (Fig. 5.17). It was important to determine whether the presence of XylE-ctDoc could influence this activity profile, therefore the first co-expression growth assays focused on the impact of linkers on the isomerase activity of XylA-ctCoh, measured through the growth of $\Delta xyIA$ (Fig. 5.21). It was found that when XylA-ctCoh fusions were expressed alongside the XylE-waldo-ctDoc fusion $\Delta xyIA$ cells grew in a pattern that was the inverse of that seen for $\Delta xyIA$ expressing the XylA-ctCoh fusion alone. In this experiment cells expressing XylA-poly(Gly-Ser)-ctCoh performed equivalently to the short Arai linker (Fig. 5.21.A) with a lag phase of 14 hours and a maximum growth rate of 0.23 h^{-1} and 0.31 h^{-1} respectively (Fig. 5.21.B). Both fusions also ultimately plateau at a final OD of ~ 0.16 .

Cells expressing the XylE-Waldo-ctDoc/XylA-Waldo-ctCoh coexpression performed slightly better, with a lag phase of 10-12 hours, equivalent with all other samples, and a maximum growth rate of 0.25 h^{-1} amongst the linkers. This rate of growth is near equivalent to expressing XylA alone, 0.26 h^{-1} , with $p=0.52$ suggesting no difference of statistical significance between the two samples. However, this linker still performed worse than the XylE and XylA co-expression, 0.44 h^{-1} . In addition, the final OD of 0.176 was lower than that seen for the XylE/XylA co-expression.

Finally, the XylA-flex-ctCoh and XylA-polypro-ctCoh performed the best out of the strains coexpressing both scaffold components. Both showed a reduce lag phase of 10-12 hours, shorter than both the XylE-ctDoc/XylA-waldo-ctCoh coexpression condition and BW pLC3+pLD3. The XylE/XylA co-expression in $\Delta xyIA$ did notably show the lowest lag phase of 9 hours, however both the XylE-waldo-ctDoc/XylA-flex-ctCoh and XylA-polypro-ctCoh showed a higher final OD of 0.330, compared to the XylE/XylA coexpression, 0.281. The cells expressing XylA-flex-ctCoh had the highest maximum growth rate at 0.577 h^{-1} , greater than even the XylE/XylA coexpression, 0.440 h^{-1} , however the p value of 0.099 was greater than 0.05, suggesting there is no statistical significance to this variance.

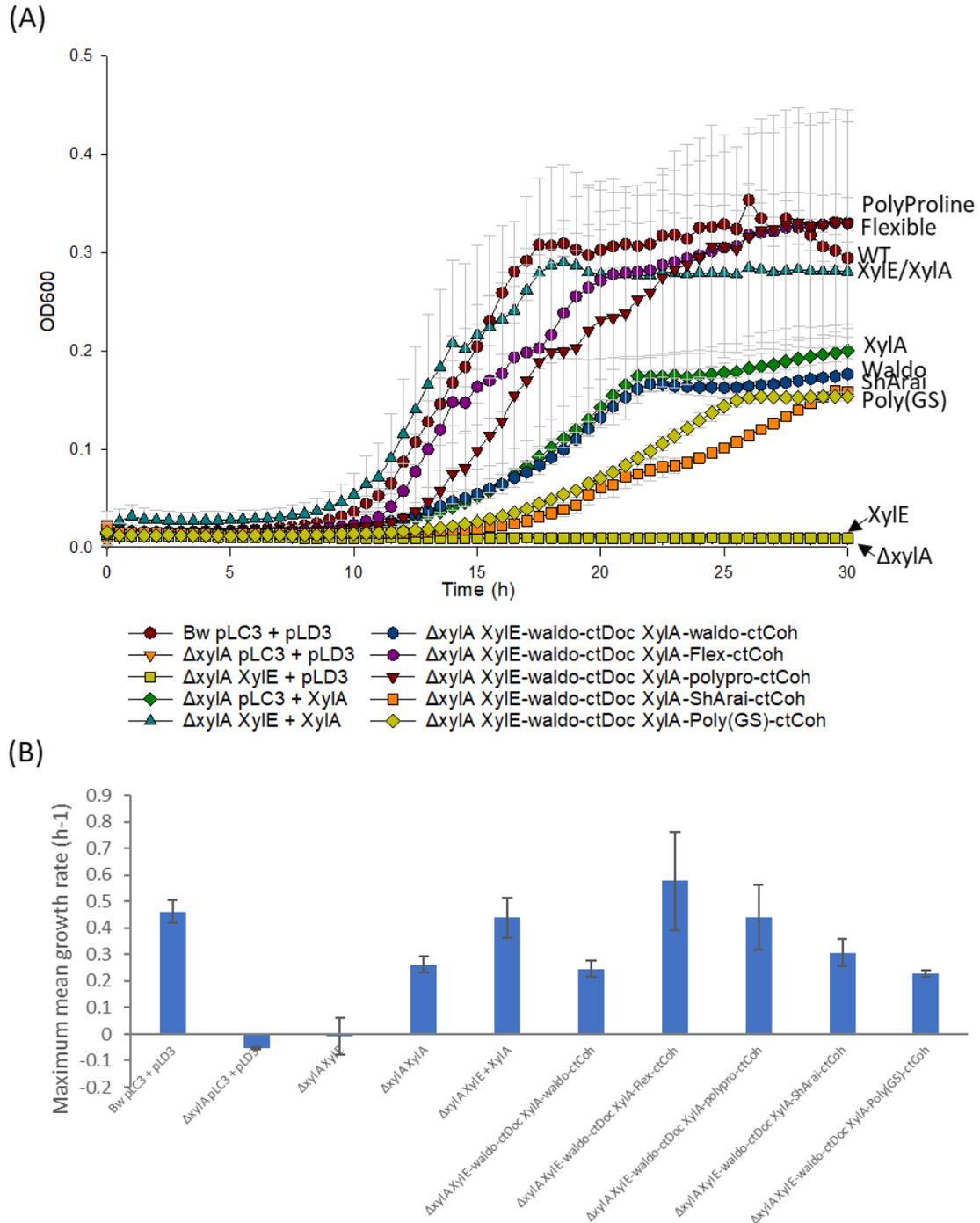


Figure 5.21. Evaluating the impact of XylA-ctCoh linker variation on scaffold formation

(A) Growth assay of $\Delta xyIA$ cells transformed with XylE-ctDoc and XylA-ctCoh fusions on a xylose deficiency background. Growth assay. Deletion mutant $\Delta xyIA$ was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native *xyIA* or an N-terminal fusion of XylA to ctCoh. The cells were grown for 30 hours on a M9+10 mM Xylose + 60 $\mu\text{g}/\text{mL}$ kanamycin + 34 $\mu\text{g}/\text{mL}$ chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. N=5. Error is standard deviation. **(B)** Maximum growth rate of $\Delta xyIA$ cells. N=5. Error is standard deviation.

Comparatively, the XylA-polyPro-ctCoh fusion had as growth rate of 0.441 h^{-1} with a p value of 0.990, suggesting near equivalence with the standard control.

Overall, the pattern seen is therefore Flexible = polyproline > Waldo > short Arai = poly(Gly-Ser). In essence, the worst performing linker within the XylA-ctCoh fusion, when considering expression in $\Delta xyIA$ (Fig. 5.17), is among best performing linkers when co-expressed with XylE-ctDoc. Therefore, while the linkers may negatively impact the efficiency of the enzymes individually, the improved cell growth in coexpression strains may suggest that scaffold formation has taken place. The XylA-polypro-ctCoh fusion showed an extended lag phase of 16 hours when expressed alone. The stronger performance of strains coexpressing XylE-ctDoc alongside XylA-polypro-ctCoh, compared to alternative conditions, may suggest that a rigid linker is needed to facilitate the scaffold formation.

To further determine the role of coexpression and to see if scaffold formation has taken place, a series of strains were produced, expressing either XylE-ctDoc/XylA-ctCoh co-expressions, XylE-ctDoc/ctCoh-XylA co-expressions or XylE-XylA fusions. The purpose of this experiment was to confirm the pattern seen in the XylE-ctDoc/XylA-ctCoh co-expressions was retained and to determine whether alternative methods showed any impact on cell growth.

This work demonstrated that the XylE-ctDoc/XylA-ctCoh co-expressions resulted in improved growth compared to both the direct fusion and the alternative ctCoh-XylA containing co-expression (Fig. 5.22.A). Each XylE-ctDoc/XylA-ctCoh co-expression showed a reduced lag phase of ~12 hours vs ~15 hours for the XylE-ctDoc/ctCoh-XylA co-expressions and >48 hours for the fusions. Furthermore, in this experiment all XylE-ctDoc/XylA-ctCoh co-expressions showed a lag phase equivalent to, or lower than, that of the XylE/XylA co-expression.

In terms of optical density, the highest OD measured was that of the coexpression containing the XylA-polyPro-ctCoh fusion, which had a final OD of 0.283, the next highest being XylA-poly(GS)-ctCoh at 0.251 and then XylA-Waldo-ctCoh at 0.197, higher than that of the XylE/XylA coexpression at 0.191 (although the peak OD of XylE/XylA was slightly higher at 0.213). All other conditions failed to reach the OD of the XylE/XylA coexpression. Overall, this meant that the XylE-ctDoc/XylA-ctCoh co-expressions containing either the polyproline linker or poly(Gly-Ser) linker performed better than the XylE/XylA control, demonstrating some variance from the data seen in Fig. 5.21.

The maximum growth rate measured for XylE/XylA coexpression was found to be 0.394 h^{-1} . For the XylE-ctDoc/XylA-ctCoh co-expressions growth rate was measured at a slower 0.300

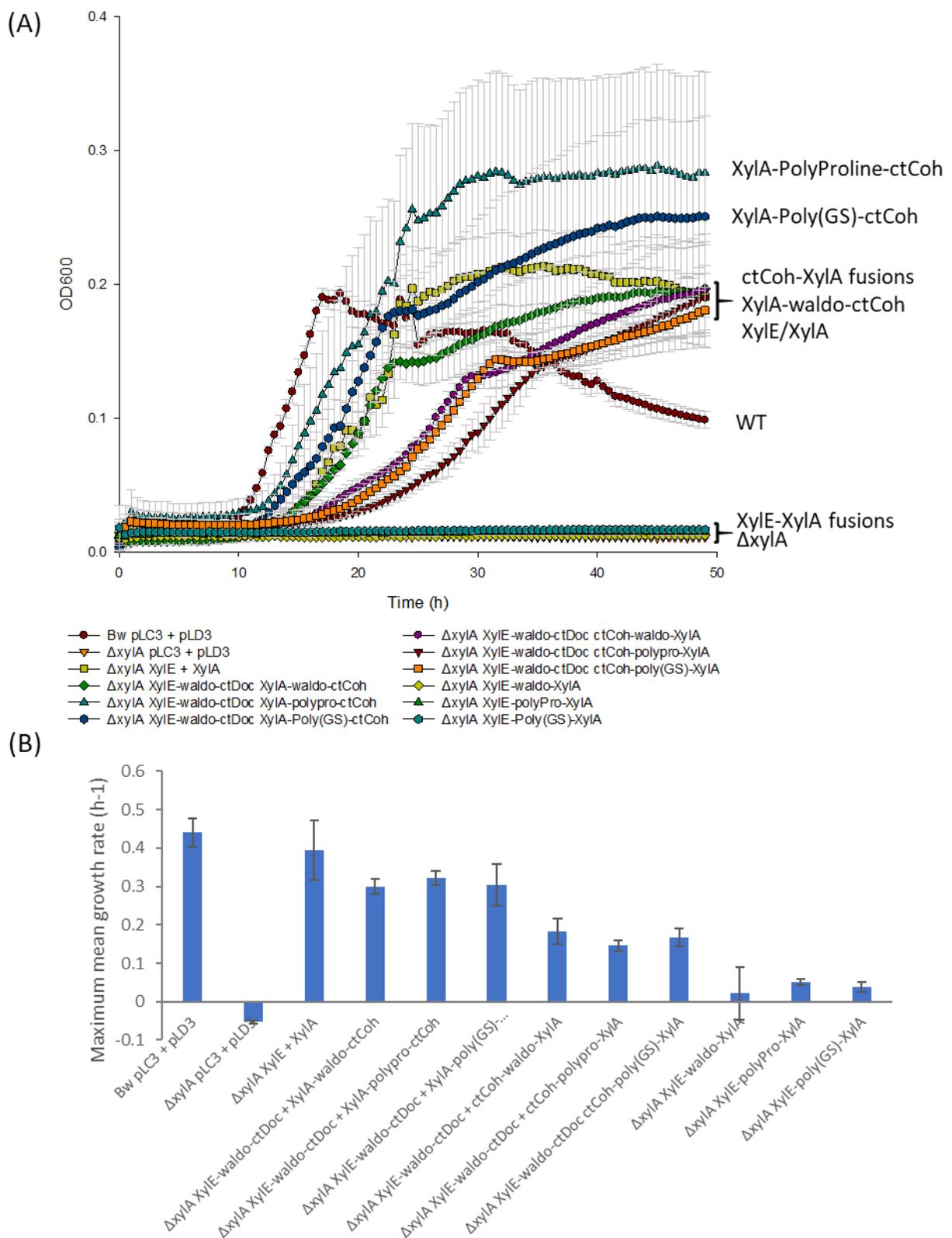


Figure 5.22. The XylA-ctCoh/XylE-ctDoc coexpression shows the best cell growth amongst other scaffold and fusion conditions (A) Growth assay of $\Delta xylA$ cells transformed with either with XylE-ctDoc and XylA-ctCoh fusions, XylE-ctDoc and ctCoh-XylA fusions or

XylE-XylA fusions on a xylose deficiency background. Deletion mutant $\Delta xyIA$ was cotransformed with pLC3 plasmids expressing either XylE-ctDoc or XylE-XylA alongside pLD3 plasmids expressing either native *xyIA*, C-terminal fusion of XylA to ctCoh, or an N-terminal fusion of XylA to ctCoh. Where XylE-XylA fusion was present, empty pLD3 plasmid was present instead. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 μ g/mL kanamycin + 34 μ g/mL chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. N=5. Error is standard deviation. **(B)** Maximum growth rate of $\Delta xyIA$ cells. N=5. Error is standard deviation

h^{-1} , $0.321 h^{-1}$ and $0.304 h^{-1}$ for the XylA-ctCoh fusions containing the Waldo, polyproline and poly(Gly-Ser) linkers respectively (Fig. 5.22.B). While slower than that of the XylE/XylA coexpression, each XylE-ctDoc/XylA-ctCoh co-expressions displayed a p value greater than 0.05, suggesting the variation was not statistically significant. The XylE-ctDoc/ctCoh-XylA co-expressions showed growth rates between $0.145 h^{-1}$ and $0.182 h^{-1}$, each slower than of the XylA-waldo-ctCoh co-expression to statistical significance ($p < 0.05$). Finally, no cells expressing any of the XylE-XylA fusions demonstrated any growth, reinforcing that direct fusion to XylE heavily reduces, or outright inhibits, XylA activity.

Overall, this suggests that the most effective linker was the polyproline linker, a curious result as when XylA-polyproline-ctCoh and ctCoh-polyproline-XylA are expressed in isolation they perform worse than any other fusion. There is some variability, however, as within this second experiment the pattern seen is polyproline > poly(Gly-Ser) > Waldo. The variability may suggest that altering the linkers within the XylE-ctDoc fusion could offer improved access to the XylA-ctCoh.

To conclude this work, we have measured three key types of growth pattern between different methods of substrate channelling, demonstrating high variability between the conditions. Direct fusion between XylE and XylA resulted in no growth of $\Delta xyIA$ cells within this experiment, suggesting loss of XylA activity. Using the ctCoh-XylA within a coexpression results in a slow growth, in line with reduced activity of XylA when N-terminally fused. Finally, the XylA-ctCoh coexpression result in fast growth comparable to the controls. Furthermore, we demonstrate that different linkers can alter the activity of the associated enzymes, reinforcing the importance of testing linkers in protein fusions. Curiously, the impact of the linkers appears to change in strains coexpressing two fusions, which may provide some evidence towards scaffold formation. Of these linkers, the polyproline linker appeared to perform the best for both tests of the XylE-ctDoc and XylA-ctCoh coexpression (Fig. 5.19, 5.22). However ultimately, we have yet to see a clear example of substrate channelling in any of the coexpression samples, as tagged controls often perform equivalent to the “scaffold” condition. This may suggest that scaffolds are not forming, or if they are, xylose utilization is not being substantially increased through substrate channelling. It is also possible this reflects performing a growth-based assay on an organism that is already performing D-xylose metabolism at a high level, with any improvements seen falling within error or being used to produce side products to cell growth. To take this into account, it is therefore important to look more directly at D-xylose consumption.

5.6. Direct measure of D-xylose concentration in initial 13 hours of cell growth demonstrates no benefit with cohesin-dockerin system.

A shortcoming of the growth assays is that cell growth doesn't provide any direct measure of D-xylose consumption. The metabolism of D-xylose by *E. coli* is capable of producing byproducts such as acetic acid, succinate and glyoxylate (Liu *et al.*, 2018). Acetic acid, however, can be toxic to the cell and negatively impact cell growth (Roe *et al.*, 2002). Taking this into account, it is possible that increased substrate channelling through XylE and XylA would result in an increased production of cell growth inhibitors, thereby slowing cell growth. To validate that the growth assays were consistent with xylose catabolism, we therefore looked directly at the concentration of D-xylose throughout the initial 13 hours of cell growth.

To determine whether scaffolds were forming and the impact of substrate channelling on xylose metabolism, 13 conditions were produced in $\Delta xylA$ cells to represent the conditions of previous growth assays. By using a deletion strain it was hoped that the recovery of xylose would be variable between the different conditions, allowing for a better understanding of when xylose metabolism begins and to see any variations.

The samples were selected to contain one of three XylE-ctDoc fusions, containing either linker waldo, flexible or poly(Gly-Ser) and one of three XylA-ctCoh fusions, containing linker waldo, polyproline or poly(Gly-Ser). Each sample was grown in triplicate. These fusions were selected as they were presumed to represent the best, worst and most average conditions of XylE and XylA fusion to ctDoc and ctCoh respectively. To further support our understanding of cell growth, 4 controls were produced expressing: empty plasmids pLC3 and pLD3, cells expressing XylE and XylA, cells expressing XylE and XylA-waldo-ctCoh and cells expressing XylE-waldo-ctDoc and XylA. Each control was grown in duplicate.

The concentration of D-xylose was measured using the Megazyme D-xylose assay kit, which is a colorimetric assay where OD_{340} absorbance correlates to the concentration of xylose. The initial absorbance of xylose level at $t=0$ h was found to be outside instrument limit and therefore could not be recorded, however each subsequent recording was diluted 5-fold to facilitate absorbance within instrument limit, with subsequent calculations of D-xylose concentration adapted for this. Because of this shortcoming, $t=2.5$ h is the first time point recorded. Based on previous growth assays, cells will still be in the early lag phase at 2.5 h, however the concentration of xylose consumed at this timepoint is unlikely to be significant. Therefore, as there is no timepoint for $t=0$ h, $t=2.5$ h is assumed to represent the initial concentration of D-xylose.

The goal of this work was to measure the concentration of D-xylose throughout the late lag phase into the early stationary phase so to determine whether there were any unexpected changes in D-xylose concentration that contrasted to the patterns seen in previous growth assays. Over the course of this experiment, 6 time points were successfully collected: 2.5 h, 4 h, 7 h, 9 h, 11 h and 13 h. The initial concentration of the M9 + 10 mM xylose solution was also measured four times, with the average concentration of D-xylose found to be 1.157 g/L. This was lower than the expected mass of 1.501 g/L which means the concentration of xylose in the M9+X solution was slightly lower than anticipated.

The worst performing strain, other than the negative control which did not consume D-xylose, was found to be the XylE+XylA control. The control demonstrated a final xylose concentration of 0.962 g/L after starting at $t=2.5$ h with a concentration of 1.259 g/L (Fig. 5.23), which meant that just 23.6% of xylose being metabolised over 10.5 hours (Fig. 5.24).

Beyond these two controls, the other strains performed very similar to one another for the first 9 hours and were largely clustered together, only really separating between 11 and 13 hours into two distinct populations: one more rapidly consuming the D-xylose present, consuming around 65% of D-xylose at 13 hours, and the second slightly slower, consuming between 40 to 50% of D-xylose at 13 hours.

Of the cluster of strains rapidly consuming D-xylose, the best performing strain was XylE-waldo-ctDoc/XylA, one of the tagged control strains measured. It outperformed all other controls reaching a final concentration of 0.367 g/L of xylose at 13 hours, having metabolised 69.7% of the available xylose remaining at $t = 2.5$ h. Furthermore, at $t=7$ h, when xylose consumption began to accelerate, this condition showed the highest percentage of xylose consumed at 14.0%, meaning that throughout the experiment, the XylE-waldo-ctDoc/XylA control consumed more D-xylose than the other conditions.

Only 3 strains coexpressing scaffold components performed within error of the XylE-waldo-ctDoc/XylA control. Of these scaffold conditions measured, the XylE-waldo-ctDoc/XylA-waldo-ctCoh was the most like the control and reached the lowest concentration of the three at 0.383 g/L, equivalent to having metabolised 65.8% of the $t = 2.5$ h xylose concentration. The other two, the XylE-flex-ctDoc/XylA-waldo-ctCoh and XylE-poly(Gly-Ser)-ctDoc/XylA-poly(Gly-ser)-ctCoh coexpression strains, metabolised similar levels of D-xylose, at 65.0% and 63.8% respectively.

For the remainder of the strains, they often performed either greater than or equivalent to the second control, XylE/XylA-ctCoh, which had a final concentration of 0.640 g/L, equivalent to 39% of xylose at 2.5 h consumed. Overall, however, the final concentrations of D-xylose in

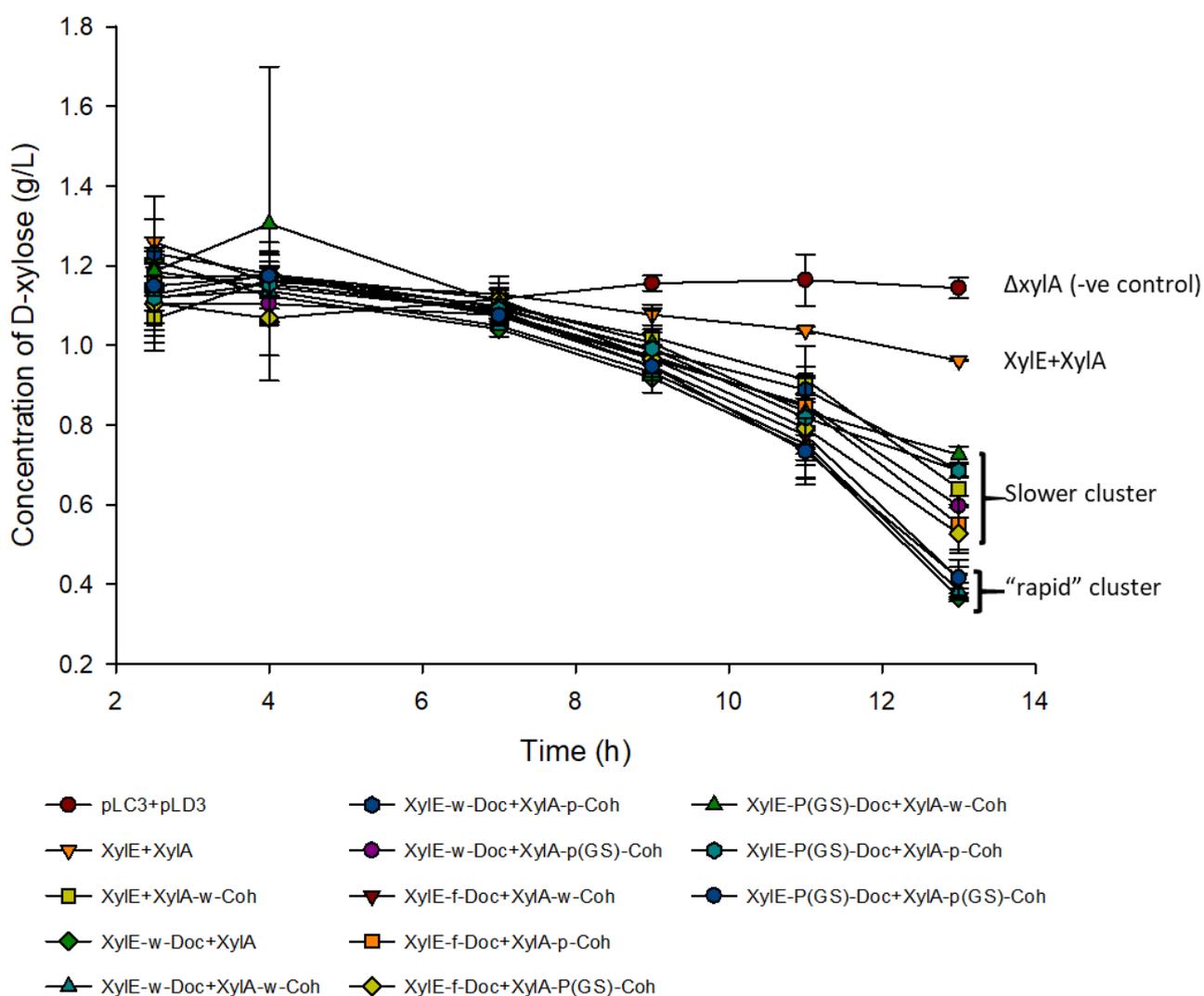


Figure 5.23. Concentration of D-xylose in cell cultures of $\Delta xyIA$ co-expressing XylE-ctDoc and XylA-ctCoh. The concentration of D-xylose in cell culture covering the period between 2.5 hours of growth and 13 hours of growth. Concentration given in g/L as this is the standard units given for sugar metabolism in the field (Huffer *et al.*, 2012; Cunha *et al.*, 2019; Amer *et al.*, 2020). D-xylose concentration was calculated through the use of a colorimetric assay. The cell cultures contained $\Delta xyIA$ cells transformed with either XylE-ctDoc/XylA-ctCoh coexpressions or a control strain featuring native enzyme or empty plasmid. Scaffold expressions were repeated in triplicate, controls in duplicate. Error bars represent standard deviation, triplicate for scaffold conditions, duplicate for control conditions.

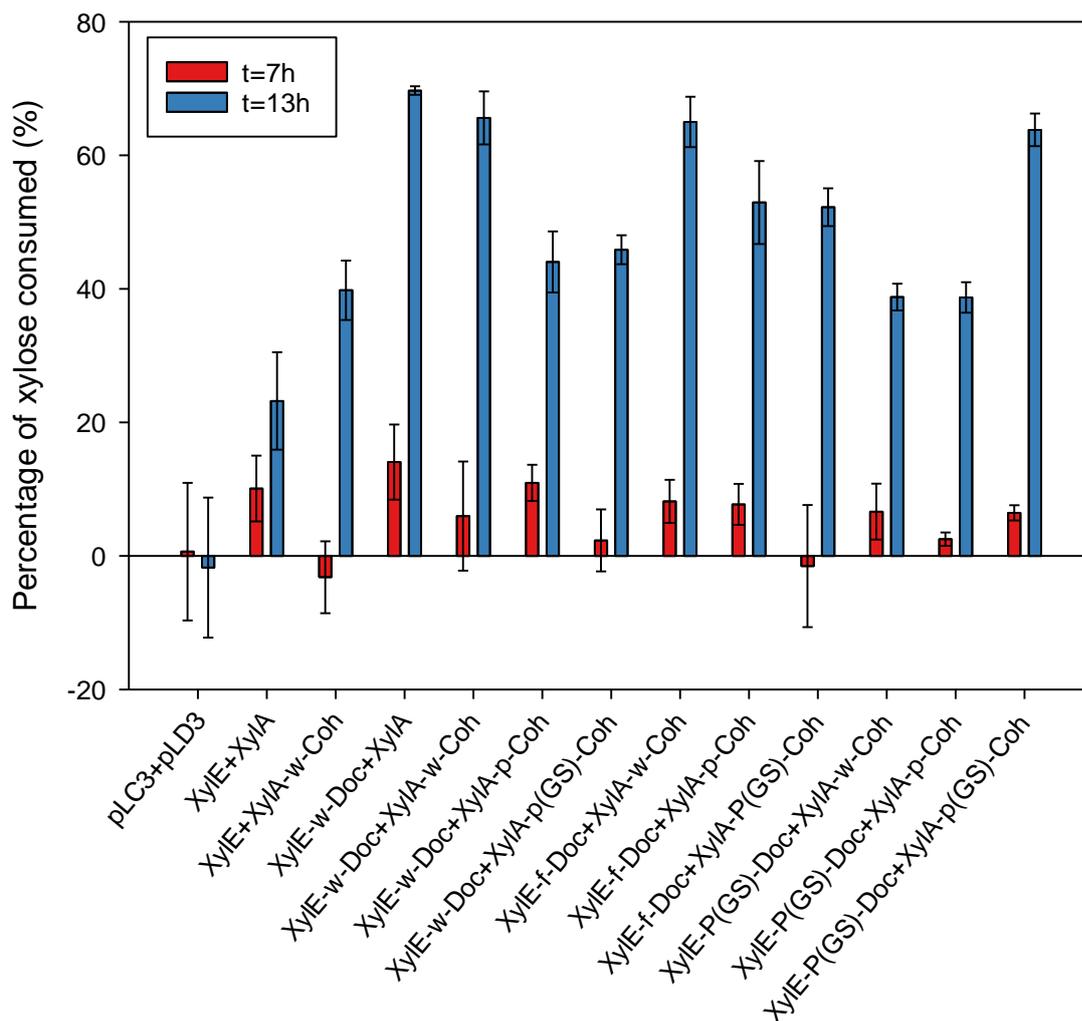


Figure 5.24. The percentage of available D-xylose consumed by scaffold conditions and their controls. The concentration of D-xylose in cell culture covering the period between 2.5 hours of growth and 13 hours of growth is shown in Fig. 5.23. The average percentage of D-xylose available at 2.5 hours was measured and used to determine the percentage of available D-xylose consumed at time points 7 hours (Red) and 13 hours (blue). Error bars represent standard deviation, triplicate for scaffold conditions, duplicate for control conditions.

the coexpression strains appeared to be within error to the control conditions. This meant that either scaffold is not able to assemble, or it has, however no substrate channelling is occurring.

In conclusion, the direct measurement of D-xylose concentration during the initial 13 hours of cell growth did not result in a significant variation compared to what was seen in previous growth assays. Curiously, tagging of XylE or XylA resulted in an increased xylose metabolism for all condition measured compared to the untagged control. As all coexpressions featuring one or more tagged proteins outperformed the untagged XylE/XylA coexpression, this suggests that tagging of the proteins has stabilised the transporter and enzyme. Perhaps a pulse-chase experiment could be used to measure the stability of the tagged proteins compared to the untagged to determine the validity of this hypothesis (Yamaguchi *et al.*, 2009). However, this also meant that while scaffold conditions demonstrated improvements in D-xylose consumption when compared to native XylE/XylA coexpression, there was no evidence that these scaffold conditions were achieving substrate channelling.

5.7. Total Internal Reflection Fluorescence Microscopy (TIRFM) to determine the localization of Dockerin tagged mCherry.

5.7.1. Designing TIRFM appropriate protein fusions.

With validation that the ctCoh-ctDoc interaction was stable via the pulldown assay, the scaffolding system had been applied to a growth and xylose utilization assay to determine if scaffolds would facilitate substrate channelling through improved growth. However, no evidence of substrate channelling was observed. There were two possible explanations for this, either the reaction pathway was not suited to substrate channelling or the scaffold had failed to form at the membrane. Therefore, it was important to demonstrate scaffold formation at the membrane. To confirm this, a TIRFM based procedure was developed to directly localise fluorescently-labelled target proteins in live bacteria.

TIRFM can be used in order to visualise the movement of individual fluorophores attached to biomolecules within a cellular body (Hern *et al.*, 2010; Mattheyses, Simon and Rappoport, 2010; Yamamura, Suzuki and Imaizumi, 2015). As opposed to light microscopy, TIRFM uses an evanescent wave that is unable to permeate the entire cell. The evanescent wave is produced at the interface between the glass coverslip and the specimen, and permeates approximately 200 nm into the cell, the exact degree being tuneable by altering the angle of incidence (Fish, 2009). This optical sectioning of the sample makes it suitable for studying the localization of fluorophores at the membrane.

A particle tracking code was developed by Alex Scott for use in MATLAB, which could follow individual fluorophores, calculating mean squared diffusion (MSD) and therefore diffusion coefficient by comparing the x,y coordinates for a fluorophore in each frame relative to its original position. A second computational filter was then applied to remove potential aggregates, which would be represented as highly fluorescent spots. Aggregates were detected by taking a complete snapshot of all intensities measured for each image's data set and then removing any trajectory with a fluorescence intensity 1.5 interquartile ranges greater or smaller than the upper or lower quartile.

It was vital to confirm the applicability of the code to compare freely diffusing cytoplasmic proteins to membrane bound proteins. Components previously produced to confirm XylE stability included a fusion to the GFP species, GFPmut3B, which had been a part of the CIDAR MoClo collection (Iverson *et al.*, 2016), and were considered suitable to test TIRFM. The two key points of comparison were between the xylulose kinase, XylB, which was selected to represent the soluble fusion (Fig. 5.25.A), and XylE, which was the transmembrane protein of

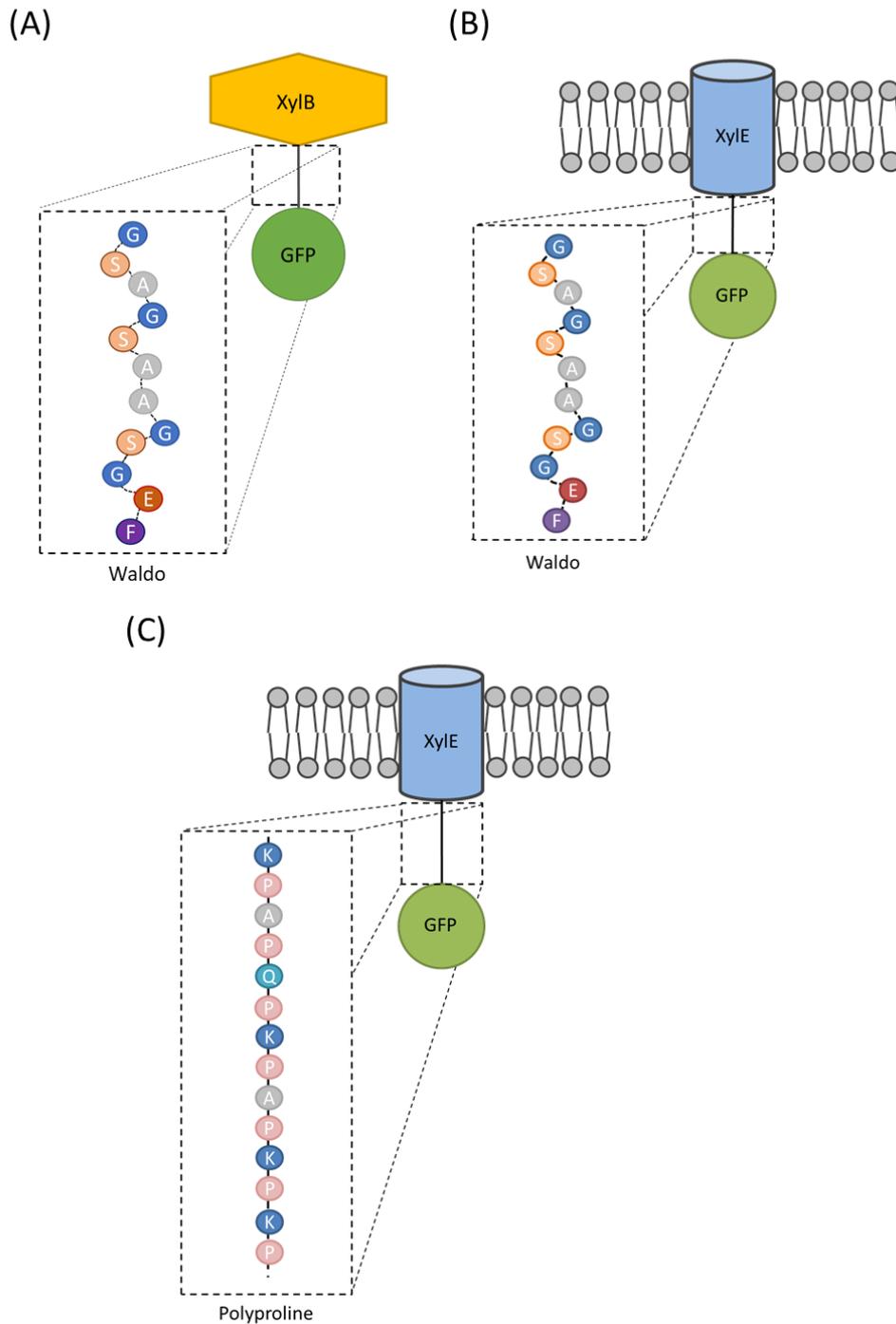


Figure 5.25. Cartoon representing the different fusions used in the initial TIRFM study of protein mobility. In order to validate that there is a distinct difference between the mobility of membrane and cytoplasmic proteins when analysed 3 fusion conditions were selected. **(A)** The fusion of cytoplasmic XylB to GFP achieved through the flexible Waldo linker. **(B)** The fusion of membrane transporter XylE to GFP by the flexible Waldo linker. **(C)** The fusion of membrane transporter XylE to GFP by rigid polyproline linker.

interest. Two XylE-GFP fusions were considered of interest to this work, containing either the flexible Waldo linker (Fig. 5.25.B) or the rigid polyproline linker (Fig. 5.25.C). These two linkers were analysed using TIRFM to determine whether fusion could impact membrane protein mobility. For this initial study proteins were tested in DH5 α , however this was changed in future experiments to BW25113.

The analysis of these particles by TIRFM and Alex Scott's code resulted in diffusion coefficients being calculated for each particle trajectory. These diffusion coefficients were compiled to obtain a box-and-whisker plot (Fig. 5.26), histogram (Fig. 5.27), the median and mean. It was found that the mobility of all conditions measured were closer than anticipated. Both XylE-GFP fusion proteins were slower than that of the XylB-GFP fusion, with means of 0.191 $\mu\text{m}^2/\text{s}$ and 0.194 $\mu\text{m}^2/\text{s}$ for XylE-p-GFP and XylE-w-GFP respectively, compared to 0.239 $\mu\text{m}^2/\text{s}$ for XylB-GFP (Fig. 5.26). In terms of distribution, the XylE-GFP fusions both showed a steeper rightward skew compared to XylB-GFP, whereas the XylB-GFP had a broader spread of diffusion coefficients (Fig. 5.27).

Using TIRFM to measure the diffusion coefficient resulted in values inconsistent with those in the literature determined by FRAP microscopy. The transporter XylE showed a diffusion coefficient of 0.19 $\mu\text{m}^2/\text{s}$ for both linkers (Fig. 5.26). Study of diffusion using FRAP suggested a much lower diffusion coefficient around 0.027 $\mu\text{m}^2/\text{s}$ (Kumar, Mommer and Sourjik, 2010). Furthermore, a soluble protein the size of XylB, 52.63 kDa, would be expected to have a significantly larger mobility, closer to 1-2 $\mu\text{m}^2/\text{s}$, based on FRAP studies (Kumar, Mommer and Sourjik, 2010), however in this work XylB-GFP had a much smaller value at 0.25 $\mu\text{m}^2/\text{s}$.

It is important to note that there are differences between protocol for FRAP completed by Kumar *et al.* and the protocol used for TIRFM. Within Kumar *et al.*'s research a different strain of *E. coli* was used, *E. coli* K-12 strain RP437 $\Delta flhC$ derivative "VS116" (Kumar, Mommer and Sourjik, 2010). The LacY-GFP fusion was also expressed by a high copy number plasmid pSN7 (Gyuris *et al.*, 1993; Wiethaus *et al.*, 2009), as opposed to a low copy number plasmids. Furthermore, Kumar *et al.* grew cells at lower temperatures, 30°C for overnight, 34°C for daytime cultures, on a tryptone broth, inducing at 3 hours into the daytime growth condition. Cells were then resuspended within tethering buffer, as opposed to supplemented M9, and incubated at room temperature for 30 minutes before adding to the poly-L-lysine treated coverslips (Kumar, Mommer and Sourjik, 2010). The combination of low growth rate and high copy number plasmid would have resulted in the increased protein production, as needed for FRAP, however, high protein production, particularly membrane protein production, increases risk of aggregation (Wagner *et al.*, 2007; Gubellini *et al.*, 2011), which would have lowered the recorded diffusion coefficient. Evidence of over-produced target proteins fused to GFP forming

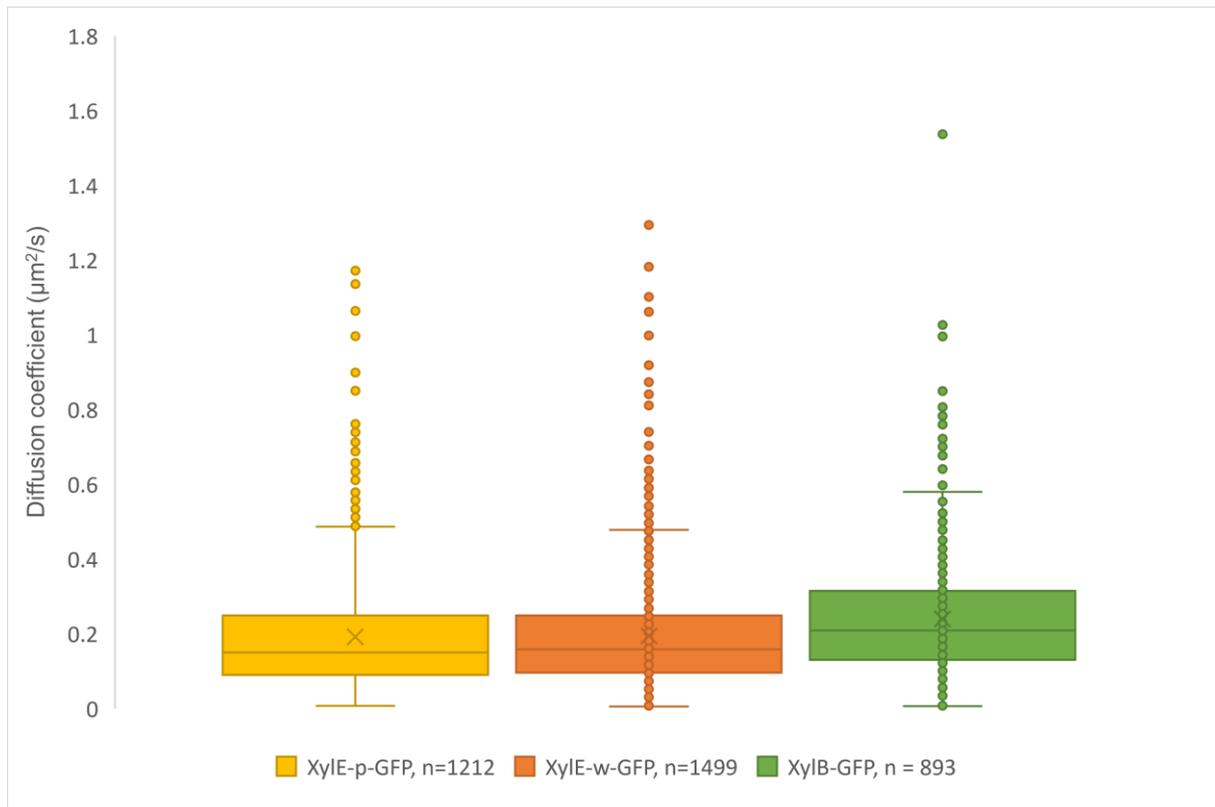


Figure 5.26. Box and whisker plot of Diffusion coefficients of each individual trajectory for the GFP fusions. Box and whisker plots calculated from total diffusion coefficients from each condition expressed in DH5α. XylE-polyproline-GFP (n = 1212, mean = 0.191 µm²/s, median = 0.150 µm²/s, yellow), XylE-waldo-GFP (n = 1499, mean = 0.194 µm²/s, median = 0.158 µm²/s, yellow), XylB-GFP (n = 893, mean = 0.239 µm²/s, median = 0.209 µm²/s). Diffusion coefficient was obtained by linear regression of this first four time intervals as described in the Methods section.

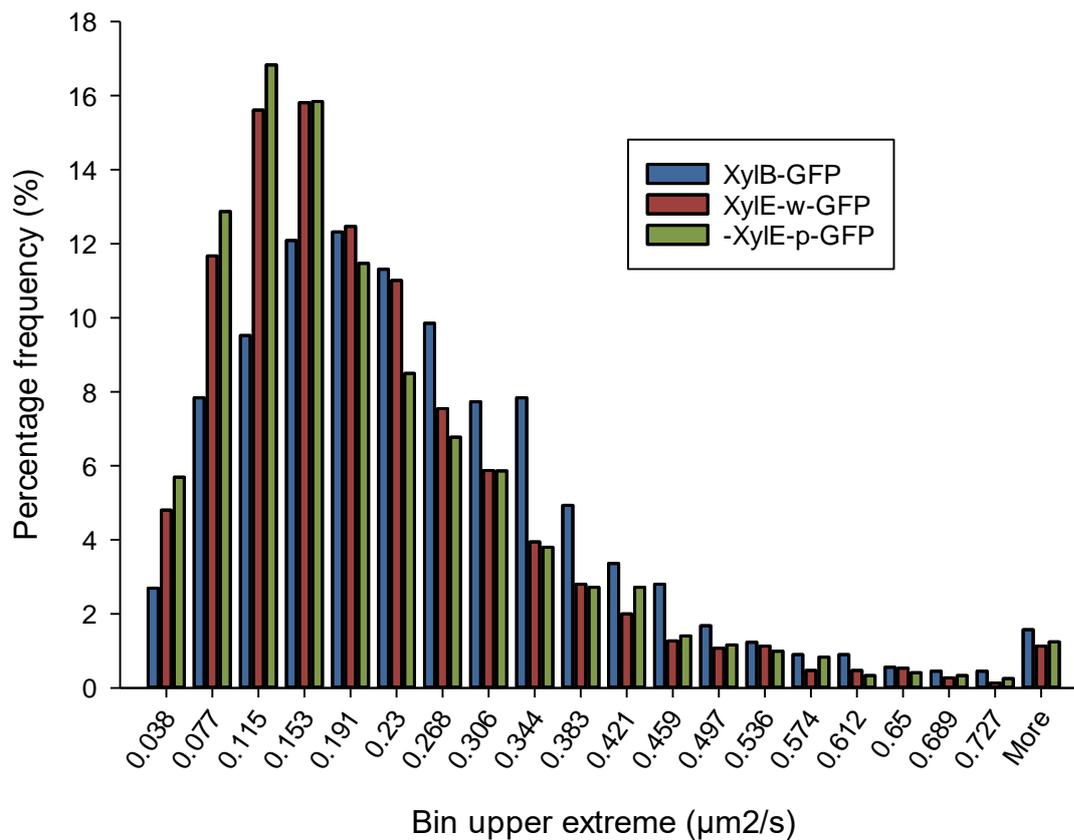


Figure 5.27. Histogram representing the distribution of particles across different diffusion coefficients for GFP fusions expressed in DH5 α . The percentage frequency for different diffusion coefficients for XylB-GFP (n=893, blue), XylE-waldo-GFP (n=1499, red) and XylE-polyproline-GFP (n=1212, green). Bin distance of 0.038 calculated from XylB-GFP using the Freedman-Diaconis rule (Freedman and Diaconis, 1981) – a total of 40 bins were suggested to reach maximum diffusion coefficient of 1.54 $\mu\text{m}^2/\text{s}$, however as many of the final bins were empty only the first 19 bins are shown here.

large immobile aggregates was obtained using TIRFM (data not shown), hence the use of low-copy plasmids in this work.

Therefore, a stronger point of comparison would be diffusion coefficients obtained through single particle tracking techniques (Table 5.5). Looking more directly into single particle tracking of inner membrane proteins, the diffusion coefficient of Xyle-mCherry appears slightly higher than anticipated, although there are potential explanations for this result. The relationship between protein size and mobility can be clearly seen when observing the integral membrane protein TatA, a 10.7 KDa protein part of the twin-arginine translocation pathway, and exists as part of a complex of around 25 TatA monomers, however the exact amount can vary (Gohlke *et al.*, 2005). A linear relationship was found between the number of TatA monomers in a complex and the diffusion coefficient when expressed, ranging from greater than $0.1 \mu\text{m}^2/\text{s}$, when 10 or fewer TatA proteins were present, to less than $0.01 \mu\text{m}^2/\text{s}$ when more than a 100 were present (Leake *et al.*, 2008). A second small membrane protein, Tsr, can give insight into the upper rates of diffusion. Tsr is a membrane localized serine chemoreceptor protein, which forms heterotrimeric complexes that cluster at the poles of the cell. Within *E. coli* K-12 BW36931 around 7% of Tsr freely diffuses in the membrane. To understand the specifics of diffusion, Tsr was fused to Venus, a YFP variant, and single molecule tracking was achieved by oblique angle fluorescence microscopy. Initial microscopy suggested that the polar Tsr-Venus fusion had a diffusion coefficient of $0.012 \mu\text{m}^2/\text{s}$, whereas freely diffusing Tsr-Venus had a diffusion coefficient of $0.40 \mu\text{m}^2/\text{s}$ (Oh *et al.*, 2014). However, at 1000 Hz frame rates, the polar Tsr proteins appear to have a higher diffusion coefficient, equivalent to diffusion in 3D, and are simply limited in the distance they can diffuse. Ultimately, this demonstrates that smaller membrane proteins appear to diffuse more rapidly than the Xyle-GFP measured – supporting the diffusion coefficient obtained.

A more comparable protein to Xyle may be the 12 TMH proton-drug antiporter AcrB, belonging to the resistance-nodulation-cell division-type efflux system. AcrB-GFP was expressed in a BW25113 derivative, and demonstrated a lower diffusion coefficient than that of Xyle-GFP, $D = 0.035 \pm 0.018 \mu\text{m}^2/\text{s}$, which was further lowered in presence of an outer membrane protein TolC. AcrB, however, is known to form a stable trimer and hypothesized to associate with the membrane fusion protein AcrA, therefore creating a larger complex which is likely to lower the rate of diffusion (Yamamoto *et al.*, 2016). A homologue of Xyle, the MFS protein GlpT, was also measured with single particle tracking (Oswald *et al.*, 2016). *E. coli* strain MC4100 was observed under a wide-field fluorescence microscope, with the individual particles tracked and their diffusion in 3D was extrapolated using IPODD (Oswald *et al.*, 2014), meaning that the diffusion coefficient was obtained in a different manner compared to the 2 dimensional calculations for this work. However, the proposed diffusion coefficient of GlpT, $0.153 \mu\text{m}^2/\text{s}$

Protein/complex	Diffusion coefficient ($\mu\text{m}^2/\text{s}$)	Reference
XylE-Waldo-GFP	0.194	This work
XylE-polyProline-GFP	0.191	This work
TatA-YFP	0.1-0.01	(Leake <i>et al.</i> , 2008)
Tsr-Venus	0.40	(Oh <i>et al.</i> , 2014)
AcrB	0.035	(Yamamoto <i>et al.</i> , 2016)
GlpT-eGFP	0.153	(Oswald <i>et al.</i> , 2016)
YedZ-eGFP	0.188	(Oswald <i>et al.</i> , 2016)
CybB-eGFP	0.175	(Oswald <i>et al.</i> , 2016)
CtsA-eGFP	0.131	(Oswald <i>et al.</i> , 2016)
PleC-eYFP	0.012	(Deich <i>et al.</i> , 2004)
GFP-MotB	0.008	(Leake <i>et al.</i> , 2006)
XylB-GFP	0.239	This work
Kaede	7.3	(Bakshi, Bratton and Weisshaar, 2011)
GFP	5.6	(Lill <i>et al.</i> , 2012)
SecA-Ypet	2.09	(Seinen <i>et al.</i> , 2021)
Atto67-CheY	1.95	(Di Paolo <i>et al.</i> , 2016)
LacI-Venus	3.0	(Elf, Li and Xie, 2007)
30S ribosome subunit (S2-YFP)	0.20-0.13	(Mohapatra and Weisshaar, 2018)

Table 5.5. The diffusion coefficient of inner membrane proteins and cytoplasmic proteins. A table displaying the diffusion coefficient, in $\mu\text{m}^2/\text{s}$, found in literature of both inner membrane proteins (upper table) and cytoplasmic proteins (lower table). XylE-GFP and XylB-GFP fusions produced within this work are also listed. References shown in margin.

(Oswald *et al.*, 2016), is close to the value of D measured for XylE-GFP fusions, supporting the result obtained.

However, the diffusion coefficient seen for the XylB-GFP is still much lower than values seen for single molecule tracking of other cytoplasmic protein. Individual fluorophores demonstrate high diffusion coefficients in the cytoplasm, as large as $7.3 \pm 1.1 \mu\text{m}^2/\text{s}$ for Kaede, or $5.6 \mu\text{m}^2/\text{s}$ for GFP (Bakshi, Bratton and Weisshaar, 2011; Lill *et al.*, 2012). While this value reduces when the fluorophore is fused to another cytoplasmic protein, it is still much larger than that of membrane proteins. For example the protein SecA was found to have a diffusion coefficient of $2.09 \mu\text{m}^2/\text{s}$ when not associated with the SecYEG complex (Seinen *et al.*, 2021). Other proteins, such as Atto67-treated CheY or LacI-Venus lacking its DNA binding demonstrated similar diffusion coefficients at $1.95 \mu\text{m}^2/\text{s}$ and $3 \mu\text{m}^2/\text{s}$ respectively (Elf, Li and Xie, 2007; Di Paolo *et al.*, 2016). These values are consistently higher than the $0.239 \mu\text{m}^2/\text{s}$ measured for XylB-GFP. The cytoplasmic, free moving 30S ribosomal subunit showed a comparable diffusion coefficient to XylB-GFP, in which $D \leq 0.2 \mu\text{m}^2/\text{s}$ (Mohapatra and Weisshaar, 2018), however it is important to note that this 850 KDa complex is composed of multiple proteins alongside RNA (Schluenzen *et al.*, 2000) and is therefore much larger than the homodimeric XylB, which in its oligomeric state is only 106 KDa (Di Luccio *et al.*, 2007). Therefore, the diffusion coefficient value obtained for XylB-GFP is anomalous when compared to other cytosolic proteins and could indicate some type of confinement in the cytosol.

This data overall suggests that the current calculations for single particle tracking in TIRFM are suitable for comparing membrane and cytoplasmic proteins to one another, however for obtaining the exact diffusion coefficient, further refinement is required. In this work a cut off of 15 frames was used, meaning that trajectories were a minimum of 0.225 s in length. Perhaps a higher stringency, such as 0.5 s, would have resulted in a clearer delineation.

One consideration that may explain the lower than anticipated diffusion coefficient for cytoplasmic proteins is the nature of the mobility. Membrane bound proteins are largely limited to 2D movement throughout the membrane, whereas the cytoplasmic protein can also move along the Z-axis. It is possible that the lower than anticipated diffusion coefficient is due to the inability to measure displacement in the Z-axis.

Ultimately, within this work we can see a small, yet distinct variation between the cytoplasmic XylB-GFP and membrane localized XylE-GFP fusions. While some overlap takes place, both the mean and median values for the soluble protein is noticeably higher than that of the membrane proteins. Furthermore, both membrane conditions appear to show high overlap with one another, both demonstrating that the linker did not impact the fusion assembly and that membrane proteins appear highly consistent.

5.7.2. Particle tracking of scaffold components did not demonstrate mobility that was comparable to membrane transporters.

Single particle tracking of TIRFM video data has enabled a difference between the mobility of cytoplasmic and membrane bound proteins to be observed, thus the methodology was then applied to compare the scaffold assembly to cytoplasmic and membrane-tagged conditions. During this time, it was decided to transition to a different fluorophore compared to GFPmut3B due to handful of associated issues. GFPmut3B proved to be subject to both high levels of background noise and rapid photobleaching, which made data analysis a challenge.

To overcome issues associated with GFPmut3B, the fluorophore mCherry was selected, which showed an improved fluorescence signal-to-noise ratio under TIRFM imaging conditions when compared to GFP. Two potential variants of mCherry were considered: an unmodified mCherry (Shaner *et al.*, 2004) or photoactivatable mCherry (PAmCherry) (Subach *et al.*, 2009).

The PAmCherry variant was of interest because it had been confirmed to be less susceptible to photobleaching compared with GFP, and furthermore, the fluorophore was in a dark non-emitting state until photoactivated – photoactivation was achieved using light selected with an excitation filter of 390/40 nm (Subach *et al.*, 2009). In addition, it was hypothesized that the low level of fluorescence prior to photoactivation may have also been suitable for TIRFM and could allow individual particles to be more easily detected and tracked.

To determine the feasibility of these two fluorophores both mCherry forms were fused to the C-terminus of XylE. Cells were then transformed with a vector expressing either a *xylE-mCherry* or *xylE-PAmCherry* fusion, induced with 0.01% (w/v) arabinose and observed under an epifluorescence microscope. It was subsequently found that PAmCherry did not fluoresce, even after photoactivation, whereas the mCherry fusion was clearly detectable (Fig 5.28). These results were then confirmed by an anti-mCherry western blot, that showed a noticeable decrease in expression of PAmCherry compared to mCherry when part of identical fusions and expressed using the same plasmid (Fig. 5.29).

For this reason, unmodified mCherry was chosen to measure scaffold assembly. Three conditions were therefore prepared. The first condition was termed “membrane,” in which BW25113 was transformed with both pLD1_31112 and empty vector pLC3. This meant cells would express only the XylE-mCherry fusion (Fig. 5.30.A). The second condition was the “soluble” condition, in which BW25113 cells were transformed with pLD1_31712 (Table 5.6) alongside empty vector pLC3. This meant cells would only express the ctCoh-mCherry fusion, which would be able to freely diffuse in the cytoplasm (Fig. 5.30.B). Finally, the “scaffold”

Plasmid identification number	Promotor	RBS	MFS	Linker	Downstream protein
pLD1_31112	I13453 (arabinose inducible)	B0034m	XylE	Waldo	mCherry
pLD1_3111B	I13453 (arabinose inducible)	B0034m	XylE	Waldo	Photoactivatable mCherry
pLD1_3171B	I13453 (arabinose inducible)	B0034m	ctCoh (Clostridium thermocellum cohesin)	Waldo	Photoactivatable mCherry

Table 5.6. The components of novel fusions developed for TIRFM. A table displaying the components making up the four novel fusion constructs produced for use in TIRFM. The plasmid used as an expression vector was pLD1, which can contain up to 5 components.

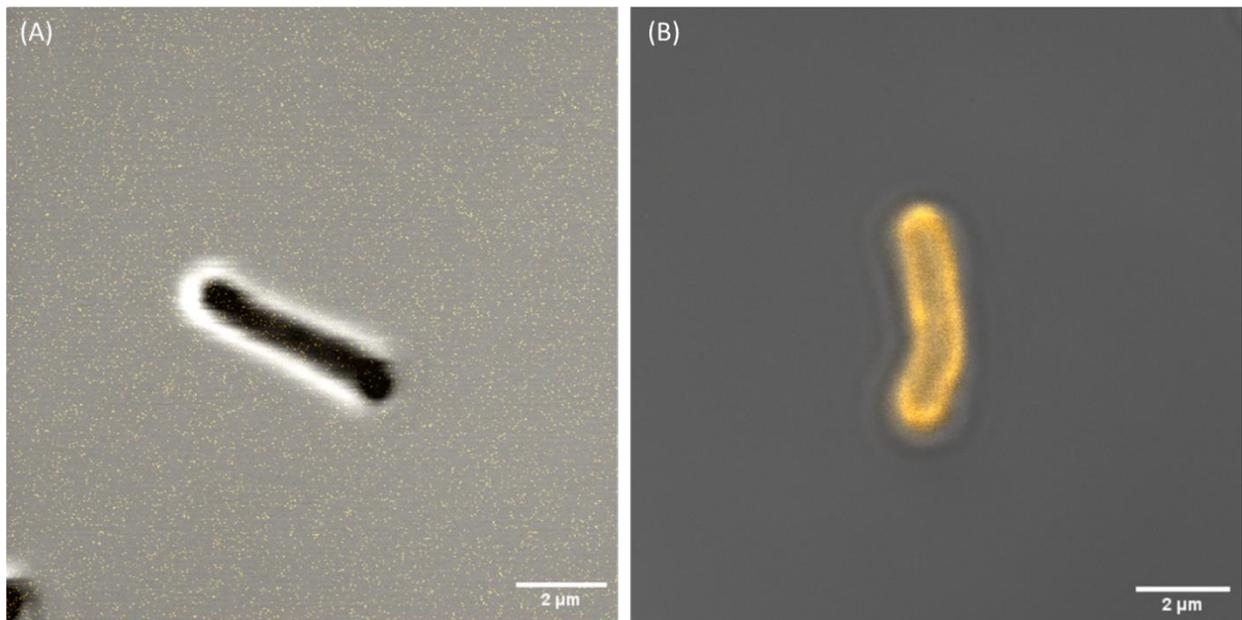


Figure 5.28. Epifluorescence microscopy of the BW25113 cells expressing XylE-mCherry fusions. BW25113 cells expressing XylE-mCherry fusions were observed by a confocal microscope. Fluorescence and phase contrast images were taken for each cell and a composite made for the image. Cells were induced with 0.01% (w/v) arabinose at 1 hour of growth and measured after 6 hours. The mCherry was excited using a 530/43 filter **(A)** Composite image from phase-contrast microscopy and epifluorescence microscopy of BW25113 cells expressing XylE-PAmCherry following photoactivation with a 399 nm laser line. **(B)** Composite image from phase-contrast microscopy and epifluorescence microscopy of BW25113 cells expressing XylE-mCherry.

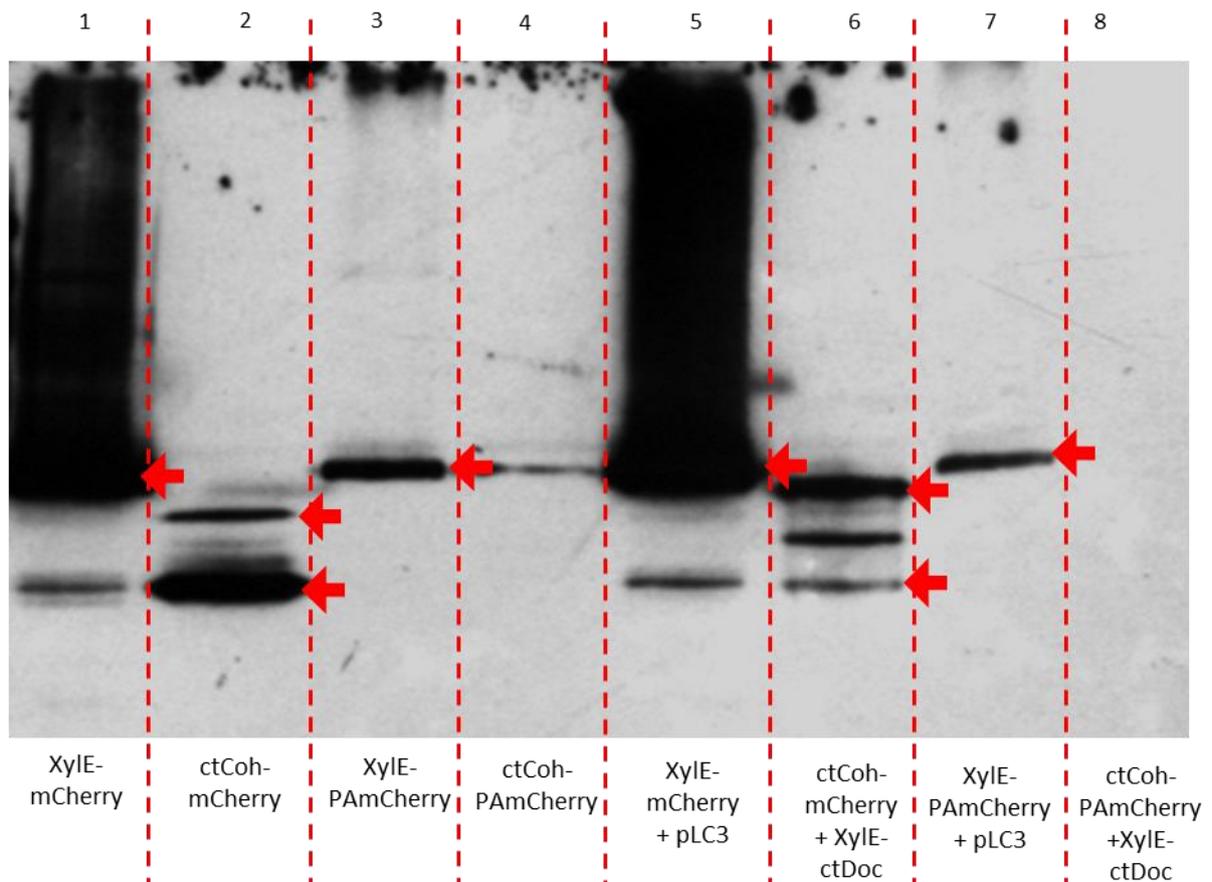


Figure 5.29. Anti-mCherry western blot of the mCherry and PAmCherry fusions. Primary antibody was mouse anti-mCherry, secondary antibody was Anti-mouse. Blocked using a combination of 10% (w/v) skimmed milk and 5% (w/v) BSA. Total cell lysate prepared using BugBuster solution. Columns 1-4 expressed by BW25113 transformed with only one plasmid, pLD1 expressing the fluorophore fusion. Columns 5-8 expressed by BW25113 transformed with two plasmids, either empty pLD1 expressing the fluorophore fusion and either empty pLC3 or pLC3 expressing XylE-ctDoc. Column 4 is expected to contain some bleed from column 5, explaining the high molecular weight blot comparable to XylE-PAmCherry.

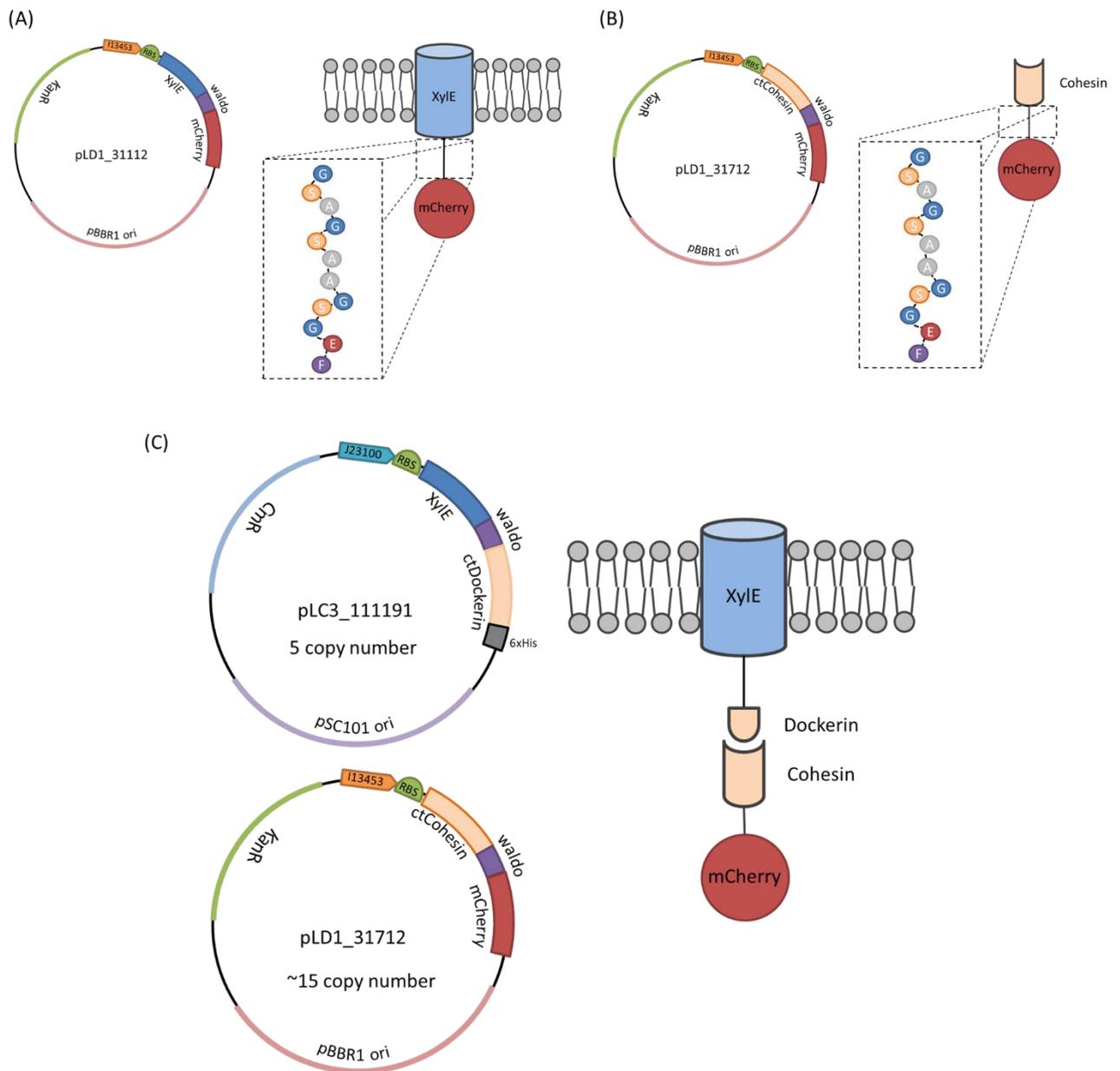


Figure 5.30. Cartoon representing the different fusions used to study scaffold mobility via TIRFM. In **(A)** The fusion XylE to mCherry, linked by the waldo linker. The protein is expressed from the pLD1_31112 vector **(B)** The fusion of ctCoh to mCherry, linked by the waldo linker. The protein is expressed from the pLD1_31712 vector. **(C)** The fusion of membrane transporter XylE to ctDoc and ctCoh to mCherry. Both fusions contain the Waldo linker. The XylE-ctDoc fusion is expressed on the pLC3_111191 vector, the ctCoh-mCherry fusion is expressed on the pLD1_31712 vector.

condition was produced so that BW25113 was transformed with both pLC3_111191 and pLD1_31712. This meant that cells would express both the XylE-ctDoc and ctCoh-mCherry fusions (Fig. 5.30.C). Under ideal conditions, these two proteins would then interact and the mCherry mobility would be more alike the “membrane” condition rather than the “soluble” condition.

Using TIRFM we found that the scaffold condition results in mCherry mobility being consistent with that seen for the cytoplasmic ctCoh-mCherry, as opposed to the membrane-localized XylE-mCherry condition. Under the conditions studied this resulted in a mean of $0.249 \mu\text{m}^2/\text{s}$ for the scaffold condition, identical to the cytoplasmic mean of $0.249 \mu\text{m}^2/\text{s}$ to 3 significant figures. Conversely, the membrane localized, XylE-mCherry fusion resulted in a diffusion coefficient of $0.200 \mu\text{m}^2/\text{s}$. While slightly higher than the rates seen for the GFP fusions, the $>0.04 \mu\text{m}^2/\text{s}$ difference in mean between soluble and membrane proteins has been retained (Fig. 5.31)

The histogram displaying the distribution of diffusion coefficients for the three conditions further reinforces the difference between the membrane condition and scaffold condition (Fig. 5.32). The membrane condition showed a far more significant right skew when compared to the scaffold and cytoplasmic conditions. Furthermore, while all samples showed a centre at the 0.103 bin for all samples, the membrane condition had a much steeper peak, whereas the other two conditions were much shallower.

The overlap of soluble and scaffold conditions suggests that the ctCoh-mCherry did not form a strong interaction with the XylE-ctDoc. There are several potential explanations for this output. The most likely explanation is that the scaffold is unable to form due to issues relating to access. From *in vitro* studies we have shown that the ctCoh-ctDoc interaction is capable of withstanding washing with 100 mL of KPI wash solution, therefore if the interaction was possible it would have been retained. Taking this into account, we did not see it, one could argue different linkers could have been a better option.

A second consideration is that the copy number of the pLD1 plasmid was higher than that of the pLC3 plasmid. This in combination with the induction of 0.01% (w/v) arabinose may have resulted in over-production of the ctCoh-mCherry. Therefore, any examples of the membrane localization that occurred in the scaffold condition could be explained by this imbalance. However, as a counter point to this argument, one would anticipate some shift towards the membrane conditions suggesting a mixed population, as opposed to the almost identical conditions between soluble and scaffold. Unfortunately, initial attempts to measure trajectories in the uninduced cells proved challenging.

Consistent with previous growth and xylose assays, this methodology has failed to demonstrate that scaffold formation has occurred. This reinforces the possibility that scaffold formation is either not as simple or not as stable as initially assumed. The cohesin-dockerin scaffold is either unable to form at the membrane in its current state, or any scaffolds formed are more transient than the expected low K_D would suggest. Going forward it would be important to look at more conditions to better establish the difference between soluble and membrane mobility, as well as to determine if by changing the variables, scaffold formation can be produced. A different orientation in the fusion could be tested, alternative scaffold components to the cohesin-dockerin system could be used, a larger library of soluble and MFS proteins could be tested for comparison, or further filtering of the data to better differentiate between soluble and membrane trajectories.

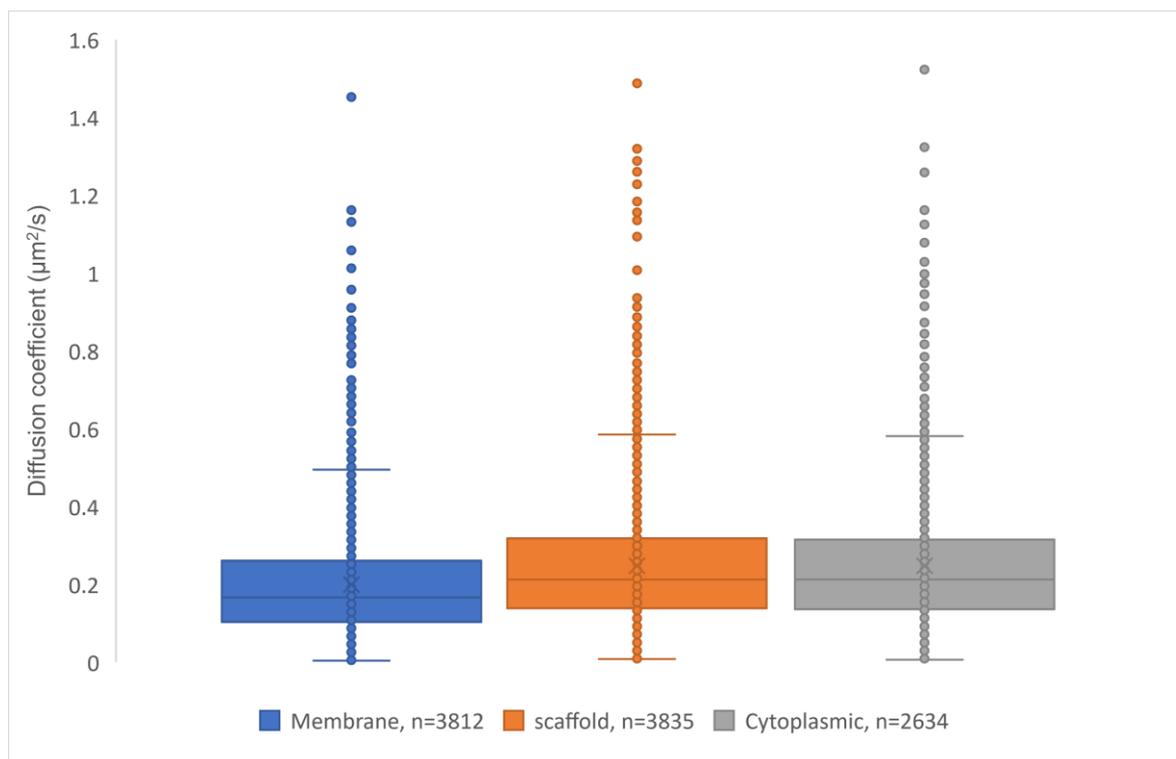


Figure 5.31. A box and whisker plot of diffusion coefficient of the conditions used to study scaffold protein mobility. The diffusion coefficient for each trajectory was calculated for the three conditions expressed in BW25113, “membrane” (n = 3812, mean = 0.200 $\mu\text{m}^2/\text{s}$, median = 0.167 $\mu\text{m}^2/\text{s}$, blue), “scaffold” (n = 3835, mean = 0.248 $\mu\text{m}^2/\text{s}$, median = 0.213 $\mu\text{m}^2/\text{s}$, orange) and “cytoplasmic” (n = 2634, mean = 0.248 $\mu\text{m}^2/\text{s}$, median = 0.213 $\mu\text{m}^2/\text{s}$, grey). Condition “membrane” represents the sole expression of XylE-mCherry, condition “cytoplasmic” represents the sole expression of ctCoh-mCherry and condition “scaffold” represents the coexpression of XylE-ctDoc and ctCoh-mCherry. Diffusion coefficient was obtained by linear regression of this first four time intervals as described in the Methods section.

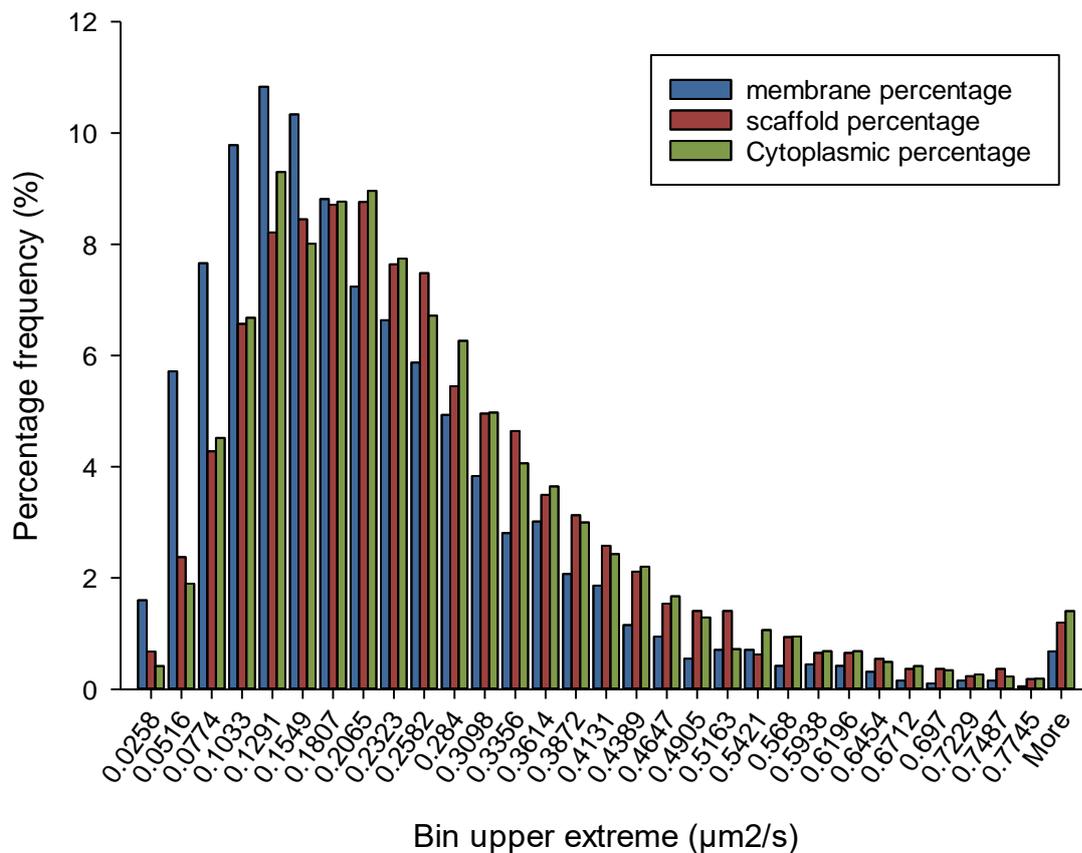


Figure 5.32. Histogram representing the distribution of particles across different diffusion coefficients for mCherry fusions in BW25113. The percentage frequency for different diffusion coefficients for tracked membrane (n = 3812, blue), scaffold (n = 3835, red) and cytoplasmic (n = 2634, green) proteins. Bin distance of 0.0258 calculated from the Cytoplasmic condition using the Freedman-Diaconis rule (Freedman and Diaconis, 1981) – a total of 59 bins were suggested to reach maximum diffusion coefficient of 1.52 μm²/s, however as many of the final bins were empty only the first 30 bins are shown here.

5.8. Summary of Chapter

The goal of this work was to increase the efficiency of the xylose utilization pathway by increasing substrate channelling through the first two steps: transport of D-xylose into the cell through the plasma membrane and isomerisation of the D-xylose to D-xylulose. As direct fusion between transporter XylE and xylose isomerase XylA demonstrated a loss in XylA activity, we therefore looked at using protein scaffolds to achieve localization without negatively impacting the activity of the enzymes.

To produce a protein scaffold, a novel plasmid family was needed that was low copy number and capable of co-expression alongside pLC series plasmids. The plasmids pLD1 and pLD3 were therefore produced and a series of XylA-ctCoh fusions were cloned within them, while XylE-ctDoc fusions were cloned within the plasmid pLC3.

The functionality of the cohesin-dockerin system was then confirmed using an *in vitro* pull down assay, which demonstrated that regardless of whether mCherry was fused to the C- or N-terminal of ctCoh, it was capable of interacting with ctDoc. Therefore, a series of growth assays and a direct xylose concentration assay was used to determine whether the scaffold had formed – hoping that scaffold formation would result in increased substrate channelling, improving cell growth and increasing the amount of D-xylose metabolised by cells.

In practise, however, we failed to record a significant increase in cell growth or xylose consumption compared to control conditions. To determine if scaffold formation was occurring *in vivo* we developed a TIRFM system to track particles and compare diffusion coefficient between fused and “scaffolded” proteins to determine if mobility was equivalent. Ultimately, we could not find any similarity between the mobility of membrane proteins and the scaffold components. It was possible the ctCoh-mCherry fusion had been present in a disproportionate amount, however some small shift would still have been expected.

Ultimately, as we could not find evidence of scaffold formation, we cannot conclude as to whether scaffold formation can increase xylose utilization. This work does demonstrate the challenges associated with scaffold formation introducing the scaffold components to the membrane appears to have prevented the cohesin-dockerin scaffold forming. Going forward it would be important to produce a greater variety of scaffold structures if this is a fault of the cohesin-dockerin system, such as applying the coiled-coil system, and to attempt fusing scaffold components the N-terminal domain of XylE.

Chapter 6

Discussion - Evaluating the role of scaffolds and protein fusions in substrate channelling.

6. Discussion - Evaluating the role of scaffolds and protein fusions in substrate channelling.

6.1. Substrate channelling – a multifaceted challenge

Within this work substrate channelling has been raised as a potential method to enhance productivity of enzyme reactions, with examples of increased metabolic flux readily found in the literature, from both naturally occurring protein fusions and scaffolds (Schweizer and Hofmann, 2004; Fontes and Gilbert, 2010; Willson, Chapman and Thomas, 2019) or synthetically introduced systems (Dueber *et al.*, 2009; Lewicka *et al.*, 2014; You and Zhang, 2014; Kim *et al.*, 2015; Aalbers and Fraaije, 2019). By increasing the local concentration of a rate limiting enzyme, it is possible to reduce the impact of that rate limiting step (Chen and Silver, 2012). This is because the intermediate is unable to enter equilibrium with the bulk cell solution (Spivey and Ovádi, 1999) (Fig. 6.1).

However, if substrate channelling has the capacity to enhance substrate yields, why is it not more prevalent in nature? Why are more examples of protein fusions or scaffold assemblies not observed in naturally occurring pathways? Wouldn't it be advantageous to the cell to make these systems ubiquitous? These were questions that were considered when planning out methods to enhance product yields. If, for example, D-xylose utilization could be improved by direct fusion between pathway enzymes, then why hadn't this process occurred naturally?

Over the course of this thesis, the challenges associated with producing a functional scaffold or fusion has offered some context as to why methods for enhancing substrate channelling are not ubiquitous in reaction pathways. It was found that attempts to increase substrate channelling through targeted fusion or non-covalent interactions, even when using well established cloning techniques, is prone to error, inhibition, and unforeseen challenges.

A major novel aspect of this work was the use of membrane transporters, which introduced additional challenges relating to copy number and localization. Channelling at the membrane can reduce sequestering of side reactions, similar to the positive effects of substrate channelling within the cytoplasm (Thomik *et al.*, 2017; Behrendorff, Borràs-Gas and Pribil, 2020), however in addition immediate catalysis of the substrates can prevent backflow and increase transporter efficiency. Antiporters which export product while importing the substrate can be enhanced even more rapidly, as product is rapidly produced to be exported, increasing transporter efficiency and potentially increasing substrate channelling further (Morales and Reithmeier, 2012). Furthermore, a number of enzymes, such as the cytochrome P450, require some associated with the membrane for function, which creates an opportunity to associate downstream soluble enzymes at the membrane to enhance metabolic flux

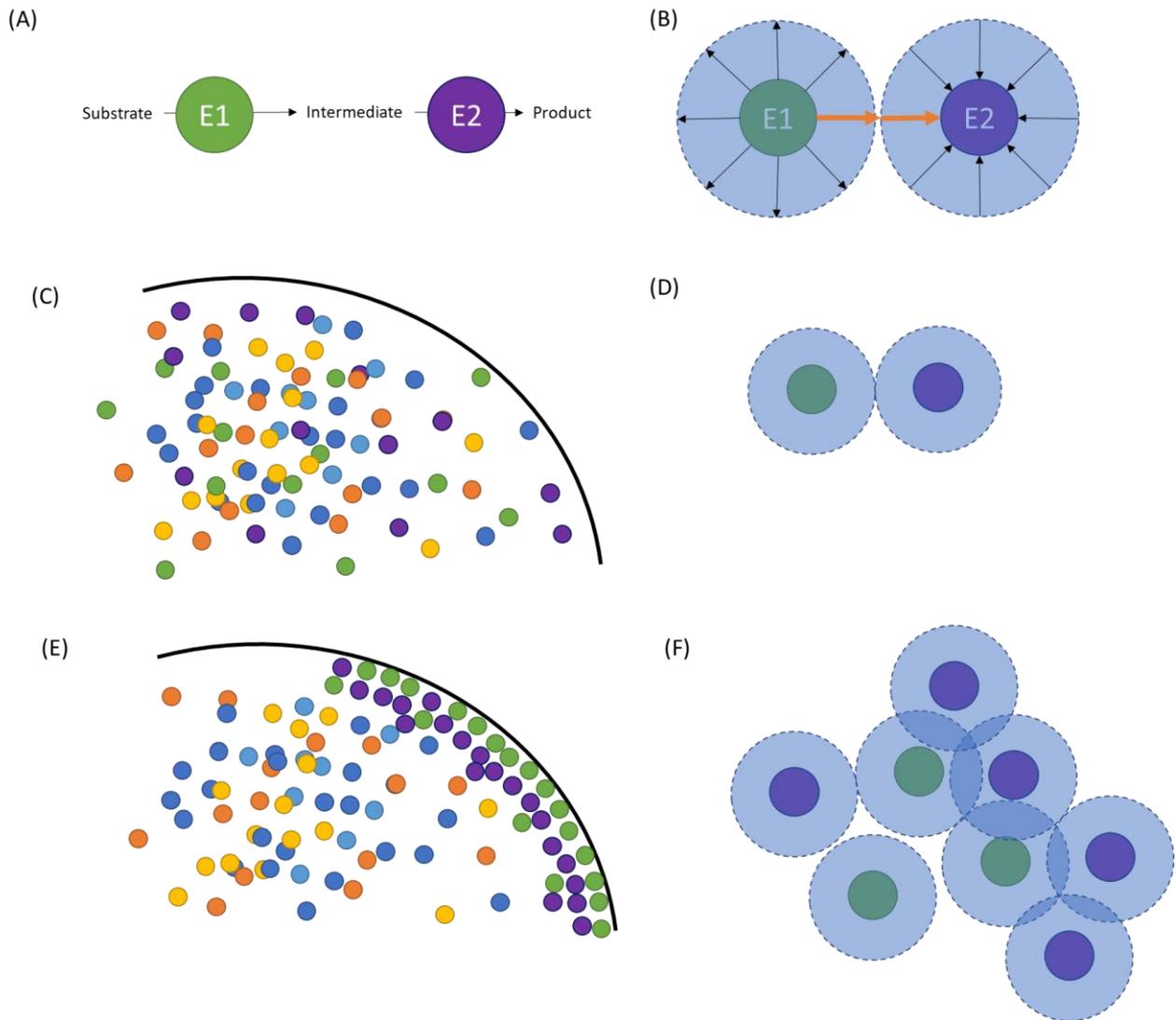


Figure 6.1. A cartoon representing changes to local enzyme concentration in presence of scaffold/fusion. (A) A hypothetical two step reaction in which Enzyme 1 converts substrate to intermediate, and Enzyme 2 converts intermediate to product. **(B)** Direction where intermediate exits Enzyme 1 and enters Enzyme 2 – where overlap takes place substrate channelling occurs. Without overlap substrate will enter equilibrium with bulk cell solution and diffuse naturally to the next enzyme. **(C)** Distribution of enzymes in absence of scaffold/fusion – low local enzyme concentration **(D)** low local enzyme concentration limits substrate channelling **(E)** Distribution of enzymes in presence of scaffold/fusion – at membrane to show high local enzyme concentration relative to one another. **(F)** High local enzyme concentration means that substrate channelling can occur either due to direct overlap or as the enzymes are surrounded by downstream enzymes – equivalent to increasing expression levels of the protein. Figure derived from (Klann, Lapin and Reuss, 2011).

(Nebel *et al.*, 2014; Behrendorff, Borràs-Gas and Pribil, 2020). Therefore, substrate channelling at the transporter stage offers significant opportunity to enhance product yield; however, it requires a fair amount of troubleshooting to achieve.

The interaction between membrane proteins and their downstream enzymes is a relatively unexplored region of substrate channelling. There are limited confirmed cases of naturally occurring interactions between a transporter and its upstream or downstream enzyme, and are usually in the form of fusion (Harvat *et al.*, 2005; Babu *et al.*, 2010; Wu, Song and Beitz, 2010). More commonly, non-covalent interactions form between a transporter and a protein essential for function, such as the ABC domain, in which the transmembrane domain forms a complex with a substrate binding protein and nucleotide binding domains (Hollenstein, Dawson and Locher, 2007; Rees, Johnson and Lewinson, 2009). However, for the ABC transporters there are also examples where some or all of the domains are fused (Fig. 6.2.) (Biemans-Oldehinkel, Doeven and Poolman, 2006). The fact that fusions are prevalent in ABC transporters in eukaryotic cells but rarer in ABC transporters prokaryotic cells, bar the multidrug efflux transporters (Igarashi, 2004), may suggest a need for variation depending on the reaction, meaning evolutionary drift towards fusion occurs for some transporters and not others. Furthermore, some scaffold structures, such as the cellulosome, are localized to the membrane although do not show any association with transporters (Artzi, Bayer and Morais, 2016). At the start of this project there were no examples of scaffolds being produced at the membrane, although over the course of the past five years that has changed. This provided little in the way of context when designing scaffold or fusion constructs, and therefore necessitated methods that allowed easy alteration of the components of any fusions produced.

Using synthetic biology design principles, the aim of this work was to enhance the activity of the D-Xylose utilization pathway through the first two steps of the reaction: transport of D-xylose into the cytoplasm and isomerisation to D-xylulose. Instead, it was demonstrated that direct fusion resulted in a loss of activity of the component enzymes, and whereas fusion of peptide tags to the enzymes did not cause activity loss, scaffold assembly was not as simple as just expressing two parts of a scaffold on separate proteins. Ultimately, these processes reveal a great deal of unexpected challenge that may require trial and error to find the optimal condition.

In this discussion the systems applied throughout this thesis will be reviewed to produce an outline for future experiments and to offer some considerations when planning to enhance substrate channelling.

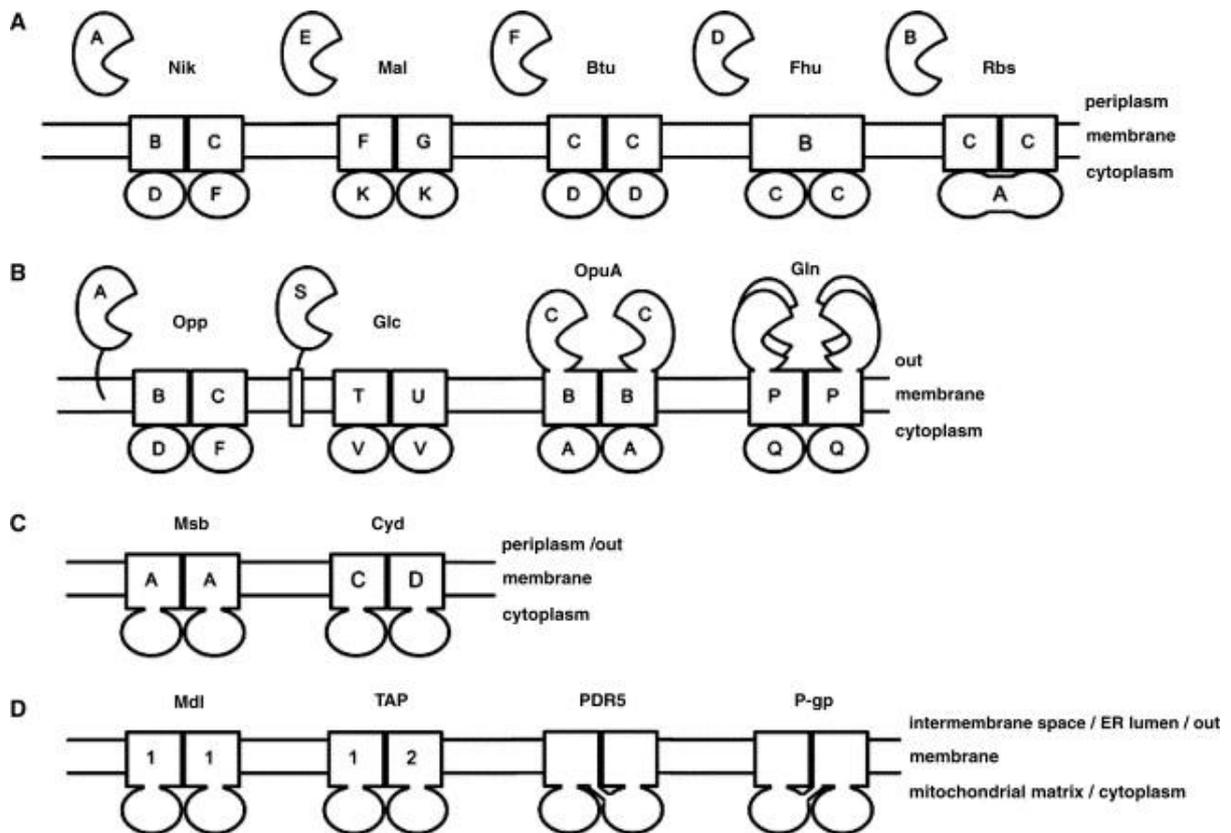


Figure 6.2. Domain architecture of the ABC transporters. A schematic of ABC transporters found in nature. ABC transporters are composed of substrate-binding proteins (pac-man shaped), transmembrane domains (rectangles), and nucleotide binding domains (ovals). (A) Gram-negative bacteria (all the examples are from *Escherichia coli*): Nik, nickel transporter; Mal, maltose/maltodextrin transporter; Btu, vitamin B12 transporter; Fhu, siderophore/haem/vitamin B12 transporter; Rbs, ribose transporter. (B) Gram-positive bacteria and Archaea: Opp, oligopeptide transporter from *Lactococcus lactis*; Glc, glucose transporter from *Sulfolobus solfataricus*; OpuA, glycine betaine transporter from *Lactococcus lactis*; Gln, glutamine/glutamic acid transporter from *Lactococcus lactis*. (C,D) Functional and structural homologues are present in all three kingdoms of life. (C) Msb, lipid flippase from *Escherichia coli*; Cyd, cysteine exporter from *Escherichia coli*. (D) Mdl, mitochondrial peptide transporter from *Saccharomyces cerevisiae*; TAP1/2 (ABCB2/3), human peptide transporter; PDR5, yeast pleiotropic drug resistance transporter; P-gp (MDR1/ABCB1), human multidrug transporter. Figure obtained from (Biemans-Oldehinkel, Doeven and Poolman, 2006)

6.2. Evaluating Golden Gate assembly as a tool for protein fusion libraries

The key focus of this work was generating a system, through design of multiple plasmids and protocols, that could assemble a functional scaffold or fusion from predesigned parts. Using this system to enhance product yield would demonstrate the applicability of a synthetic biology approach to rapidly designing scaffold assemblies going forward for a variety of pathways. In addition, due to the limited examples of synthetic association, be they fusions or scaffolds, between membrane transporters and the downstream enzymes (Behrendorff, Borràs-Gas and Pribil, 2020), producing an easily applied system offered the potential to enhance the product yield of numerous novel pathways. While this work ultimately failed to enhance product yield, it can be argued that the system developed is suited towards enhancing substrate channelling in the membrane, however some considerations must be taken.

Golden Gate assembly facilitates the cloning of multiple components in a fixed order and orientation within a single reaction mixture (Engler *et al.*, 2009; Iverson *et al.*, 2016). Compared to traditional cloning of genes within a plasmid using multiple digestion and ligation steps, this methodology facilitates a great deal of variety without requiring multiple steps. This methodology could be comparable to Gibson assembly (Gibson *et al.*, 2009), which also can assemble multiple amplicons into a new plasmid within a single reaction, however within this work we found Golden Gate assembly showed a superior success rate, with a higher amount of colonies expressing the desired product. For production of constructs such as the pLC or pLD family of plasmids, Gibson assembly methodology produced only a small number of colonies expressing the target plasmids, despite only involving three components, compared to the up to 7 parts of larger Golden Gate assembly reactions. It is important to note that high enzyme activity is vital to Golden Gate success and may necessitate replacing enzymes more frequently than the expiration date would suggest.

As a result of Golden Gate assembly, it was possible to easily produce multiple fusions in quick succession by altering the composition of the reaction media. Over the course of this thesis multiple unique fusions, from XylE-GFP fusions to scaffold components, were generated from just a handful of pre-designed parts – demonstrating the applicability of this technique. For example, during the study of the XylE-XylA fusions, each reaction mixture was identical bar the linker involved, allowing research into the impact of linkers on the XylE-XylA fusion (Fig. 4.13-4.16). Furthermore, elements such as the promoter can also be adjusted according with the demands of the experiment.

With automation, this methodology would also be appropriate for high throughput tests to fuse multiple transporters in an organism to their downstream enzymes. Following PCR

amplification with desired overhangs, each component would simply need to be added to a reaction mixture containing the otherwise fixed components (Fig. 6.3). To improve reliability, and validate all the components were correct, however, it would be preferable to first insert the amplicon into a holding vector.

Throughout this work, when larger assemblies of multiple parts were produced, each component of the fusion would first need to be inserted into a DVA series holding vector. This allowed for confirmation that correct gene had been correctly amplified and that no mutations were present. It also improved the consistency between reaction mixtures, as there was no risk of a mixed population.

This however highlighted one key shortcoming of the Golden Gate assembly methodology – if a small change was needed, such as changing a single gene in the final construct, rather than simply using a restriction enzyme digest to excise an undesired gene and insert a new one, the process must instead start from scratch. New holding vectors must be produced and sequenced to confirm amplification was correct, then cloned into the expression vector. This introduced multiple stopping points that slowed progress when a new component was needed. Furthermore, during the last stages of lab work, some experiments such as determining the role of the C- and N-terminal domain in XylE, could not be attempted due to the time constraints associated with Golden Gate assembly.

Therefore, Golden Gate assembly requires a great deal of forward planning – while it is a high throughput method capable of generating multiple products in quick succession, it is not a flexible process and requires returning to the starting point to make changes.

Evaluating the plasmids generated within this work, both pLC and pLD plasmids are suited towards membrane protein fusions. Both plasmids are low copy number and have demonstrated the ability to overexpress membrane proteins without resulting in toxicity to the cell.

The pLC and pLD family plasmids have great potential for developing protein fusions beyond impacting substrate channelling. For example, they are suited towards fusing membrane proteins to fluorophores to track mobility. Within this work constructs were produced with XylE, XylA, NanT or ExuT fused to GFP or mCherry, with the expression levels suitable to track the mobility of the attached protein without the need for excessive photobleaching. Golden Gate assembly facilitated the ability to test multiple different fluorophores when background proved to be an issue, facilitating fine tuning of the fluorescent fusion, with the fused proteins expressed by the same plasmid.

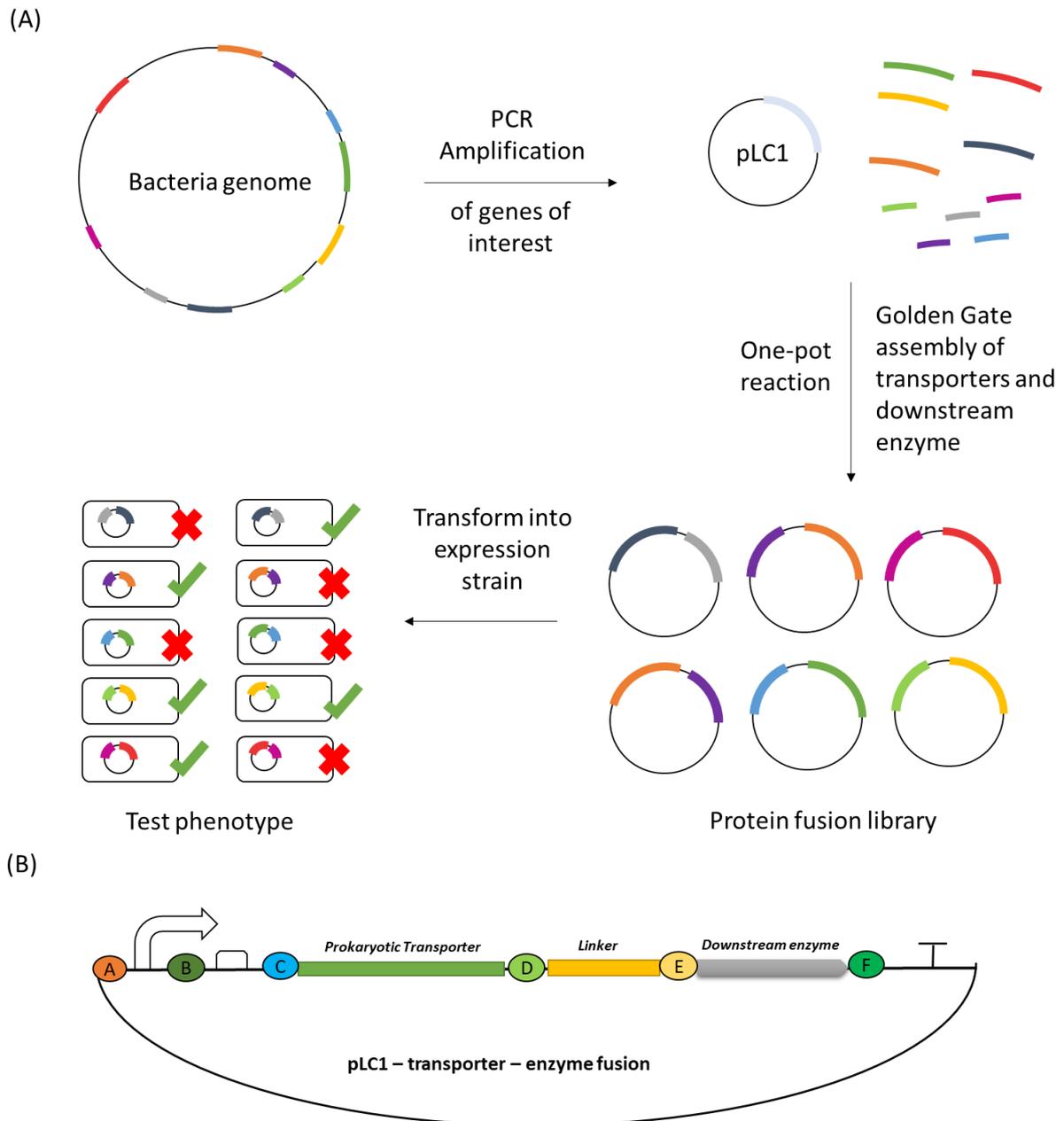


Figure 6.3. High throughput Golden Gate assembly of 2 gene constructs composed of a transporter gene fused to downstream enzymes. (A) A series of genes are PCR amplified with suitable overhangs to facilitate Golden Gate assembly. Complimentary genes are then inserted into a one-pot reaction in the desired orientation alongside a pLC1 or pLD1 expression vector. In order to confirm activity, the resultant plasmids (selected through blue white screening and sequencing) are then transformed into expression strains and a phenotype confirmed. **(B)** Schematic of a pLC1 plasmid expressing a hypothetical transporter-enzyme C-terminal fusion as produced in part A. The first insert represents a promoter and second an RBS.

In addition, the plasmids may be suited for fusing membrane proteins to enzymes responsible for measuring interprotein interactions. An example would be the non-specific biotin protein ligases (BPL). Whereas unmodified BPL will bind specific partners, promiscuous BPL enzymes have been generated by mutating *E. coli* BirA so it is capable of biotinylation of all neighbouring proteins (Cronan, 2005; Tron *et al.*, 2009). This fusion would be of interest as it would allow for discovery of unknown protein-protein interactions between membrane proteins – highly biotinylated proteins could be selected for to discover such interactions (Varnaité and MacNeill, 2016; Gingras, Abe and Raught, 2019).

Overall, the Golden Gate assembly system is highly suited to generating libraries of protein fusions – it however features limitation when looking at a targeted fusion. For high throughput studies it would be a favourable technique – with plasmids pLC and pLD highly suited to membrane protein fusions or scaffolds. For single changes, however, restriction enzyme digest to insert a gene into a known structure reproduced on a plasmid backbone may be preferable.

6.3. Fusion – a question of sterics

Synthetic fusion can show the capacity to enhance product yield. The mechanism of this is debated - fusion of two proteins has the ability bring active sites within proximity to facilitate channelling directly between them (Meynial Salles *et al.*, 2007; Yu *et al.*, 2015). Alternatively, others attribute the improved yield from to external effects such as pH optima, enzyme concentration and enzyme multimerization could also play a role in the substrate channelling effects seen (Pettersson and Pettersson, 2001; Zhang, 2011). In addition, there are suggestions it may achieve clustering effects similar to scaffolds (Castellana *et al.*, 2014). Incidents where enzyme activity has been reduced by fusion but product yield improved suggests that some form of channelling takes place in protein fusions (Kim *et al.*, 2015).

The fusion of transporters that is explored in this work confers the opportunity to increase transporter efficiency through immediate substrate metabolism (Behrendorff, Borràs-Gas and Pribil, 2020) in addition to facilitating enhanced substrate channelling through the downstream reactions (Elleuche, 2015). The opportunity, however, is not universal for all pathways, as fusion can introduce sterical constraints that negatively impact the activity or expression of the proteins of interest (Dueber *et al.*, 2009).

Steric hinderance describes potential collisions between two proteins or peptide structures that can limit degrees of freedom and challenge protein functionality (Fitzkee, 2004; Yu *et al.*, 2015). Sterical hinderance in protein fusion can be associated with protein aggregation, misfolding, formation of inclusion bodies, low yield in protein productivity and loss of activity (Yu *et al.*, 2015). Furthermore, it has been demonstrated that oligomeric proteins often cause sterical hinderance within protein fusions (Yeom *et al.*, 2017; Aalbers and Fraaije, 2019).

The terminus at which fusion takes place plays a key role in determining the impact of steric hinderance. In this work we found that XylA was not suited to N-terminal fusion. Both fusion of the bulky, membrane-bound XylE and small, cytoplasmic *Clostridium thermocellum* cohesin (ctCoh) resulted in a loss of XylA activity when fused to the N-terminal. Comparatively, XylA activity was equivalent with the wild-type when ctCoh was fused to the C-terminal.

The role of the terminus is also reinforced in literature. The astaxanthin producing enzymes CrtZ and CrtW, β -carotene hydroxylase and ketolase respectively, both contain multiple transmembrane helices. Fusions of these two proteins have been applied in organisms such as *E. coli*, *Nicotiana benthamiana* and *Corynebacterium glutamicum* to enhance yields of astaxthin (Henke and Wendisch, 2019; Nogueira *et al.*, 2019; Wu *et al.*, 2019). These fusions, however, only demonstrate an improved yield when the orientation is in a CrtZ-W formation,

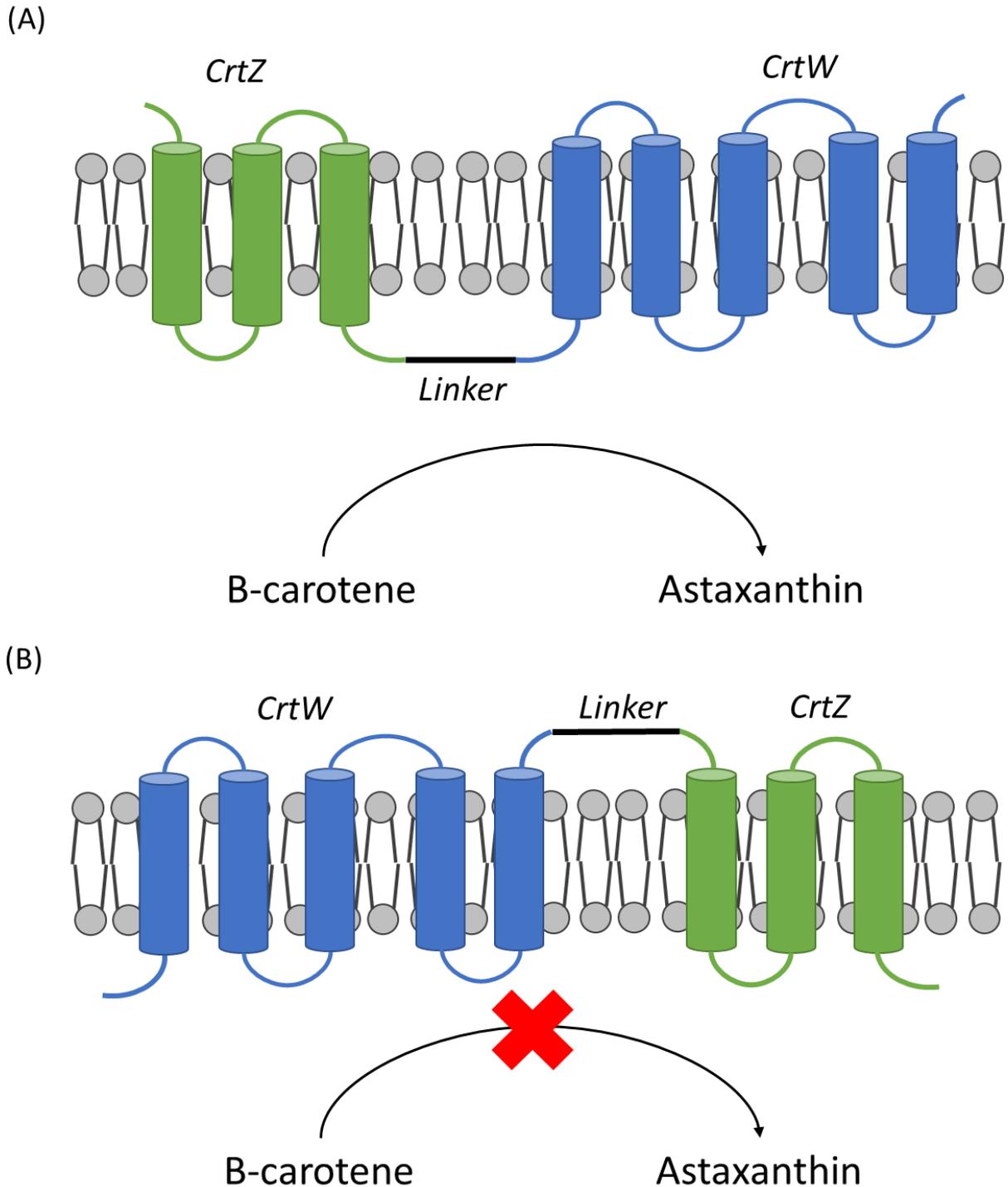


Figure 6.4. A fusion between membrane bound CrtW and CrtZ to increase Astaxanthin production. A schematic of the synthetic fusion between CrtW (β -carotene ketolase) and CrtZ (β -carotene hydroxylase), each localized to the membrane by a series of TMH. Figure adapted from (Henke and Wendisch, 2019) **(A)** Functional fusion at the C-terminal of CrtZ and CrtW. **(B)** Non-functional fusion between the N-terminal of CrtZ and CrtW.

demonstrating a complete loss of activity when a CrtW-Z fusion is produced instead (Fig. 6.4) (Henke and Wendisch, 2019; Nogueira *et al.*, 2019).

Due to time constraints it was not validated whether a N-terminal fusion of XylA to XylE could have resulted in a retained xylose isomerase activity, however considering the large size of the proteins involved it can be hypothesized that activity loss would remain. This is due to the size of the proteins involved, as the fusion to a large protein can negatively impact enzyme activity. For the xylose isomerase of *Clostridium phytofermentans* there was a direct correlation between the size of a fused partner and the loss of enzyme activity, with the effect only partially alleviated by changing termini (Thomik *et al.*, 2017).

Within the literature, different linkers are applied to space the two proteins to lessen the negative impacts of fusion. Various linker architectures have been applied to protein fusions, with flexible linkers allowing greater degrees of freedom so that active sites can have closer proximity (Chen, Zaro and Shen, 2013; Li *et al.*, 2016), whereas rigid linkers can separate enzymes associated with high steric constraint when fused (Arai *et al.*, 2001). To reflect this variance, a mixture of rigid and flexible linkers were applied in this work, with the rigid linkers being divided into helical linkers and peptide rulers (Table 3.5).

In some cases, the length of the linker can be of vital importance towards enzyme functionality (Agapakis *et al.*, 2010). In the case of a fusion protein consisting of an enzyme and a chaperone, the larger linkers resulted in the enzyme being able to retain activity, while the protein stability was inversely found to be highest in the smaller linker (Bergeron *et al.*, 2009). Alternatively, within the CrtZ-W fusion, three flexible linkers of variable length (10, 20 or 29 amino acids) were used within the fusion. It was found that within *E. coli* there was little impact in changing the linker length, however this result altered in *N. benthamiana*, where the longest linker did result in a reduced product yield compared to the other two samples (Nogueira *et al.*, 2019).

For the XylE-XylA fusion the linkers were between 12 and 20 amino acids in length. The long Arai linker, a 20 amino acid rigid linker, had issues associated with assembly resulting in few-to-no cells expressing the complete fusion, whereas the shorter Arai linker resulted in no measurable growth (Fig. 4.15), comparatively, the flexible linkers “flexible” and “poly(Gly-Ser),” 14 and 20 amino acids respectively, outperformed the shorter Waldo linker (12 amino acids). Finally, a polyproline linker (14 amino acid) performed equivalent to the best performing flexible linkers, however, was also the least stable fusion (Fig. 4.12), suggesting high accumulation of cleaved XylA, resulting in the high activity.

Ultimately, linkers can aid in alleviating some of the steric impact of fusion, however it appears that this can be highly limited. An obstructed fusion, such as the XylE-XylA fusion produced within this work, could not be overcome by changing the linkers alone, which meant that enzyme activity was still reduced when compared to the enzymes expressed individually. Furthermore, while enhanced length can help stabilise the fusion or facilitate equivalent activity to the native fusion (McCormick, Thomas and Heath, 2001; Bergeron *et al.*, 2009), it also increases separation between active sites, which could reduce substrate channelling effects (Nogueira *et al.*, 2019).

Considering the risk of steric hinderance, there are limited examples of a synthetic fusion between a downstream or upstream enzyme and its associated transporter within literature. In terms of membrane localized enzymes, the membrane-bound β -carotene hydroxylase and ketolase, CrtZ and CrtW are a key example (Henke and Wendisch, 2019; Nogueira *et al.*, 2019; Wu *et al.*, 2019), however as both enzymes are innately membrane bound this is not an example of a previously soluble enzyme being localized to the membrane. Alternatively, an arsenic transporter ArsB has been fused to SNARE proteins in order to produce transgenic *Arabidopsis* capable of accumulating arsenic in the shoot (Deromachi *et al.*, 2020), which demonstrated the role of fusion in organelle targeting but once more was not a fusion to a pathway enzyme. At the time of writing, a published fusion of a cytoplasmic enzyme to a membrane transporter through a lab-based approach could not be found in the literature.

Curiously, however, there is evidence of fusions occurring between an enzyme and transporter in nature. Novel fusions were found in nature of an aquaglyceroporin-derived arsenite channel with a C-terminal arsenate reductase domain, present in soil and marine actinobacteria. These domains were homologous to *E. coli* arsenate reductase (ArsC) and arsenite transporter (ArsB), with a novel ~40 amino acid linker. Beyond the actinobacteria *Mycobacterium tuberculosis* also encoded an analogous fusion between ACR3 (yeast type arsenite transporter) and ArsC (Wu, Song and Beitz, 2010) (Fig. 6.5). As these organisms are capable of metabolising arsenate, this system facilitates rapid export of arsenite from the cell and therefore is a key example of the role of substrate channelling in preventing the accumulation of toxic metabolites.

In addition, it is important to consider the group 2 MFS fusions, covered in 3.3.1. Group 2 fusions entail fusion of an MFS, homologous to *E. coli* LpIT, to a phospholipid acyltransferase domain of the *E. coli* acyltransferase/acyl-ACP synthetase (Aas) protein with some fusions also containing the AMP-dependant synthase that makes up the remainder of the *E. coli* Aas (Harvat *et al.*, 2005). This system is responsible for acylating the lysophospholipid (LPL) and

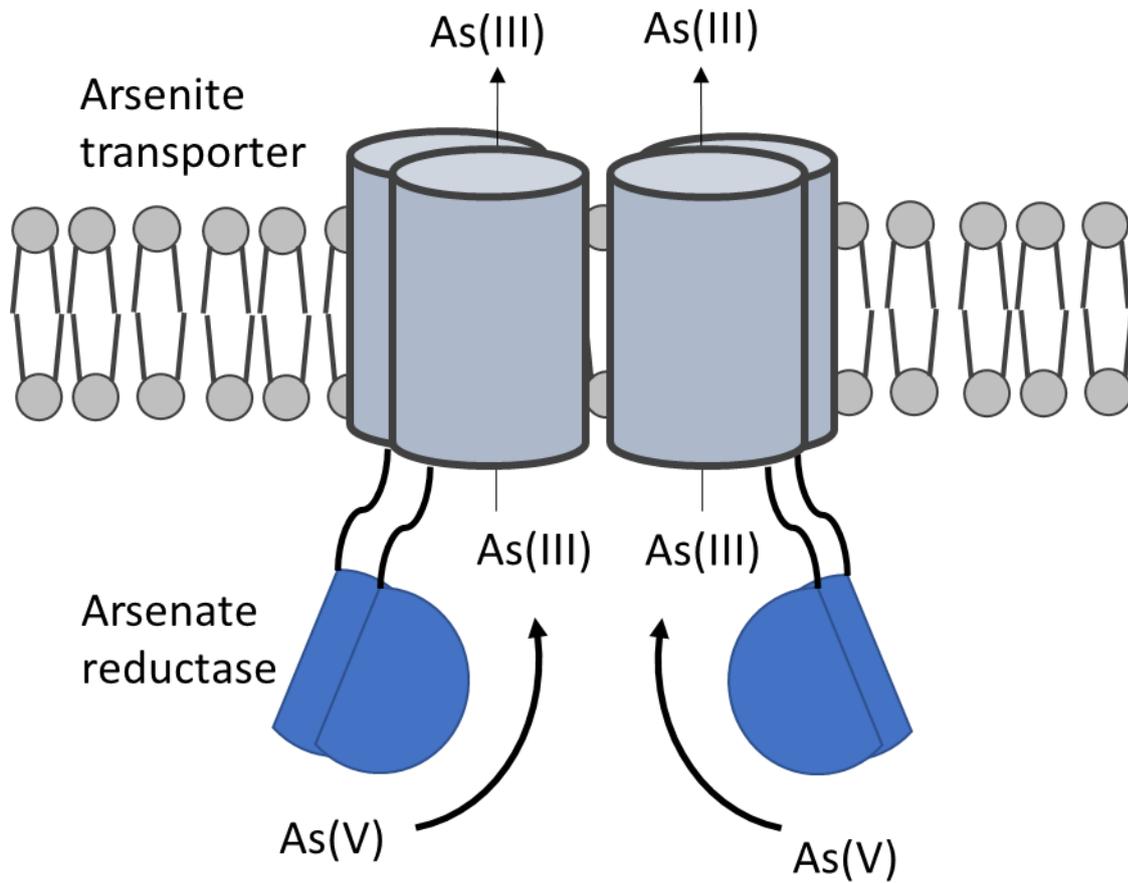


Figure 6.5. Natural fusion from *Salinispora tropica* between an N-terminal aquaporin-derived arsenite channel and a C-terminal arsenate reductase domain. A cartoon representing the natural fusion between arsenite transporter and upstream arsenate reductase. The aquaporin-derived channel is believed to be tetrameric, as represented here.

is associated with phospholipid repair (Lin *et al.*, 2018). Therefore, fusion of the downstream enzyme would be expected to increase channelling of LPL metabolites into the plasma membrane by immediately transferring an acyl-group onto the LPL, reducing the concentration of the mono-acyl phospholipids present in the membrane.

This system is ubiquitous within bacteria, with 3,307 organisms found following an interpro domain architecture search that expressed either a full or partial MFS-Aas fusion (Willson, Chapman and Thomas, 2019), suggesting a strong advantage towards forming this fusion. In addition, within cell lines that do not contain the fusion, such as *E. coli*, the Aas protein is itself an example of a fusion that localizes an otherwise soluble enzyme, PlsC, to the membrane (Harvat *et al.*, 2005). Overall, it appears that the acyltransferase domain, PlsC, is often found fused to a membrane localized partner, suggesting a need for membrane localization. An explanation towards why membrane localization is vital within this reaction pathway is that the LPL substrates are themselves situated within the membrane following transport by LplT and its homologues (Lin *et al.*, 2018). This means the advantages towards fusion at the membrane are likely to be more pronounced – it does however suggest pathways with substrates linked to the membrane are strong targets for synthetic protein fusion.

The fact that MFS proteins are stable to fusion at the C-terminal is also seen within this work, as XylE was consistently amenable to fusions at its C-terminal. Over the course of this work XylE was fused to various proteins, including fluorophores GFP and mCherry, the short peptide ctDoc and XylA, and at no point did fusion of a protein to the C-terminal of XylE result in a negative impact of the activity of XylE. This, however, failed to account for the impact the MFS protein may have on its fusion partner.

Considering the XylE-XylA fusion once more, there is evidence that simply localizing a freely diffusing enzyme such as XylA to the membrane alone would account for a notable increase in product yield. Work by Lv *et al.* demonstrated that localizing one component of the GlcNAc assembly pathway to a functional membrane microdomains (FMM), achieved by fusion of the enzyme to a membrane protein associated with FMM formation, resulted in an increase in GlcNAc yield (Lv *et al.*, 2020). This was despite other components of the pathway being freely diffusing and suggested that high local concentration of a single enzyme can benefit yield alone.

Proteins have been synthetically localised to the membrane for both inter and extracellular reaction process. The β -ionone producing enzyme PhCCD1 has been localized to the inner membrane in *E. coli* through fusion to a glycerol-conducting channel GlpF. Localization was considered important as the PhCCD1 substrate, β -carotene is located in the membrane (Ye *et al.*, 2018). The outer membrane of cell itself has been used as a “scaffold” on which multiple

enzymes can be displayed by fusing the enzymes to membrane anchors. The enzymes AmyAF, an α -amylase and CALB, a Lipase B enzyme, have both been fused to PgsA, a poly- γ -glutamate synthetase complex, to achieve localization. The PgsA fusions showed minimal loss of enzyme activity when fused and demonstrated high stability, with CALB otherwise unable to be expressed within the cytoplasm (Narita *et al.*, 2006).

Fusion of a soluble protein to an associated transporter has multiple benefits. The fusion can increase substrate channelling and enhance transporter activity, in addition to the increase in activity seen when soluble proteins are localized to the membrane despite no other proteins being localized there. Therefore, the failure of the XylE-XylA fusion to result in an improved activity highlights the role of steric hinderance in fusion. Ultimately, despite the variety of potential benefits, high steric hindrance can introduce significant limitations to protein folding or functionality, as seen by the reduced activity of the xylose isomerase. The risk of this high steric hinderance has resulted in recalcitrance to develop novel protein fusions at the membrane.

6.4. Scaffolds as an alternative to direct protein fusion

This work attempted to use protein scaffolds to overcome the steric hindrance introduced by protein fusion between transporter XylE and downstream XylA. The fusion of small peptide tags to XylE and XylA as opposed to bulky proteins was expected to reduce the negative impact of fusion. The scaffolding procedure, however, increased the complexity in localizing XylA to XylE, as a new plasmid was needed, scaffold components had to be selected and alternative methodologies were needed to confirm assembly.

Naturally occurring non-covalent metabolons in *E. coli* that form at transporters, include the ABC transporters, which form complexes with substrate binding proteins (Hollenstein, Frei and Locher, 2007), the bicarbonate transporter YchM, which is found in complex with an Acyl Carrier Protein (Babu *et al.*, 2010), and the regulatory binding of GlnK to AmtB (Javelle *et al.*, 2004). In addition some hypothesize that polycistronic genes may suggest metabolon formation (Moraes and Reithmeier, 2012). Beyond prokaryotes, we find complexes forming between cytochrome P450 enzymes associated with dhurrin production and their immediate downstream enzyme in sorghum (Laursen *et al.*, 2016) and the localization of three cytochrome P450 monooxygenases by two membrane steroid binding proteins to increase lignin production in *Arabidopsis* (Gou *et al.*, 2018). Arguably, the cellulosome could also be included, as enzymes including cellulases, β -glucanases, xylanases, mannanases, galactanases and xyloglucanases are localized to the outer membrane, so that sugars are produced at associated transporters, although the cellulosome does not directly interact with any specific transporters (Fontes and Gilbert, 2010).

Synthetic biology has introduced a greater diversity of scaffold archetypes, including various ligand/binding protein domains (Dueber *et al.*, 2009), cohesin-dockerin pairs (Liu *et al.*, 2013; Kim and Hahn, 2014), coiled coil domains (Thomas *et al.*, 2013; Thomik *et al.*, 2017), affibodies (Tippmann *et al.*, 2016) and the TatBC complex (Henriques de Jesus *et al.*, 2017). In addition atypical scaffolds can be formed on complex structures, such as the use of coiled coils to localize enzymes to the large filamentous PduA scaffold (Lee *et al.*, 2017). Ultimately, synthetic biology offers multiple opportunities to facilitate scaffold formation as an alternative to direct protein fusion.

The scaffold assemblies in this work used the cohesin-dockerin system due to both its involvement in previous synthetic scaffolds (You, Myung and Zhang, 2012; Jindou *et al.*, 2014; You and Zhang, 2014) and due to the cellulosome already demonstrating association with cell membranes. The dockerin was fused to the C-terminal of XylE, as it was assumed the MFS protein would be more amenable to a C-terminal fusion as opposed to an N-terminal fusion,

due to the propensity of MFS proteins to form C-terminal fusions with downstream proteins (Willson *et al.*, 2019; Willson, Chapman and Thomas, 2019).

The hypothesized scaffolding structure is seen in Fig. 6.6.A, in which the downstream XylA would be localized to the membrane through the scaffold. Over the course of the study, however, evidence suggested neither scaffold formation nor substrate channelling.

Through the course of this project, novel scaffolds continued to be published demonstrating a role in enhancing product yields. Of particular interest to this work were two examples of scaffolding at membrane proteins that were published in 2017. Through looking at these successful scaffold assemblies in detail it is possible to evaluate the applicability of scaffolds and their potential as an alternative to protein fusion.

The first scaffold assembly focused on increasing substrate channelling through a synthetic glutamate decarboxylase pathway expressed in an *E. coli* cell. The first protein of the reaction was the *E. coli* glutamate/gamma-aminobutyric acid (GABA) antiporter (GadC), which exports GABA and imports glutamate, and was fused to a SH3 ligand. A complimenting SH3 domain was fused to the C-terminus of the *Pyrococcus horikoshii* monomeric glutamate decarboxylase (GadB), bringing GadB to the membrane (Fig. 6.6.B) (Somasundaram *et al.*, 2017). The SH3 scaffold had previously been used in enhancing yields of mevalonate and malic acid (Dueber *et al.*, 2009; Somasundaram, Tae Eom and Ho Hong, 2010), and within this work, by directly fusing the scaffold components to the proteins of interest, the scaffold was assembled in a manner similar to the cohesin-dockerin system applied in this project (Fig. 6.6.A).

Within Somasundaram *et al.*'s work, the GadC-GadB scaffold demonstrated a capacity to increase GABA production by up to 3.5-fold, which contrasts with the application of scaffolds in the XylE pathway attempted within this thesis. GadC is a 12 transmembrane helix transporter active at acidic pH and part of the amino-acid-polyamine-organocation superfamily (Ma *et al.*, 2012; Ma, Lu and Shi, 2013). In addition, the mode action by which substrate is transported is through the movement of a 4 TMH gate domain (Ma *et al.*, 2012), as opposed to the rocker switch motif seen in MFS proteins such as XylE (Wisedchaisri *et al.*, 2014). This variation could go some way to explain the difference in substrate channelling observed, however another factor is of particular interest.

Somasundaram *et al.* produced three different fusions between the SH3 ligand and GadC, where the ligand was fused either to the N-terminal, C-terminal, or a loop between TMH 6 and TMH 7, which present in the cytoplasm. This allowed for testing the impact of the membrane protein terminal on substrate channelling. Within this work it was demonstrated that fusion to

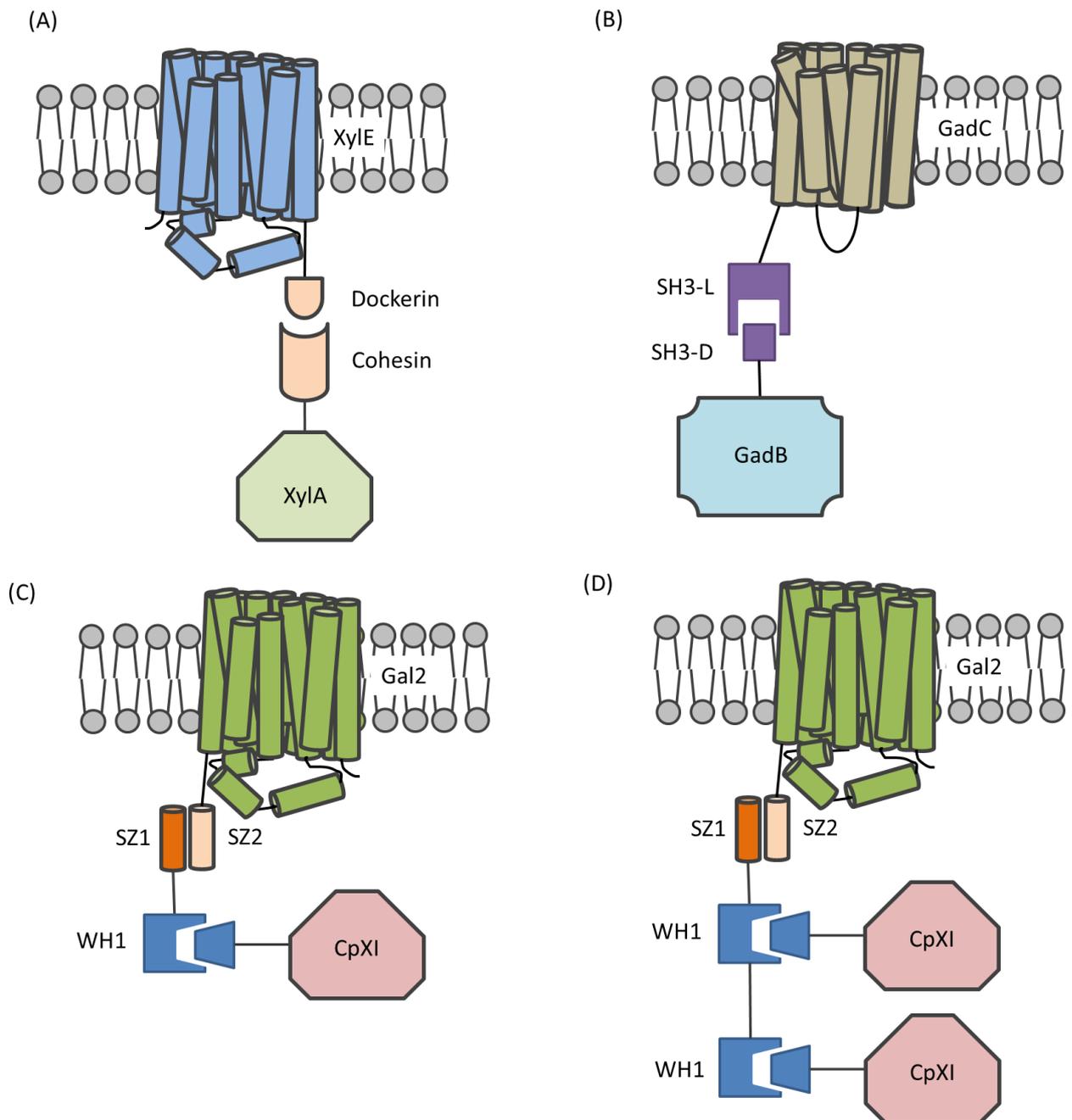


Figure 6.6. A cartoon representing the different scaffold assemblies seen both within this work and in literature. Cartoons representing different scaffold assemblies at the membrane either produced in literature or in this work – cytoplasmic proteins are presented as monomers for space purposes, but do not reflect actual oligomeric structure. **(A)** The hypothetical scaffold produced in this thesis between *E. coli* XylE and XylA (tetrameric protein represented as a single protein monomer in this image). The C-terminal of the 12 TMH MFS XylE was fused to ctDockerin and was expected to form a scaffold due to the cohesin-dockerin interaction. **(B)** The scaffold assembly of the GABA pathway in produced by Somasundaram *et al.* between *E. coli* GadC and the monomeric *P. horikoshii* GadB, produced through the

interaction of the SH3 domain and SH3 ligand. Displayed here is the N-terminal fusion between GadC and SH3 ligand, however within this work a C-terminal fusion and fusion at a cytoplasmic loop between TMH 6 and 7 were also produced. **(C)** The scaffold assembly of the D-xylose utilization in *S. cerevisiae*, produced through interaction between 12 TMH MFS Gal2 and *Clostridium phytofermentans* xylose isomerase (homotetrameric but represented by a single protein in this image). The coiled coil SZ2 was fused to the N-terminal of Gal2, and interacted with complement SZ1. As SZ1 fusion to CpXI resulted in loss of activity, it instead was fused to WH1 ligand at the N-terminal. A scaffold was therefore produced of SZ1, which would interact with the SZ2-Gal2 fusion, fused to a WH1 domain, which would interact with WH1L-CpXI. **(D)** Addition of a second WH1 domain to the scaffold in Fig. 6.6.C, producing SZ1-[WH1D]₂ to localize additional CpXI.

the C-terminal did not result in any improvement to substrate channelling (Somasundaram *et al.*, 2017) – which is consistent with work completed in chapter 5 of this thesis. It is possible, as in this work with SH3, that the fusion of ctDoc to the XylE C-terminal prevented scaffold formation, which could have been circumvented by fusion to another domain.

Our decision to focus on the C-terminal of XylE was derived from the domain architecture of MFS fusions in nature, and the rarity of N-terminal fusion. Considering this natural propensity for fusion, we hypothesized that the C-terminal would be better suited to fusion, both in terms of the stability of the membrane transporter and the activity of the downstream enzyme. It appears, however, that this hypothesis was imperfect.

There is however a need for hesitancy when declaring such a conclusion from Somasundaram *et al.*'s work. This is due to the lack of control tests on the activity of the various GadC-SH3 fusion in absence of the scaffold partner (Somasundaram *et al.*, 2017). There is a history of protein fusions to the peptide tag both positively impacting the activity of the enzyme, such as the affibody tags increasing sesquiterpene production in absence of a scaffold partner (Tippmann *et al.*, 2016), or lowering activity, such as ctCoh-XylA produced in this work. This means there is no negative control of GadC-SH3 activity to determine whether GadC activity had been negatively impacted by the fusion of SH3 to the C-terminal domain. Furthermore, there is no direct evidence that the SH3 scaffold was produced at the membrane, which could have been achieved through fluorescence study, such as particle tracking or FRAP. This is an important consideration as we do not know whether the C-terminal fusion of GadC to SH3 has negatively impacted GadC activity or has prevented scaffold formation – which makes it difficult to apply to the XylE-ctDoc result.

The second construct produced is of particular interest to this work, as it was an application of scaffolds to increase xylose utilization in *Saccharomyces cerevisiae* (Thomik *et al.*, 2017). Thomik *et al.*'s work also focused on a pathway synthetically introduced to the organism, increasing channelling through the *S. cerevisiae* Gal2, a homologue to XylE, and *Clostridium phytofermentans* xylose isomerase (CpXI). As both enzymes are homologous to the pathway enzymes used in XylE, this work was of great interest towards understanding the impact of this thesis.

First Gal2 was fused to an SZ1 domain, a coiled coil, and the formation of the scaffold was confirmed by demonstrating fluorophore localization through epifluorescence microscopy, with GFP fused to the complementary coiled coil SZ2. The GFP-SZ2 fusions were demonstrated to accumulate at the membrane, causing a distinct membrane fluorescence. While this result confirmed that an SZ1-SZ2 scaffold would result in localization at the membrane, direct fusion of SZ2 to either the C- or N-terminal of CpXI resulted in a loss of activity (Thomik *et al.*, 2017).

To achieve membrane localization, a variety of alternative scaffold proteins were considered. CpXI was fused to three alternative domains associated with protein-protein interaction: the SH3 domain, the WW domain from human YAP protein and WASP-homology 1 (WH1). Each was tested against activity of the wild type, finding that the WH1 ligand had low impact on xylose isomerase activity when fused to the N-terminal, making it suitable for the scaffold. Therefore, Thomik *et al.* produced an independent scaffold structure of SZ2 fused to a WH1 domain. This scaffold was localized to the membrane by the SZ2 domain, with the WH1D interacting with the WH1L-CpXI, bringing CpXI to the xylose transporter Gal2 and ultimately increasing yield by approximately 2-fold (Fig. 6.6.C) (Thomik *et al.*, 2017).

A particular advantage of this technique was the ability to introduce multiple copies of the WH1 domain into the scaffold structure to increase local CpXI concentration (Fig. 6.6.D). Expressing a scaffold with two WH1 domains, SZ1-[WH1]₂, was anticipated to increase substrate channelling by further increasing local CpXI concentration at the membrane, however instead the larger scaffold resulted in a decreased fermentation (Thomik *et al.*, 2017). This was due to a reduced transport efficiency for Gal2, likely due to sterical hinderance associated with the large construct at the N-terminal.

An important element to consider about this work was the decision to fuse the SZ2 molecule to the N-terminal. There was no justification applied to this decision, nor tests shown of any C-terminal fusions. This means that one cannot confirm that SZ2 would need to be fused to the N-terminal to be functional, just that this was the only condition tested (Thomik *et al.*, 2017).

The final consideration of Thomik *et al.*'s work that is relevant to this project was that linkers were also considered. In contrast to the work in this thesis, only rigidity was considered, with the options were selected being a 10 amino acid flexible linker, and a 10 amino acid rigid helical linker. In addition, these linkers were only applied to the SZ2-Gal2 fusion. It was found that using the rigid linker as opposed to the flexible linker in the SZ2-Gal2 fusion increased the negative impact of the SZ1-[WH1]₂ scaffold, which was demonstrated to be due to a reduced uptake of D-xylose by the *S. cerevisiae* cell, suggesting steric hinderance at the Gal2 was negatively impacting transporter functionality. Comparatively, linker rigidity was not a factor in the more successful SZ1-WH1 scaffold – once more demonstrating that linkers reduce unfavourable effects but have little impact when conditions are stable (Thomik *et al.*, 2017).

From both examples, the potential of scaffold proteins to increase substrate channelling at the membrane has been demonstrated – with the capacity to increase product yield by localizing a single downstream enzyme. There are key differences between the design of scaffolds in these experiments and the design of scaffolds in this thesis which offers some illumination as

to why the XylE XylA scaffold failed to assemble. Within this work, it was assumed that linker structure would play a significant role in achieving scaffold formation, resulting in five conditions tested, however based off the results seen, it appears that other considerations were of greater importance.

In both experiments an N-terminal fusion was used to bind the scaffold molecule to the transporter. In addition, the GadC fusion demonstrated that changing the position of the fused scaffold domain at the transporter can impact degree of substrate channelling (Somasundaram *et al.*, 2017). This suggests that the localization of ctDoc at the C-terminal of XylE may have played a role in preventing scaffold formation within this thesis. It is possible that the ctDoc peptide was not accessible to its complimentary ctCoh, regardless of whether 12 or 20 amino acid linker architecture were used.

Another consideration, however, was the choice of the cohesin-dockerin system – neither Thomik's nor Somasundaram's experiments attempted to use cohesin-dockerin pairs at the membrane. It is possible that the high affinity cohesin-dockerin pair selected does not function at the membrane and instead needs to be fused to a soluble partner. It may have been of more interest to this work, as opposed to testing multiple different types of linkers during the scaffold studies, to have used fewer linkers and instead vary the scaffold components. The parts for an AN4-BN4 coiled coil system had been developed for this work, however due to time constraints it was not applied beyond the planning stage. In addition, a lack of consideration was given to the impact of fusion to the XylE C-terminal on the functionality of attached partner, instead focusing on XylE function by using domain architecture to guide decision making.

It is important to acknowledge that linkers will still likely play a role in scaffold formation, as seen by Thomik *et al.* a poor choice of linker may cause a negative impact to become more significant (Thomik *et al.*, 2017). Unlike fusion, however, there are a greater choice of options to overcome negative effects of one or more considerations of substrate channelling.

Ultimately it has been demonstrated that synthetic scaffolds can be produced at the membrane, however, requires consideration of multiple factors beyond simply the stability of the enzymes involved.

6.5. To scaffold or fuse – what is the best choice for substrate channelling?

Substrate channelling can be enhanced at the membrane (Somasundaram *et al.*, 2017; Thomik *et al.*, 2017). While this project has failed to demonstrate such a positive impact in D-xylose utilization when XylE and XylA interact, examples of success in literature reveal that this is more a fault of design as opposed to fault of theory.

This work instead demonstrates the challenge associated with membrane protein fusions and scaffolds. The production of the Golden Gate assembly system facilitates the production of large assemblies of fusion proteins and scaffold components; however, time constraints and access to equipment can introduce limits on what can feasibly be tested. One key restriction of this work was in attempting to both evaluate protein fusions and scaffolds simultaneously, there was not enough time to truly troubleshoot all the components of either condition which limited the ability to produce an efficient system for substrate channeling.

Considering the work produced in this thesis, as well as the results in literature, it is important to compare and contrast the role of protein fusion and protein scaffolds in substrate channelling and offer insight into where each technique may be best applied.

The first consideration is the convenience of the two methods. In order to assemble a scaffold, at least two separate protein fusions must be expressed representing the enzymes of interest and complementary peptide tags. To achieve this requires expressing the genetic fusions on two distinct plasmids. Comparatively, the direct fusion requires only a single plasmid to express the desired construct.

Furthermore, direct tandem fusion has only a few variants to consider – such as the linker used and whether to fuse at the C- or N-domain. Comparatively, the scaffold contains a greater number of variants to consider. For each protein involved both the N- and C- terminal domain and linker used can play a role, however in addition the scaffold component can be altered as there are a variety of potential choices, including coiled coil domains, SH3 ligands, WHL and the cohesin dockerin system (Behrendorff, Borràs-Gas and Pribil, 2020). Furthermore, it may be of interest to produce a separate structure as a scaffold, that facilitates finer control over the scaffold components used (Thomik *et al.*, 2017).

This complexity also necessitates additional testing to confirm that the scaffold has correctly assembled, which creates additional consideration when applying a protein scaffold. For protein fusion by introducing a His-tag it is possible to use a combination of SDS-PAGE and a western blot to confirm assembly. Conversely, the scaffold requires a more complex

methodology. SDS-PAGE can disrupt the scaffold assembly, making it difficult to confirm interaction.

Within this work a TIRFM protocol was developed that could demonstrate *in vivo* association of scaffold components to the membrane. Alternatively others have used techniques such as FRET between two complementary fluorophores (Lin, Zhu and Wheeldon, 2017) or Native-PAGE (Willson *et al.*, 2016). Techniques such as native-PAGE can make it difficult to confirm the exact mass of the construct but can be achieved through comparisons with the unfused components and by looking for large smudges (Willson *et al.*, 2016). In addition, activity assays may also require additional testing to confirm scaffolds are having an effect and it is not simply the result of tagging, requiring production of additional controls to compare enzyme activity (Tippmann *et al.*, 2016).

The complexity of the scaffold system, with its numerous variables to consider, does offer greater options to alleviate problems with activity or expression when compared to a direct fusion. When producing a scaffold system between Gal2 and CbXI, Thomik *et al.* attempted multiple different scaffold components to reduce the impact of tagging on the xylose isomerase activity. In addition, they increased the number of repeat units in the scaffold to localize higher concentrations of the CbXI to the membrane to determine if this could increase yields higher (Thomik *et al.*, 2017). Comparatively, fusions have more limited choice because of their convenience.

A key advantage of the use of protein scaffolds is the ability to control the relative stoichiometry of the enzymes involved. Direct fusion is often limited to a single copy of each gene, as large protein fusions show high risks of misfolding, meaning that the stoichiometry of hypothetical proteins A and B is 1:1 (Dueber *et al.*, 2009). Scaffold tags can be easily duplicated within a separate structure with a reduced risk of misfolding due to their small size, allowing for stoichiometries ranging from 1:1 to 1:4 (Fig. 6.7) (Agapakis *et al.*, 2010; Thomik *et al.*, 2017), therefore increasing the amount of enzyme B localized at each unit of enzyme A.

This control over stoichiometry can aid in increasing enzyme activity (Siu *et al.*, 2015). Deuber *et al.* demonstrated control over the stoichiometry of three components can result in yields increasing from less than 5-fold up to 77-fold, with clear variation depending on the stoichiometry used (Dueber *et al.*, 2009). Comparatively, however, increasing stoichiometry at the membrane has actually resulted in a loss of transporter activity (Thomik *et al.*, 2017), demonstrating that this methodology is reaction specific. Furthermore, in some cases enforcing a strict 1:1 stoichiometry through fusion is desired (Kleinlogel *et al.*, 2011; Elleuche, 2015; Aalbers and Fraaije, 2019).

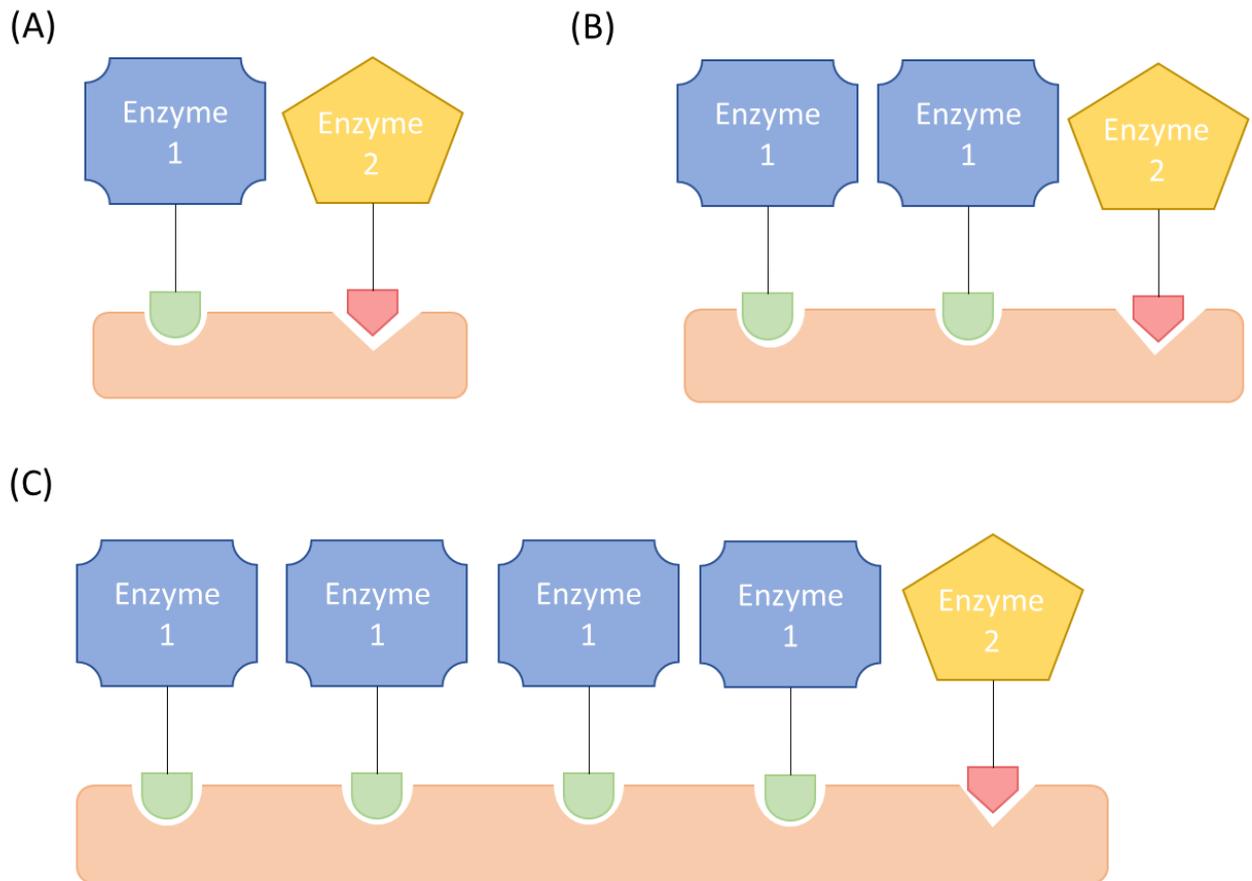


Figure 6.7. Different scaffold compositions to achieve altered local enzyme stoichiometry. A cartoon schematic of scaffold architectures to achieve altered enzyme stoichiometry. **(A)** a scaffold that achieves a 1:1 stoichiometry of enzyme 1 to enzyme 2. **(B)** a scaffold that achieves a 1:2 stoichiometry **(C)** a scaffold that achieves a 1:4 stoichiometry of enzyme 1 to enzyme 2.

Sterical hinderance is a key factor when considering whether to apply a protein fusion or scaffold (Vanderstraeten and Briers, 2020). For larger and oligomeric proteins, there is a high risk of aggregation or misfolding (Yu *et al.*, 2015), which undermines the goals of direct fusion.

Comparatively, scaffold proteins involve tagging each enzyme individually with small peptides, which are less likely to be subject to hinderance (Dueber *et al.*, 2009) or cause issues with oligomerization (Thomik *et al.*, 2017). Furthermore, when a scaffold design results in aggregation, the scaffold systems used can easily be adjusted so that the enzyme is fused to a smaller tag (Thomik *et al.*, 2017).

The oligomerization of tagged proteins is also believed to be a key factor towards increasing local enzyme concentration through enzyme clustering (Fig. 6.8), and permitting oligomerisation is considered a key advantage of protein scaffolding (Sweetlove and Fernie, 2018). However, the complete impact of highly oligomeric enzymes in protein scaffolds is relatively unexplored. For the scaffolding of GadC to GadB (Fig. 6.6.B), a monomeric *P. horikoshii* GadB was used in place of the hexameric *E. coli* GadB (Somasundaram *et al.*, 2017). This decision was based on the hypothesis that polymeric proteins were less efficiently synthesised than monomeric proteins (Tsung, Inouye and Inouye, 1989), however this also means the monomeric GadB-GadC scaffold cannot produce the enzyme cluster formation represented in Fig. 6.8. But is enzyme clustering always desirable? As previously discussed, within the Gal2 scaffold to CpXI, increasing enzyme stoichiometry from 1:1 to 1:2 decreased transporter functionality. Thomik *et al.* produced protein scaffolds because of concerns that direct fusion would undermine tetrameric CpXI oligomerisation (Thomik *et al.*, 2017), however taking into account the impact of increasing stoichiometry, it is possible the tetrameric CpXI enhanced enzyme clustering to the point of creating steric hinderance at Gal2. Therefore, it can be hypothesized that while scaffolds are better able to tolerate oligomeric proteins compared to direct fusion, under conditions such as increased stoichiometry, highly oligomeric proteins are more susceptible to introducing steric hinderance to a system than less oligomeric proteins.

As a result of the high steric hinderance associated, the synthetic fusions of more than two proteins are rare and are associated with loss of activity or complete misfolding of the enzymes (Conrado *et al.*, 2012). Comparatively, scaffolds have been produced containing three enzyme components, with the smaller peptide tags less impacted by steric pressure, even facilitating changes in stoichiometry in three enzyme reactions and perhaps suggesting larger assemblies as a possibility (Dueber *et al.*, 2009; You, Myung and Zhang, 2012).

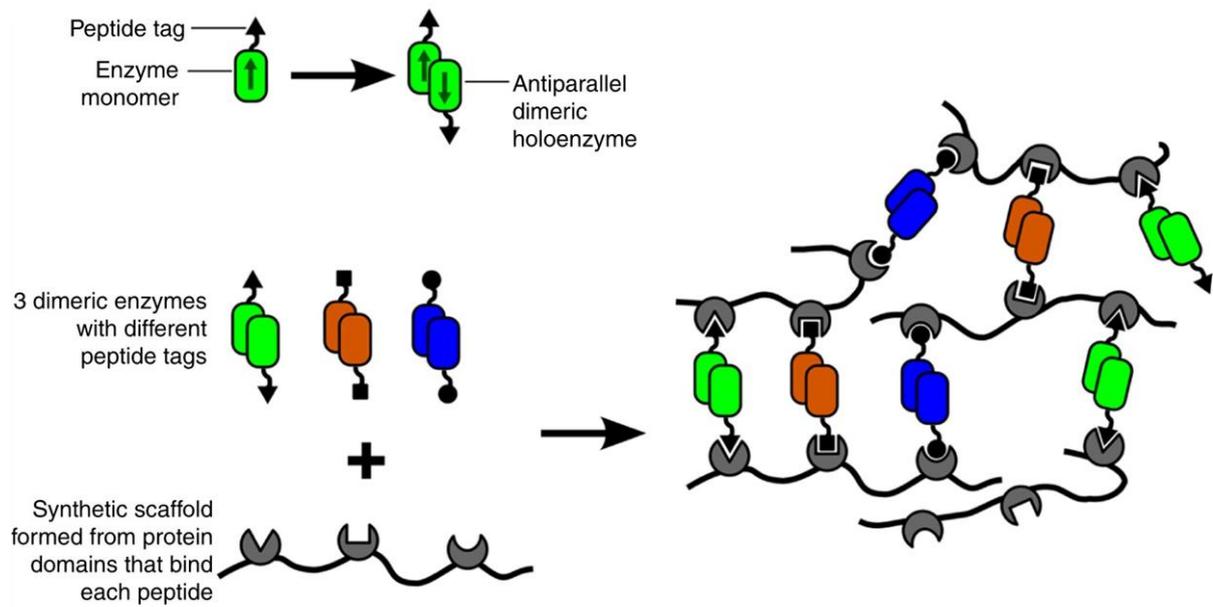


Figure 6.8. Schematic of oligomeric proteins in a scaffold. A cartoon demonstrating the mechanism of enzyme cluster formation in synthetic scaffold-enzyme assemblies for dimeric enzymes. When oligomeric enzymes are docked onto synthetic protein scaffolds via peptide tags, then interaction with more than one scaffold molecule are possible leading to the formation of a large aggregation of scaffolded enzyme complexes. Figure obtained from (Sweetlove and Fernie, 2018).

Scaffolds are usually applied where direct fusion has failed, however in some cases both systems have been applied. This gives insight into the impact of either system in terms of enhancing product yield. There is evidence, where both systems have been successfully applied, that fusions have the capacity to outperform the scaffold. In a pyruvate metabolism pathway, direct fusion resulted in a 4.4 fold-increase in product yield compared to a 3-fold increase seen in scaffolds (Agapakis *et al.*, 2010). This likely reflects a closer interaction between the two active sites when physically held in proximity compared to the looser association in scaffold complexes. It is not a unanimous situation, however, as Wang and Yu demonstrated that scaffolds improved yields of resveratrol 5-fold compared to the 2.7-fold of direct fusion (Wang and Yu, 2012). Overall, this means whether either technique is better suited to a pathway is not known without some testing.

To conclude – protein fusion is a convenient technique for bringing two enzymes in close proximity, however, has shown a propensity towards aggregation and loss of enzyme activity. Protein scaffolds can offer greater control over bringing the enzymes together and reduce negative impacts such as misfolding, however is a more complex technique with multiple variables to consider and that requires significant testing to confirm assembly and improvements in activity.

Protein fusion**Protein scaffold**

	Protein fusion	Protein scaffold
Complexity	Only variables to consider are the linkers used and the terminal to fuse at.	More variables – which ligands to be used for scaffolds, the orientation for each protein-scaffold fusion and the linkers used for each fusion. In addition, is the scaffold fused to another protein (Fig. 6.6.A and B) or is it a separate structure (Fig. 6.6.C and D)?
Ease of confirming assembly	Assembly can be confirmed by a simple western blot against a His tag following SDS-PAGE.	SDS-PAGE breaks non-covalent bonds therefore, alternative methods must be used which make determining the scaffold formation more of a challenge. Native PAGE may be an option or to confirm in vivo assembly fluorescence studies can be attempted.
Options to change conditions	Can only change the linker or terminal if activity is lost.	Greater flexibility – different types of ligands can be used to overcome issues with enzyme or transporter activity alongside different linkers and terminals.
Enzyme stoichiometry	Less control over enzyme stoichiometry.	Using a separate scaffold, it is possible to add multiple repeats of the same scaffold component to alter stoichiometry of the complementary enzyme.
Sterical hinderance	High risk of sterical hinderance – large proteins are fused to one another with risk of activity loss.	Reduced steric hinderance as small peptide tags less likely to cause negative impacts to enzyme activity.

Number of components	Largely limited to two proteins	Can include multiple enzymes or transporters using a separate scaffold structure.
Conclusion	Protein fusion produces an easy method to attach two proteins together to produce a synthetic metabolon structure. However, there is less flexibility and a greater risk of sterical constraints limiting enzyme activity. Testing this system for two component reactions could be advantageous.	Overall: Protein scaffolds introduce complex structures into enzymatic reactions to produce the metabolon. While offering great flexibility, allowing to change multiple factors to facilitate metabolon formation, it is harder to confirm assembly and the high complexity can make the system inconvenient. Furthermore, this system can link more reaction steps than direct fusion.

Table 6.1. Comparison of direct fusion and protein scaffolding for substrate channelling. A summary of the different advantages and disadvantages associated with using either direct fusion or protein scaffolding in order to achieve substrate channelling through a reaction pathway.

6.6. Concluding remarks.

Within this work it was attempted to apply both a direct fusion between XylE and XylA and facilitate scaffold formation through use of a cohesin-dockerin system. A series of growth assays were used to determine whether fusions could improve functionality of the native proteins, however there were no noticeable improvements that would suggest substrate channelling was taking place. The XylE-XylA fusion resulted in a loss of xylose isomerase activity, whereas the scaffolds retained the functionality of both enzymes however there was no evidence of scaffold formation. While neither technique resulted in an improved rate of substrate channelling, a system for rapid assembly of libraries of membrane protein fusions was developed.

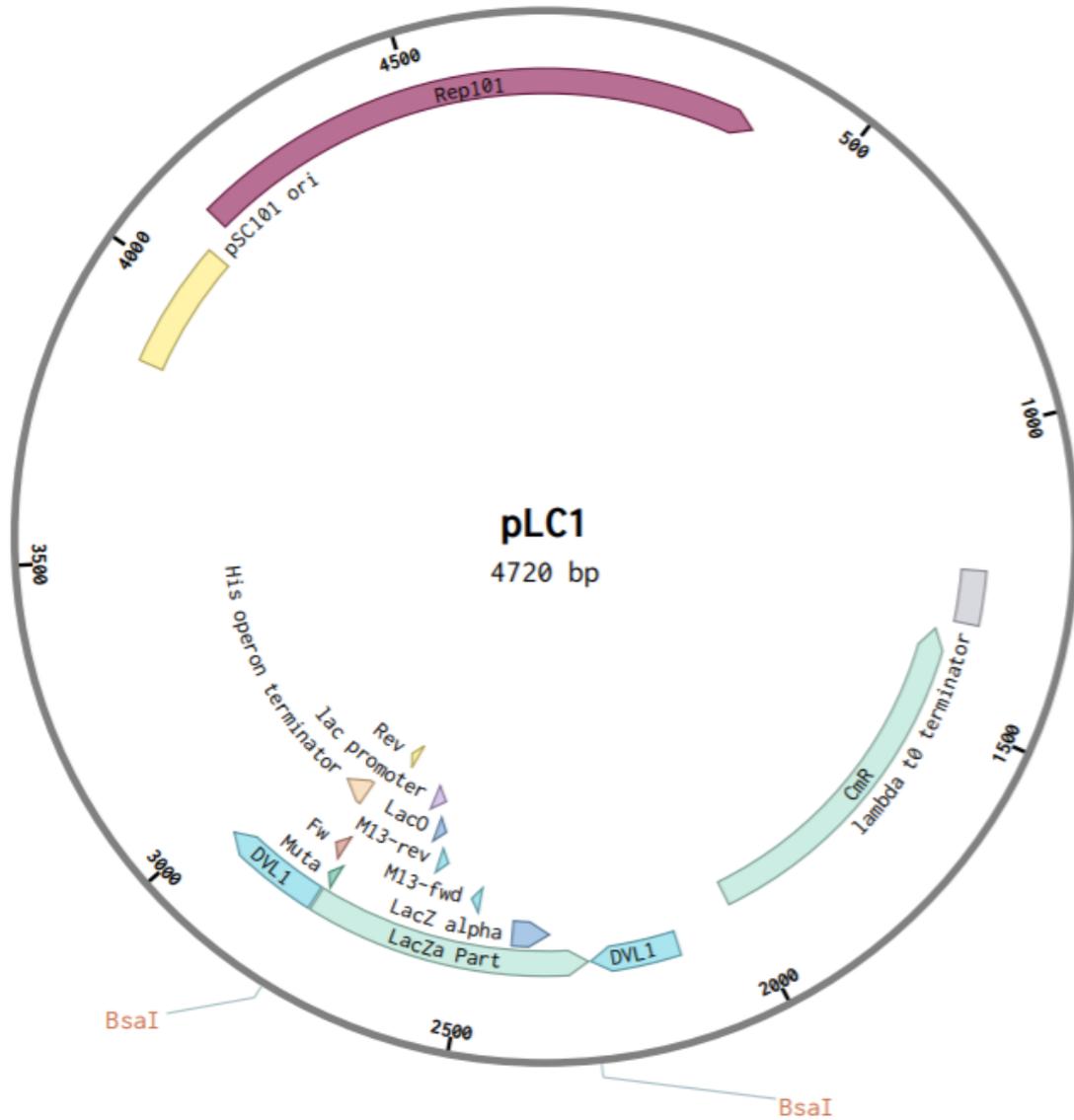
To understand the results of fusion a review of literature was completed focused on fusions and scaffolds at the membrane. A significant challenge is associated with direct fusion, with fusions between transporters and their downstream enzymes largely isolated to nature, however localization of soluble proteins to membranes or microdomains can result in an improvement in product yield. In comparison, two successful scaffold assemblies at the membrane allowed for comparison of the scaffold generated in this work to the successful results, finding variation in the domain the scaffold was attached to and the scaffold peptides used. Ultimately these results illuminated some shortcomings in the design of this experiment.

Finally, an evaluation was made of the potential of both protein fusion and scaffolds, evaluating the potential of each technique and the challenges associated. Protein fusion is a convenient method that introduces high steric to a system, whereas scaffolds require more consideration towards multiple variables, however, is less likely to face enzyme inhibition through steric hinderance.

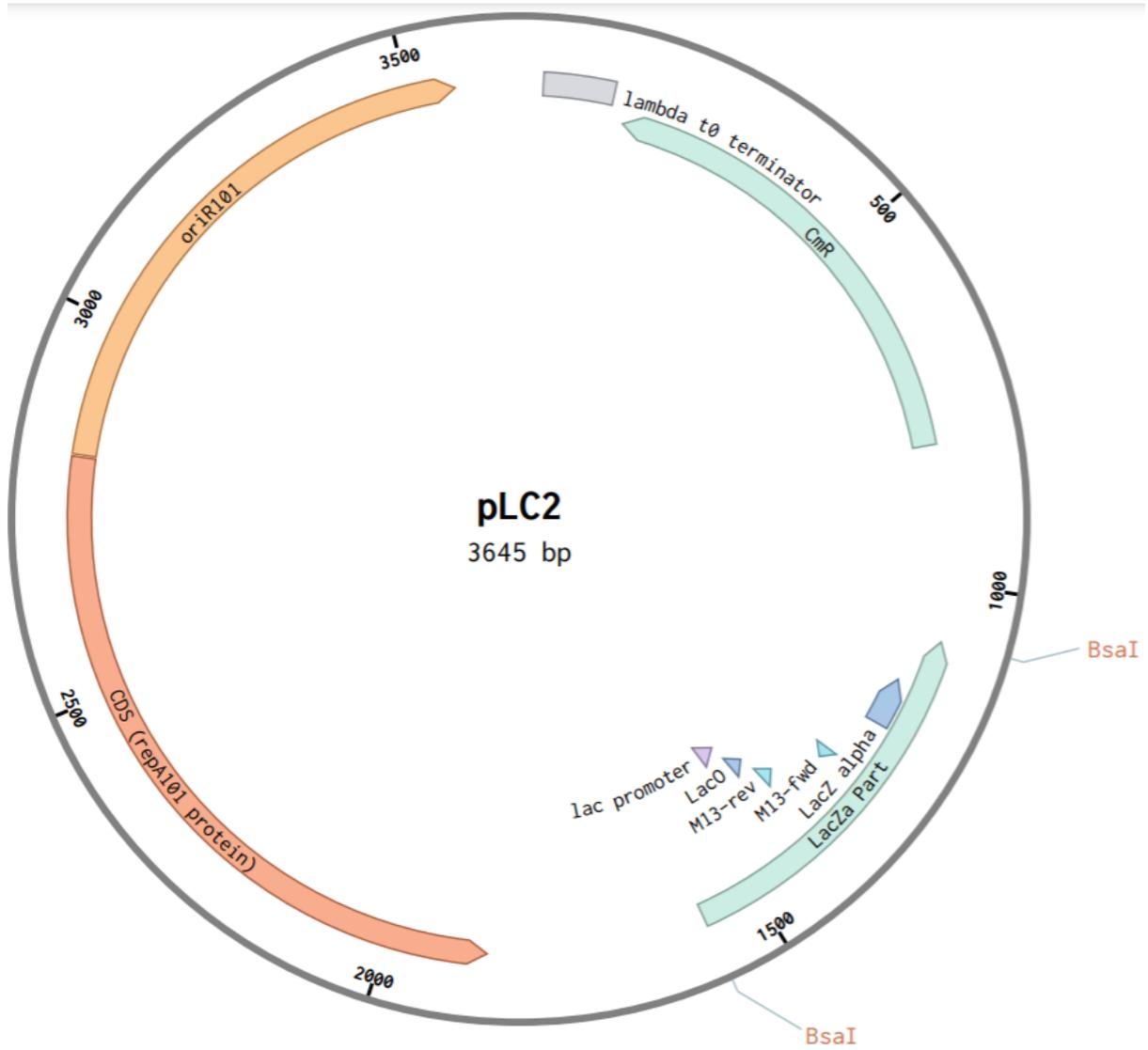
Ultimately while this work did not demonstrate substrate channelling through the xylose utilization pathway, the systems developed would be suitable for establishing large libraries of scaffolds or fusions. Evidence from literature suggests that the xylose utilization pathway is a suitable system to exploit through substrate channelling, suggesting that with more time it would have been possible to find conditions suitable for scaffold formation. Due to the high number of variables, time constraints will ultimately be a key factor when developing constructs going forwards.

Appendices

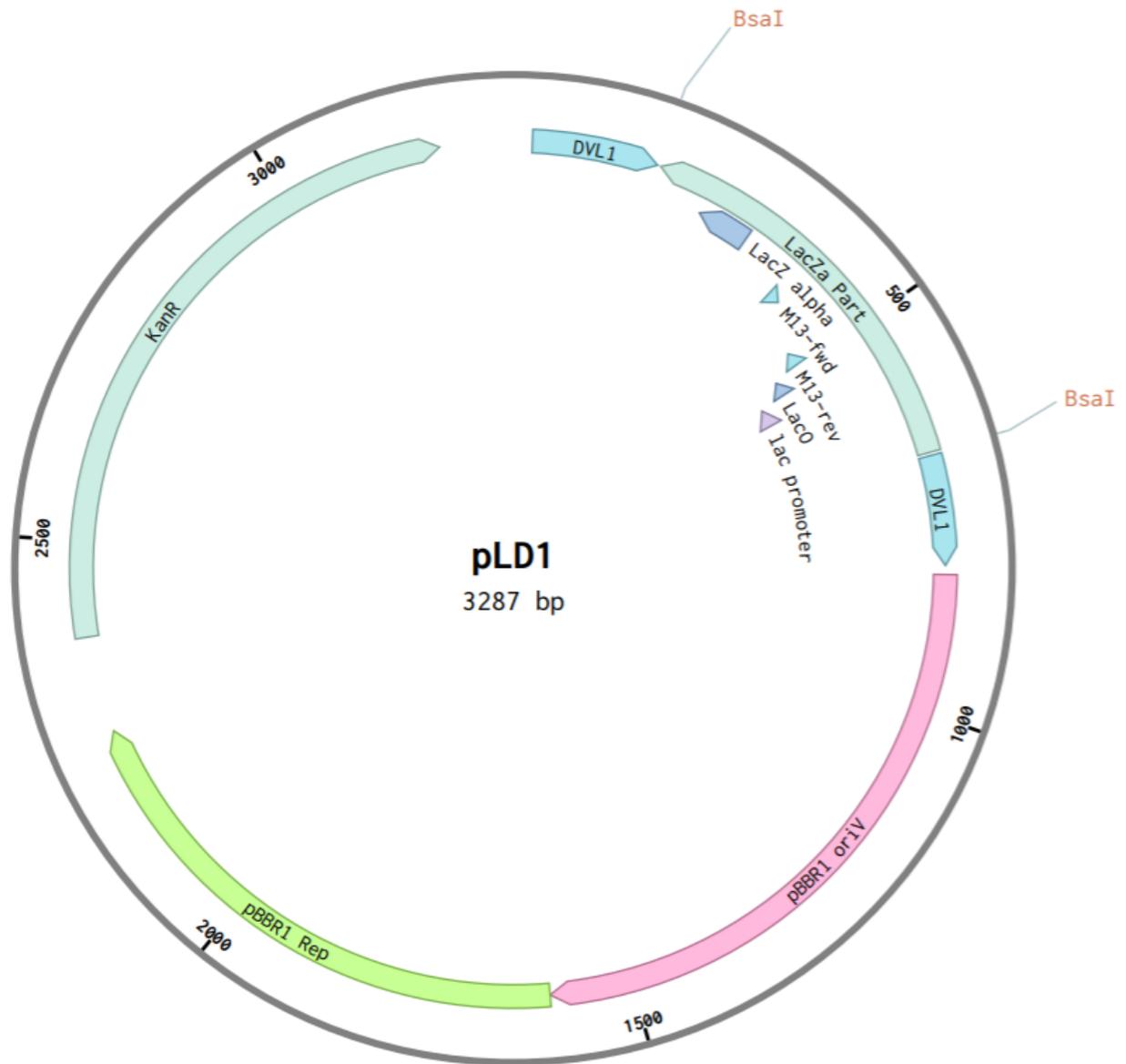
Appendix A – plasmid maps of novel plasmids



Plasmid map of pLC1 – a low copy number, chloramphenicol resistant level 1 golden gate plasmid.



Plasmid map of pLC2 – a low copy number, heat-sensitive chloramphenicol resistant level 1 golden gate plasmid.



Plasmid map of pLD1 – a low copy number, kanamycin resistant level 1 golden gate plasmid.

Appendix B - Synthetic gene sequences.

Nucleotide sequences of synthetic gBlocks ordered for codon-optimised scaffold peptides.

gBlock for *Clostridium thermocellum* dockerin and cohesin

GTACAAGAGCATAGTACACCTACGCGCGGGGAACCTCGTTCGTATGGCCTCTGATGGCG
TGGTAGTAGAAATTGGCAAAGTTACGGGATCTGTTGGAACCTACAGTTGAAATACCTGTA
TATTCAGAGGAGTTCCATCCAAAGGAATAGCAAACCTGCGACTTTGTGTTTCAGATATGAT
CCGAATGTATTGGAAATTATAGGGATAGATCCCGGAGACATAATAGTTGACCCGAATCC
TACCAAGAGCTTTGATACTGCAATATATCCTGACAGAAAGATAATAGTATTCCTGTTTGC
GGAAGATAGCGGAACAGGAGCGTATGCAATAACTAAAGACGGAGTATTTGCAAAAATAA
GAGCAACTGTAAAATCAAGTGCTCCGGGCTATATTACTTTTCGACGAAGTAGGTGGATTT
GCAGATAATGACCTGGTAGAACAGAAGGTATCATTATAGACGGTGGTGTAAACGTTGG
CAATGCAACACCGACCAAGGGAGCAACACCAACAAATACAGCTACGCCGACAAAAGTA
GCCTTAGCTGCAACCGGTGCAGCGTTAATTCCTAAATATGTATTAGATGGGCGATGTAA
ACGGTGATGGAACCATTAACCTCAACTGACTTGACAATGTAAAGAGAAGCGTGTTGAGG
GCAATCACCTTACCGACGATGCAAAGGCTAGAGCAGACGTTGACAAGAATGGATCGA
TAAACAGCACTGATGTTTTACTTCTTTACGCTACCTTTAAGAGTAATCGCAGTTATCC
GATGATCATTATAGCTATGCCAG

gBlock for *Clostridium cellulolyticum* dockerin and cohesin

GTACAAGAGCATAGTACACCTACGCGCGGGGAACCTCGTTCGTGGATCCGCTGGCTCCG
CTGCTGGTTCTGGCGAATTCATGATCACAGTGCTGCCGAAGGACATCCCCGGAGACAG
TTTGAAAGTCACGGTAGGTACCGCAAATGGCAAACCGGGTGATACGGTTACAGTTCCT
GTAACCTTTGCGGATGTCGCAAAGATGAAGAACGTGGGGACCTGCAATTTTTACTTAGG
CTATGACGCTTCCCTTTTAGAGGTCGTAAGCGTTGACGCAGGCCCCATCGTGAAAAAC
GCTGCCGTAAACTTCTCGTCCAGTGCGTCCAACGGCACAATTCGTTCTGTTTTTGGTA
TAACACTATTACAGACGAATTGATCACTGCCGACGGAGTGTGGCAATATCAAATTTAA
GCTGAAGTCCGTTACGGCTAAAACGACGACCCAGTTACTTTCAAAGATGGTGGGGCA
TTCGGGGATGGGACTATGTCCAAAATCGCTTCAGTTACTAAAACCAACGGCTCTGTCAC
GATCGACCCTTAGCCTTAGCTGCAACCGGTGCAGCGTTAATTCCTAAATATGTATTAGA
TGATTGTTTATGGTGACTACAATAACGATGGGAACGTCGATGCCCTTGATTTGCGAGGG
TTGAAGAAATACATTATGGCAGCAGACCACGCTTATGTAAAAATTTAGATGTCAATCTT
GACAACGAGGTAAACGCCTTTGACTTAGCTATTTTAAAGAAATACCTTCTGGGAATGGC
AGTTATCCGATGATCATTATAGCTATGCCAG

Appendix C - Golden Gate assembly Code

Golden gate code used in this thesis to identify follows the following protocol

Part 1:

Name	Designation
1	J23100_AB (constitutive promoter)
2	R0010_AB (IPTG induced promoter)
3	I3453_AB (Arabinose induced Promoter)

Part 2:

Name	Designation
1	B0034_BC (strong RBS)
2	B0032_BC (Weak RBS)

Part 3:

Name	Designation
1	<i>E coli</i> XylE mutated
2	<i>E coli</i> ExuT
3	<i>E coli</i> NanT mutated
4	<i>E coli</i> XylB
5	DjlA
6	BASU
7	<i>C thermocellum</i> Cohesin
8	AN4

9	BN4
A	E0040m
B	<i>E coli</i> XylA
C	mCherry
D	<i>C thermocellum</i> Dockerin

Part 4:

Name	Designation
1	Waldo linker (wld)
2	Flexible linker (flex)
3	Proline rich linker (pro)
4	Poly(eak) linker (Long aria) (EAK)
5	Empirical rigid linker(Short Aria) (EAK)2
6	Poly(glycine serine) linker (poly GS)

Part 5:

Name	Designation
1	E0040m
2	mCherry
3	E0040m-His
4	BirA-His
5	BASU-His
6	<i>E coli</i> XylA

7	AN4
8	BN4
9	<i>C thermocellum</i> Dockerin
A	<i>C thermocellum</i> Cohesin
B	PAmCherry
C	<i>C cellulolyticum</i> Dockerin
D	<i>E coli</i> XylE

Part 6 (For pLC3/pLD3 an additional part is needed):

Name	Designation
1	His6
2	AN4
3	BN4
4	His10
5	His5

Appendix D – Growth rates of D-xylose growth assays

Attached are the p-values representing difference between samples and controls for growth rates measured for D-xylose growth assays.

	BW GFP	TDX aXylE	TDX waldo	TDX flex	TDX polyPro	TDX ShArai	TDX poly(GS)
BW GFP	-	0.290138	0.940466	0.612488	0.585897	0.577226	0.113975
TDX GFP	0.006831	0.020413	0.016757	0.012305	0.017987	0.072458	7.68E-05
TDX aXylE	0.290138	-	0.031989	0.054378	0.075014	0.65835	0.876059

Table of student T test p-values for Figure 4.14. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values.

	Δ xylA pLC3	BW pLC3	Δ xylA aXylA	Δ xylA waldo	Δ xylA flex	Δ xylA polyPro	Δ xylA ShArai	Δ xylA Poly(GS)
Δ xylA pLC3	-	0.000643	0.006951	0.986448	0.005898	0.185717	0.061777	0.005397
BW pLC3	0.000643	-	0.010319	0.006675	0.000711	0.259393	0.000475	0.007221
Δ xylA aXylA	0.006951	0.010319	-	0.125112	0.152158	0.742805	0.000348	0.290097

Table of student T test p-values for figure 4.16. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values.

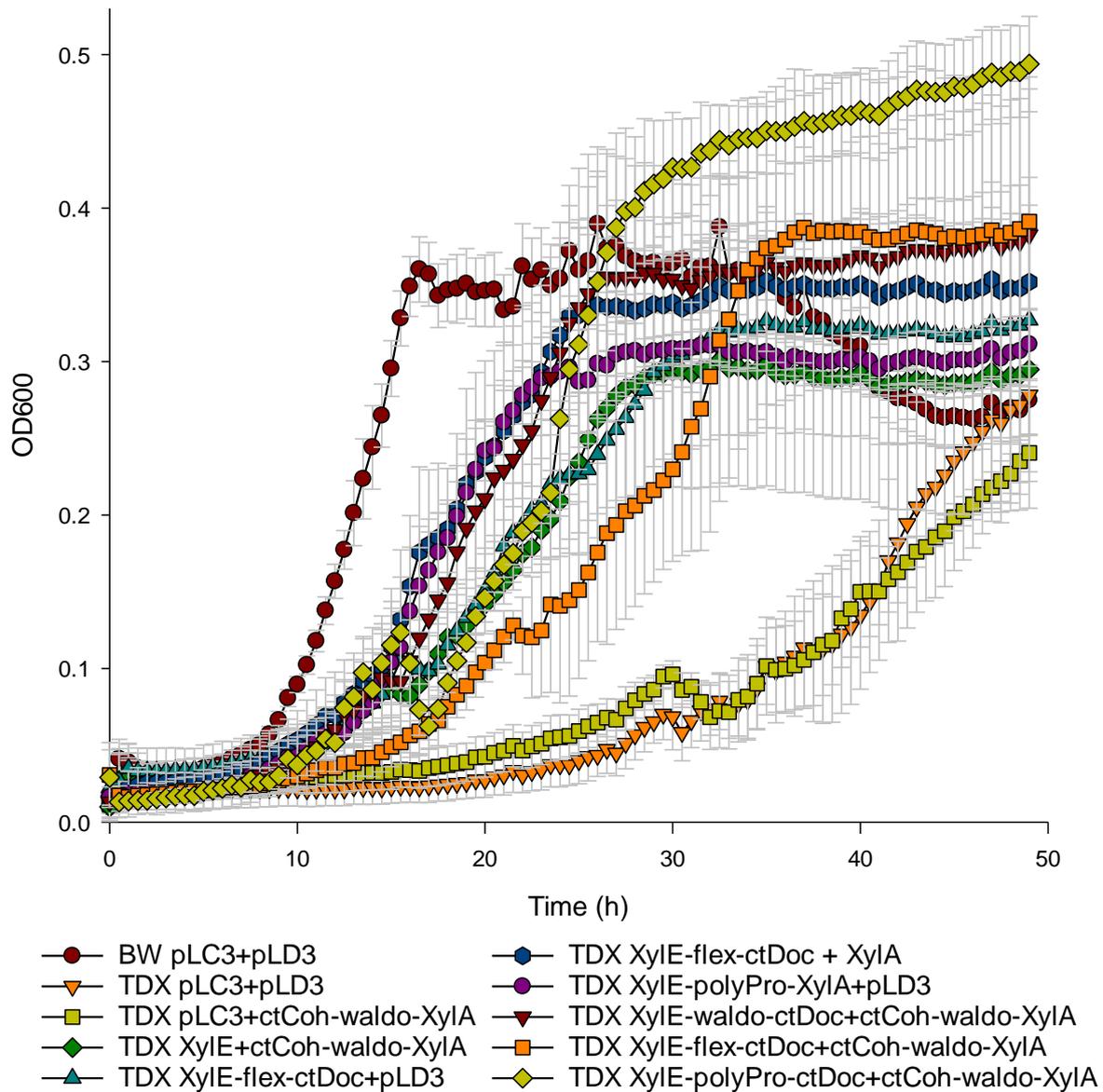
	BW pLD3	Δ xylA pLD3	Δ xylA XylA	Δ xylA ctCoh-Waldo- XylA	Δ xylA ctCoh-Flex- XylA	Δ xylA ctCoh-polyPro- XylA	Δ xylA ctCoh-ShArai- XylA	Δ xylA ctCoh-LnArai- XylA	Δ xylA ctCoh-Poly(GS)- XylA
BW pLD3	-	2.00327E-05	0.378598049	0.000764992	0.078420053	6.43119E-05	0.000201056	0.363088433	0.086665896
Δ xylA pLD3	2.00327E-05	-	0.000347183	6.40758E-05	0.00070358	1.02601E-05	4.57363E-06	0.000601515	7.7047E-05
Δ xylA XylA	0.378598049	0.000347183	-	0.007490512	0.077481526	0.00355099	0.004993712	0.318160331	0.197319707

Table of student T test p-values for maximum growth rates measured in Figure 5.16. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values.

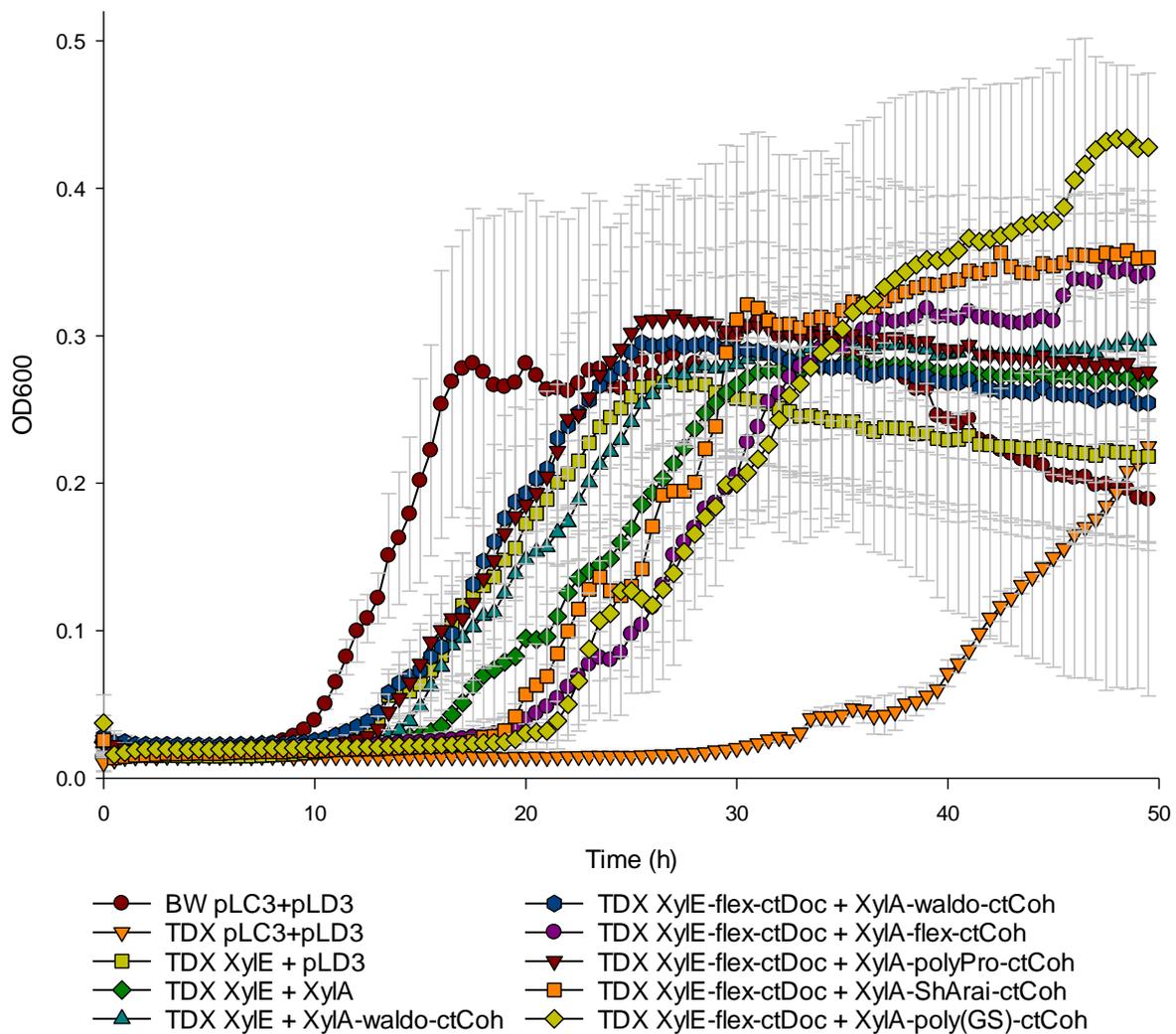
	BW pLD3	Δ xylA pLD3	Δ xylA XylA	Δ xylA XylA- Waldo- ctCoh	Δ xylA XylA-Flex- ctCoh	Δ xylA XylA- polyPro- ctCoh	Δ xylA XylA- ShArai- ctCoh	Δ xylA XylA- Poly(GS)- ctCoh
BW pLD3	-	0.002113	0.001689614	0.000695	0.000152	0.000102	0.011419	0.303947
Δ xylA pLD3	0.002113	-	0.006800174	0.010422	0.007472	0.012708	0.002899	0.000182

ΔxylA	0.00169	0.0068	-	0.183254	0.500106	0.08748	0.148894	0.006963
xylA								

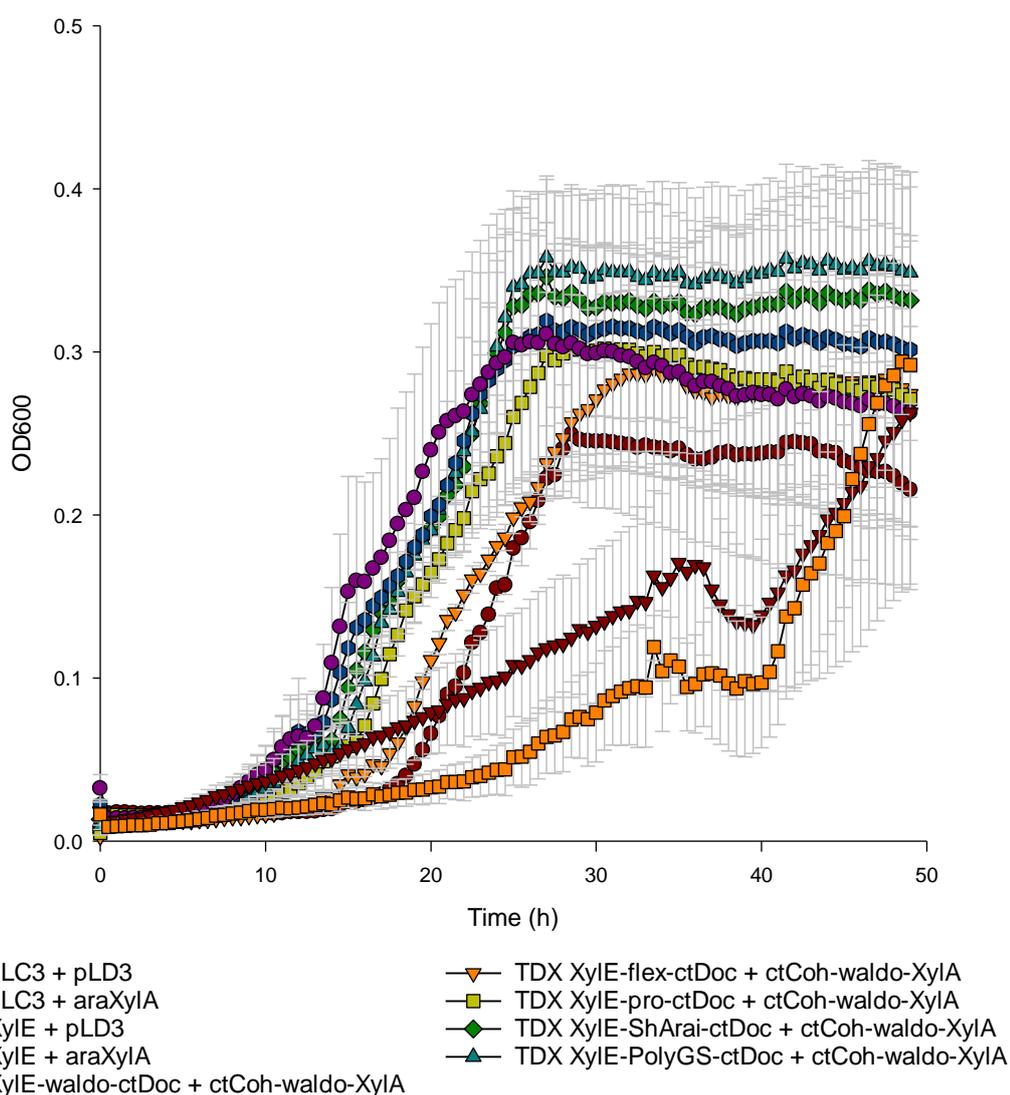
Table of student T test p-values for maximum growth rates measured in Figure 5.17. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values



The impact of linkers in XylE-ctDoc on a growth assay of TDXYlose cells on a low Xylose background when cotransformed alongside XylA-waldo-ctCoh fusions The growth assay shown in Figure 5.19. with all conditions present on a single graph. Deletion mutant TDXYlose was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native XylA or an C-terminal XylA-ctCoh fusion. Variation of the linker in the XylE-ctDoc fusions were expressed within this work. In addition, multiple controls were also expressed to determine the impact of individual components. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 $\mu\text{g}/\text{mL}$ kanamycin + 34 $\mu\text{g}/\text{mL}$ chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. N=5.



The impact of linkers in XylA-ctCoh on recovery of growth of TDXylose cells on a low Xylose background when cotransformed with both XylE-flex-ctDoc fusions The growth assay shown in Figure 5.20. with all conditions present on a single graph. Deletion mutant TDXylose was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native XylA or a C-terminal XylA-ctCoh fusion. Variation of the linker in the XylA-ctCoh fusions were expressed within this work. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 μ g/mL kanamycin + 34 μ g/mL chloramphenicol growth media. A BW25113 strain expressing pLC3 + pLD3 alone was present as a control. N=5.



The impact of linkers in XylE-ctDoc on a growth assay of TDXylose cells cotransformed with both XylE-ctDoc and ctCoh-XylA fusions on a low Xylose background Deletion mutant TDXylose was cotransformed with pLC3 plasmids expressing XylE-ctDoc and pLD3 plasmids expressing either native XylA or an N-terminal ctCoh-XylA fusion. Variation of the linker in the XylE-ctDoc fusions were expressed within this work. The cells were grown for 48 hours on a M9+10 mM Xylose + 60 µg/mL kanamycin + 34 µg/mL chloramphenicol growth media. N=5.

The controls of this figure were used to produce Fig. 5.18, however it was not used beyond this as the figure did not demonstrate any significant variation in the role of linkers compared to Fig. 5.19, therefore was not included in the main text.

	BW pLC3+pLD3	TDX pLC3+pLD3	TDX pLC3+ ctCoh- waldo-XylA	TDX XylE + ctCoh- waldo-XylA	TDX XylE-flex- ctDoc + pLD3	TDX XylE-flex- ctDoc + XylA	TDX XylE- polyPro- XylA + pLD3	TDX XylE-waldo- ctDoc + ctCoh- waldo-XylA	TDX XylE-flex- ctDoc + ctCoh- waldo-XylA	TDX XylE- polypro- ctDoc + ctCoh- waldo-XylA
BW pLC3 + pLD3	-	0.061580772	0.00575949	0.065295876	0.202861854	0.781274271	0.064937841	0.037318785	0.740944536	0.016175
TDX pLC3 + pLD3	0.061580772	-	0.239425146	0.320963093	0.094492317	0.064161002	0.013788935	0.037814303	0.023352425	0.02764
TDX XylE + ctCoh- waldo-XylA	0.065295876	0.320963093	0.02129529	-	0.621433808	0.068012658	0.050912595	0.011907353	0.368728806	0.01953
TDX XylE-flex- ctDoc + XylA	0.781274271	0.064161002	0.018387788	0.068012658	0.157239329	-	0.244843432	0.266561782	0.494152889	0.447659

Table of student T test p-values for maximum growth rates measured in Figure 5.19. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values.

Bw	TDX	TDX	TDX	TDX	TDX	TDX	TDX	TDX	TDX	TDX
pLC3	+ pLC3	+ XylE	XylE	+ XylE	+ XylE-flex-	XylE-flex-	XylE-flex-	XylE-flex-	XylE-flex-	XylE-flex-
pLD3	pLD3		XylA	XylA-	waldo-	ctCoh	XylA-	XylA-Flex-	XylA-	XylA-
				ctCoh	waldo-	ctCoh	ctDoc	ctDoc	polypro-	ctCoh
					ctCoh	ctCoh	ctDoc	ctDoc	ctCoh	ShArai-
							ctCoh	ctCoh	ctCoh	ctCoh
										Poly(GS)-
										ctCoh

<i>Bw pLC3 + pLD3</i>	-	0.0003241 99	0.8928763 42	0.0997039 83	0.6390311 88	0.0933207 18	0.0779204 37	0.6987891 73	0.3845976 62	0.160664
<i>TDX pLC3 + pLD3</i>	0.0003241 99	-	0.0001318 29	0.0002856 5	0.0001699 12	0.0003732 48	0.0073012 36	0.0001982 37	8.8444E-06	9.63E-05
<i>TDX XylE</i>	0.8928763 42	0.0001318 29	-	0.0337185 46	0.5655486 4	0.0485293 94	0.0352939 42	0.5375394 07	0.2504652 42	0.100687
<i>TDX XylE + XylA</i>	0.0997039 83	0.0002856 5	0.0337185 46	-	0.0022408 36	0.9069761 91	0.1103246 87	0.1184497 31	0.0037221 72	0.001333
<i>TDX XylE + XylA-waldo-ctCoh</i>	0.6390311 88	0.0001699 12	0.5655486 4	0.0022408 36	-	0.0945763 46	0.0174975 72	0.9310627 47	0.1121455 4	0.004205

Table of student T test p-values for Figure 5.20. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values.

Bw pLC3 + pLD3	$\Delta xyIA$ pLC3 pLD3	+	$\Delta xyIA$ XylE	$\Delta xyIA$ XylA	$\Delta xyIA$ XylE + XylA	$\Delta xyIA$ XylE- waldo- ctDoc XylA- waldo- ctCoh	$\Delta xyIA$ XylE- waldo- ctDoc XylA-Flex- ctCoh	$\Delta xyIA$ XylE- waldo- ctDoc XylA- polypro- ctCoh	$\Delta xyIA$ XylE- waldo- ctDoc XylA- ShArai- ctCoh	$\Delta xyIA$ XylE- waldo- ctDoc XylA- Poly(GS)- ctCoh
---------------------------	---	----------	--------------------------------------	--	---	---	---	---	--	--

<i>Bw pLC3 + pLD3</i>	-	1.05671E- 05	3.43869E- 05	0.0006895 85	0.6754069 81	0.0019878 04	0.3192522 59	0.7898562 46	0.0105365 2	0.00013
<i>$\Delta xyIA$ pLC3 + pLD3</i>	1.05671E- 05	-	0.2396905 77	3.71596E- 05	0.0001392 75	3.15423E- 05	0.0017299 24	0.0009452 75	0.0001086 38	1.84E-06
<i>$\Delta xyIA$ XylE</i>	3.43869E- 05	0.2396905 77	-	0.0004054 52	0.0007635 61	0.0023649 19	0.0042600 57	0.0036901 85	0.0016417 04	0.000985
<i>$\Delta xyIA$ XylA</i>	0.0006895 85	3.71596E- 05	0.0004054 52		0.0129477 32	0.5253111 72	0.0211219 39	0.0324038 3	0.1606962 27	0.031679
<i>$\Delta xyIA$ XylE + XylA</i>	0.6754069 81	0.0001392 75	0.0007635 61	0.0129477 32	-	0.0030666 74	0.0992303 64	0.9901460 92	0.0132977 18	0.005475

Table of student T test p-values for figure 5.21. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values

	Bw pLC3 + pLD3	ΔxylA pLC3 + pLD3	ΔxylA XylE + XylA	ΔxylA XylE-waldo-ctDoc + XylA-waldo-ctCoh	ΔxylA XylE-waldo-ctDoc + ctCoh-waldo-XylA
<i>Bw pLC3 + pLD3</i>	-	9.13643E-06	0.389081051	0.000803595	0.000336322
<i>ΔxylA pLC3 + pLD3</i>	9.13643E-06	-	0.000198464	3.62437E-06	9.67841E-05
<i>ΔxylA XylE + XylA</i>	0.389081051	0.000198464	-	0.080010188	0.001222412
<i>ΔxylA XylE-waldo-ctDoc + XylA-waldo-ctCoh</i>	0.000803595	3.62437E-06	0.080010188		0.004983292
<i>ΔxylA XylE-waldo-ctDoc + XylA-polypro-ctCoh</i>	0.007550147	1.19472E-06	0.096945758	0.183085603	0.002110506
<i>ΔxylA XylE-waldo-ctDoc + XylA-poly(GS)-ctCoh</i>	0.013236932	0.000118129	0.148819009	0.870347541	0.029103636
<i>ΔxylA XylE-waldo-ctDoc + ctCoh-waldo-XylA</i>	0.000336322	9.67841E-05	0.001222412	0.004983292	-

$\Delta xyIA$ <i>XylE-waldo-ctDoc</i> + <i>ctCoh-polypro-XylA</i>	2.56882E-05	7.47677E-06	0.003553377	9.68286E-05	0.129184601
$\Delta xyIA$ <i>XylE-waldo-ctDoc</i> + <i>ctCoh-poly(GS)-XylA</i>	8.91556E-05	2.17051E-05	0.002468063	0.001000705	0.29821883
$\Delta xyIA$ <i>XylE-waldo-XylA</i>	0.00018402	0.076142258	0.003479035	0.00042528	0.29821883
$\Delta xyIA$ <i>XylE-polyPro-XylA</i>	1.93501E-05	0.020302232	0.000579028	8.93971E-06	0.0203772
$\Delta xyIA$ <i>XylE-poly(GS)-XylA</i>	1.13041E-05	3.52635E-05	0.000909793	5.03432E-06	0.002014954

Table of student T test p-values for figure 5.22. A series of two-tailed student T tests were applied to the mean growth rates of each fusion compared to the control values

Appendix E – SEC-MALLS data with time in X-axis

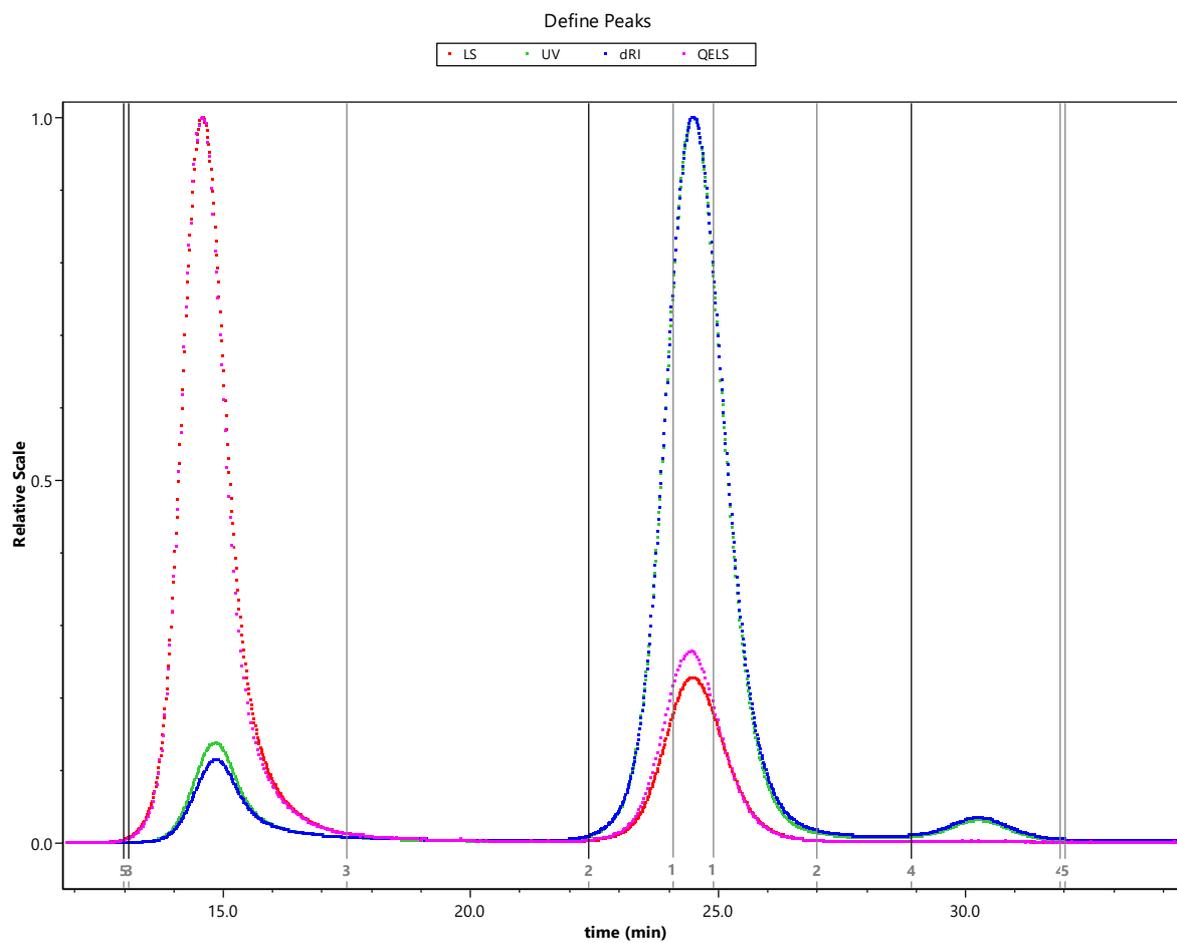


Figure 4.2. with original-axis of time vs relative scale. Figure 4.2 was adjusted to show the elution volume (mL) as opposed elution time (min). This image displays the original file, which was used to calculate the peaks analysed. These peaks are represented as the lines across the image, and correspond with 1-5 in the text of section 4.1.1.

Appendix F – Data collection and refinement table of *E. coli* XylA

Data Collection

Source	Diamond Light Source, UK
Beamline	i03
Wavelength (Å)	0.9763
Type	Synchrotron
Detector	Eiger2 XE 16M

Data Statistics

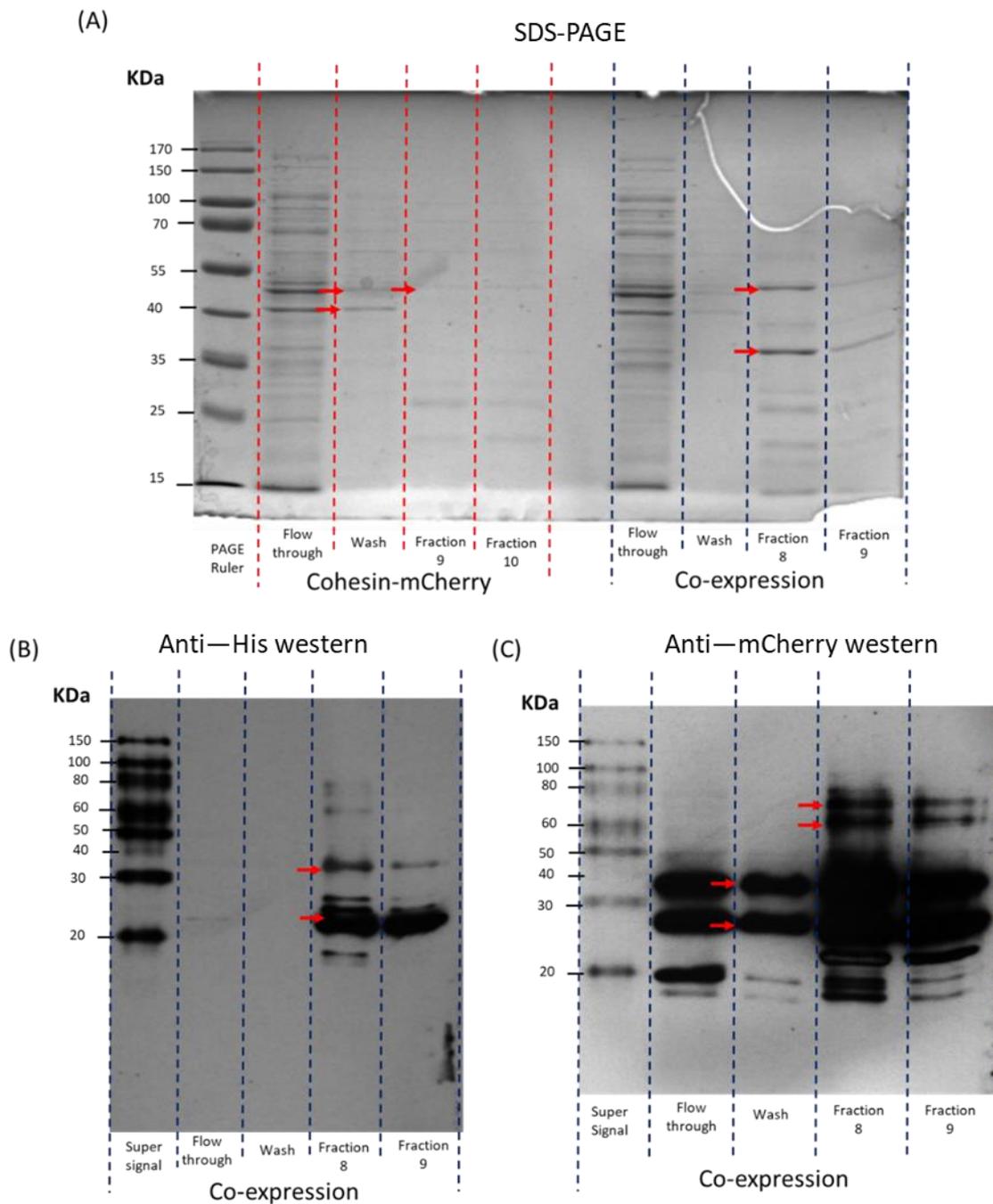
Space group	P 1 2 ₁ 1
Cell constants a, b, c (Å)	96.17, 91.64, 111.17
Cell constants α , β , γ (°)	90.00, 101.57, 90.00
Overall resolution range (Å)	54.46 - 2.13
Inner shell resolution range (Å)	54.46 - 11.67
Outer shell resolution range (Å)	2.17 - 2.13
Overall completeness (%)	83.9
R_{merge} for all I+ and I- ^a	0.104
$\langle I/\sigma(I) \rangle$	0.41 (at 2.12Å)
Overall multiplicity	6.9

Refinement Statistics

Program	REFMAC
Number of observations	733523
Number of reflections, Free R set	105714, 5287
R, Rfree ^{b,c}	0.22, 0.28
Total number of atoms	26155
Average B, all atoms (Å ²)	60
RMS Deviations – Bonds ^d	0.0083
RMS Deviations – Angles ^d	1.726
Ramachandran plot ^e	0.60%

Data collection and refinement table for Crystal structure of XylA Table of data collection, data statistics and refinement statistics for XylA. Superscript footnotes: **(a)** $R_{merge} = \frac{\sum_{HKL} \sum_i |I_i - \langle I \rangle|}{\sum_{HKL} \sum_i \langle I \rangle}$ where I_i is the intensity of the i th measurement of a reflection with indexes hkl and $\langle I \rangle$ is the Statistically Weighted Average Reflection Intensity. **(b)** R-factor = $\frac{\sum ||F_o| - |F_c||}{\sum |F_o|}$ where F_o and F_c are the observed and calculated structure factor amplitudes, respectively. **(c)** R-free is the R-factor calculated with 5 % of the reflections chosen at random and omitted from refinement. **(d)** Root-mean-square deviation of bond lengths and bond angles from ideal geometry. **(e)** Percentage of residues in outlier regions.

Appendix G – SDS-PAGE and western blots of a test of the coexpression pulldown assay.



SDS-PAGE and Western Blot of cells expressing mCherry-ctCoh and GFP-ctDoc-His following nickel affinity chromatography. Cells transformed with pLC3_11A191 (GFP-ctDoc-His) + pLD1_31712 (encoding mCherry-waldo-ctCoh) were sonicated and the supernatant run through a nickel affinity column, with 1 mL eluate fraction collected. Regions of eluate were then obtained and run through an SDS-PAGE gel alongside an aliquot of the flowthrough and wash. Samples were diluted 10-fold due to high concentration of the mCherry

in flowthrough overpowering results. **(A)** SDS-PAGE gel of both ctCoh-mCherry expressed in isolation (seen in figure 5.9A) and the coexpression of ctCoh-mCherry and GFP-ctDoc-His – ladder was PAGE Ruler **(B)** Anti-tetraHis western blot – ladder was SuperSignal™ Molecular Weight Protein Ladder, which has since been determined to be inaccurate in terms of mass.- Red arrows identify bands expected to be the GFP-ctDoc-His (lower mass) and the ctCoh-mCherry (upper mass) which was highly overexpressed and therefore may have increased background **(C)** Anti-mCherry western blot - was SuperSignal™ Molecular Weight Protein Ladder, which has since been determined to be inaccurate in terms of mass. Was sample believed to contain both the complete ctCoh-mCherry fusion (upper arrow) as well as lysed mCherry (lower arrow). Within the fractions there are large bands selected by the red arrow that were either background or a complete scaffold complex that did not degrade during SDS-PAGE. At this time we believe it to be the prior, as such behaviour has not been replicated in future pulldown assays.

References

- Aalbers, F. S. and Fraaije, M. W. (2017) 'Coupled reactions by coupled enzymes: alcohol to lactone cascade with alcohol dehydrogenase–cyclohexanone monooxygenase fusions', *Applied Microbiology and Biotechnology*, 101(20), pp. 7557–7565. doi: 10.1007/s00253-017-8501-4.
- Aalbers, F. S. and Fraaije, M. W. (2019) 'Enzyme Fusions in Biocatalysis: Coupling Reactions by Pairing Enzymes', *ChemBioChem*, 20(1), pp. 20–28. doi: 10.1002/cbic.201800394.
- Adlakha, N. *et al.* (2011) 'Synthesis and characterization of chimeric proteins based on cellulase and xylanase from an insect gut bacterium', *Applied and Environmental Microbiology*, 77(14), pp. 4859–4866. doi: 10.1128/AEM.02808-10.
- Adlakha, N. *et al.* (2012) 'Specific fusion of β -1,4-endoglucanase and β -1,4-glucosidase enhances cellulolytic activity and helps in channeling of intermediates', *Applied and Environmental Microbiology*, 78(20), pp. 7447–7454. doi: 10.1128/AEM.01386-12.
- Agapakis, C. M. *et al.* (2010) 'Insulation of a synthetic hydrogen metabolism circuit in bacteria', *Journal of Biological Engineering*, 4(1), p. 3. doi: 10.1186/1754-1611-4-3.
- Agrawal, M., Mao, Z. and Chen, R. R. (2011) 'Adaptation yields a highly efficient xylose-fermenting *Zymomonas mobilis* strain', *Biotechnology and Bioengineering*, 108(4), pp. 777–785. doi: 10.1002/bit.23021.
- Ahmad, S. and Scopes, R. K. (2002) 'Isolation and properties of a constitutive D-xylulokinase from a novel thermophilic *Saccharococcus caldxylosilyticus* DSM 12041 (ATCC 700356)', *Enzyme and Microbial Technology*, 30(5), pp. 627–632. doi: 10.1016/S0141-0229(01)00518-X.
- Aidelberg, G. *et al.* (2014) 'Hierarchy of non-glucose sugars in *Escherichia coli*', *BMC Systems Biology*, 8(1), p. 133. doi: 10.1186/s12918-014-0133-z.
- Albertsen, L. *et al.* (2011) 'Diversion of flux toward sesquiterpene production in *Saccharomyces cerevisiae* by fusion of host and heterologous enzymes', *Applied and Environmental Microbiology*, 77(3), pp. 1033–1040. doi: 10.1128/AEM.01361-10.
- Alm, E. J. *et al.* (2005) 'The MicrobesOnline Web site for comparative genomics', *Genome Research*, 15(7), pp. 1015–1022. doi: 10.1101/gr.3844805.
- Amer, M. *et al.* (2020) 'Low carbon strategies for sustainable bio-alkane gas production and renewable energy', *Energy and Environmental Science*, 13(6), pp. 1818–1831. doi:

10.1039/d0ee00095g.

Amet, N., Lee, H. F. and Shen, W. C. (2009) 'Insertion of the designed helical linker led to increased expression of Tf-based fusion proteins', *Pharmaceutical Research*, 26(3), pp. 523–528. doi: 10.1007/s11095-008-9767-0.

Ammar, E. M., Wang, X. and Rao, C. V. (2018) 'Regulation of metabolism in *Escherichia coli* during growth on mixtures of the non-glucose sugars: Arabinose, lactose, and xylose', *Scientific Reports*, 8(1), pp. 1–11. doi: 10.1038/s41598-017-18704-0.

Anandharaj, M. *et al.* (2020) 'Constructing a yeast to express the largest cellulosome complex on the cell surface', *Proceedings of the National Academy of Sciences of the United States of America*, 117(5), pp. 2385–2394. doi: 10.1073/pnas.1916529117.

Anderlund, M., Rådström, P. and Hahn-Hägerdal, B. (2001) 'Expression of bifunctional enzymes with xylose reductase and xylitol dehydrogenase activity in *Saccharomyces cerevisiae* alters product formation during xylose fermentation', *Metabolic Engineering*, 3(3), pp. 226–235. doi: 10.1006/mben.2001.0190.

Arai, R. *et al.* (2001) 'Design of the linkers which effectively separate domains of a bifunctional fusion protein.', *Protein engineering*, 14(8), pp. 529–32. doi: 10.1093/PROTEIN/14.8.529.

De Araujo, C. *et al.* (2014) 'Identification and characterization of a carboxysomal γ -carbonic anhydrase from the cyanobacterium *Nostoc sp.* PCC 7120', in *Photosynthesis Research*. Kluwer Academic Publishers, pp. 135–150. doi: 10.1007/s11120-014-0018-4.

Argos, P. (1990) 'An investigation of oligopeptides linking domains in protein tertiary structures and possible candidates for general gene fusion', *Journal of Molecular Biology*, 211(4), pp. 943–958. doi: 10.1016/0022-2836(90)90085-Z.

Aristidou, A. and Penttilä, M. (2000) 'Metabolic engineering applications to renewable resource utilization', *Current Opinion in Biotechnology*, 11(2), pp. 187–198. doi: 10.1016/S0958-1669(00)00085-9.

Artzi, L., Bayer, E. A. and Moraïs, S. (2016) 'Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides', *Nature Reviews Microbiology*, 15, pp. 83–95. doi: 10.1038/nrmicro.2016.164.

Baba, T. *et al.* (2006) 'Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection', *Molecular Systems Biology*, 2, p. 2006.0008. doi: 10.1038/msb4100050.

Babu, M. *et al.* (2010) 'Structure of a SLC26 Anion Transporter STAS Domain in Complex with Acyl Carrier Protein: Implications for *E. coli* YchM in Fatty Acid Metabolism', *Structure*, 18(11), pp. 1450–1462. doi: 10.1016/j.str.2010.08.015.

Bakshi, S., Bratton, B. P. and Weisshaar, J. C. (2011) 'Subdiffraction-limit study of Kaede diffusion and spatial distribution in live *Escherichia coli*', *Biophysical Journal*, 101(10), pp. 2535–2544. doi: 10.1016/j.bpj.2011.10.013.

Bandhu, S. *et al.* (2018) 'Single Cell Oil from Oleaginous Yeast Grown on Sugarcane Bagasse-Derived Xylose: An Approach toward Novel Biolubricant for Low Friction and Wear', *ACS Sustainable Chemistry and Engineering*, 6(1), pp. 275–283. doi: 10.1021/acssuschemeng.7b02425.

Van Bastelaere, P. B. M. *et al.* (1992) Binding characteristics of Mn²⁺, Co²⁺ and Mg²⁺ ions with several D-xylose isomerases, *Biochem. J.* Available at: <https://portlandpress.com/biochemj/article-pdf/286/3/729/607939/bj2860729.pdf> (Accessed: 31 January 2020).

Bateman, A. (2019) 'UniProt: A worldwide hub of protein knowledge', *Nucleic Acids Research*, 47(D1), pp. D506–D515. doi: 10.1093/nar/gky1049.

Batt, C. A., Jamieson, A. C. and Vandeyar, M. A. (1990) 'Identification of essential histidine residues in the active site of *Escherichia coli* xylose (glucose) isomerase', *Biochemistry*, 87(2), pp. 618–622. doi: 10.1073/pnas.87.2.618.

Bauler, P. *et al.* (2010) 'Channeling by proximity: The catalytic advantages of active site colocalization using brownian dynamics', *Journal of Physical Chemistry Letters*, 1(9), pp. 1332–1335. doi: 10.1021/jz1002007.

Bayer, E. A. *et al.* (2004) 'The Cellulosomes: Multienzyme Machines for Degradation of Plant Cell Wall Polysaccharides', *Annual Review of Microbiology*, 58(1), pp. 521–554. doi: 10.1146/annurev.micro.57.030502.091022.

Bayer, E. A., Morag, E. and Lamed, R. (1994) 'The cellulosome — A treasure-trove for biotechnology', *Trends in Biotechnology*, 12(9), pp. 379–386. doi: 10.1016/0167-7799(94)90039-6.

Beale, J. H. *et al.* (2015) 'Crystal Structures of the Extracellular Domain from PepT1 and PepT2 Provide Novel Insights into Mammalian Peptide Transport', *Structure*, 23(10), pp. 1889–1899. doi: 10.1016/j.str.2015.07.016.

Bebenek, K. and Kunkel, T. A. (1995) '[18] Analyzing fidelity of DNA polymerases', *Methods in Enzymology*, 262, pp. 217–232. doi: 10.1016/0076-6879(95)62020-6.

Beedessee, G. *et al.* (2015) 'Multifunctional polyketide synthase genes identified by genomic survey of the symbiotic dinoflagellate, *Symbiodinium minutum*', *BMC Genomics*, 16(1), p. 941. doi: 10.1186/s12864-015-2195-8.

Behrendorff, J. B. Y. H., Borràs-Gas, G. and Pribil, M. (2020) 'Synthetic Protein Scaffolding at Biological Membranes', *Trends in Biotechnology*, 38(4), pp. 432–446. doi: 10.1016/j.tibtech.2019.10.009.

Belardinelli, J. M. and Jackson, M. (2017) 'Green Fluorescent Protein as a protein localization and topological reporter in mycobacteria', *Tuberculosis*, 105, pp. 13–17. doi: 10.1016/j.tube.2017.04.001.

Benner, S. A. and Sismour, A. M. (2005) 'Synthetic biology', *Nature Reviews Genetics*, 6(7), pp. 533–543. doi: 10.1038/nrg1637.

Bergeron, L. M. *et al.* (2009) 'Self-renaturing enzymes: Design of an enzyme-chaperone chimera as a new approach to enzyme stabilization', *Biotechnology and Bioengineering*, 102(5), pp. 1316–1322. doi: 10.1002/bit.22254.

Berman, H. M. *et al.* (2000) 'The Protein Data Bank', *Nucleic Acids Research*. Oxford University Press, pp. 235–242. doi: 10.1093/nar/28.1.235.

Berry, A. *et al.* (2002) 'Application of metabolic engineering to improve both the production and use of biotech indigo', *Journal of Industrial Microbiology & Biotechnology*, 28(3), pp. 127–133. doi: 10.1038/sj/jim/7000228.

Bhatia, S. K. *et al.* (2017) 'Current status and strategies for second generation biofuel production using microbial systems', *Energy Conversion and Management*. Elsevier Ltd, pp. 1142–1156. doi: 10.1016/j.enconman.2017.06.073.

Bhosale, S. H., Rao, M. B. and Deshpande, V. V (1996) *Molecular and Industrial Aspects of Glucose Isomerase*, *MICROBIOLOGICAL REVIEWS*.

Biemans-Oldehinkel, E., Doeven, M. K. and Poolman, B. (2006) 'ABC transporter architecture and regulatory roles of accessory domains', *FEBS Letters*, 580(4), pp. 1023–1035. doi: 10.1016/j.febslet.2005.11.079.

Bird, R. *et al.* (1988) 'Single-chain antigen-binding proteins', *Science*, 242(4877), pp. 423–426. doi: 10.1126/science.3140379.

- Black, C. B., Huang, H.-W. and Cowan, J. A. (1994) 'Biological coordination chemistry of magnesium, sodium, and potassium ions. Protein and nucleotide binding sites', *Coordination Chemistry Reviews*, 135–136(C), pp. 165–202. doi: 10.1016/0010-8545(94)80068-5.
- Boado, R. J. *et al.* (2008) 'IgG-Paraoxonase-1 fusion protein for targeted drug delivery across the human blood-brain barrier', *Molecular Pharmaceutics*, 5(6), pp. 1037–1043. doi: 10.1021/mp800113g.
- Boado, R. J. *et al.* (2014) 'Insulin receptor antibody-sulfamidase fusion protein penetrates the primate blood-brain barrier and reduces glycosaminoglycans in sanfilippo type a cells', *Molecular Pharmaceutics*, 11(8), pp. 2928–2934. doi: 10.1021/mp500258p.
- Boer, H. *et al.* (2019) 'In vitro reconstitution and characterisation of the oxidative d-xylose pathway for production of organic acids and alcohols', *AMB Express*, 9(1), pp. 1–13. doi: 10.1186/s13568-019-0768-7.
- Bolhuis, A. *et al.* (1998) 'SecDF of *Bacillus subtilis*, a molecular Siamese twin required for the efficient secretion of proteins.', *The Journal of biological chemistry*, 273(33), pp. 21217–24. doi: 10.1074/JBC.273.33.21217.
- Bond-Watts, B. B., Bellerose, R. J. and Chang, M. C. Y. (2011) 'Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways', *Nature Chemical Biology*, 7(4), pp. 222–227. doi: 10.1038/nchembio.537.
- Boretti, A. (2012) 'Towards 40% efficiency with BMEP exceeding 30 bar in directly injected, turbocharged, spark ignition ethanol engines', *Energy Conversion and Management*, 57, pp. 154–166. doi: 10.1016/j.enconman.2011.12.011.
- Bouvier, J. *et al.* (2000) 'Characterization of OpuA, a glycine-betaine uptake system of *Lactococcus lactis*.' , *Journal of molecular microbiology and biotechnology*, 2(2), pp. 199–205. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10939245> (Accessed: 20 May 2019).
- Brett, M. (1989) 'A novel gonococcal β -lactamase plasmid', *Journal of Antimicrobial Chemotherapy*, pp. 653–654. doi: 10.1093/jac/23.4.653.
- Brilli, M. and Fani, R. (2004) 'The origin and evolution of eucaryal HIS7 genes: From metabolon to bifunctional proteins?', *Gene*, 339(1–2), pp. 149–160. doi: 10.1016/j.gene.2004.06.033.
- Brown, M. D. and Sacks, D. B. (2009) 'Protein scaffolds in MAP kinase signalling', *Cellular Signalling*, 21(4), pp. 462–469. doi: 10.1016/j.cellsig.2008.11.013.

Bülow, L., Ljungcrantz, P. and Mosbach, K. (1985) 'Preparation of a soluble bifunctional enzyme by gene fusion', *Bio/Technology*, 3(9), pp. 821–823. doi: 10.1038/nbt0985-821.

Bülow, L. and Mosbach, K. (1991) 'Multienzyme systems obtained by gene fusion', *Trends in Biotechnology*, 9(1), pp. 226–231. doi: 10.1016/0167-7799(91)90075-S.

Bunker, R. D. *et al.* (2013) 'Structure and function of human xylulokinase, an enzyme with important roles in carbohydrate metabolism', *Journal of Biological Chemistry*, 288(3), pp. 1643–1652. doi: 10.1074/jbc.M112.427997.

Butterfield, G. L. *et al.* (2017) 'Evolution of a designed protein assembly encapsulating its own RNA genome', *Nature*, 552(7685), pp. 415–420. doi: 10.1038/nature25157.

Cam, Y. *et al.* (2016) 'Engineering of a Synthetic Metabolic Pathway for the Assimilation of (d)-Xylose into Value-Added Chemicals', *ACS Synthetic Biology*, 5(7), pp. 607–618. doi: 10.1021/acssynbio.5b00103.

Cannon, K. A. *et al.* (2020) 'Design and Characterization of an Icosahedral Protein Cage Formed by a Double-Fusion Protein Containing Three Distinct Symmetry Elements', *ACS Synthetic Biology*, 9(3), pp. 517–524. doi: 10.1021/acssynbio.9b00392.

Care, A. *et al.* (2017) 'Solid-binding peptides for immobilisation of thermostable enzymes to hydrolyse biomass polysaccharides', *Biotechnology for Biofuels*, 10(1), p. 29. doi: 10.1186/s13068-017-0715-2.

Carriquiry, M. A., Du, X. and Timilsina, G. R. (2011) 'Second generation biofuels: Economics and policies', *Energy Policy*, 39(7), pp. 4222–4234. doi: 10.1016/j.enpol.2011.04.036.

Carvalho, A. L. *et al.* (2003) 'Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex.', *Proceedings of the National Academy of Sciences of the United States of America*, 100(24), pp. 13809–14. doi: 10.1073/pnas.1936124100.

Castellana, M. *et al.* (2014) 'Enzyme clustering accelerates processing of intermediates through metabolic channeling', *Nature Biotechnology*, 32(10), pp. 1011–1018. doi: 10.1038/nbt.3018.

Chandel, A. K. *et al.* (2018) 'The path forward for lignocellulose biorefineries: Bottlenecks, solutions, and perspective on commercialization', *Bioresource Technology*. Elsevier Ltd, pp. 370–381. doi: 10.1016/j.biortech.2018.06.004.

Chandel, A. K. and Singh, O. V. (2011) 'Weedy lignocellulosic feedstock and microbial metabolic engineering: Advancing the generation of "Biofuel"', *Applied Microbiology and*

Biotechnology. Springer, pp. 1289–1303. doi: 10.1007/s00253-010-3057-6.

Chang, A. *et al.* (2021) 'BRENDA, the ELIXIR core data resource in 2021: New developments and updates', *Nucleic Acids Research*, 49(D1), pp. D498–D508. doi: 10.1093/nar/gkaa1025.

Chaves, J. E. *et al.* (2017) 'Engineering Isoprene Synthase Expression and Activity in Cyanobacteria', *ACS Synthetic Biology*, 6(12), pp. 2281–2292. doi: 10.1021/acssynbio.7b00214.

Chen, A. H. and Silver, P. A. (2012) 'Designing biological compartmentalization', *Trends in Cell Biology*, 22(12), pp. 662–670. doi: 10.1016/j.tcb.2012.07.002.

Chen, C. and Beattie, G. A. (2007) 'Characterization of the osmoprotectant transporter OpuC from *Pseudomonas syringae* and demonstration that cystathionine- β -synthase domains are required for its osmoregulatory function', *Journal of Bacteriology*, 189(19), pp. 6901–6912. doi: 10.1128/JB.00763-07.

Chen, H. *et al.* (2017) 'Effect of Linker Length and Flexibility on the *Clostridium thermocellum* Esterase Displayed on *Bacillus subtilis* Spores', *Applied Biochemistry and Biotechnology*, 182(1), pp. 168–180. doi: 10.1007/s12010-016-2318-y.

Chen, X., Zaro, J. L. and Shen, W.-C. (2013) 'Fusion protein linkers: Property, design and functionality', *Advanced Drug Delivery Reviews*, 65(10), pp. 1357–1369. doi: 10.1016/J.ADDR.2012.09.039.

Chen, X., Zhu, P. and Liu, L. (2016) 'Modular optimization of multi-gene pathways for fumarate production', *Metabolic Engineering*, 33, pp. 76–85. doi: 10.1016/j.ymben.2015.07.007.

Chiasson, D. *et al.* (2019) 'A unified multi-kingdom Golden Gate cloning platform', *Scientific Reports*, 9(1), pp. 1–12. doi: 10.1038/s41598-019-46171-2.

Chin, W. C. *et al.* (2013) 'Improvement of n-butanol tolerance in *Escherichia coli* by membrane-targeted tilapia metallothionein', *Biotechnology for Biofuels*, 6(1), pp. 1–9. doi: 10.1186/1754-6834-6-130.

Chockalingam, K. *et al.* (2020) 'Golden Gate assembly with a bi-directional promoter (GBid): A simple, scalable method for phage display Fab library creation', *Scientific Reports*, 10(1), pp. 1–14. doi: 10.1038/s41598-020-59745-2.

Choudhary, V., Caratzoulas, S. and Vlachos, D. G. (2013) 'Insights into the isomerization of xylose to xylulose and lyxose by a Lewis acid catalyst', *Carbohydrate Research*, 368, pp. 89–95. doi: 10.1016/j.carres.2012.12.019.

Collyer, C. A., Henrick, K. and Blow, D. M. (1990) 'Mechanism for aldose-ketose interconversion by d-xylose isomerase involving ring opening followed by a 1,2-hydride shift', *Journal of Molecular Biology*, 212(1), pp. 211–235. doi: 10.1016/0022-2836(90)90316-E.

Colpa, D. I. *et al.* (2017) 'Creating Oxidase-Peroxidase Fusion Enzymes as a Toolbox for Cascade Reactions', *ChemBioChem*, 18(22), pp. 2226–2230. doi: 10.1002/cbic.201700478.

Conant, G. C. and Wagner, A. (2005) 'The rarity of gene shuffling in conserved genes.', *Genome biology*, 6(6), p. R50. doi: 10.1186/gb-2005-6-6-r50.

Conrado, R. J. *et al.* (2012) 'DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency', *Nucleic Acids Research*, 40(4), pp. 1879–1889. doi: 10.1093/nar/gkr888.

Conrado, R. J., Varner, J. D. and DeLisa, M. P. (2008) 'Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy', *Current Opinion in Biotechnology*, 19(5), pp. 492–499. doi: 10.1016/j.copbio.2008.07.006.

Cronan, J. E. (2005) 'Targeted and proximity-dependent promiscuous protein biotinylation by a mutant *Escherichia coli* biotin protein ligase', in *Journal of Nutritional Biochemistry*. Elsevier, pp. 416–418. doi: 10.1016/j.jnutbio.2005.03.017.

Cunha, J. T. *et al.* (2019) 'Xylose fermentation efficiency of industrial *Saccharomyces cerevisiae* yeast with separate or combined xylose reductase/xylitol dehydrogenase and xylose isomerase pathways', *Biotechnology for Biofuels*, 12(1), p. 20. doi: 10.1186/s13068-019-1360-8.

Datsenko, K. A. and Wanner, B. L. (2000) 'One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products', *Proceedings of the National Academy of Sciences*, 97(12), pp. 6640–6645. doi: 10.1073/pnas.120163297.

David, J. D. and Weismeyer, H. (1970) 'Control of xylose metabolism in *Escherichia coli*', *BBA - General Subjects*, 201(3), pp. 497–499. doi: 10.1016/0304-4165(70)90171-6.

Deich, J. *et al.* (2004) 'Visualization of the movement of single histidine kinase molecules in live *Caulobacter* cells', *Proceedings of the National Academy of Sciences of the United States of America*, 101(45), pp. 15921–15926. doi: 10.1073/pnas.0404200101.

Delebecque, C. J. *et al.* (2011) 'Organization of intracellular reactions with rationally designed RNA assemblies', *Science*, 333(6041), pp. 470–474. doi: 10.1126/science.1206938.

Delebecque, C. J., Silver, P. A. and Lindner, A. B. (2012) 'Designing and using RNA scaffolds

to assemble proteins in vivo', *Nature Protocols*, 7(10), pp. 1797–1807. doi: 10.1038/nprot.2012.102.

DeLisa, M. P. and Conrado, R. J. (2009) 'Synthetic metabolic pipelines', *Nature Biotechnology*, 27(8), pp. 728–729. doi: 10.1038/nbt0809-728.

Deromachi, Y. *et al.* (2020) 'Stable expression of bacterial transporter ArsB attached to SNARE molecule enhances arsenic accumulation in *Arabidopsis*', *Plant Signaling and Behavior*, 15(11), p. 1802553. doi: 10.1080/15592324.2020.1802553.

Desai, T. A. and Rao, C. V (2010) 'Regulation of arabinose and xylose metabolism in *Escherichia coli*.', *Applied and environmental microbiology*, 76(5), pp. 1524–32. doi: 10.1128/AEM.01970-09.

Dien, B. S., Cotta, M. A. and Jeffries, T. W. (2003) 'Bacteria engineered for fuel ethanol production: Current status', *Applied Microbiology and Biotechnology*. Springer, pp. 258–266. doi: 10.1007/s00253-003-1444-y.

Doolittle, R. F. (1995) 'The Multiplicity of Domains in Proteins', *Annual Review of Biochemistry*, 64(1), pp. 287–314. doi: 10.1146/annurev.bi.64.070195.001443.

Dosset, P. *et al.* (2016) 'Automatic detection of diffusion modes within biological membranes using back-propagation neural network'. doi: 10.1186/s12859-016-1064-z.

Drozdetskiy, A. *et al.* (2015) 'JPred4: A protein secondary structure prediction server', *Nucleic Acids Research*, 43(W1), pp. W389–W394. doi: 10.1093/nar/gkv332.

Du, D. *et al.* (2015) 'Structure, mechanism and cooperation of bacterial multidrug transporters', *Current Opinion in Structural Biology*. Elsevier Ltd, pp. 76–91. doi: 10.1016/j.sbi.2015.07.015.

Dueber, J. E. *et al.* (2009) 'Synthetic protein scaffolds provide modular control over metabolic flux', *Nature Biotechnology*, 27(8), pp. 753–759. doi: 10.1038/nbt.1557.

Durrens, P., Nikolski, M. and Sherman, D. (2008) 'Fusion and fission of genes define a metric between fungal genomes', *PLoS Computational Biology*, 4(10). doi: 10.1371/journal.pcbi.1000200.

Edelstein, A. *et al.* (2010) 'Computer control of microscopes using manager', *Current Protocols in Molecular Biology*. Curr Protoc Mol Biol. doi: 10.1002/0471142727.mb1420s92.

Elf, J., Li, G. W. and Xie, X. S. (2007) 'Probing transcription factor dynamics at the single-molecule level in a living cell', *Science*, 316(5828), pp. 1191–1194. doi:

10.1126/science.1141967.

Elleuche, S. (2015) 'Bringing functions together with fusion enzymes—from nature's inventions to biotechnological applications', *Applied Microbiology and Biotechnology*. Springer Verlag, pp. 1545–1556. doi: 10.1007/s00253-014-6315-1.

Engler, C. *et al.* (2009) 'Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes', *PLoS ONE*. Edited by J. Peccoud, 4(5), p. e5553. doi: 10.1371/journal.pone.0005553.

Epting, K. L. *et al.* (2005) 'Influence of divalent cations on the structural thermostability and thermal inactivation kinetics of class II xylose isomerases', *FEBS Journal*, 272(6), pp. 1454–1464. doi: 10.1111/j.1742-4658.2005.04577.x.

Fan, Y. *et al.* (2009) 'A combined theoretical and experimental study of the ammonia tunnel in carbamoyl phosphate synthetase', *Journal of the American Chemical Society*, 131(29), pp. 10211–10219. doi: 10.1021/ja902557r.

Farber, G. K. *et al.* (1989) 'Crystallographic studies of the mechanism of xylose isomerase', *Biochemistry*, 28(18), pp. 7289–7297. doi: 10.1021/bi00444a022.

Fatima, B., Aftab, M. N. and Haq, I. (2016) 'Cloning, purification, and characterization of xylose isomerase from *Thermotoga naphthophila* RKU-10', *Journal of Basic Microbiology*, 56(9), pp. 949–962. doi: 10.1002/jobm.201500589.

Fatima, B. and Javed, M. M. (2020) 'Production, purification and physicochemical characterization of D-xylose/glucose isomerase from *Escherichia coli* strain BL21', *3 Biotech*, 10(2). doi: 10.1007/s13205-019-2036-6.

Fell, D. A. (1998) 'Increasing the flux in metabolic pathways: A metabolic control analysis perspective', *Biotechnology and Bioengineering*, 58(2–3), pp. 121–124. doi: 10.1002/(SICI)1097-0290(19980420)58:2/3<121::AID-BIT2>3.0.CO;2-N.

Feng, Q. *et al.* (2018) 'Signature pathway expression of xylose utilization in the genetically engineered industrial yeast *Saccharomyces cerevisiae*', *PLoS ONE*, 13(4). doi: 10.1371/journal.pone.0195633.

Fiencke, C. and Bock, E. (2006) 'Immunocytochemical localization of membrane-bound ammonia monooxygenase in cells of ammonia oxidizing bacteria', *Archives of Microbiology*, 185(2), pp. 99–106. doi: 10.1007/s00203-005-0074-4.

Fierobe, H.-P. *et al.* (2001) 'Design and Production of Active Cellulosome Chimeras

SELECTIVE INCORPORATION OF DOCKERIN-CONTAINING ENZYMES INTO DEFINED FUNCTIONAL COMPLEXES*. doi: 10.1074/jbc.M102082200.

Fischbach, M. A. and Walsh, C. T. (2006) 'Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms', *Chemical Reviews*. American Chemical Society, pp. 3468–3496. doi: 10.1021/cr0503097.

Fischer, M. *et al.* (2015) 'Tripartite ATP-independent Periplasmic (TRAP) Transporters use an arginine-mediated selectivity filter for high affinity substrate binding', *Journal of Biological Chemistry*, 290(45), pp. 27113–27123. doi: 10.1074/jbc.M115.656603.

Fish, K. N. (2009) 'Total internal reflection fluorescence (TIRF) microscopy', *Current Protocols in Cytometry*. John Wiley and Sons Inc., p. Unit12.18. doi: 10.1002/0471142956.cy1218s50.

Fitzkee, N. C. (2004) 'Steric restrictions in protein folding: An α -helix cannot be followed by a contiguous α -strand', *Protein Science*, 13(3), pp. 633–639. doi: 10.1110/ps.03503304.

Flores Bueso, Y. and Tangney, M. (2017) 'Synthetic Biology in the Driving Seat of the Bioeconomy', *Trends in Biotechnology*. Elsevier Ltd, pp. 373–378. doi: 10.1016/j.tibtech.2017.02.002.

Fontes, C. M. G. A. and Gilbert, H. J. (2010) 'Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates', *Annual Review of Biochemistry*, 79(1), pp. 655–681. doi: 10.1146/annurev-biochem-091208-085603.

Formighieri, C. and Melis, A. (2015) 'A phycocyanin-phellandrene synthase fusion enhances recombinant protein expression and β -phellandrene (monoterpene) hydrocarbons production in *Synechocystis* (cyanobacteria)', *Metabolic Engineering*, 32, pp. 116–124. doi: 10.1016/j.ymben.2015.09.010.

Formighieri, C. and Melis, A. (2016) 'Sustainable heterologous production of terpene hydrocarbons in cyanobacteria', *Photosynthesis Research*, 130(1–3), pp. 123–135. doi: 10.1007/s11120-016-0233-2.

Francois, J. M., Alkim, C. and Morin, N. (2020) 'Engineering microbial pathways for production of bio-based chemicals from lignocellulosic sugars: Current status and perspectives', *Biotechnology for Biofuels*. BioMed Central, pp. 1–23. doi: 10.1186/s13068-020-01744-6.

Freedman, D. and Diaconis, P. (1981) 'On the histogram as a density estimator:L2 theory', *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 57(4), pp. 453–476. doi: 10.1007/BF01025868.

Frieden, C. (2019) 'Protein oligomerization as a metabolic control mechanism: Application to apoE', *Protein Science*, 28(4), pp. 837–842. doi: 10.1002/pro.3583.

Fu, J. *et al.* (2012) 'Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures', *Journal of the American Chemical Society*, 134(12), pp. 5516–5519. doi: 10.1021/ja300897h.

Fujiwara, R. *et al.* (2018) 'Muconic Acid Production Using Gene-Level Fusion Proteins in *Escherichia coli*', *ACS Synthetic Biology*, 7(11), pp. 2698–2705. doi: 10.1021/acssynbio.8b00380.

Fukasawa, Y. and Tomii, K. (2019) 'Accurate Classification of Biological and non-Biological Interfaces in Protein Crystal Structures using Subtle Covariation Signals', *Scientific Reports*, 9(1), pp. 1–12. doi: 10.1038/s41598-019-48913-8.

Fulyani, F. *et al.* (2013) 'Functional Diversity of Tandem Substrate-Binding Domains in ABC Transporters from Pathogenic Bacteria', *Structure*, 21(10), pp. 1879–1888. doi: 10.1016/j.str.2013.07.020.

Gabaldón, T., Ginger, M. L. and Michels, P. A. M. (2016) 'Peroxisomes in parasitic protists', *Molecular and Biochemical Parasitology*, 209(1–2), pp. 35–45. doi: 10.1016/j.molbiopara.2016.02.005.

Gao, J. and Zhou, Y. J. (2019) 'Repurposing peroxisomes for microbial synthesis for biomolecules', in *Methods in Enzymology*. Academic Press Inc., pp. 83–111. doi: 10.1016/bs.mie.2018.12.004.

Geertsma, E. R. *et al.* (2008) 'Quality control of overexpressed membrane proteins', *Proceedings of the National Academy of Sciences of the United States of America*, 105(15), pp. 5722–5727. doi: 10.1073/pnas.0802190105.

Germer, K., Leonard, M. and Zhang, X. (2013) 'RNA aptamers and their therapeutic and diagnostic applications.', *International journal of biochemistry and molecular biology*, 4(1), pp. 27–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23638319> (Accessed: 19 January 2017).

Gibson, D. G. *et al.* (2009) 'Enzymatic assembly of DNA molecules up to several hundred kilobases', *Nature Methods*, 6(5), pp. 343–345. doi: 10.1038/nmeth.1318.

Gingras, A.-C., Abe, K. T. and Raught, B. (2019) 'Getting to know the neighborhood: using proximity-dependent biotinylation to characterize protein complexes and map organelles',

Current Opinion in Chemical Biology, 48, pp. 44–54. doi: 10.1016/j.cbpa.2018.10.017.

Gogoi, P. and Kanaujia, S. P. (2019) 'Role of Structural Features in Oligomerization, Active-Site Integrity and Ligand Binding of Ribose-1,5-Bisphosphate Isomerase', *Computational and Structural Biotechnology Journal*, 17, pp. 333–344. doi: 10.1016/j.csbj.2019.02.009.

Gohlke, U. *et al.* (2005) 'The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter', *Proceedings of the National Academy of Sciences of the United States of America*, 102(30), pp. 10482–10486. doi: 10.1073/pnas.0503558102.

Gokhale, R. S. and Khosla, C. (2000) 'Role of linkers in communication between protein modules', *Current Opinion in Chemical Biology*. Current Biology Ltd, pp. 22–27. doi: 10.1016/S1367-5931(99)00046-0.

Gora, A., Brezovsky, J. and Damborsky, J. (2013) 'Gates of enzymes', *Chemical Reviews*. American Chemical Society, pp. 5871–5923. doi: 10.1021/cr300384w.

Gou, M. *et al.* (2018) 'The scaffold proteins of lignin biosynthetic cytochrome P450 enzymes', *Nature Plants*, 4(5), pp. 299–310. doi: 10.1038/s41477-018-0142-9.

Greening, C. and Lithgow, T. (2020) 'Formation and function of bacterial organelles', *Nature Reviews Microbiology*. Nature Research, p. 677. doi: 10.1038/s41579-020-0413-0.

Gronenberg, L. S., Marcheschi, R. J. and Liao, J. C. (2013) 'Next generation biofuel engineering in prokaryotes', *Current Opinion in Chemical Biology*. Elsevier Current Trends, pp. 462–471. doi: 10.1016/j.cbpa.2013.03.037.

Guamán, L. P. *et al.* (2018) 'xylA and xylB overexpression as a successful strategy for improving xylose utilization and poly-3-hydroxybutyrate production in *Burkholderia sacchari*', *Journal of Industrial Microbiology and Biotechnology*, 45(3), pp. 165–173. doi: 10.1007/s10295-018-2007-7.

Gubellini, F. *et al.* (2011) 'Physiological response to membrane protein overexpression in *E. coli*', *Molecular and Cellular Proteomics*, 10(10). doi: 10.1074/mcp.M111.007930.

Guo, H. *et al.* (2017) 'Effect of flexible linker length on the activity of fusion protein 4-coumaroyl-CoA ligase::stilbene synthase', *Molecular BioSystems*, 13(3), pp. 598–606. doi: 10.1039/c6mb00563b.

Guo, M., Song, W. and Buhain, J. (2015) 'Bioenergy and biofuels: History, status, and perspective', *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, pp. 712–725. doi:

10.1016/j.rser.2014.10.013.

Gustavsson, M. and Lee, S. Y. (2016) 'Prospects of microbial cell factories developed through systems metabolic engineering', *Microbial Biotechnology*, 9(5), pp. 610–617. doi: 10.1111/1751-7915.12385.

Gyuris, J. *et al.* (1993) 'Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2', *Cell*, 75(4), pp. 791–803. doi: 10.1016/0092-8674(93)90498-F.

Hahn-Hägerdal, B. *et al.* (2007) 'Towards industrial pentose-fermenting yeast strains', *Applied Microbiology and Biotechnology*. Springer, pp. 937–953. doi: 10.1007/s00253-006-0827-2.

Hahn, M. and Stachelhaus, T. (2006) 'Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides', *Proceedings of the National Academy of Sciences of the United States of America*, 103(2), pp. 275–280. doi: 10.1073/pnas.0508409103.

Hall, B. G. *et al.* (2014) 'Growth rates made easy', *Molecular Biology and Evolution*, 31(1), pp. 232–238. doi: 10.1093/molbev/mst187.

Hammer, S. K. and Avalos, J. L. (2017) 'Harnessing yeast organelles for metabolic engineering', *Nature Chemical Biology*, 13(8), pp. 823–832. doi: 10.1038/nchembio.2429.

Handelsman, T. *et al.* (2004) 'Cohesin–dockerin interaction in cellulosome assembly: a single Asp-to-Asn mutation disrupts high-affinity cohesin–dockerin binding', *FEBS Letters*, 572(1–3), pp. 195–200. doi: 10.1016/J.FEBSLET.2004.07.040.

Harikumar, A. *et al.* (2017) 'An Endogenously Tagged Fluorescent Fusion Protein Library in Mouse Embryonic Stem Cells', *Stem Cell Reports*, 9(4), pp. 1304–1314. doi: 10.1016/j.stemcr.2017.08.022.

Hartley, B. S. *et al.* (2000) 'Glucose isomerase: Insights into protein engineering for increased thermostability', *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, pp. 294–335. doi: 10.1016/S0167-4838(00)00246-6.

Harvat, E. M. *et al.* (2005) 'Lysophospholipid flipping across the *Escherichia coli* inner membrane catalyzed by a transporter (LpIT) belonging to the major facilitator superfamily.', *The Journal of biological chemistry*, 280(12), pp. 12028–34. doi: 10.1074/jbc.M414368200.

Hassan, S. S., Williams, G. A. and Jaiswal, A. K. (2019) 'Moving towards the second generation of lignocellulosic biorefineries in the EU: Drivers, challenges, and opportunities', *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, pp. 590–599. doi:

10.1016/j.rser.2018.11.041.

Håvarstein, L. S., Diep, D. B. and Nes, I. F. (1995) 'A family of bacteriocin ABC transporters carry out proteolytic processing of their substrates concomitant with export.', *Molecular microbiology*, 16(2), pp. 229–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7565085> (Accessed: 20 May 2019).

Heide, T. van der and Poolman, B. (2002) 'ABC transporters: one, two or four extracytoplasmic substrate-binding sites?', *EMBO reports*, 3(10), pp. 938–943. doi: 10.1093/embo-reports/kvf201.

Von Heijne, G. (1986) The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology, *The EMBO Journal*.

Hemmerlin, A. *et al.* (2006) 'A cytosolic arabidopsis D-xylulose kinase catalyzes the phosphorylation of 1-deoxy-D-xylulose into a precursor of the plastidial isoprenoid pathway', *Plant Physiology*, 142(2), pp. 441–457. doi: 10.1104/pp.106.086652.

Henderson, P. J. F. (1990) MINI-REVIEW Proton-Linked Sugar Transport Systems in Bacteria, *Journal of Bioenergetics and Biomembranes*.

Henke, N. A. and Wendisch, V. F. (2019) 'Improved Astaxanthin Production with *Corynebacterium glutamicum* by Application of a Membrane Fusion Protein', *Marine Drugs*, 17(11), p. 621. doi: 10.3390/md17110621.

Henrick, K., Collyer, C. A. and Blow, D. M. (1989) 'Structures of d-xylose isomerase from *Arthrobacter* strain B3728 containing the inhibitors xylitol and d-sorbitol at 2.5 Å and 2.3 Å resolution, respectively', *Journal of Molecular Biology*, 208(1), pp. 129–157. doi: 10.1016/0022-2836(89)90092-2.

Henriques de Jesus, M. P. R. *et al.* (2017) 'Tat proteins as novel thylakoid membrane anchors organize a biosynthetic pathway in chloroplasts and increase product yield 5-fold', *Metabolic Engineering*, 44, pp. 108–116. doi: 10.1016/j.ymben.2017.09.014.

Hern, J. A. *et al.* (2010) 'Formation and dissociation of M1 muscarinic receptor dimers seen by total internal reflection fluorescence imaging of single molecules', *Proceedings of the National Academy of Sciences of the United States of America*, 107(6), pp. 2693–2698. doi: 10.1073/pnas.0907915107.

Hess, J. M. *et al.* (1998) '*Thermotoga neapolitana* homotetrameric xylose isomerase is expressed as a catalytically active and thermostable dimer in *Escherichia coli*.', *Applied and*

environmental microbiology, 64(7), pp. 2357–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9647799> (Accessed: 16 July 2019).

Hilario, E. *et al.* (2016) 'Visualizing the tunnel in tryptophan synthase with crystallography: Insights into a selective filter for accommodating indole and rejecting water', *Biochimica et Biophysica Acta - Proteins and Proteomics*, 1864(3), pp. 268–279. doi: 10.1016/j.bbapap.2015.12.006.

Hollenstein, K., Dawson, R. J. and Locher, K. P. (2007) 'Structure and mechanism of ABC transporter proteins', *Current Opinion in Structural Biology*. Elsevier Current Trends, pp. 412–418. doi: 10.1016/j.sbi.2007.07.003.

Hollenstein, K., Frei, D. C. and Locher, K. P. (2007) 'Structure of an ABC transporter in complex with its binding protein', *Nature*, 446(7132), pp. 213–216. doi: 10.1038/nature05626.

Hölsch, K. and Weuster-Botz, D. (2010) 'Enantioselective reduction of prochiral ketones by engineered bifunctional fusion proteins', *Biotechnology and Applied Biochemistry*, 56(4), pp. 131–140. doi: 10.1042/ba20100143.

Hopkins, A. P., Hawkhead, J. A. and Thomas, G. H. (2013) 'Transport and catabolism of the sialic acids *N*-glycolylneuraminic acid and 3-keto-3-deoxy- D-glycero-D-galactononic acid by *Escherichia coli* K-12', *FEMS Microbiology Letters*, 347(1), pp. 14–22. doi: 10.1111/1574-6968.12213.

Hou, J. *et al.* (2016) 'Characterization and evolution of xylose isomerase screened from the bovine rumen metagenome in *Saccharomyces cerevisiae*', *Journal of Bioscience and Bioengineering*, 121(2), pp. 160–165. doi: 10.1016/j.jbiosc.2015.05.014.

Hu, T. *et al.* (2020) 'Engineering chimeric diterpene synthases and isoprenoid biosynthetic pathways enables high-level production of miltiradiene in yeast', *Metabolic Engineering*. doi: 10.1016/j.ymben.2020.03.011.

Huang, H. M., Stephan, P. and Kries, H. (2020) 'Engineering DNA-Templated Nonribosomal Peptide Synthesis', *Cell Chemical Biology*, 28(2), pp. 221–227.e7. doi: 10.1016/j.chembiol.2020.11.004.

Huang, X., Holden, H. M. and Raushel, F. M. (2001) 'Channeling of Substrates and Intermediates in Enzyme-Catalyzed Reactions', *Annual Review of Biochemistry*, 70(1), pp. 149–180. doi: 10.1146/annurev.biochem.70.1.149.

Huang, Z. *et al.* (2016) 'A study on the effects of linker flexibility on acid phosphatase PhoC-

GFP fusion protein using a novel linker library', *Enzyme and Microbial Technology*, 83, pp. 1–6. doi: 10.1016/j.enzmictec.2015.11.002.

Huffer, S. *et al.* (2012) 'Escherichia coli for biofuel production: Bridging the gap from promise to practice', *Trends in Biotechnology*. Elsevier Current Trends, pp. 538–545. doi: 10.1016/j.tibtech.2012.07.002.

Huttanus, H. M. and Feng, X. (2017) 'Compartmentalized metabolic engineering for biochemical and biofuel production', *Biotechnology Journal*, 12(6), p. 1700052. doi: 10.1002/biot.201700052.

Hwang, M. P. *et al.* (2013) 'Think modular: A simple apoferritin-based platform for the multifaceted detection of pancreatic cancer', *ACS Nano*, 7(9), pp. 8167–8174. doi: 10.1021/nn403465a.

Hyde, C. C. and Miles, E. W. (1990) 'The tryptophan synthase multienzyme complex: Exploring structure-function relationships with x-ray crystallography and mutagenesis', *Bio/Technology*, 8(1), pp. 27–32. doi: 10.1038/nbt0190-27.

Idan, O. and Hess, H. (2013) 'Engineering enzymatic cascades on nanoscale scaffolds', *Current Opinion in Biotechnology*, 24(4), pp. 606–611. doi: 10.1016/J.COPBIO.2013.01.003.

Igarashi, Y. (2004) 'The Evolutionary Repertoires of the Eukaryotic-Type ABC Transporters in Terms of the Phylogeny of ATP-binding Domains in Eukaryotes and Prokaryotes', *Molecular Biology and Evolution*, 21(11), pp. 2149–2160. doi: 10.1093/molbev/msh226.

iGEM (2015) *Part:pSB1C3* @ *parts.igem.org*. Available at: <https://parts.igem.org/Part:pSB1C3> (Accessed: 6 February 2020).

Ingram, L. O. *et al.* (1987) 'Genetic engineering of ethanol production in *Escherichia coli*.', *Applied and environmental microbiology*, 53(10), pp. 2420–2425. doi: 10.1128/aem.53.10.2420-2425.1987.

Inokuma, K. *et al.* (2010) 'Improvement of isopropanol production by metabolically engineered *Escherichia coli* using gas stripping', *Journal of Bioscience and Bioengineering*, 110(6), pp. 696–701. doi: 10.1016/j.jbiosc.2010.07.010.

Isikgor, F. H. and Becer, C. R. (2015) 'Lignocellulosic biomass: a sustainable platform for the production of bio-based chemicals and polymers', *Polymer Chemistry*, 6(25), pp. 4497–4559. doi: 10.1039/c5py00263j.

Iturrate, L. *et al.* (2009) 'Substrate channelling in an engineered bifunctional aldolase/kinase

enzyme confers catalytic advantage for C-C bond formation', *Chemical Communications*, (13), pp. 1721–1723. doi: 10.1039/b822345a.

Iverson, S. V. *et al.* (2016) 'CIDAR MoClo: Improved MoClo Assembly Standard and New *E. coli* Part Library Enable Rapid Combinatorial Design for Synthetic and Traditional Biology', *ACS Synthetic Biology*, 5(1), pp. 99–103. doi: 10.1021/acssynbio.5b00124.

Jagtap, S. S. and Rao, C. V. (2018) 'Microbial conversion of xylose into useful bioproducts', *Applied Microbiology and Biotechnology*. Springer Verlag, pp. 9015–9036. doi: 10.1007/s00253-018-9294-9.

Javelle, A. *et al.* (2004) 'Ammonium sensing in *Escherichia coli*: Role of the ammonium transporter AmtB and AmtB-GlnK complex formation', *Journal of Biological Chemistry*, 279(10), pp. 8530–8538. doi: 10.1074/jbc.M312399200.

Jazini, M. H., Fereydouni, E. and Karimi, K. (2017) 'Microbial xanthan gum production from alkali-pretreated rice straw', *RSC Advances*, 7(6), pp. 3507–3514. doi: 10.1039/C6RA26185J.

Jenkins, J. *et al.* (1992) 'Protein engineering of xylose (glucose) isomerase from *Actinoplanes missouriensis*. 1. Crystallography and site-directed mutagenesis of metal binding sites', *Biochemistry*, 31(24), pp. 5449–5458. doi: 10.1021/bi00139a005.

Jeon, E.-Y. *et al.* (2015) 'Enzyme fusion for whole-cell biotransformation of long-chain sec-alcohols into esters', *Applied Microbiology and Biotechnology*, 99(15), pp. 6267–6275. doi: 10.1007/s00253-015-6392-9.

Jeppsson, M. *et al.* (2002) 'Reduced oxidative pentose phosphate pathway flux in recombinant xylose-utilizing *Saccharomyces cerevisiae* strains improves the ethanol yield from xylose', *Applied and Environmental Microbiology*, 68(4), pp. 1604–1609. doi: 10.1128/AEM.68.4.1604-1609.2002.

Jin, Y.-S. *et al.* (2002) 'Molecular cloning of XYL3 (D-xylulokinase) from *Pichia stipitis* and characterization of its physiological function.', *Applied and environmental microbiology*, 68(3), pp. 1232–9. doi: 10.1128/aem.68.3.1232-1239.2002.

Jindou, S., Soda, A., *et al.* (2004) 'Cohesin-dockerin interactions within and between *clostridium josui* and *clostridium thermocellum*: Binding selectivity between cognate dockerin and cohesin domains and species specificity', *Journal of Biological Chemistry*, 279(11), pp. 9867–9874. doi: 10.1074/jbc.M308673200.

Jindou, S., Kajino, T., *et al.* (2004) 'Interaction between a type-II dockerin domain and a Type-

II cohesin domain from *Clostridium thermocellum* cellulosome', *Bioscience, Biotechnology and Biochemistry*, 68(4), pp. 924–926. doi: 10.1271/bbb.68.924.

Jindou, S. *et al.* (2014) 'Engineered Platform for Bioethylene Production by a Cyanobacterium Expressing a Chimeric Complex of Plant Enzymes', *ACS Synthetic Biology*, 3(7), pp. 487–496. doi: 10.1021/sb400197f.

Jo, S. *et al.* (2017) 'Modular pathway engineering of *Corynebacterium glutamicum* to improve xylose utilization and succinate production', *Journal of Biotechnology*, 258, pp. 69–78. doi: 10.1016/j.jbiotec.2017.01.015.

Jojima, T. *et al.* (2010) 'Sugar transporters in efficient utilization of mixed sugar substrates: Current knowledge and outlook', *Applied Microbiology and Biotechnology*. Springer, pp. 471–480. doi: 10.1007/s00253-009-2292-1.

Kanekiyo, M. *et al.* (2013) 'Self-assembling influenza nanoparticle vaccines elicit broadly neutralizing H1N1 antibodies', *Nature*, 499(7456), pp. 102–106. doi: 10.1038/nature12202.

Kang, M. K. *et al.* (2014) 'Synthetic biology platform of CoryneBrick vectors for gene expression in *Corynebacterium glutamicum* and its application to xylose utilization', *Applied Microbiology and Biotechnology*, 98(13), pp. 5991–6002. doi: 10.1007/s00253-014-5714-7.

Kaplan, A. and Reinhold, L. (1999) 'CO₂ concentrating mechanisms in photosynthetic microorganisms', *Annual Review of Plant Biology*, 50(1), pp. 539–570. doi: 10.1146/annurev.arplant.50.1.539.

Kawaguchi, H. *et al.* (2016) 'Bioprocessing of bio-based chemicals produced from lignocellulosic feedstocks', *Current Opinion in Biotechnology*. Elsevier Ltd, pp. 30–39. doi: 10.1016/j.copbio.2016.02.031.

Ke, M. *et al.* (2017) 'Molecular determinants for the thermodynamic and functional divergence of uniporter GLUT1 and proton symporter Xyle', *PLOS Computational Biology*. Edited by B. L. de Groot, 13(6), p. e1005603. doi: 10.1371/journal.pcbi.1005603.

Kenri, T. *et al.* (2004) 'Use of fluorescent-protein tagging to determine the subcellular localization of *Mycoplasma pneumoniae* proteins encoded by the cytoadherence regulatory locus', *Journal of Bacteriology*, 186(20), pp. 6944–6955. doi: 10.1128/JB.186.20.6944-6955.2004.

Kenworthy, A. K. (2001) 'Imaging protein-protein interactions using fluorescence resonance energy transfer microscopy', *Methods*, 24(3), pp. 289–296. doi: 10.1006/meth.2001.1189.

- Keppler, A. *et al.* (2004) 'Labeling of fusion proteins with synthetic fluorophores in live cells', *Proceedings of the National Academy of Sciences of the United States of America*, 101(27), pp. 9955–9959. doi: 10.1073/pnas.0401923101.
- Khalil, A. S. and Collins, J. J. (2010) 'Synthetic biology: applications come of age', *Nature Reviews Genetics*, 11(5), pp. 367–379. doi: 10.1038/nrg2775.
- Kim, B.-J. *et al.* (2010) 'Transferrin fusion technology: a novel approach to prolonging biological half-life of insulinotropic peptides.', *The Journal of pharmacology and experimental therapeutics*, 334(3), pp. 682–92. doi: 10.1124/jpet.110.166470.
- Kim, G. *et al.* (2020) 'Aldehyde-alcohol dehydrogenase undergoes structural transition to form extended spiroosomes for substrate channeling', *Communications Biology*, 3(1), pp. 1–9. doi: 10.1038/s42003-020-1030-1.
- Kim, H. M. *et al.* (2015) 'Improving lignocellulose degradation using xylanase–cellulase fusion protein with a glycine–serine linker', *International Journal of Biological Macromolecules*, 73(1), pp. 215–221. doi: 10.1016/j.ijbiomac.2014.11.025.
- Kim, S. and Hahn, J.-S. (2014) Synthetic scaffold based on a cohesin–dockerin interaction for improved production of 2,3-butanediol in *Saccharomyces cerevisiae*, *Journal of Biotechnology*. doi: 10.1016/j.jbiotec.2014.10.015.
- Kizer, L. *et al.* (2008) 'Application of functional genomics to pathway optimization for increased isoprenoid production.', *Applied and environmental microbiology*, 74(10), pp. 3229–41. doi: 10.1128/AEM.02750-07.
- Klann, M. T., Lapin, A. and Reuss, M. (2011) 'Agent-based simulation of reactions in the crowded and structured intracellular environment: Influence of mobility and location of the reactants', *BMC Systems Biology*, 5(1), p. 71. doi: 10.1186/1752-0509-5-71.
- Kleanthous, C., Rassam, P. and Baumann, C. G. (2015) 'Protein-protein interactions and the spatiotemporal dynamics of bacterial outer membrane proteins', *Current Opinion in Structural Biology*. Elsevier Ltd, pp. 109–115. doi: 10.1016/j.sbi.2015.10.007.
- Kleinlogel, S. *et al.* (2011) 'A gene-fusion strategy for stoichiometric and co-localized expression of light-gated membrane proteins', *Nature Methods*, 8(12), pp. 1083–1091. doi: 10.1038/nmeth.1766.
- Klesney-Tait, J. *et al.* (1997) The Major Outer Membrane Protein of *Haemophilus ducreyi* Consists of Two OmpA Homologs, *JOURNAL OF BACTERIOLOGY*. Available at:

<http://jb.asm.org/> (Accessed: 21 May 2020).

Knighton, D. R. *et al.* (1994) 'Structure of and kinetic channelling in bifunctional dihydrofolate reductase-thymidylate synthase', *Nature Structural Biology*, 1(3), pp. 186–194. doi: 10.1038/nsb0394-186.

Kobayashi, N. and Arai, R. (2017) 'Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks', *Current Opinion in Biotechnology*. Elsevier Ltd, pp. 57–65. doi: 10.1016/j.copbio.2017.01.001.

Koirala, S., Wang, X. and Rao, C. V. (2016) 'Reciprocal regulation of L-arabinose and D-xylose metabolism in *Escherichia coli*', *Journal of Bacteriology*, 198(3), pp. 386–393. doi: 10.1128/JB.00709-15.

de Kok, A. *et al.* (1998) 'The pyruvate dehydrogenase multi-enzyme complex from Gram-negative bacteria', *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1385(2), pp. 353–366. doi: 10.1016/S0167-4838(98)00079-X.

Koppolu, V. and Vasigala, V. K. (2016) 'Role of *Escherichia coli* in Biofuel Production', *Microbiology Insights*, 9, p. MBI.S10878. doi: 10.4137/mbi.s10878.

Kosuri, S. and Church, G. M. (2014) 'Large-scale de novo DNA synthesis: Technologies and applications', *Nature Methods*. Nature Publishing Group, pp. 499–507. doi: 10.1038/nmeth.2918.

Kötter, P. and Ciriacy, M. (1993) 'Xylose fermentation by *Saccharomyces cerevisiae*', *Applied Microbiology and Biotechnology*, 38(6), pp. 776–783. doi: 10.1007/BF00167144.

Kovach, M. E. *et al.* (1995) 'Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes.', *Gene*, 166(1), pp. 175–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8529885> (Accessed: 21 November 2018).

Kumar, M., Mommer, M. S. and Sourjik, V. (2010) 'Mobility of Cytoplasmic, Membrane, and DNA-Binding Proteins in *Escherichia coli*', *Biophysical Journal*, 98(4), pp. 552–559. doi: 10.1016/j.bpj.2009.11.002.

Kumar, S. and Nussinov, R. (2002) 'Close-Range Electrostatic Interactions in Proteins', *ChemBioChem*, 3(7), pp. 604–617. doi: 10.1002/1439-7633(20020703)3:7<604::AID-CBIC604>3.0.CO;2-X.

Kuyper, M. *et al.* (2005) 'Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation', *FEMS Yeast*

Research, 5(4–5), pp. 399–409. doi: 10.1016/j.femsyr.2004.09.010.

Lai, Y. T., Cascio, D. and Yeates, T. O. (2012) 'Structure of a 16-nm cage designed by using protein oligomers', *Science*. American Association for the Advancement of Science, p. 1129. doi: 10.1126/science.1219351.

Laskowski, R. A., Gerick, F. and Thornton, J. M. (2009) 'The structural basis of allosteric regulation in proteins', *FEBS Letters*, 583(11), pp. 1692–1698. doi: 10.1016/j.febslet.2009.03.019.

Lass, A. *et al.* (2011) 'Lipolysis – A highly regulated multi-enzyme complex mediates the catabolism of cellular fat stores', *Progress in Lipid Research*, 50(1), pp. 14–27. doi: 10.1016/J.PLIPRES.2010.10.004.

Latysheva, N. S. and Babu, M. M. (2019) 'Molecular signatures of fusion proteins in cancer', *ACS Pharmacology and Translational Science*, 2(2), pp. 122–133. doi: 10.1021/acspsci.9b00019.

Laursen, T. *et al.* (2016) 'Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum', *Science*, 354(6314), pp. 890–893. doi: 10.1126/science.aag2347.

Leake, M. C. *et al.* (2006) 'Stoichiometry and turnover in single, functioning membrane protein complexes', *Nature*, 443(7109), pp. 355–358. doi: 10.1038/nature05135.

Leake, M. C. *et al.* (2008) 'Variable stoichiometry of the TatA component of the twin-arginine protein transport system observed by in vivo single-molecule imaging', *Proceedings of the National Academy of Sciences of the United States of America*, 105(40), pp. 15376–15381. doi: 10.1073/pnas.0806338105.

Lee, H., DeLoache, W. C. and Dueber, J. E. (2012) 'Spatial organization of enzymes for metabolic engineering', *Metabolic Engineering*, 14(3), pp. 242–251. doi: 10.1016/j.ymben.2011.09.003.

Lee, H. L. *et al.* (2011) 'Construction and characterization of different fusion proteins between cellulases and β -glucosidase to improve glucose production and thermostability', *Bioresource Technology*, 102(4), pp. 3973–3976. doi: 10.1016/j.biortech.2010.11.114.

Lee, J. H. *et al.* (2013) 'Improved Production of L-Threonine in *Escherichia coli* by Use of a DNA Scaffold System', *Applied and Environmental Microbiology*, 79(3), pp. 774–782. doi: 10.1128/AEM.02578-12.

Lee, J., Sands, Z. A. and Biggin, P. C. (2016) 'A Numbering System for MFS Transporter Proteins', *Frontiers in Molecular Biosciences*, 3(JUN), p. 21. doi: 10.3389/fmolb.2016.00021.

Lee, M. E. *et al.* (2013) 'Expression-level optimization of a multi-enzyme pathway in the absence of a high-throughput assay', *Nucleic Acids Research*, 41(22), pp. 10668–10678. doi: 10.1093/nar/gkt809.

Lee, M. J. *et al.* (2017) 'Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm', *Nature Chemical Biology*, 14(2), pp. 142–147. doi: 10.1038/nchembio.2535.

Lee, R. A. and Lavoie, J.-M. (2013) 'From first- to third-generation biofuels: Challenges of producing a commodity from a biomass of increasing complexity', *Animal Frontiers*, 3(2), pp. 6–11. doi: 10.2527/af.2013-0010.

Lee, S. K. *et al.* (2008) 'Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels', *Current Opinion in Biotechnology*. Elsevier Current Trends, pp. 556–563. doi: 10.1016/j.copbio.2008.10.014.

Leonard, S. P. *et al.* (2018) 'Genetic Engineering of Bee Gut Microbiome Bacteria with a Toolkit for Modular Assembly of Broad-Host-Range Plasmids', *ACS Synthetic Biology*, 7(5), pp. 1279–1290. doi: 10.1021/acssynbio.7b00399.

Letunic, I. and Bork, P. (2019) 'Interactive Tree of Life (iTOL) v4: Recent updates and new developments', *Nucleic Acids Research*, 47(W1), p. W256. doi: 10.1093/nar/gkz239.

Letunic, I. and Bork, P. (2021) 'Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation', *Nucleic Acids Research*. doi: 10.1093/nar/gkab301.

Lewicka, A. J. *et al.* (2014) 'Fusion of pyruvate decarboxylase and alcohol dehydrogenase increases ethanol production in *Escherichia coli*', *ACS Synthetic Biology*. American Chemical Society, pp. 976–978. doi: 10.1021/sb500020g.

Li, G. *et al.* (2016) 'Construction of a linker library with widely controllable flexibility for fusion protein design', *Applied Microbiology and Biotechnology*, 100(1), pp. 215–225. doi: 10.1007/s00253-015-6985-3.

Liang, J. and Blumenthal, R. M. (2013) 'Naturally-occurring, dually-functional fusions between restriction endonucleases and regulatory proteins', *BMC Evolutionary Biology*, 13(1), p. 218. doi: 10.1186/1471-2148-13-218.

Liao, L. *et al.* (2021) 'Construction and characterization of a novel glucose dehydrogenase-

leucine dehydrogenase fusion enzyme for the biosynthesis of l-tert-leucine', *Microbial Cell Factories*, 20(1), p. 3. doi: 10.1186/s12934-020-01501-2.

Lill, Y. *et al.* (2012) 'Single-molecule study of molecular mobility in the cytoplasm of *Escherichia coli*', *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 86(2), p. 021907. doi: 10.1103/PhysRevE.86.021907.

Lim, S. I. *et al.* (2016) 'Controlled Orientation of Active Sites in a Nanostructured Multienzyme Complex', *Scientific Reports*, 6(1), pp. 1–9. doi: 10.1038/srep39587.

Lin, C. Y. and Lu, C. (2021) 'Development perspectives of promising lignocellulose feedstocks for production of advanced generation biofuels: A review', *Renewable and Sustainable Energy Reviews*. Elsevier Ltd, p. 110445. doi: 10.1016/j.rser.2020.110445.

Lin, J.-L., Zhu, J. and Wheeldon, I. (2017) 'Synthetic Protein Scaffolds for Biosynthetic Pathway Co-Localization on Lipid Droplet Membranes', *ACS Synthetic Biology*, p. acssynbio.7b00041. doi: 10.1021/acssynbio.7b00041.

Lin, Y. *et al.* (2018) 'The phospholipid-repair system LpIT/Aas in Gram-negative bacteria protects the bacterial membrane envelope from host phospholipase A2 attack', *Journal of Biological Chemistry*, 293(9), pp. 3386–3398. doi: 10.1074/jbc.RA117.001231.

Lindsay, M. R. *et al.* (2001) 'Cell compartmentalisation in planctomycetes: Novel types of structural organisation for the bacterial cell', *Archives of Microbiology*, 175(6), pp. 413–429. doi: 10.1007/s002030100280.

Lioliou, G. and Barnett, A. M. (2018) 'Prototype GaAs X-ray detector and preamplifier electronics for a deep seabed mineral XRF spectrometer', *X-Ray Spectrometry*, 47(3), pp. 201–214. doi: 10.1002/xrs.2818.

Liu, F. *et al.* (2013) 'Functional assembly of a multi-enzyme methanol oxidation cascade on a surface-displayed trifunctional scaffold for enhanced NADH production', *Chemical Communications*, 49(36), p. 3766. doi: 10.1039/c3cc40454d.

Liu, H. and Naismith, J. H. (2008) 'An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol', *BMC Biotechnology*, 8(1), p. 91. doi: 10.1186/1472-6750-8-91.

Liu, J. *et al.* (2017) 'Improving xylose utilization of defatted rice bran for nisin production by overexpression of a xylose transcriptional regulator in *Lactococcus lactis*', *Bioresource Technology*, 238, pp. 690–697. doi: 10.1016/j.biortech.2017.04.076.

- Liu, M. *et al.* (2016) 'Programming a topologically constrained DNA nanostructure into a sensor', *Nature Communications*, 7, p. 12074. doi: 10.1038/ncomms12074.
- Liu, M. *et al.* (2018) 'Metabolic engineering of a xylose pathway for biotechnological production of glycolate in *Escherichia coli*', *Microbial Cell Factories*, 17(1). doi: 10.1186/s12934-018-0900-4.
- Liu, X., Ding, W. and Jiang, H. (2017) 'Engineering microbial cell factories for the production of plant natural products: From design principles to industrial-scale production', *Microbial Cell Factories*. BioMed Central Ltd., p. 125. doi: 10.1186/s12934-017-0732-7.
- Liu, Y. *et al.* (2014) 'Spatial modulation of key pathway enzymes by DNA-guided scaffold system and respiration chain engineering for improved N-acetylglucosamine production by *Bacillus subtilis*', *Metabolic Engineering*, 24, pp. 61–69. doi: 10.1016/j.ymben.2014.04.004.
- Ljungcrantz, P. *et al.* (1989) 'Construction of an Artificial Bifunctional Enzyme, β -Galactosidase/Galactose Dehydrogenase, Exhibiting Efficient Galactose Channeling', *Biochemistry*, 28(22), pp. 8786–8792. doi: 10.1021/bi00448a016.
- Löfblom, J. *et al.* (2010) 'Affibody molecules: Engineered proteins for therapeutic, diagnostic and biotechnological applications', *FEBS Letters*, 584(12), pp. 2670–2680. doi: 10.1016/j.febslet.2010.04.014.
- Long, M. (2000) 'A new function evolved from gene fusion', *Genome Research*. Cold Spring Harbor Laboratory Press, pp. 1655–1657. doi: 10.1101/gr.165700.
- Loomis, W. F. and Magasanik, B. (1967) 'Glucose-lactose diauxie in *Escherichia coli*.', *Journal of Bacteriology*, 93(4), pp. 1397–1401. doi: 10.1128/jb.93.4.1397-1401.1967.
- Lu, J. Z. *et al.* (2010) 'Genetic engineering of a bifunctional IgG fusion protein with iduronate-2-sulfatase', *Bioconjugate Chemistry*, 21(1), pp. 151–156. doi: 10.1021/bc900382q.
- Lu, P. and Feng, M.-G. (2008) 'Bifunctional enhancement of a β -glucanase-xylanase fusion enzyme by optimization of peptide linkers', *Applied Microbiology and Biotechnology*, 79(4), pp. 579–587. doi: 10.1007/s00253-008-1468-4.
- Di Luccio, E. *et al.* (2007) 'Structural and Kinetic Studies of Induced Fit in Xylulose Kinase from *Escherichia coli*', *Journal of Molecular Biology*, 365(3), pp. 783–798. doi: 10.1016/j.jmb.2006.10.068.
- Lv, X. *et al.* (2020) 'Synthetic metabolic channel by functional membrane microdomains for compartmentalized flux control', 59, pp. 106–118. doi: 10.1016/j.ymben.2020.02.003.

Ma, D. *et al.* (2012) 'Structure and mechanism of a glutamate-GABA antiporter', *Nature*, 483(7391), pp. 632–636. doi: 10.1038/nature10917.

Ma, D., Lu, P. and Shi, Y. (2013) 'Substrate selectivity of the acid-activated glutamate/ γ -aminobutyric acid (GABA) antiporter GadC from *Escherichia coli*', *Journal of Biological Chemistry*, 288(21), pp. 15148–15153. doi: 10.1074/jbc.M113.474502.

Madsen, M. A. *et al.* (2018) 'Engineering Mannitol Biosynthesis in *Escherichia coli* and *Synechococcus* sp. PCC 7002 Using a Green Algal Fusion Protein', *ACS Synthetic Biology*, 7(12), pp. 2833–2840. doi: 10.1021/acssynbio.8b00238.

Maeda, Y. *et al.* (1997) 'Engineering of functional chimeric protein G-Vargula luciferase', *Analytical Biochemistry*, 249(2), pp. 147–152. doi: 10.1006/abio.1997.2181.

Maier, T., Leibundgut, M. and Ban, N. (2008) 'The crystal structure of a mammalian fatty acid synthase', *Science*, 321(5894), pp. 1315–1322. doi: 10.1126/science.1161269.

Malik, A. (2016) 'Protein fusion tags for efficient expression and purification of recombinant proteins in the periplasmic space of *E. coli*', *3 Biotech*. Springer Verlag, pp. 1–7. doi: 10.1007/s13205-016-0397-7.

Mamoun, C. Ben *et al.* (1998) 'Identification and characterization of an unusual double serine/threonine protein phosphatase 2C in the malaria parasite *Plasmodium falciparum*', *Journal of Biological Chemistry*, 273(18), pp. 11241–11247. doi: 10.1074/jbc.273.18.11241.

Mangala, S. L. *et al.* (2003) 'Fusion of family VI cellulose binding domains to *Bacillus halodurans* xylanase increases its catalytic activity and substrate-binding capacity to insoluble xylan', *Journal of Molecular Catalysis B: Enzymatic*, 21(4–6), pp. 221–230. doi: 10.1016/S1381-1177(02)00226-6.

Marcelino, P. R. F. *et al.* (2019) 'Biosurfactants production by yeasts using sugarcane bagasse hemicellulosic hydrolysate as new sustainable alternative for lignocellulosic biorefineries', *Industrial Crops and Products*, 129, pp. 212–223. doi: 10.1016/j.indcrop.2018.12.001.

Marillonnet, S. and Grütznér, R. (2020) 'Synthetic DNA Assembly Using Golden Gate Cloning and the Hierarchical Modular Cloning Pipeline', *Current Protocols in Molecular Biology*, 130(1), p. e115. doi: 10.1002/cpmb.115.

Marqusee, S. and Baldwin, R. L. (1987) 'Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design.', *Proceedings of the National Academy of Sciences of the United States of America*, 84(24), pp. 8898–902. Available at:

<http://www.ncbi.nlm.nih.gov/pubmed/3122208> (Accessed: 20 February 2018).

Martin, C. *et al.* (2018) 'Creating a more robust 5-hydroxymethylfurfural oxidase by combining computational predictions with a novel effective library design', *Biotechnology for Biofuels*, 11(1), p. 56. doi: 10.1186/s13068-018-1051-x.

Martins, G. M. *et al.* (2018) 'The isolation of pentose-assimilating yeasts and their xylose fermentation potential', *Brazilian Journal of Microbiology*, 49(1), pp. 162–168. doi: 10.1016/j.bjm.2016.11.014.

Mattheyses, A. L., Simon, S. M. and Rappoport, J. Z. (2010) 'Imaging with total internal reflection fluorescence microscopy for the cell biologist', *Journal of Cell Science*. The Company of Biologists Ltd, pp. 3621–3628. doi: 10.1242/jcs.056218.

McCormick, A. L., Thomas, M. S. and Heath, A. W. (2001) 'Immunization with an Interferon- γ -gp120 Fusion Protein Induces Enhanced Immune Responses to Human Immunodeficiency Virus gp120', *The Journal of Infectious Diseases*, 184(11), pp. 1423–1430. doi: 10.1086/324371.

McDonald, P. H. *et al.* (2000) ' β -Arrestin 2: A Receptor-Regulated MAPK Scaffold for the Activation of JNK3', *Science*, 290(5496).

Mechaly, A. *et al.* (2001) 'Cohesin-dockerin interaction in cellulosome assembly: a single hydroxyl group of a dockerin domain distinguishes between nonrecognition and high affinity recognition.', *The Journal of biological chemistry*, 276(13), pp. 9883–8. doi: 10.1074/jbc.M009237200.

Meijnen, J. P., De Winde, J. H. and Ruijssenaars, H. J. (2008) 'Engineering *Pseudomonas putida* S12 for efficient utilization of D-xylose and L-arabinose', *Applied and Environmental Microbiology*, 74(16), pp. 5031–5037. doi: 10.1128/AEM.00924-08.

Melo, J. V. (1996) 'The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype', *Blood*. American Society of Hematology, pp. 2375–2384. doi: 10.1182/blood.v88.7.2375.bloodjournal8872375.

Meng, M., Bagdasarian, M. and Zeikus, J. G. (1993) The role of active-site aromatic and polar residues in catalysis and substrate discrimination by xylose isomerase (glucose isomerase/site-directed mutagenesis/enzyme-active site/protein engineering/catalytic efficiency), *Proc. Natl. Acad. Sci. USA*.

Merlin, C., McAteer, S. and Masters, M. (2002) 'Tools for characterization of *Escherichia coli*

genes of unknown function', *Journal of Bacteriology*, 184(16), pp. 4573–4581. doi: 10.1128/JB.184.16.4573-4581.2002.

Mestrom, L. *et al.* (2019) 'Artificial fusion of mCherry enhances trehalose transferase solubility and stability', *Applied and Environmental Microbiology*, 85(8), pp. 3084–3102. doi: 10.1128/AEM.03084-18.

Meyer, S. and Dutzler, R. (2006) 'Crystal Structure of the Cytoplasmic Domain of the Chloride Channel ClC-0', *Structure*, 14(2), pp. 299–307. doi: 10.1016/j.str.2005.10.008.

Meynial Salles, I. *et al.* (2007) 'Evolution of a *Saccharomyces cerevisiae* metabolic pathway in *Escherichia coli*', *Metabolic Engineering*, 9(2), pp. 152–159. doi: 10.1016/j.ymben.2006.09.002.

Miles, B. W., Banzon, J. A. and Raushel, F. M. (1998) 'Regulatory control of the amidotransferase domain of carbamoyl phosphate synthetase', *Biochemistry*, 37(47), pp. 16773–16779. doi: 10.1021/bi982018g.

Miles, E. W., Rhee, S. and Davies, D. R. (1999) 'The molecular basis of substrate channeling.', *The Journal of biological chemistry*, 274(18), pp. 12193–6. doi: 10.1074/JBC.274.18.12193.

Moes, C. J., Pretorius, I. S. and Van Zyl, W. H. (1996) 'Cloning and expression of the *Clostridium thermosulfurogenes* d-xylose isomerase gene (*xylA*) in *Saccharomyces cerevisiae*', *Biotechnology Letters*, 18(3), pp. 269–274. doi: 10.1007/BF00142943.

Mohapatra, S. and Weisshaar, J. C. (2018) 'Functional mapping of the *E. coli* translational machinery using single-molecule tracking', *Molecular Microbiology*, 110(2), pp. 262–282. doi: 10.1111/mmi.14103.

Moon, H. G. *et al.* (2016) 'One hundred years of clostridial butanol fermentation', *FEMS Microbiology Letters*, 363(3), p. fnw001. doi: 10.1093/femsle/fnw001.

Moon, T. S. *et al.* (2010) 'Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli*', *Metabolic Engineering*, 12(3), pp. 298–305. doi: 10.1016/j.ymben.2010.01.003.

Moraes, T. F. and Reithmeier, R. A. F. (2012) 'Membrane transport metabolons', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818(11), pp. 2687–2706. doi: 10.1016/j.bbamem.2012.06.007.

Murat, D., Byrne, M. and Komeili, A. (2010) 'Cell biology of prokaryotic organelles.', *Cold Spring Harbor perspectives in biology*. Cold Spring Harbor Laboratory Press. doi:

10.1101/cshperspect.a000422.

Myhrvold, C., Polka, J. K. and Silver, P. A. (2016) 'Synthetic Lipid-Containing Scaffolds Enhance Production by Colocalizing Enzymes', *ACS Synthetic Biology*, p. acssynbio.6b00141. doi: 10.1021/acssynbio.6b00141.

Nagradova, N. K. (2001) 'Interdomain interactions in oligomeric enzymes: Creation of asymmetry in homo-oligomers and role in metabolite channeling between active centers of hetero-oligomers', *FEBS Letters*. No longer published by Elsevier, pp. 327–332. doi: 10.1016/S0014-5793(00)02338-3.

Naik, S. N. *et al.* (2010) 'Production of first and second generation biofuels: A comprehensive review', *Renewable and Sustainable Energy Reviews*. Pergamon, pp. 578–597. doi: 10.1016/j.rser.2009.10.003.

Narita, J. *et al.* (2006) 'Display of active enzymes on the cell surface of *Escherichia coli* using PgsA anchor protein and their application to bioconversion', *Applied Microbiology and Biotechnology*, 70(5), pp. 564–572. doi: 10.1007/s00253-005-0111-x.

Nebel, B. A. *et al.* (2014) 'Biooxidation of n-butane to 1-butanol by engineered P450 monooxygenase under increased pressure', *Journal of Biotechnology*, 191, pp. 86–92. doi: 10.1016/j.jbiotec.2014.08.022.

Neuberger, M. S., Hartley, B. S. and Walker, J. E. (1981) 'Purification and properties of D-ribulokinase and D-xylulokinase from *Klebsiella aerogenes*', *Biochemical Journal*, 193(2), pp. 513–524. doi: 10.1042/bj1930513.

Ni, L. *et al.* (2013) 'Structures of the *Escherichia coli* transcription activator and regulator of diauxie, XylR: An AraC DNA-binding family member with a LacI/GalR ligand-binding domain', *Nucleic Acids Research*, 41(3), pp. 1998–2008. doi: 10.1093/nar/gks1207.

Nidetzky, B. *et al.* (2003) 'Multiple Forms of Xylose Reductase in *Candida intermedia*: Comparison of Their Functional Properties Using Quantitative Structure–Activity Relationships, Steady-State Kinetic Analysis, and pH Studies', *Journal of Agricultural and Food Chemistry*, 51(27), pp. 7930–7935. doi: 10.1021/jf034426j.

Nies, D. H. and Herzberg, M. (2013) 'A fresh view of the cell biology of copper in *enterobacteria*', *Molecular Microbiology*, 87(3), pp. 447–454. doi: 10.1111/mmi.12123.

Nogueira, M. *et al.* (2019) 'Construction of a fusion enzyme for astaxanthin formation and its characterisation in microbial and plant hosts: A new tool for engineering ketocarotenoids',

Metabolic Engineering, 52, pp. 243–252. doi: 10.1016/j.ymben.2018.12.006.

Notredame, C., Higgins, D. G. and Heringa, J. (2000) 'T-coffee: A novel method for fast and accurate multiple sequence alignment', *Journal of Molecular Biology*, 302(1), pp. 205–217. doi: 10.1006/jmbi.2000.4042.

Oh, D. *et al.* (2014) 'Dynamics of the serine chemoreceptor in the *Escherichia coli* inner membrane: A high-speed single-molecule tracking study', *Biophysical Journal*, 106(1), pp. 145–153. doi: 10.1016/j.bpj.2013.09.059.

Okamoto, T. *et al.* (1998) 'Caveolins, a family of scaffolding proteins for organizing "preassembled signaling complexes" at the plasma membrane.', *The Journal of biological chemistry*, 273(10), pp. 5419–22. doi: 10.1074/JBC.273.10.5419.

Oldham, P., Hall, S. and Burton, G. (2012) 'Synthetic biology: Mapping the Scientific landscape', *PLoS ONE*, 7(4). doi: 10.1371/journal.pone.0034368.

Oliva, J. M. *et al.* (2006) 'Effects of acetic acid, furfural and catechol combinations on ethanol fermentation of *Kluyveromyces marxianus*', *Process Biochemistry*, 41(5), pp. 1223–1228. doi: 10.1016/j.procbio.2005.12.003.

Oreb, M. (2020) 'Construction of artificial membrane transport metabolons – an emerging strategy in metabolic engineering', *FEMS Microbiology Letters*, 367(1), p. 27. doi: 10.1093/femsle/fnaa027.

Ortiz de Montellano, P. (2005) *Cytochrome P450*. Edited by P. R. Ortiz de Montellano. Boston, MA: Springer US. doi: 10.1007/b139087.

Oswald, F. *et al.* (2014) 'Imaging and quantification of trans-membrane protein diffusion in living bacteria', *Physical Chemistry Chemical Physics*. Royal Society of Chemistry, pp. 12625–12634. doi: 10.1039/c4cp00299g.

Oswald, F. *et al.* (2016) 'MreB-Dependent Organization of the *E. coli* Cytoplasmic Membrane Controls Membrane Protein Diffusion', *Biophysical Journal*, 110(5), pp. 1139–1149. doi: 10.1016/j.bpj.2016.01.010.

Paddon, C. J. and Keasling, J. D. (2014) 'Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development', *Nature Reviews Microbiology*, 12(5), pp. 355–367. doi: 10.1038/nrmicro3240.

Di Paolo, D. *et al.* (2016) 'Single-molecule imaging of electroporated dye-labelled cheY in live *Escherichia coli*', *Philosophical Transactions of the Royal Society B: Biological Sciences*,

371(1707). doi: 10.1098/rstb.2015.0492.

Papaleo, E. *et al.* (2016) 'The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery', *Chemical Reviews*. American Chemical Society, pp. 6391–6423. doi: 10.1021/acs.chemrev.5b00623.

Park, H.-S. (2006) 'Design and Evolution of New Catalytic Activity with an Existing Protein Scaffold', *Science*, 311(5760), pp. 535–538. doi: 10.1126/science.1118953.

Park, S. H. *et al.* (2019) 'Crystal structure and functional characterization of a Xylose isomerase (PbXI) from the psychrophilic soil microorganism, *Paenibacillus* sp.', *Journal of Microbiology and Biotechnology*, 29(2), pp. 244–255. doi: 10.4014/jmb.1810.10057.

Parsonage, D. *et al.* (2005) 'Analysis of the link between enzymatic activity and oligomeric state in AhpC, a bacterial peroxiredoxin', *Biochemistry*, 44(31), pp. 10583–10592. doi: 10.1021/bi050448i.

Pasek, S., Risler, J.-L. and Brezellec, P. (2006) 'Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins', *Bioinformatics*, 22(12), pp. 1418–1423. doi: 10.1093/bioinformatics/btl135.

Pauly, M. *et al.* (2013) 'Hemicellulose biosynthesis', *Planta*. Springer, pp. 627–642. doi: 10.1007/s00425-013-1921-1.

Peng, B. *et al.* (2015) 'Bacterial xylose isomerases from the mammal gut *Bacteroidetes* cluster function in *Saccharomyces cerevisiae* for effective xylose fermentation', *Microbial Cell Factories*, 14(1), p. 70. doi: 10.1186/s12934-015-0253-1.

Pérez-Arellano, I. *et al.* (2010) 'Pyrroline-5-carboxylate synthase and proline biosynthesis: From osmotolerance to rare metabolic disease', *Protein Science*, 19(3), p. NA-NA. doi: 10.1002/pro.340.

Pettersson, H. and Pettersson, G. (2001) 'Kinetics of the coupled reaction catalysed by a fusion protein of β -galactosidase and galactose dehydrogenase', *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1549(2), pp. 155–160. doi: 10.1016/S0167-4838(01)00252-7.

Phane Richard, S. *et al.* (1995) 'Association of p62, a Multifunctional SH2-and SH3-Domain-Binding Protein, with src Family Tyrosine Kinases, Grb2, and Phospholipase C α -1', *MOLECULAR AND CELLULAR BIOLOGY*, 15(1), pp. 186–197.

Postle, K. *et al.* (2010) 'The TonB dimeric crystal structures do not exist in vivo', *mBio*, 1(5),

pp. 307–317. doi: 10.1128/mBio.00307-10.

Price, G. D. *et al.* (2008) 'Advances in understanding the cyanobacterial CO₂-concentrating-mechanism (CCM): Functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants', in *Journal of Experimental Botany*. Oxford Academic, pp. 1441–1461. doi: 10.1093/jxb/erm112.

Pröschel, M. *et al.* (2015) 'Engineering of Metabolic Pathways by Artificial Enzyme Channels.', *Frontiers in bioengineering and biotechnology*, 3, p. 168. doi: 10.3389/fbioe.2015.00168.

Qi, X. *et al.* (2015) 'Heterologous xylose isomerase pathway and evolutionary engineering improve xylose utilization in *Saccharomyces cerevisiae*', *Frontiers in Microbiology*, 6(OCT). doi: 10.3389/fmicb.2015.01165.

Quin, M. B. *et al.* (2017) 'Spatial organization of multi-enzyme biocatalytic cascades', *Org. Biomol. Chem.*, 15(20), pp. 4260–4271. doi: 10.1039/C7OB00391A.

Rahman, M. H. and Bhoi, P. R. (2021) 'An overview of non-biodegradable bioplastics', *Journal of Cleaner Production*. Elsevier Ltd, p. 126218. doi: 10.1016/j.jclepro.2021.126218.

Rapin, N. and Porse, B. T. (2014) 'Oncogenic fusion proteins expressed in immature hematopoietic cells fail to recapitulate the transcriptional changes observed in human AML', *Oncogenesis*, 3(6), pp. e106–e106. doi: 10.1038/oncsis.2014.22.

Raschdorf, O. *et al.* (2013) 'The magnetosome proteins MamX, MamZ and MamH are involved in redox control of magnetite biomineralization in *Magnetospirillum gryphiswaldense*.', *Molecular microbiology*, 89(5), pp. 872–86. doi: 10.1111/mmi.12317.

Rassam, P. *et al.* (2015) 'Supramolecular assemblies underpin turnover of outer membrane proteins in bacteria', *Nature*, 523(7560), pp. 333–336. doi: 10.1038/nature14461.

Rassam, P. *et al.* (2018) 'Intermembrane crosstalk drives inner-membrane protein organization in *Escherichia coli*', *Nature Communications*, 9(1), pp. 1–8. doi: 10.1038/s41467-018-03521-4.

Rath, A. *et al.* (2009) 'Detergent binding explains anomalous SDS-PAGE migration of membrane proteins.', *Proceedings of the National Academy of Sciences of the United States of America*, 106(6), pp. 1760–5. doi: 10.1073/pnas.0813167106.

Reddy, V. S. *et al.* (2012) 'The major facilitator superfamily (MFS) revisited', *FEBS Journal*, 279(11), pp. 2022–2035. doi: 10.1111/j.1742-4658.2012.08588.x.

- Rees, D. C., Johnson, E. and Lewinson, O. (2009) 'ABC transporters: The power to change', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, pp. 218–227. doi: 10.1038/nrm2646.
- Reits, E. A. J. and Neefjes, J. J. (2001) 'From fixed to FRAP: Measuring protein mobility and activity in living cells', *Nature Cell Biology*. doi: 10.1038/35078615.
- Reuten, R. *et al.* (2016) 'Maltose-binding protein (MBP), a secretion-enhancing tag for mammalian protein expression systems', *PLoS ONE*, 11(3). doi: 10.1371/journal.pone.0152386.
- Reuter, L. J. *et al.* (2017) 'Coating Nanoparticles with Plant-Produced Transferrin-Hydrophobin Fusion Protein Enhances Their Uptake in Cancer Cells', *Bioconjugate Chemistry*, 28(6), pp. 1639–1648. doi: 10.1021/acs.bioconjchem.7b00075.
- Rey, F. *et al.* (1988) 'Structural analysis of the 2.8 Å model of xylose isomerase from *Actinoplanes missouriensis*', *Proteins: Structure, Function, and Genetics*, 4(3), pp. 165–172. doi: 10.1002/prot.340040303.
- Robinson, C. R. and Sauer, R. T. (1998) 'Optimizing the stability of single-chain proteins by linker length and composition mutagenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 95(11), pp. 5929–5934. doi: 10.1073/pnas.95.11.5929.
- Roderer, K. *et al.* (2014) 'Functional Mapping of Protein-Protein Interactions in an Enzyme Complex by Directed Evolution', *PLoS ONE*. Edited by N. Srinivasan, 9(12), p. e116234. doi: 10.1371/journal.pone.0116234.
- Rodionova, M. V. *et al.* (2017) 'Biofuel production: Challenges and opportunities', *International Journal of Hydrogen Energy*. Elsevier Ltd, pp. 8450–8461. doi: 10.1016/j.ijhydene.2016.11.125.
- Rodriguez-Pena, J. M. *et al.* (1998) 'The *YGR194c* (*XKS1*) gene encodes the xylulokinase from the budding yeast *Saccharomyces cerevisiae*', *FEMS Microbiology Letters*, 162(1), pp. 155–160. doi: 10.1111/j.1574-6968.1998.tb12993.x.
- Roe, A. J. *et al.* (2002) 'Inhibition of *Escherichia coli* growth by acetic acid: A problem with methionine biosynthesis and homocysteine toxicity', *Microbiology*, 148(7), pp. 2215–2222. doi: 10.1099/00221287-148-7-2215.
- Rong Fu Wang and Kushner, S. R. (1991) 'Construction of versatile low-copy-number vectors

for cloning, sequencing and gene expression in *Escherichia coli*, *Gene*, 100, pp. 195–199. doi: 10.1016/0378-1119(91)90366-J.

Rothemund, P. W. K. (2006) 'Folding DNA to create nanoscale shapes and patterns', *Nature*, 440(7082), pp. 297–302. doi: 10.1038/nature04586.

Sachdeva, G. *et al.* (2014) 'In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner.', *Nucleic acids research*, 42(14), pp. 9493–503. doi: 10.1093/nar/gku617.

Saha, B. C. (2003) 'Hemicellulose bioconversion', in *Journal of Industrial Microbiology and Biotechnology*. Springer, pp. 279–291. doi: 10.1007/s10295-003-0049-x.

Saier, M. H. *et al.* (1999) 'The major facilitator superfamily.', *Journal of molecular microbiology and biotechnology*, 1(2), pp. 257–79. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10943556> (Accessed: 30 April 2020).

Saier, M. H. (2003) 'Tracing pathways of transport protein evolution', *Molecular Microbiology*, 48(5), pp. 1145–1156. doi: 10.1046/j.1365-2958.2003.03499.x.

Saier, M. H. (2016) 'Transport protein evolution deduced from analysis of sequence, topology and structure', *Current Opinion in Structural Biology*, 38, pp. 9–17. doi: 10.1016/j.sbi.2016.05.001.

Sakharkar, K. R., Sakharkar, M. K. and Chow, V. T. K. (2006) 'Gene fusion in *Helicobacter pylori*: Making the ends meet', *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, 89(1), pp. 169–180. doi: 10.1007/s10482-005-9021-2.

Sánchez, C. (2009) 'Lignocellulosic residues: Biodegradation and bioconversion by fungi', *Biotechnology Advances*. Elsevier, pp. 185–194. doi: 10.1016/j.biotechadv.2008.11.001.

Sarthy, A. V. *et al.* (1987) 'Expression of the *Escherichia coli* xylose isomerase gene in *Saccharomyces cerevisiae*.', *Applied and environmental microbiology*, 53(9), pp. 1996–2000. doi: 10.1128/aem.53.9.1996-2000.1987.

Schellenberg, G. D. *et al.* (1984) 'Xylose isomerase from *Escherichia coli*. Characterization of the protein and the structural gene.', *The Journal of biological chemistry*, 259(11), pp. 6826–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6327696> (Accessed: 27 February 2017).

Schindelin, J. *et al.* (2012) 'Fiji: An open-source platform for biological-image analysis', *Nature Methods*. Nature Publishing Group, pp. 676–682. doi: 10.1038/nmeth.2019.

- Schlegel, S. *et al.* (2010) 'Revolutionizing membrane protein overexpression in bacteria', *Microbial Biotechnology*, pp. 403–411. doi: 10.1111/j.1751-7915.2009.00148.x.
- Schlunzen, F. *et al.* (2000) 'Structure of functionally activated small ribosomal subunit at 3.3 Å resolution', *Cell*, 102(5), pp. 615–623. doi: 10.1016/S0092-8674(00)00084-2.
- Schoene, C. *et al.* (2014) 'SpyTag/Spycatcher cyclization confers resilience to boiling on a mesophilic enzyme', *Angewandte Chemie - International Edition*, 53(24), pp. 6101–6104. doi: 10.1002/anie.201402519.
- Scholtz, J. M. and Baldwin, R. L. (1992) The Mechanism Of A-Helix Formation By Peptides, *Annu. Rev. Biophys. Biomol. Struct.* Available at: www.annualreviews.org (Accessed: 26 March 2020).
- Schötzig, U. (2000) 'Half-life and X-ray emission probabilities of ⁵⁵Fe', *Applied Radiation and Isotopes*, 53(3), pp. 469–472. doi: 10.1016/S0969-8043(00)00166-4.
- Schuler, B. *et al.* (2005) 'Polyproline and the "spectroscopic ruler"; revisited with single-molecule fluorescence.', *Proceedings of the National Academy of Sciences of the United States of America*, 102(8), pp. 2754–9. doi: 10.1073/pnas.0408164102.
- Schuurman-Wolters, G. K. *et al.* (2018) 'Protein Linkers Provide Limits on the Domain Interactions in the ABC Importer GlnPQ and Determine the Rate of Transport.', *Journal of molecular biology*, 430(8), pp. 1249–1262. doi: 10.1016/j.jmb.2018.02.014.
- Schweizer, E. and Hofmann, J. (2004) 'Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems.', *Microbiology and molecular biology reviews : MMBR*, 68(3), pp. 501–17, table of contents. doi: 10.1128/MMBR.68.3.501-517.2004.
- Seinen, A. B. *et al.* (2021) 'Cellular dynamics of the SecA ATPase at the single molecule level', *Scientific Reports*, 11(1), p. 1433. doi: 10.1038/s41598-021-81081-2.
- Severi, E. *et al.* (2005) 'Sialic acid transport in *Haemophilus influenzae* is essential for lipopolysaccharide sialylation and serum resistance and is dependent on a novel tripartite ATP-independent periplasmic transporter', *Molecular Microbiology*, 58(4), pp. 1173–1185. doi: 10.1111/j.1365-2958.2005.04901.x.
- Severi, E. and Thomas, G. H. (2019) 'Antibiotic export: transporters involved in the final step of natural product production', *Microbiology*. doi: 10.1099/mic.0.000794.
- Shamriz, S., Ofoghi, H. and Moazami, N. (2016) 'Effect of linker length and residues on the structure and stability of a fusion protein with malaria vaccine application', *Computers in*

Biology and Medicine, 76, pp. 24–29. doi: 10.1016/j.compbiomed.2016.06.015.

Shaner, N. C. *et al.* (2004) 'Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein', *Nature Biotechnology*, 22(12), pp. 1567–1572. doi: 10.1038/nbt1037.

Sibbesen, O. *et al.* (1994) 'Isolation of the heme-thiolate enzyme cytochrome P-450TYR, which catalyzes the committed step in the biosynthesis of the cyanogenic glucoside dhurrin in *Sorghum bicolor* (L.) Moench.', *Proceedings of the National Academy of Sciences of the United States of America*, 91(21), pp. 9740–4. doi: 10.1073/pnas.91.21.9740.

Sievers, F. *et al.* (2011) 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular Systems Biology*, 7(1), p. 539. doi: 10.1038/msb.2011.75.

Sims, R. E. H. *et al.* (2010) 'An overview of second generation biofuel technologies', *Bioresource Technology*. Elsevier Ltd, pp. 1570–1580. doi: 10.1016/j.biortech.2009.11.046.

Singh, D. *et al.* (2020) 'A review on feedstocks, production processes, and yield for different generations of biodiesel', *Fuel*. Elsevier Ltd, p. 116553. doi: 10.1016/j.fuel.2019.116553.

Singh, V. *et al.* (2017) 'Strategies for fermentation medium optimization: An in-depth review', *Frontiers in Microbiology*. Frontiers Research Foundation. doi: 10.3389/fmicb.2016.02087.

Siu, K.-H. *et al.* (2015) 'Synthetic scaffolds for pathway enhancement', *Current Opinion in Biotechnology*, 36, pp. 98–106. doi: 10.1016/j.copbio.2015.08.009.

Smolke, C. D. (2018) *Synthetic biology: parts, devices and applications*. John Wiley & Sons.

Somasundaram, S. *et al.* (2017) 'Introduction of synthetic protein complex between *Pyrococcus horikoshii* glutamate decarboxylase and *Escherichia coli* GABA transporter for the improved production of GABA', *Biochemical Engineering Journal*, 120, pp. 1–6. doi: 10.1016/j.bej.2016.12.020.

Somasundaram, S., Tae Eom, G. and Ho Hong, S. (2010) 'Efficient Malic Acid Production in *Escherichia coli* Using a Synthetic Scaffold Protein Complex'. doi: 10.1007/s12010-017-2629-7.

Song, S. and Park, C. (1997) Organization and Regulation of the D-Xylose Operons in *Escherichia coli* K-12: XylR Acts as a Transcriptional Activator, *JOURNAL OF BACTERIOLOGY*. Available at: <http://jlb.asm.org/> (Accessed: 1 May 2020).

Spivey, H. O. and Ovádi, J. (1999) 'Substrate Channeling', *Methods*, 19(2), pp. 306–321. Available at: <https://www.sciencedirect.com/science/article/pii/S1046202399908587?via%3Dihub> (Accessed: 30 November 2016).

Srivastava, D. K. and Bernhard, S. A. (1986) 'Metabolite transfer via enzyme-enzyme complexes', *Science*, 234(4780), pp. 1081–1086. doi: 10.1126/science.3775377.

Stahl, S. W. *et al.* (2012) 'Single-molecule dissection of the high-affinity cohesin-dockerin complex', *Proceedings of the National Academy of Sciences of the United States of America*, 109(50), pp. 20431–20436. doi: 10.1073/pnas.1211929109.

Stressler, T. *et al.* (2016) 'A fusion protein consisting of the exopeptidases PepN and PepX—production, characterization, and application', *Applied Microbiology and Biotechnology*, 100(17), pp. 7499–7515. doi: 10.1007/s00253-016-7478-8.

Stroud, R. M. (1994) 'An electrostatic highway', *Nature Structural Biology*. Presse Dienstleistungsgesellschaft mbH und Co. KG, pp. 131–134. doi: 10.1038/nsb0394-131.

Stuurman, N., Amdodaj, N. and Vale, R. (2007) 'µManager: Open Source Software for Light Microscope Imaging', *Microscopy Today*, 15(3), pp. 42–43. doi: 10.1017/s1551929500055541.

Subach, F. V. *et al.* (2009) 'Photoactivatable mCherry for high-resolution two-color fluorescence microscopy', *Nature Methods*, 6(2), pp. 153–159. doi: 10.1038/nmeth.1298.

Sun, L. *et al.* (2012) 'Crystal structure of a bacterial homologue of glucose transporters GLUT1–4', *Nature*, 490(7420), pp. 361–366. doi: 10.1038/nature11524.

Sutter, M. *et al.* (2008) 'Structural basis of enzyme encapsulation into a bacterial nanocompartment', *Nature Structural and Molecular Biology*, 15(9), pp. 939–947. doi: 10.1038/nsmb.1473.

Svedružić, Ž. M. *et al.* (2020) 'Substrate Channeling via a Transient Protein-Protein Complex: The case of D-Glyceraldehyde-3-Phosphate Dehydrogenase and L-Lactate Dehydrogenase', *Scientific Reports*, 10(1), pp. 1–12. doi: 10.1038/s41598-020-67079-2.

Sweetlove, L. J. and Fernie, A. R. (2018) 'The role of dynamic enzyme assemblies and substrate channelling in metabolic regulation', *Nature Communications*, 9(1), p. 2136. doi: 10.1038/s41467-018-04543-8.

Tanaka, S. *et al.* (2008) 'Atomic-level models of the bacterial carboxysome shell', *Science*,

319(5866), pp. 1083–1086. doi: 10.1126/science.1151458.

Terfrüchte, M. *et al.* (2014) 'Establishing a versatile Golden Gate cloning system for genetic engineering in fungi', *Fungal Genetics and Biology*, 62, pp. 1–10. doi: 10.1016/j.fgb.2013.10.012.

Teunissen, M. J. *et al.* (1991) 'Comparison of growth characteristics of anaerobic fungi isolated from ruminant and non-ruminant herbivores during cultivation in a defined medium', *Journal of General Microbiology*, 137(6), pp. 1401–1408. doi: 10.1099/00221287-137-6-1401.

Thoden, J. B. *et al.* (1997) 'Structure of Carbamoyl Phosphate Synthetase: A Journey of 96 Å from Substrate to Product † , ‡', *Biochemistry*, 36(21), pp. 6305–6316. doi: 10.1021/bi970503q.

Thomas, F. *et al.* (2013) 'A set of *de novo* designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime', *Journal of the American Chemical Society*, 135(13), pp. 5161–5166. doi: 10.1021/ja312310g.

Thomik, T. *et al.* (2017) 'An artificial transport metabolon facilitates improved substrate utilization in yeast', *Nature Chemical Biology*, 13(11), pp. 1158–1163. doi: 10.1038/nchembio.2457.

Tippmann, S. *et al.* (2016) 'Affibody Scaffolds Improve Sesquiterpene Production in *Saccharomyces cerevisiae*', *ACS Synthetic Biology*, p. acssynbio.6b00109. doi: 10.1021/acssynbio.6b00109.

Tomita, A. *et al.* (2017) 'ATP-dependent modulation of MgtE in Mg²⁺ homeostasis.', *Nature communications*, 8(1), p. 148. doi: 10.1038/s41467-017-00082-w.

Torres Pazmiño, D. E. *et al.* (2008) 'Self-Sufficient Baeyer–Villiger Monooxygenases: Effective Coenzyme Regeneration for Biooxygenation by Fusion Engineering', *Angewandte Chemie International Edition*, 47(12), pp. 2275–2278. doi: 10.1002/anie.200704630.

Torres Pazmiño, D. E. *et al.* (2009) 'Efficient Biooxidations Catalyzed by a New Generation of Self-Sufficient Baeyer-Villiger Monooxygenases', *ChemBioChem*, 10(16), pp. 2595–2598. doi: 10.1002/cbic.200900480.

Trifinopoulos, J. *et al.* (2016) 'W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis', *Nucleic acids research*, 44(W1), pp. W232–W235. doi: 10.1093/nar/gkw256.

Tritsch, D. *et al.* (2004) 'A sensitive radiometric assay to measure D-xylulose kinase activity',

Journal of Biochemical and Biophysical Methods, 58(1), pp. 75–83. doi: 10.1016/j.jbbm.2003.08.001.

Tron, C. M. *et al.* (2009) 'Structural and Functional Studies of the Biotin Protein Ligase from *Aquifex aeolicus* Reveal a Critical Role for a Conserved Residue in Target Specificity', *Journal of Molecular Biology*, 387(1), pp. 129–146. doi: 10.1016/J.JMB.2008.12.086.

Truffi, M. *et al.* (2016) 'Ferritin nanocages: A biological platform for drug delivery, imaging and theranostics in cancer', *Pharmacological Research*. Academic Press, pp. 57–65. doi: 10.1016/j.phrs.2016.03.002.

Tsai, S.-L. *et al.* (2009) 'Functional Assembly of Minicellulosomes on the *Saccharomyces cerevisiae* Cell Surface for Cellulose Hydrolysis and Ethanol Production', *Applied and Environmental Microbiology*, 75(19), pp. 6087–6093. doi: 10.1128/AEM.01538-09.

Tsung, K., Inouye, S. and Inouye, M. (1989) 'Factors affecting the efficiency of protein synthesis in *Escherichia coli*. Production of a polypeptide of more than 6000 amino acid residues', *Journal of Biological Chemistry*, 264(8), pp. 4428–4433. doi: 10.1016/S0021-9258(18)83760-7.

Tucker, M. Y. *et al.* (1988) 'Properties of genetically overproduced *E. coli* xylose isomerase', *Biotechnology Letters*, 10(2), pp. 79–84. doi: 10.1007/BF01024630.

Turmo, A., Gonzalez-Esquer, C. R. and Kerfeld, C. A. (2017) 'Carboxysomes: metabolic modules for CO₂ fixation', *FEMS Microbiology Letters*, 364(18), p. 176. doi: 10.1093/femsle/fnx176.

Uo, M., Wada, T. and Sugiyama, T. (2015) 'Applications of X-ray fluorescence analysis (XRF) to dental and medical specimens', *Japanese Dental Science Review*. Elsevier Ltd, pp. 2–9. doi: 10.1016/j.jdsr.2014.07.001.

Uthandi, S. *et al.* (2021) 'Microbial biodiesel production from lignocellulosic biomass: New insights and future challenges', *Critical Reviews in Environmental Science and Technology*. doi: 10.1080/10643389.2021.1877045.

Vanderstraeten, J. and Briers, Y. (2020) 'Synthetic protein scaffolds for the colocalisation of co-acting enzymes', *Biotechnology Advances*. Elsevier Inc., p. 107627. doi: 10.1016/j.biotechadv.2020.107627.

Varnaité, R. and MacNeill, S. A. (2016) 'Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID', *Proteomics*. Wiley-VCH Verlag, pp.

2503–2518. doi: 10.1002/pmic.201600123.

Vertès, A. A., Inui, M. and Yukawa, H. (2006) 'Implementing biofuels on a global scale', *Nature Biotechnology*. Nature Publishing Group, pp. 761–764. doi: 10.1038/nbt0706-761.

Vieille, C. *et al.* (1995) 'xylA cloning and sequencing and biochemical characterization of xylose isomerase from *Thermotoga neapolitana*.', *Applied and environmental microbiology*, 61(5), pp. 1867–75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7646024> (Accessed: 27 February 2017).

Vieille, C. *et al.* (2001) 'Bivalent cations and amino-acid composition contribute to the thermostability of *Bacillus licheniformis* xylose isomerase', *European Journal of Biochemistry*, 268(23), pp. 6291–6301. doi: 10.1046/j.0014-2956.2001.02587.x.

Voronovsky, A. Y. *et al.* (2005) 'Expression of xylA genes encoding xylose isomerases from *Escherichia coli* and *Streptomyces coelicolor* in the methylotrophic yeast *Hansenula polymorpha*', *FEMS Yeast Research*, 5(11), pp. 1055–1062. doi: 10.1016/j.femsyr.2005.09.001.

Wagner, S. *et al.* (2006) 'Rationalizing membrane protein overexpression', *Trends in Biotechnology*, pp. 364–371. doi: 10.1016/j.tibtech.2006.06.008.

Wagner, S. *et al.* (2007) 'Consequences of membrane protein overexpression in *Escherichia coli*', *Molecular and Cellular Proteomics*, 6(9), pp. 1527–1550. doi: 10.1074/mcp.M600431-MCP200.

Waldo, G. S. *et al.* (1999) 'Rapid protein-folding assay using green fluorescent protein', *Nature Biotechnology*, 17(7), pp. 691–695. doi: 10.1038/10904.

Walfridsson, M. *et al.* (1996) 'Ethanol fermentation of xylose with *Saccharomyces cerevisiae* harboring the *Thermus thermophilus* xylA gene, which expresses an active xylose (glucose) isomerase', *Applied and Environmental Microbiology*, 62(12), pp. 4648–4651. doi: 10.1128/aem.62.12.4648-4651.1996.

Wallace, W. and Nicholas, D. J. (1969) 'The biochemistry of nitrifying microorganisms.', *Biological reviews of the Cambridge Philosophical Society*, pp. 359–391. doi: 10.1111/j.1469-185x.1969.tb01216.x.

Waltman, M. J. *et al.* (2014) 'Engineering acidic *Streptomyces rubiginosus* D-xylose isomerase by rational enzyme design', *Protein Engineering Design and Selection*, 27(2), pp. 59–64. doi: 10.1093/protein/gzt062.

Wang, G.-Z., Lercher, M. J. and Hurst, L. D. (2011) 'Transcriptional coupling of neighboring genes and gene expression noise: evidence that gene orientation and noncoding transcripts are modulators of noise.', *Genome biology and evolution*, 3, pp. 320–31. doi: 10.1093/gbe/evr025.

Wang, R. F. *et al.* (1991) 'Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*', *Gene*, 100, pp. 195–199. doi: 10.1016/0378-1119(91)90366-J.

Wang, S. C. *et al.* (2020) 'Expansion of the Major Facilitator Superfamily (MFS) to include novel transporters as well as transmembrane-acting enzymes', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, p. 183277. doi: 10.1016/j.bbamem.2020.183277.

Wang, T. D. *et al.* (2015) 'Spatially programmed assembling of oxidoreductases with single-stranded DNA for cofactor-required reactions', *Applied Microbiology and Biotechnology*, 99(8), pp. 3469–3477. doi: 10.1007/s00253-014-6172-y.

Wang, Y. *et al.* (2011) 'Structural and Kinetic Analysis of the Unnatural Fusion Protein 4-Coumaroyl-CoA Ligase::Stilbene Synthase', *Journal of the American Chemical Society*, 133(51), pp. 20684–20687. doi: 10.1021/ja2085993.

Wang, Y. *et al.* (2019) 'Re-designing ferritin nanocages for mercuric ion detection', *Analyst*, 144(19), pp. 5890–5897. doi: 10.1039/c9an01110b.

Wang, Y. and Yu, O. (2012) 'Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells', *Journal of Biotechnology*, 157(1), pp. 258–260. doi: 10.1016/J.JBIOTECH.2011.11.003.

Wang, Z. *et al.* (2019) 'Ferritin nanocage-based antigen delivery nanoplatfoms: Epitope engineering for peptide vaccine design', *Biomaterials Science*, 7(5), pp. 1794–1800. doi: 10.1039/c9bm00098d.

Waterhouse, A. M. *et al.* (2009) 'Jalview Version 2-A multiple sequence alignment editor and analysis workbench', *Bioinformatics*, 25(9), pp. 1189–1191. doi: 10.1093/bioinformatics/btp033.

Wente, S. R. and Schachman, H. K. (1987) 'Shared active sites in oligomeric enzymes: Model studies with defective mutants of aspartate transcarbamoylase produced by site-directed mutagenesis', *Proceedings of the National Academy of Sciences of the United States of America*, 84(1), pp. 31–35. doi: 10.1073/pnas.84.1.31.

Wheeldon, I. *et al.* (2016) 'Substrate channelling as an approach to cascade reactions', *Nature Chemistry*, 8(4), pp. 299–309. doi: 10.1038/nchem.2459.

Whitlow, M. *et al.* (1991) 'A metal-mediated hydride shift mechanism for xylose isomerase based on the 1.6 Å *Streptomyces rubiginosus* structure with xylitol and D-xylose', *Proteins: Structure, Function, and Bioinformatics*, 9(3), pp. 153–173. doi: 10.1002/prot.340090302.

Wieczorek, A. S. and Martin, V. J. J. (2012) 'Effects of synthetic cohesin-containing scaffold protein architecture on binding dockerin-enzyme fusions on the surface of *Lactococcus lactis*', *Microbial Cell Factories*, 11, p. 160. doi: 10.1186/1475-2859-11-160.

Wiethaus, J. *et al.* (2009) 'Specific interactions between four molybdenum-binding proteins contribute to Mo-dependent gene regulation in *Rhodobacter capsulatus*', *Journal of Bacteriology*, 191(16), pp. 5205–5215. doi: 10.1128/JB.00526-09.

Wijaya, C. J., Ismadji, S. and Gunawan, S. (2021) 'A Review of Lignocellulosic-Derived Nanoparticles for Drug Delivery Applications: Lignin Nanoparticles, Xylan Nanoparticles, and Cellulose Nanocrystals', *Molecules (Basel, Switzerland)*. NLM (Medline). doi: 10.3390/molecules26030676.

Wilhelm, M. and Hollenberg, C. P. (1985) 'Nucleotide sequence of the *Bacillus subtilis* xylose isomerase gene: extensive homology between the *Bacillus* and *Escherichia coli* enzyme', *Nucleic Acids Research*, 13(15), pp. 5717–5722. doi: 10.1093/nar/13.15.5717.

Williams, C. L. *et al.* (2016) 'Sources of Biomass Feedstock Variability and the Potential Impact on Biofuels Production', *Bioenergy Research*. Springer New York LLC, pp. 1–14. doi: 10.1007/s12155-015-9694-y.

Willson, B. J. *et al.* (2016) 'Production of a functional cell wall-anchored minicellulosome by recombinant *Clostridium acetobutylicum* ATCC 824', *Biotechnology for Biofuels*, 9(1), p. 109. doi: 10.1186/s13068-016-0526-x.

Willson, B. J. *et al.* (2019) 'Enhanced functionalisation of major facilitator superfamily transporters via fusion of C-terminal protein domains is both extensive and varied in bacteria', *Microbiology*, 165(4), pp. 419–424. doi: 10.1099/mic.0.000771.

Willson, B. J., Chapman, L. N. and Thomas, G. H. (2019) 'Evolutionary dynamics of membrane transporters and channels: enhancing function through fusion', *Current Opinion in Genetics & Development*, 58–59, pp. 76–86. doi: 10.1016/j.gde.2019.07.017.

Wisedchaisri, G. *et al.* (2014) 'Proton-coupled sugar transport in the prototypical major

facilitator superfamily protein XyleE', *Nature Communications*, 5(1), p. 4521. doi: 10.1038/ncomms5521.

Wu, B., Song, J. and Beitz, E. (2010) 'Novel channel enzyme fusion proteins confer arsenate resistance.', *Journal of Biological Chemistry*, 285(51), pp. 40081–40087. doi: 10.1074/jbc.M110.184457.

Wu, C.-C. *et al.* (2013) 'Systems biology Identification of cancer fusion drivers using network fusion centrality', *Systems biology*, 29(9), pp. 1174–1181. doi: 10.1093/bioinformatics/btt131.

Wu, Y. *et al.* (2019) 'Combinatorial expression of different β -carotene hydroxylases and ketolases in *Escherichia coli* for increased astaxanthin production', *Journal of Industrial Microbiology & Biotechnology*, 46(11), pp. 1505–1516. doi: 10.1007/s10295-019-02214-1.

Wungsintaweekul, J. *et al.* (2001) 'Phosphorylation of 1-deoxy-D-xylulose by D-xylulokinase of *Escherichia coli*', *European Journal of Biochemistry*, 268(2), pp. 310–316. doi: 10.1046/j.1432-1033.2001.01875.x.

Xin, F., Wu, Y. R. and He, J. (2014) 'Simultaneous fermentation of glucose and xylose to butanol by *Clostridium sp.* strain BOH3', *Applied and Environmental Microbiology*, 80(15), pp. 4771–4778. doi: 10.1128/AEM.00337-14.

Xu, Q. *et al.* (2018) 'Expression of an endoglucanase-cellobiohydrolase fusion protein in *Saccharomyces cerevisiae*, *Yarrowia lipolytica*, and *Lipomyces starkeyi*', *Biotechnology for Biofuels*, 11(1), p. 322. doi: 10.1186/s13068-018-1301-y.

Yamaguchi, K. *et al.* (2009) 'Pulse-chase experiment for the analysis of protein stability in cultured mammalian cells by covalent fluorescent labeling of fusion proteins.', *Methods in molecular biology (Clifton, N.J.)*, 577, pp. 121–131. doi: 10.1007/978-1-60761-232-2_10.

Yamamoto, K. *et al.* (2016) 'Substrate-dependent dynamics of the multidrug efflux transporter AcrB of *Escherichia coli*', *Scientific Reports*, 6(1), pp. 1–10. doi: 10.1038/srep21909.

Yamamura, H., Suzuki, Y. and Imaizumi, Y. (2015) 'New light on ion channel imaging by total internal reflection fluorescence (TIRF) microscopy', *Journal of Pharmacological Sciences*. Japanese Pharmacological Society, pp. 1–7. doi: 10.1016/j.jphs.2015.04.004.

Yazdani, S. S. and Gonzalez, R. (2007) 'Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry', *Current Opinion in Biotechnology*. Elsevier Current Trends, pp. 213–219. doi: 10.1016/j.copbio.2007.05.002.

Ye, L. *et al.* (2018) 'Optimizing the localization of astaxanthin enzymes for improved

productivity', *Biotechnology for Biofuels*, 11(1), p. 278. doi: 10.1186/s13068-018-1270-1.

Yeates, T. O. *et al.* (2008) 'Protein-based organelles in bacteria: Carboxysomes and related microcompartments', *Nature Reviews Microbiology*. Nature Publishing Group, pp. 681–691. doi: 10.1038/nrmicro1913.

Yeates, T. O., Crowley, C. S. and Tanaka, S. (2010) 'Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution', *Annual Review of Biophysics*, 39(1), pp. 185–205. doi: 10.1146/annurev.biophys.093008.131418.

Yeom, S.-J. *et al.* (2017) 'Controlled Aggregation and Increased Stability of β -Glucuronidase by Cellulose Binding Domain Fusion', *PLOS ONE*. Edited by M. Oberer, 12(1), p. e0170398. doi: 10.1371/journal.pone.0170398.

Yoon, H. *et al.* (2011) 'Discovery of *Salmonella* Virulence Factors Translocated via Outer Membrane Vesicles to Murine Macrophages †', *INFECTION AND IMMUNITY*, 79(6), pp. 2182–2192. doi: 10.1128/IAI.01277-10.

You, C., Myung, S. and Zhang, Y.-H. H. P. (2012) 'Facilitated substrate channeling in a self-assembled trifunctional enzyme complex', *Angewandte Chemie - International Edition*, 51(35), pp. 8787–8790. doi: 10.1002/anie.201202441.

You, C. and Zhang, Y.-H. H. P. (2014) 'Annexation of a high-activity enzyme in a synthetic three-enzyme complex greatly decreases the degree of substrate channeling', *ACS Synthetic Biology*, 3(6), pp. 380–386. doi: 10.1021/sb4000993.

Yu, K. *et al.* (2015) 'Synthetic fusion protein design and applications', *Biotechnology Advances*, 33(1), pp. 155–164. doi: 10.1016/j.biotechadv.2014.11.005.

Zakharova, N. *et al.* (1999) 'Fused and Overlapping *rpoB* and *rpoC* Genes in *Helicobacters*, *Campylobacters*, and Related Bacteria', *Journal of Bacteriology*, 181(12), pp. 3857–3859.

Zaldivar, J., Martinez, A. and Ingram, L. O. (1999) 'Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*', *Biotechnology and Bioengineering*, 65(1), pp. 24–33. doi: 10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2.

Zhang, G., Quin, M. B. and Schmidt-Dannert, C. (2018) 'Self-Assembling Protein Scaffold System for Easy in Vitro Coimmobilization of Biocatalytic Cascade Enzymes', *ACS Catalysis*, 8(6), pp. 5611–5620. doi: 10.1021/acscatal.8b00986.

Zhang, Xuejun C. *et al.* (2015) 'Energy coupling mechanisms of MFS transporters', *Protein Science*, 24(10), pp. 1560–1579. doi: 10.1002/pro.2759.

Zhang, Xuejun C *et al.* (2015) 'Energy coupling mechanisms of MFS transporters', *PROTEIN SCIENCE*, 24, pp. 1560–1579. doi: 10.1002/pro.2759.

Zhang, Y.-H. P. (2011) 'Substrate channeling and enzyme complexes for biotechnological applications', *Biotechnology Advances*, 29(6), pp. 715–725. doi: 10.1016/j.biotechadv.2011.05.020.

Zhang, Y. *et al.* (2006) 'Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and mammalian cells', *Journal of the American Chemical Society*, 128(40), pp. 13030–13031. doi: 10.1021/ja0622094.

Zhang, Yonghui *et al.* (2017) 'Engineering bi-functional enzyme complex of formate dehydrogenase and leucine dehydrogenase by peptide linker mediated fusion for accelerating cofactor regeneration', *Engineering in Life Sciences*, 17(9), pp. 989–996. doi: 10.1002/elsc.201600232.

Zhang, Youjun *et al.* (2017) 'Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle', *Nature Communications*, 8. doi: 10.1038/ncomms15212.

Zhang, Y., Tsitkov, S. and Hess, H. (2016) 'ARTICLE Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade', *Nature Communications*, 7. doi: 10.1038/ncomms13982.

Zhang, Z. and Chen, H. (2010) 'Fermentation performance and structure characteristics of xanthan produced by *Xanthomonas campestris* with a glucose/xylose mixture', *Applied Biochemistry and Biotechnology*, 160(6), pp. 1653–1663. doi: 10.1007/s12010-009-8668-y.

Zhao, H. L. *et al.* (2008) 'Increasing the homogeneity, stability and activity of human serum albumin and interferon- α 2b fusion protein by linker engineering', *Protein Expression and Purification*, 61(1), pp. 73–77. doi: 10.1016/j.pep.2008.04.013.

Zhao, W. *et al.* (2019) 'An efficient expression tag library based on self-assembling amphipathic peptides', *Microbial Cell Factories*, 18(1), p. 91. doi: 10.1186/s12934-019-1142-9.

Zheng, S. *et al.* (2018) 'ATPase activity regulation by leader peptide processing of ABC transporter maturation and secretion protein, NukT, for lantibiotic nukacin ISK-1', *Applied Microbiology and Biotechnology*, 102(2), pp. 763–772. doi: 10.1007/s00253-017-8645-2.

Zheng, Z. *et al.* (2012) 'Engineering *Escherichia coli* for succinate production from hemicellulose via consolidated bioprocessing', *Microbial Cell Factories*, 11, p. 37. doi:

10.1186/1475-2859-11-37.