Towards the cell-free expression of glycoproteins

Kealan Peter Exley

Imperial College London

Department of Life Sciences

Centre for Synthetic Biology and Innovation

Submitted for the Degree of Doctor in Philosophy

Declaration of Originality

All work shown in this thesis is my own except where referenced.

Copyright Declaration

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

Abstract

N-linked glycosylation, is the covalent linkage of an oligosaccharide to an amido-group of an asparagine in a consensus sequence motif. The biosynthesis and transfer of a carbohydrate chain onto a nascent protein occurs within the ER and these initial steps are highly conserved throughout eukaryotes. Further processing of the carbohydrate chain occurs in the Golgi apparatus, where differences in processing between species and cell-types arise, resulting in different saccharide composition of the end-product. These differences are important in nature as N-linked glycosylation can influence a protein's function and interaction. However, this trait can be problematic when manufacturing a biopharmaceutical, due to the potential pharmacokinetic influence of the glycan structure. Consequently, it is important to ensure a suitable N-linked glycan is produced.

The removal of the N-linked glycosylation processing from the constraints of a cell will allow for greater control of the end saccharide composition of the glycan chain. A cell-free protein expression system will reduce heterogeneity allowing for the production of a homogenesis glycan chain.

With this in mind, a project was conceived to create a cell free protein synthesis system for the expression of glycoproteins. The project commenced with implementation of a glycoengineering approach on *P. pastoris* to generate pentamannosyl oligosaccharide. This oligosaccharide is seen as substrate for the synthesis of a complex-type glycoprofile. To process the pentamannosyl oligosaccharide the expression of a glycosidase and two glycotransferases was attempted. The feasibility of producing a cell-free protein synthesis (CFPS) system from yeast cell lysate was also tested.

This project resulted in a knockout *P. pastoris* strain which prevented the hypermannosylation glycoprofile. This strain can viably be used to generate the pentamannosyl oligosaccharide. A recombinant human GlcNAcT-I was produced which demonstrated the *in vitro* modification of an oligosaccharide chain. Finally, the project developed an active CFPS with a *S. cerevisiae* cell lysate but not with *P. pastoris,* suggesting that not all cell lysates can be utilised to generate an active CFPS.

Together, these results highlight the difficulties in utilising a non-conventional yeast for genetic and metabolic engineering. This research has suggested targets for future work which could yet see the generation the first cell free glycoprotein synthesis with a complex-type glycoprofile.

Acknowledgements

Firstly, I would like to show my gratitude and appreciation to my supervisors' Dr Karen Polizzi and Dr Cleo Kontoravdi, and to BBSRC for funding this research. I would like to specially thank Karen and Dr Kate Royle for their continual support and guidance throughout this project, especially during the thesis construction. Their enthusiasm, knowledge and perseverance has helped build and progress this project and help compose a thesis I can be proud of.

I am greatly appreciative to the guidance and support I have received from many researchers across Imperial College. To my colleagues in the CSynBI, where there is truly too many to mention, I thank you for discussions both scientifically or weird and helping me get to the finish line. My appreciation in particular goes out to Felix, Charlie and Will whose constant help, knowledge and friendship have really helped over the years. Special thanks goes to Dr Rochelle Aw and Kate Royle (again), whose expertise's in *Pichia pastoris* and protein expression where greatly appreciated. I'd also like to thank Dr Chris Reynolds for your help.

I would like to thank my brothers whose constant support and love have helped me throughout this PhD. With special thanks going to my little brother Liam for your moola, cheers bruv. To all my friends old and new, across the country and world whose continual love, support, beats and jokes made this a lot more bearable experience. A special thanks to Colin and Sam for lending of floor space to sleep and their wisdom. An incredible thank you to Sam for your persistence in document checking. A big thanks to Walid whose mastery of Excel helped greatly.

And finally, a warm hearted thank you to my Mum and Dad for your love and support, and hope you both enjoy reading this.

Contents

Abstract 2

Acknowledgements 3

List of figures 11

List of tables 13

List of abbreviations 14

Preface; Monosaccharide symbols 19

Chapter 1. Introduction 20

1.1 Heterologous protein production overview 20

1.2 Heterologous expression from prokaryotic organisms 21

1.3 Heterologous expression from Gram-positive bacterium 22

1.4 Heterologous expression from Gram-negative bacterium 23

1.5 Heterologous expression from eukaryotic organisms 25

1.6 Heterologous expression from fungi 25

1.7 Heterologous expression from yeast 26

1.8 Heterologous expression from nonconventional yeast 27

1.9 Heterologous expression from filamentous fungi 29

1.10 Heterologous expression from multi-cellular eukaryotic species 29

1.11 Heterologous expression from plants 30

1.12 Heterologous expression from insect cells 31

1.13 Heterologous glycoprotein production from mammalian cells 31

1.14 Heterologous glycoprotein production overview in Transgenic Animals 32

1.15 Considerations for selecting a host 33

1.16 Timescale and cell growth 33

1.17 Secretion and protein purification 34

1.18 Heterologous protein production yields 34

1.19 Media and production 35

1.20 Post translation modifications 36

1.21 Glycosylation 36

1.22 N-linked glycosylation overview in Eukaryotes 38
1.23 N-linked glycosylation overview in Prokaryotes 46
1.24 O-linked glycosylation overview 48
1.25 A glycoengineering approach in bioprocessing 48
1.26 Glycoengineering in yeast 53
1.27 Insect cell-line glycoengineering 57
1.28 Glycoengineering in E. coli 61
1.29 Cell Free Protein Synthesis 64
1.30 Biopharmaceutical production in CFPS systems 67
1.31 Summary and Project Overview 68
1.32 Project aims 69
1.33 Thesis overview 70
Chapter 2. Materials and Methods 71
2.1 Strains and plasmids 71
2.2 Media 72
2.2.1 Miller Lysogeny broth (LB) media 72
2.2.2 2XYT media 72
2.2.3 Terrific broth (TB) media 72
2.2.4 Yeast extract peptone medium (YPD) 72
2.2.5 Buffered glycerol-complex medium (BMGY) or buffered methanol-complex medium (BMMY) 72
2.2.3.1 Potassium phosphate buffer, pH6.0 73
2.3 Molecular biology methods 73
2.3.1 Nucleic acid agarose gel electrophoresis 73
2.3.2 Nucleic Acid isolation methods 73
2.3.2.1 DNA agarose gel and PCR isolation 73
2.3.2.2 Plasmid DNA isolation 73
2.3.2.3 Isolation of yeast genomic DNA 74
2.3.2.4 Isolation of RNA 74

2.3.2.5 cDNA preparation 74
2.3.3 Restriction enzymes reaction 74
2.3.4 Ligation 74
2.3.5 Polymerase Chain Reaction 74
2.3.6 DNA assembly methods 75
2.4 Microbiology techniques 75
2.4.1 Measuring optical density 75
2.4.2 Transformation and preparation of chemically competent *E. coli* 75
2.4.3 Approximating the ATP concentration 76
2.4.4 Flow Cytometry Analysis 76
2.4.5 Transformation and preparation of *P. pastoris* for electroporation 76
2.4.5.1 Preparation of DNA for Transformation 76
2.4.5.2 Electroporation Protocol 77
2.5 Heterologous Protein Expression 77
2.5.1 Heterologous protein expression from recombinant *E. coli* strains 77
2.5.2 Heterologous protein expression from recombinant *P. pastoris* 77
2.6 Protein Preparation from *E. coli* lysates 78
2.6.1 Protein preparation buffers 78
2.6.2 *E. coli* cell disruption 78
2.6.3 Protein purification of 6xHis-tagged proteins 78
2.7 Protein analytical methods 79
2.7.1 Protein Gel Electrophoresis 79
2.7.1.1 Gel preparation 79
2.7.1.2 Precipitation of protein sample 79
2.7.1.3 Protein sample preparation 79
2.7.1.4 Protein gel electrophoresis 80
2.7.2 Protein Gel Visualisation 80
2.7.3 Protein immunoblotting 80
2.7.3.1 Protein transfer by dot-blotting 80

2.7.3.2 Protein transfer by electroblotting 80
2.7.3.3 Western Blotting 80
2.7.3.4 Measuring fluorescent proteins 81
2.8 Carbohydrate analytical methods 81
2.8.1 Carbohydrate isolation 81
2.8.2 Carbohydrate Electrophoresis 81
2.8.2.1 Staining derivatisation of oligosaccharide chains 81
2.8.2.2 Gel electrophoresis preparation 82
2.8.2.3 Fluorophore Assisted Carbohydrate Electrophoresis (FACE) 82
2.8.3 Measuring of nucleotides by HPLC 82
2.9 Extract preparation and reaction conditions for yeast cell-free protein synthesis 83
2.9.1 Yeast cell culture 83
2.9.2 Preparing spheroplasts 83
2.9.3 Cell lysis by homogenisation 83
2.9.3.1 Staining of ER and microsomes 84
2.9.4 Preparation of energy and co-factor supplements 84
2.9.5 Protein translation reaction 85
2.9.5.1 T7 RNA polymerase 85
2.10 Enzymatic activity assays 85
2.10.1 N-Acetylglucosaminyltransferase I (GlcNAcT-I) activity assays (A) 85
2.10.2 N-Acetylglucosaminyltransferase I (GlcNAcT-I) activity assays (B) 86
2.10.3 Type II α-mannosidase (Man-II) activity assays 86
2.11 Inverted microscope image capture 86
Chapter 3. A glycoengineering approach in *P. pastoris* for the synthesis of a recombinant glycoprotein with a core pentamannosyl glycoprofile 87
3.1 Introduction 87
3.2 Creating a *P. pastoris OCH1* mannosyl transferase knock-out through homologous recombination with a mammalian mannosidase 90
3.3 The knock-out of a recombinant glycoprotein producing *P. pastoris* strain with Golgi Mannosidase I construct 94

3.4 Genomic analysis of the *P. pastoris* strain transformed with Golgi Mannosidase I construct..... 94

3.5 Yeast humanization in *P. pastoris Δku70* strain 98

3.6 Redesign and construction of a plasmid for homologous recombination of *OCH1* gene with Nat1 resistance gene in *P. pastoris (Δku70)*..... 99

3.7 *P. pastoris (Δku70)* knock-out of OCH1 gene with pJAN OCH1:nat1 plasmid..... 100

3.8 Phenotypical characterisation of the *P. pastoris Δku70:och1* 101

3.9 Transformation of the *P. pastoris (Δku70:och1)* with a recombinant glycoprotein kringle-3 domain 102

3.10 The expression of recombinant glycoprotein Kringle-3 domain in *P. pastoris (Δoch1)* and assessing the N-glycosylation profile 103

3.11 The transformation of *P. pastoris* (Δoch1) with a mouse Golgi-Mannosidase for production of core N-glycan structure 104

3.12 Redesign of plasmid for introducing the mammalian Golgi Mannosidase to the *P. pastoris* (*Δoch1*) strain..... 106

3.13 The transformation of *P. pastoris (Δoch1)* with pKAN-B Man1a plasmid..... 108

3.14 Detecting the *TEF1* transcript in *P. pastoris* (*Δku70*) and *P. pastoris* (*Δoch)* strains 109

3.15 Results Summary 110

Chapter 4. Towards the development of an *in vitro* metabolic pathway for the processing of a pentamannosyl glycoprofile..... 113

4.1 Introduction..... 113

4.2 The synthesis of N-acetylglucosaminyltransferase-I, N-acetylglucosaminyltransferase-II and type II α-Mannosidase genes..... 116

4.3 Preliminary expression of N-acetylglucosaminyltransferase-I, N-acetylglucosaminyltransferase-II and type II α-Mannosidase in *E. coli* strain BL21 (DE3) 118

4.4 The attempted expression of N-acetylglucosaminyltransferase-I and type II α-Mannosidase in *E. coli* strain SHuffle®..... 119

4.5 The expression, purification and characterisation of N-acetylglucosaminyltransferase-I 121

4.6 Assessing the activity of N-acetylglucosaminyltransferase-I 122

4.7 The expression, purification and characterisation of type II α-mannosidase 126

4.8 Assessing the activity of type II α-mannosidase..... 127

4.9 The attempted expression of active type II α-Mannosidase in *Pichia pastoris*..... 128

4.10 The generation of an expression vector for type II α-Mannosidase in *Pichia pastoris* 128

4.11 The expression of a type II α-Mannosidase from *Caenorhabditis elegans* in *P. pastoris* 130

4.12 Transcriptional analysis of the type II α-Mannosidase from *Homo sapiens* and *Caenorhabditis elegans* in *Pichia pastoris* 131

4.13 Summary 132

Chapter 5. The development of a cell-free protein synthesis system based on *P. pastoris* cell lysate 135

5.1 Introduction..... 135

5.2 Plasmid construction for use in *P. pastoris* CFPS 138

5.3 Developing a methodology for *Pichia pastoris* CFPS 142

5.4 *P. pastoris* cell preparation for generation of a cell lysate 145

5.5 *P. pastoris* cell disruption method for generation of a cell lysate 147

5.6 From *P. pastoris* cell lysate to generation of a CFPS 150

5.7 Transcriptional analysis of GFP expression vectors..... 150

5.8 Testing the expression of GFP in a *P. pastoris* CFPS..... 151

5.9 Comparison of *P. pastoris* CFPS against *S. cerevisiae* CFPS 152

5.10 Energy limitations with a CFPS system..... 154

5.11 Summary 157

Chapter 6. Discussion 160

6.1 Summary of the intended work 160

6.2 Improvements and Future Work 160

6.2.1 A glycoengineering approach in *P. pastoris* for the generation of a core pentamannosyl glycoprofile 161

6.2.2 Towards the development of an *in vitro* metabolic pathway for the processing of a core glycoprofile 162

6.2.3 The development of a Cell-Free Protein Synthesis system based on *P. pastoris* cell lysate 164
References 166
Appendix 208
Section 1 Primer sequences 208
Section 2 Plasmid sequencing information 212
Section 3 SDS-PAGE assessment of Kringle-3 219
Section 4 SDS-PAGE and Western blot analysis of Golgi Enzymes 220
Section 5 GlcNAcT-I activity analysis 223
Section 6 Man-II activity analysis 224
Section 7 RT-PCR of Man-II 225
Section 8 pET-28a Vector 226
Section 9 Characterisation of GFP expression from *P. pastoris* CFPS and S. cerevisiae CFPS 227
Section 10 ATP Standard curve produced CellTiter-Glo® 228
Section 11 Copyright licenses 229

List of figures

Figure 0.1 Symbolic representation of monosaccharides. 19
Figure 1.1 Overview of glycoprotein folding and quality control process in the ER..... 42
Figure 1.2 Schematic representation of N-linked oligosaccharides processing in the Golgi of a vertebrate species 45
Figure 1.3 the N-linked glycosylation pathway in the prokaryote *C. jejuni* 47
Figure 1.4 the biosynthetic pathway of GDP-L-fucose 59
Figure 1.5 . Schematic overview of a Cell-free protein synthesis system. 65
Figure 3.1 typical N-linked oligosaccharide structures from selected number of eukaryotic species. 87
Figure 3.2 A schematic of the plasmid pGEM-OCH1:Man1a:G418. 91
Figure 3.3 visualisation of the colony PCR which was performed on P. pastoris strains which were transformed with the exogenous DNA pGEM-OCH1:Man1a:G418. 95
Figure 3.4 PCR analysis of the *P. pastoris* strain transformed with pGEM-OCH1:Man1a:G418. 96
Figure 3.5 PCR analysis to detect for an intact *OCH1* gene in *P. pastoris* strains transformed with pGEM-OCH1:Man1a:G418. 97
Figure 3.6 a proposed integration event by a single crossover event. 98
Figure 3.7 a proposed multiple tandem integration event of the Man1a construct at the OCH1 locus. 98
Figure 3.8 Schematic of the pJAN OCH1:nat1 plasmid. 100
Figure 3.9 PCR analysis of the *P. pastoris* (*Δku70*) transformed with pJAN OCH1:nat1 plasmid 101
Figure 3.10 the morphological differences between *P. pastoris* (*Δku70*) and *P. pastoris* (*Δ ku70::och1*). 102
Figure 3.11 Assessing the protein expression of K3 in wild-type *P. pastoris* (GS115) and *P. pastoris* (*Δoch1*). 104
Figure 3.12 PCR analysis of the *P. pastoris Δoch1* strain transformed with pGEM-OCH1:Man1a:G418. 105
Figure 3.13 RT-PCR analysis of the Man1a transcript from a *P. pastoris* (*Δoch1*) transformed with pGEM OCH1:Man1a:G418 plasmid 105
Figure 3.14 Schematic of the pKANB TefP:Man1a plasmid. 107
Figure 3.15 PCR analysis of the *P. pastoris* (Δoch1) transformed with pKAN-B Man1a plasmid 108
Figure 3.16 RT-PCR of the Man1a transcript from a *P. pastoris* (Δoch1) transformed with pKAN-B Man1a plasmid 109

Figure 3.17 RT-PCR analysis of the Tef1 transcript from a *P. pastoris* (Δku70) and *P. pastoris* (Δoch1)..... 110
Figure 4.1 Diagram of an enzymatic reaction network. 113
Figure 4.2 A sequential pathway for the generation of $GlcNAc_2Man_3GlcNAc_2$ glycoprofile 115
Figure 4.3 Expression of GlcNAcT-II...... 118
Figure 4.4 expression of GlcNAcT-I...... 120
Figure 4.5 expression of Man-II...... 120
Figure 4.6 SDS-PAGE of a His-tag purification of GlcNAcT-I...... 121
Figure 4.7 The RP-HPLC analysis of UDP concentration, from an enzyme reaction containing GlcNAcT-I and a reaction with no enzyme. 124
Figure 4.8 The RP-HPLC analysis of UDP-GlcNAc concentration, from an enzyme reaction containing GlcNAcT-I and a reaction with no enzyme...... 124
Figure 4.9 FACE analysis of $Man_5GlcNAc_2$ with enzyme GlcNAcT-I and donor sugar UDP-GlcNAc. 126
Figure 4.10 SDS-PAGE of a His-tag purification of Man-II..... 127
Figure 4.11 Dot Blot assessing Man-II expression from *Pichia pastoris* (GS115) 129
Figure 4.12 transcript analysis of *Homo sapiens* Man-II (A) and C. elegans Man-II (B) synthesised in *P. pastoris*. 132
Figure 5.1 the enzyme reaction for Creatine-Kinase. 135
Figure 5.2 Schematic of the 12 plasmids constructed to test CFPS showing the different DNA elements...... 141
Figure 5.3 Growth curve of *Pichia pastoris* X33...... 145
Figure 5.4 Culture profile of *P. pastoris X33.* 146
Figure 5.5 Cell viability of *P. pastoris* X33...... 146
Figure 5.6 *P. pastoris* CFPS stained with ER-Tracker™ Blue-White DPX. 149
Figure 5.7 the transcription analysis of pGPR-GFP-pA and pCrPV-GFP-pA with T7 RNA Polymerase...... 151
Figure 5.8 Analyses of GFP expression from the pCrPV-GFP-pA plasmid within a CFPS. 153
Figure 5.9 Analyses of GFP expression from the pGPR-GFP-pA plasmid within a CFPS. 154
Figure 5.10 comparison of ATP concentration during *P. pastoris* growth profile..... 155
Figure 5.11 comparison of ATP concentration between *P. pastoris* and *S. cerevisiae* cell lysate. 156
Figure 5.12 comparison of ATP concentration during *S. cerevisiae* growth profile. . 156

List of tables

Table 1.1 The reported lengths of hyper-mannose structures from nonconventional yeast species..... 27
Table 1.2 Summarising the key features of expression hosts..... 33
Table 1.3 Production of recombinant monoclonal antibody IgG by expression host.. 35
Table 1.4 Genes for enzymes of LLO synthesis 39
Table 1.5 list of glycoengineering strategies employed on CHO cells...... 50
Table 2.1 *E. coli* strains utilised in this project..... 71
Table 2.2 Yeast strains utilised in this project..... 71
Table 2.3 Plasmids utilised in this project 71
Table 2.4 Table of utilised primary antibodies..... 81
Table 2.5 Reagents required in CFPS system..... 85
Table 4.1 truncation details of GlcNAcT-I, GlcNAcT-I and Man-II. 117
Table 4.2 the chromatogram retention times of the donor sugar nucleotide UDP-GlcNAc and the nucleotide diphosphate UDP...... 123
Table 5.1 a literature review of procedural steps utilised to generate an active CFPS system. 143
Table 5.2 cell viability of *P. pastoris* X33...... 147
Table 5.3 Chemical composition of the energy mix utilised in a CFPS system based on *P. pastoris* cell lysate...... 150
Table 6.1 A selection of human glycosyltransferases and glycosidases which could be utilised for the processing/maturation of N-linked glycan chain...... 163

List of abbreviations

AOX1	Alcohol Oxidase Promoter
AA	Amino Acids
ADCC	Antibody Dependent Cell-Mediated Cytotoxicity
ADCs	Antibody Drug-Conjugates
ASGPR	Asialoglycoprotein Receptor
ATP	Adenosine Triphosphate
Bac	Bacillosamine
BMGY	Buffered Glycerol-Complex Medium
BMMY	Buffered Methanol-Complex Medium; Induces Paox1
CDC	Complement Dependent Cytotoxicity
CFG	Consortium For Functional Glycomics
CFPS	Cell Free Protein Synthesis
CHO	Chinese Hamster Ovary
COPII	Coat Protein Complex II
DNA	Deoxyribonucleic Acid
DsbC	Disulphide Bond Isomerase
DTT	Dithiothreitol
EDEM	ER Degradation-Enhancing Mannosidase
EPO	Erythropoietin
ERAD	ER Associated Degradation
ERES	ER Exit Sites
FDA	Food And Drug Administration
FPLC	Fast Protein Liquid Chromatography
Gal	Galactose
GalNac	Acetylgalactosamine
GAP	Glyceraldehyde-3-Phosphate Dehydrogenase
GlcNAc	A-N-Acetylgalactosamine
GlcNAcT-I	N-Acetylglucosaminyltransferase-I
GlcNAcT-II	N-Acetylglucosaminyltransferase-Ii
GMD	GDP-Mannose Dehydratase
GNPE	Glcnac-6-Phosphate-2'-Epimerase

GRAS	Generally Regarded As Safe
GT	Glycosyltransferases
HEK293	Human Embryonic Kidney 293 Cells
IgG	Immunoglobulin G
IRES	Internal Ribosome Entry Site
K3	Kringle 3 Domain Of Human Plasminogen
LcrV	Low-Calcium Response V
LLO	Lipid-Linked Oligosaccharide
LPS	Lipopolysaccharides
mAbs	Monoclonal Antibodies
Man-II	Type II A-Mannosidase
MBP	Maltose Binding Protein
NAGLU	N-Acetylglucosaminidase
O-Gal	O-Galactose
O-GalNAc	O-N-Acetylgalactosamine
O-GlcNAc	O-N-Acetylglucosamine
OST	Oligosaccharyltransferase
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
pgl	Protein Glycosylation
PNGaseF	Peptide-N-Glycosidase
PTM	Post-Translational Modification
RmD	GDP-4-Dehydro-6-Deoxy-D-Mannose Reductase
RNA	Ribonucleic Acid
RRL	Rabbit Reticulocytes
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SAS	Sialic Acid Synthase
Stt3	Staurosporine And Temperature Sensitivity 3
Tm	Melting Temperature Of Oligonucleotides
tPA	Tissue-Type Plasminogen Activator
UDP	Uridine Diphosphate
UDP-GlcNAc	Uridine Diphosphate A-N-Acetylgalactosamine

UGE	UDP-Galactose 4-Epimerase
UGGT	UDP-Glucose:Glycoprotein Glucosyltransferase
UOXF	Oxford Glycobiology Institute
WGE	Wheat Germ Extract
AOX1	Alcohol Oxidase Promoter
AA	Amino Acids
ADCC	Antibody Dependent Cell-Mediated Cytotoxicity
ADCs	Antibody Drug-Conjugates
ASGPR	Asialoglycoprotein Receptor
ATP	Adenosine Triphosphate
Bac	Bacillosamine
BMGY	Buffered Glycerol-Complex Medium
BMMY	Buffered Methanol-Complex Medium; Induces Paox1
CDC	Complement Dependent Cytotoxicity
CFG	Consortium For Functional Glycomics
CFPS	Cell Free Protein Synthesis
CHO	Chinese Hamster Ovary
COPII	Coat Protein Complex II
DNA	Deoxyribonucleic Acid
DsbC	Disulphide Bond Isomerase
DTT	Dithiothreitol
EDEM	ER Degradation-Enhancing Mannosidase
EPO	Erythropoietin
ERAD	ER Associated Degradation
ERES	ER Exit Sites
FDA	Food And Drug Administration
FPLC	Fast Protein Liquid Chromatography
Gal	Galactose
GalNac	Acetylgalactosamine
GAP	Glyceraldehyde-3-Phosphate Dehydrogenase
GlcNAc	A-N-Acetylgalactosamine
GlcNAcT-I	N-Acetylglucosaminyltransferase-I

GlcNAcT-II	N-Acetylglucosaminyltransferase-Ii
GMD	GDP-Mannose Dehydratase
GNPE	Glcnac-6-Phosphate-2'-Epimerase
GRAS	Generally Regarded As Safe
GT	Glycosyltransferases
HEK293	Human Embryonic Kidney 293 Cells
IgG	Immunoglobulin G
IRES	Internal Ribosome Entry Site
K3	Kringle 3 Domain Of Human Plasminogen
LcrV	Low-Calcium Response V
LLO	Lipid-Linked Oligosaccharide
LPS	Lipopolysaccharides
mAbs	Monoclonal Antibodies
Man-II	Type II A-Mannosidase
ManNAc	N-acetylmannosamine
MBP	Maltose Binding Protein
NAGLU	N-Acetylglucosaminidase
O-Gal	O-Galactose
O-GalNAc	O-N-Acetylgalactosamine
O-GlcNAc	O-N-Acetylglucosamine
OST	Oligosaccharyltransferase
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
pgl	Protein Glycosylation
PNGaseF	Peptide-N-Glycosidase
PTM	Post-Translational Modification
RmD	GDP-4-Dehydro-6-Deoxy-D-Mannose Reductase
RNA	Ribonucleic Acid
RRL	Rabbit Reticulocytes
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
SAS	Sialic Acid Synthase
Stt3	Staurosporine And Temperature Sensitivity 3

Tm	Melting Temperature Of Oligonucleotides
tPA	Tissue-Type Plasminogen Activator
UDP	Uridine Diphosphate
UDP-GlcNAc	Uridine Diphosphate A-N-Acetylgalactosamine
UGE	UDP-Galactose 4-Epimerase
UGGT	UDP-Glucose:Glycoprotein Glucosyltransferase
UOXF	Oxford Glycobiology Institute
WGE	Wheat Germ Extract

Preface; Monosaccharide symbols

Monosaccharides are class of carbohydrates that cannot be hydrolysed into a simpler carbohydrate[1]. Monosaccharides are the building block of oligosaccharides and polysaccharides. To aid in identification of monosaccharides within a chain, the structures are represented as symbol[2]. The symbolic representation of monosaccharides with geometric shapes and colours match the annotations used by the Consortium for Functional Glycomics (CFG) and the Oxford Glycobiology Institute (UOXF)[2]. The symbolic representation of each monosaccharide present in this study is displayed in Figure 0.1. Each monosaccharide class has the same shape, and isomers are differentiated by colour, to minimize variations only one sialic acid type is represented. The symbols will be utilised continually throughout this document.

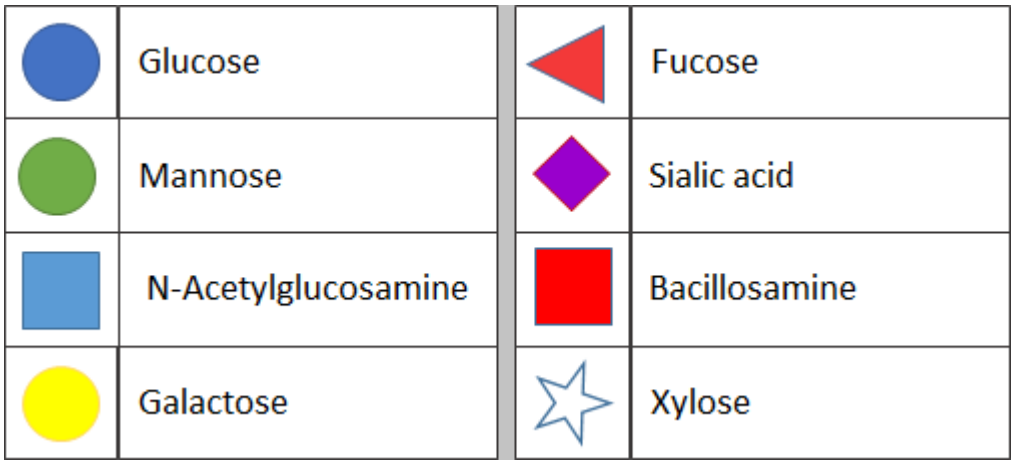


Figure 0.1 Symbolic representation of monosaccharides.
Representation of monosaccharides with geometric shapes

Chapter 1. Introduction

1.1 Heterologous protein production overview

Recombinant DNA technology, specifically the expression of foreign genes within a selected host, has revolutionised biology, creating a multibillion-dollar biotechnology industry[3]. Where once proteins were harvested from their natural source, the heterologous expression of a protein within a foreign host allows for greater processing control, saves time, increases yields and lowers the cost[3]. The classic example often cited is that of the blood-sugar regulating hormone, insulin. Initially insulin, which is used in the treatment of diabetes, was extracted from animal pancreases mostly porcine sources[4]. The development of recombinant DNA technology has allowed for the synthesis of recombinant human insulin within a bacterial expression host (*E. coli)*[5]. The drug Humulin® (human insulin) became the first FDA approved therapeutic protein produced from a recombinant source in 1982[6].

Since the commercialisation of insulin the number of approved protein therapeutics or biopharmaceuticals marketed in the United States and/or EU now stands at 212[7]. A biopharmaceutical or biologic is a medical product extracted or synthesised from a biological source. This covers a range of products including hormones, enzymes, vaccines and monoclonal antibodies (mAbs). mAbs are the predominant biopharmaceutical representing 27% of all market approved biopharmaceuticals with a total cumulative sales value of $140 billion in 2013[7]; to date only IgG-type mAb products have been licensed. The success of mAbs can be attributed to their high target specificity and varied functionality, such as binding and sequestering a pathogenic agent or completely removing the pathogenic agent. mAbs can also bind to the cell surface proteins of diseased cells and recruit immunological molecules or immune cells (macrophages, neutrophils or Natural Killer cells) to specifically kill the mAb-bound cell by complement dependent cytotoxicity (CDC) or Antibody Dependent Cell-mediated Cytotoxicity (ADCC), respectively[8]. mAb-based drug conjugates have also be developed, which deliver a toxic payload to a targeted a cell for destruction[9]. Consequently, mAbs are used in the treatment of a wide number of diseases including cancer, inflammatory and infectious diseases.

The majority of biopharmaceuticals are peptide or protein based and therefore can be expressed in heterologous expression systems. Recombinant insulin Humulin® for instance, is expressed using a prokaryotic host, however examples of heterologous expression systems can be found in all three domains of life. Although there are examples of archaeal expression hosts[10], the majority of expression hosts are prokaryotic or eukaryotic (therefore archaeal expression hosts will not be a subject of review in this thesis). The ideal expression host should

be easy to genetically manipulate, stably assimilate a recombinant gene and most importantly express a fully folded soluble protein. However, if insoluble aggregates (known as inclusion bodies) of the recombinant protein are formed, a soluble recombinant protein can sometimes be recovered[11,12]. This involves the *in vitro* solubilisation of an inclusion body and refolding of the protein to a bioactive form[12]. In general, inclusion bodies are solubilized by utilising high concentration of denaturant chemicals such as urea or guanidine hydrochloride, along with a reducing agent such as dithiothreitol[11]. The solubilised proteins are then refolded by slow removal of the denaturant, usually by dialysis, in the presence of oxidizing agent[11]. Strategies have also been employed to decrease the formation of inclusion bodies and will be discussed later (Section 1.4).

In the case of proteins which perform a function these can be expressed either in an active or an inactive form. The inactivity of a protein is only problematic if an active protein is essential for its intended use, for instance if an enzyme is required for a reaction, an active state is vital. The functionality of a protein could be dependent on a post-translational modification (PTM) which can include acetylation, glycosylation, methylation, phosphorylation, sulphation, ubiquitination and disulphide bond formation[13]. Not all expression hosts have the ability to perform these PTMs; therefore, consideration of the properties of a protein will be needed before choosing an expression host. Many biopharmaceuticals require a PTM, such as N-linked glycosylation, for efficient functioning and drug efficacy[14]. The N-linked glycan chains can alter the binding efficiency, and therefore in a clinical setting (i.e. as biopharmaceutical), may alter the therapeutic efficacy of the product[14]. Therefore, when manufacturing a biopharmaceutical via a recombinant source, it important to understand the characteristics which promote drug efficacy and whether these can be achieved in a given expression host.

Expression hosts are often single-cell organisms such as bacteria, yeast, and filamentous fungi or have been propagated from multicellular eukaryotic to a single cell culture, examples of which can be found from plants, insects and mammals[15]. Heterologous expression can also occur in a multi-cellular organism such as plants and animals. Although, the use of GM crops and animals is not without its controversies[16]. The next subsection will introduce and review a number of different expression hosts.

1.2 Heterologous expression from prokaryotic organisms

Prokaryotes are single cell organisms classified by their lack of membrane-bound organelles such as the nucleus[17]. Therefore, transcription and translation occurs solely within the cytoplasm. Most bacterial species contain a peptidoglycan cell wall surrounding the cellular membrane, the thickness of which is used for classification[18]. The ease of genetic manipulation, cheap feedstock, fast growth and potentially high production yields, make

bacteria a favourable recombinant protein expression host and examples can be found in both gram positive and gram negative bacterial species[19]. However, despite the positives the major drawback to bacterial recombinant expression is in differences in PTMs, which can prevent the translation of some eukaryotic recombinant proteins. Nevertheless, they remain a dominant expression host in biotechnology.

1.3 Heterologous expression from Gram-positive bacterium

Gram-positive bacteria have had millennia-long association with human 'biotechnology' activity, through the fermentation of food by *Lactococcus* strains[20]. *Lactococci* produce lactic acid as a by-product of glucose metabolism[21]. In cheese production, for example, the lactic acid produced by *L. lactis* is used to curdle milk by acid-associated denaturing of proteins[22]. Although its use in food biotechnology does not require recombinant protein production, *Lactococci* have been utilised as an expression host.

The identification, isolation and manipulation of plasmids within *Lactococcus* species has enabled recombinant protein expression[22,23]. *L. lactis* is capable of surviving gut passage and has been classified as Generally Regarded as Safe (GRAS) by the FDA. As such, it has been exploited as an *in vivo* delivery system for therapeutic drugs[24,25]. This is where recombinant protein antigen is either secreted from or localised to the cell wall of a recombinant expression host. For instance a *L. lactis* strain secreting the *Yersinia pseudotuberculosis* low-calcium response V (LcrV) antigen was able to protect mice against *Yersinia* infection[26].

A major reason for using *L. lactis* as an *in vivo* delivery system is due to its capacity to secrete proteins, simplifying protein purification[27]. This valuable feature has led to the development of many Gram-positive expression hosts, such as GRAS soil dwelling bacteria species of *Bacillus subtilis*[28], *Corynebacterium glutamicum*[29] and *Streptomyces lividans*[30]. These expression hosts have been utilised in the biotechnology industry to produce chemicals, organic acids, higher alcohols, polymers, amino acids and vitamins[3]. Although, a drawback to recombinant protein secretion can be the occurrence of extracellular hydrolysis activity by the naturally secreted host proteases, the engineering of protease deficient strains can minimise this risk[31]. For instance, the engineered *B. subtilis* strain WB800N has had seven secreted proteases and one membrane bound protease knocked out from the hosts genome[32]. *C. glutamicum*, and *S. lividans* show little to no extracellular hydrolysis activity.

These three species as well as other gram-positive bacteria are useful in industrial biotechnology. However, their use for the production of protein therapeutic is still in its infancy, and as of yet, no regulatory approved biopharmaceutical has been synthesised within a gram-positive expression host. Nevertheless, with the increasing utilisation of gram-positive

expression hosts and improving knowledge about promoters, plasmids, fermentation techniques, and proteases etc[33], it should not be long before a regulatory approved pharmaceutical protein is produced in a gram-positive host.

1.4 Heterologous expression from Gram-negative bacterium

Gram-negative bacteria are classified due to the presence of an outer membrane. This outer membrane can often restrict the ability of gram-negative organisms to secrete proteins. On top of this, a number of gram-negative bacteria are known to be pathogenic to humans. This perhaps should limit gram-negative bacteria as cell factories. However, the gram-negative bacteria species *Escherichia coli* is one of most popular hosts and is known as the workhorse of recombinant protein expression[34,35]. Its fast growth kinetics (doubling time is about 20 min[36]), even within a simple and cheap glucose-salts media, and its ease in genetic manipulation have made this organism a popular choice for both academic research and within an industrial setting[35]. Although other gram-negative bacteria are biotechnologically relevant, such as *Pseudomonas putida* for bioremediation[37], *E. coli* is by far the most dominant heterologous expression system, responsible for the expression of thousands of different recombinant proteins.

The commercial use of *E. coli* is dominated by biopharmaceutical production, with 30% of all protein drugs being synthesise from an *E. coli* host[38]. As was previously mentioned, *E coli* became the first heterologous expression system to produce a biopharmaceutical, with the translation of recombinant insulin. Since then *E. coli* has been used to produce a number of different biopharmaceuticals, ranging from cytokines such as interferon in 1986[39], hormones such as human growth hormone in 1987[40], and therapeutic enzymes such as urate oxidase for the treatment of gout in 2010[41].

Despite its dominance *E. coli* has its drawbacks; as previously stated *E. coli* is incapable of secreting recombinant proteins into the surrounding environment[27] and similar to other prokaryotic expression hosts *E. coli* is unable to perform PTMs hindering effective expression of eukaryotic proteins. This can lead to an over accumulation of misfolded recombinant protein, forming aggregates within the host cell known as inclusion bodies[35,42]. However due to its popularity as an expression host considerable effort has been placed to overcome these obstacles. One such example is in the generation of *E. coli* strains which can generate disulphide bridges between two cysteine residues[43].

The cytoplasm is a reductive environment which restricts oxidative folding. Although bacteria do possess an oxidising environment within the periplasmic space[44,45], which can allow for the disulphide bond formation, the periplasmic translocation of recombinant protein often leads to

cellular toxicity[46] and generally overall lower yields compared to cytoplasmic expression[47]. For instance, a 10-fold yield difference was noted between cytoplasmic (2 mg/mL) and periplasmic (0.2mg/mL) translation when expressing a recombinant single-chain variable fragment (scFv)[48]. The lack of cytosolic oxidative folding in *E. coli* has been overcome with genetic modifications and protein engineering, reducing the formation of inclusion bodies.

This was achieved by forming a higher oxidising cytoplasmic environment, this is the basis of two commercially available *E. coli* strains; *E. coli* SHuffle® T7 strain (NEB) and *E coli* Origami™ 2 (DE3) pLysS (Novagen). The removal of glutaredoxin and thioredoxin activity (*gor- trxB-* knockout) led to a less reducing environment in both strains[49,50]. The *E. coli* SHuffle® T7 strain has been further improved through the overexpression of periplasmic Disulphide Bond Isomerase (DsbC) within the host cytoplasm. The DsbC catalyses disulphide bond formation and also has been shown to act as a chaperone to assist in the folding of proteins that do not require disulphide bonds[51]. These modifications allow for efficient formation of disulphide bonds with the increased product yields associated with cytosolic expression.

Although disulphide bridge formation is important for correct protein folding, a fully folded protein does not necessarily have functionality. Other PTMs such as phosphorylation and glycosylation are essential for activity in many proteins, neither of which can be performed in native *E. coli*[52]. The production of biopharmaceuticals with the absence of PTMs is potentially troublesome, as it can lead to instability, inactivity and immunogenicity. However, as before a number of methods have been employed to add PTMs to recombinant proteins expressed in *E. coli.* This can be achieved either by chemical addition, such as in the case of PEGylation[53], or by genetically engineering *E. coli* strains. The attachment of synthetic polyethylene glycol (PEG) molecule to a biopharmaceutical has been shown to increase resistance to proteolysis, reduce immunogenicity and increase the protein half-life[54,55]. PEGylation has been employed in 20 FDA approved biopharmaceutical products[56], with the immunotherapy drug Oncaspar® being the first FDA approved PEGylated biopharmaceutical produced recombinantly in *E. coli* in 1990[57]. Strain improvement programmes, which have been so effective in producing disulphide bonds, have too been employed to engineer further eukaryotic processes into engineered *E. coli*. For instance, the co-expression of mammalian methylases and acetylases, with their particular substrates within the *E. coli* host resulted in the production of recombinant proteins with the PTMs of methylation and acetylation, respectively[58]. Likewise the introduction of the *C. jejuni* glycosylation machinery into *E. coli* allowed the organism to perform N-linked glycosylation[59]. The *E. coli* was engineered to synthesise a trisaccharide of $GlcNAc_2Man_3$, which was then transferred *en bloc* to a fully folded glycoprotein within the periplasm. Further details on this engineered strain will be discussed later (Section 1.28).

Despite the considerable effort being placed to overcome these limitations the proportion of proteins produced commercially in *E. coli* has been steadily decreasing from 55% in 2000[60] to 19% in 2014[7]. In 2000 34 products (55%) of FDA approved biopharmaceuticals were produced in *E. coli*[60] this is compared to 45 products (30%) in 2009[34] and 40 products (19%) in 2014[7]. This decline is mostly because of the increase in complexity of the biopharmaceuticals, in particular the requirement of PTMs, which cannot be performed in native *E. coli*. Although the strain improvements described above aim to enhance PTMs capacity, as of yet no *E. coli* strain is commercially available which can perform uniquely eukaryotic PTMs. This could be due to the need for regulatory approval of new engineered hosts or possible due to low efficiency in creating PTMs and in product yield. Therefore, in many cases eukaryotic expression systems can be the more favourable expression host.

1.5 Heterologous expression from eukaryotic organisms

Eukaryotes are defined by the presence of membrane bound organelles and a nucleus, which contains the genetic material. Eukaryotic cell types normally fall into one of three categories: (1) animal, (2) plant, or (3) fungal. There are many structural and organisational differences between the cell types, for instance animal cells lack a cell wall, which can be found in plants and fungal cells. Eukaryotes have been effectively utilised as recombinant expression hosts and examples can be found from each of the cell types. As expression hosts are often single-cell organisms, expression hosts from multicellular sources need to be propagated to a single cell culture.

Where prokaryotic expression hosts could have difficulties in translating eukaryotic proteins and multimeric proteins, this problem understandably is lessened within eukaryotic expression host. This due to eukaryotes containing compartments to aid protein translation, having machinery to enhance protein folding (such as chaperones), and the ability to perform PTMs (although with differing interspecies results). Some eukaryotic expression systems also benefit from ease in genetic manipulation, cheap feedstock and fast growth however other eukaryotic expression hosts do not share these attributes.

1.6 Heterologous expression from fungi

Many different fungal organisms have been utilised as heterologous expression systems. Fungal species have many of the benefits of prokaryotic hosts such as the ease in genetic manipulation, cheap feedstock and fast growth, although not as fast as *E. coli*[61,62]. Typically, the yeast organism *Saccharomyces cerevisiae* has a doubling time of 90 minutes[63] compared to a doubling time of about 20 min in *E. coli*[36]. However, fungi have the potential to produce much higher protein yields than *E. coli* due to their ability to grow to a much higher cell

concentration. Most fungal strains used for heterologous expression are classified as GRAS[61]. Other advantages fungal expression hosts have over *E. coli* and other prokaryotes is in their ability for protein secretion, and in their native ability to produce PTMs[61].

Despite the ability to perform PTMs the end product can differ greatly from other eukaryotes. For instance, the N-linked glycosylation profile for fungi is high-mannose type, this is compared to the complex and/or hybrid structures observed in higher eukaryotic species[64]. More will be discussed on species differences of N-linked glycosylation later (Section 1.22). These differences in PTMs can be problematic when a specific PTM is needed for correct functioning of a recombinant protein.

1.7 Heterologous expression from yeast

S. cerevisiae is one of the most popular heterologous protein expression systems among yeast species. By the same token as the gram-positive prokaryote *L. lactis*, *S. cerevisiae* has had millennia long association with human 'biotechnology' activity, through fermentation[65]. In more recent history it has become an important model system for the study of eukaryotic pathways and processes such as the cell cycle[66]. This consequently has led to a vast wealth of genetic and biochemical knowledge, including the whole annotated genome sequence published in 1996[67].

One of major advantages of yeast systems over a prokaryotic host is in the ability to secrete recombinant protein[68]. The secretion of a protein allows for simpler purification from the growth medium, lowering the cost and time compared to intracellular protein extraction[68,69]. A further advantage is that in yeast species it is possible to perform many of the PTMs which are lacking or insufficient in prokaryotic species, thus enabling the production and proper folding of many human proteins. However, differences in the end product do exist between the fungal lower eukaryotes and multicellular higher eukaryotic species. For instance, the O-glycosylation chain in humans is often mucin-type, characterized by α-N-acetylgalactosamine (GalNAc) sugar, while yeast have a shorter mannosyl-glycan chain[70,71]. Similarly to the O-glycosylation chain, the N-linked glycosylation profile of *S. cerevisiae* is also characterised by mannosyl-glycan chain[72].

S. cerevisiae have a number of different mannosyltransferases which add mannose via α1,2, α1,3 or α1,6 linkages this can result in an oligosaccharide in excess of 200 mannose residues[73]. This extreme hyperglycosylation can hinder secretion, lowering protein yields[65]. Furthermore, the hypermannosylated glycoprofile in *S. cerevisiae* possess α1,3-linked mannose residues[74], which are considered to be immunogenic in humans[75]. Therefore, *S.*

cerevisiae is limited in the production of biopharmaceuticals, which require a more mammalian-like N-linked glycosylation.

S. cerevisiae was also one of the first recombinant expression hosts to produce an FDA approved biopharmaceutical, in the form of human insulin. Since then *S. cerevisiae* has been used for the production of a number of different biopharmaceuticals, such as growth hormone for treatment of dwarfism, vaccines and virus-like particles such as hepatitis virus vaccines and human papilloma virus[76,77]. However, like *E. coli, S. cerevisiae* too has seen an overall decline in the proportion of approved biopharmaceuticals produced in this host. In 2000, 11 products (18%) of FDA approved biopharmaceuticals were produced in *S. cerevisiae*[60], with 17 (11.2%) products in 2009[34] and 33 (15.5%) products in 2014[7]. This decrease in the overall percentage of biopharmaceuticals produced in *S. cerevisiae,* is due in part to the differences in the end-product of PTMs, in particular the hypermannosylated N-linked glycoprofile. Also, the emergence of other microbial eukaryotic hosts has resulted in a decline in *S. cerevisiae* usage.

1.8 Heterologous expression from nonconventional yeast

The decreasing cost of genome sequencing, which has led to the availability of well-characterised host strains, has seen an increase in the uptake and emergence of alternative eukaryotic microbial systems such as *Hansenula polymorpha, Kluyveromyces lactis, Pichia pastoris, and Yarrowia lipolytica* (Table 1.1). These alternative hosts benefit from shorter length mannose structures and some lack the immunogenic α1,3-linked mannose residues (Table 1.1). For instance, *Y. lipolytica* contains only mannosyltransferases which add α1,6 linkages resulting in an oligosaccharide of between 8 and 12 mannose sugars.

Table 1.1 The reported lengths of hyper-mannose structures from nonconventional yeast species

A number of nonconventional yeast species have been developed for heterologous protein expression. A nonconventional yeast is defined as any species that is not *S. cerevisiae.* The increasing use of nonconventional yeast has witnessed an increase in genetic information and characterisation. A benefit often attributed to utilising nonconventional yeast is in the lesser extent to which they mannosylated their N-linked oligosaccharides and that many do not harbour an immunogenic α1,3-linked mannose.

Organism	Published genome	Mannose –profile	Presence of α1,3-linked mannose
S. pombe	2002[78]	$Gal_{2-6}Man_{8-12}GlcNAc_2$[79]	No[79]
H. polymorpha	2003[80]	$Man_{8-12}GlcNAc_2$ [81]	No [81]
K. lactis	2004[82]	$Man_{>30}GlcNAc_2$[83]	Yes[83]
Y. lipolytica	2004[82]	$Man_{8-12}GlcNAc_2$[84]	No[84]
P. pastoris	2009[85]	$Man_{12-16}GlcNAc_2$[86]	No[86]

The methylotrophic yeast *P. pastoris* is by far the most popular of the nonconventional yeast expression systems. This organism utilises methanol as a sole carbon and energy source through a metabolic pathway involving several unique enzymes[87]. The first enzyme in this pathway, Alcohol oxidase 1, is under the regulation of a strong methanol-induced alcohol oxidase promoter (AOX1)[88]. The up-regulation of Alcohol Oxidase, in the absence of glucose and in the presence of methanol, can see the Alcohol Oxidase enzyme constitute as much as 35% of the soluble protein in the cell[87,89]. The strong-inducible AOX1 promoter has been utilised for controlling recombinant protein expression.

Nonconventional yeast can grow to much higher cell densities than *S. cerevisiae.* For instance, methylotrophic yeasts such as *P. pastoris* can be grown to a density as high as 130 g L^{-1} and 200 g L^{-1} dry weight[90,91], this is compared to between 10 g L^{-1} and 20 g L^{-1} in *S. cerevisiae*[92,93]*.* This increased biomass accumulation is a result of many nonconventional yeast being 'Crabtree-negative' and therefore they do not process waste carbon to aerobic ethanol production[94]. The strong-inducible promoter AOX1 coupled with increased biomass enables *P. pastoris* to produce recombinant proteins at much higher titres. As illustrated in the production of blood transfusion biopharmaceutical serum albumin, where *S. cerevisiae* yielded 0.15 g L^{-1} compared to *P. pastori*s which produced 10 g L^{-1} [95]. *P. pastoris* has been demonstrated to express mammalian proteins at high protein titres with yields over 14 g L^{-1} for mouse collagen and gelatin[96]. *P. pastoris* has visibly been demonstrated to secrete much lower levels of native proteins than *S. cerevisiae* therefore allowing for an even simpler purification[69,97]*.*

These advantages have seen *P. pastoris* surpass *S. cerevisiae* for the proportion of recombinant genes expressed between the two strains. In 1990 the proportion of recombinant genes expressed in *S. cerevisiae* stood at 8%, in 2013 this number was 2%. This is compared to *P. pastoris* which saw no recombinant genes expressed in 1990 to a proportion of 11% in 2013. This increased popularity has culminated in the FDA approval of two biopharmaceuticals produced from *P. pastoris*. Despite this, the high mannose structure can often limit *P. pastoris* use as production host for biopharmaceuticals, which require a complex glycan chain for full functioning. Hence, most biopharmaceuticals, which possess an N-linked glycosylation chain are produced in mammalian cell lines. However, the many advantages over mammalian cells namely lower cost and higher product titres, have seen *P. pastoris* subjected to strain improvement programmes. For instance, a *P. pastoris* strain has been generated which is able to synthesise a complex oligosaccharide with sialylated ends[98]. This is achieved by the removal of a α1,6 mannosyltransferase enzyme *OCH1,* halting the elongation of oligosaccharides, resulting in the glycan motif of $Man_8GlcNac_2$. This glycan motif is the substrate for mammalian Golgi based α1,2-mannosidases and the addition of a Golgi α1,2-

mannosidase 1A into the genome of *P. pastoris* results in the generation of $Man_5GlcNac_2$ glycan motif[99]. In total 14 heterologous genes have been inserted in the *P. pastoris* genome, including genes involved in sugar synthesis and transport and a number of mammalian glycosyltransferases which have led to the synthesis of therapeutically relevant glycan chains. Further details on the modification of N-linked glycosylation pathway of *P. pastoris* will be discussed below (Section 1.26). Despite these developments as of yet no biopharmaceutical has been licensed from the engineered strain and currently only mammalian expression hosts have produced biopharmaceuticals with a complex type N-linked glycosylation.

1.9 Heterologous expression from filamentous fungi

Filamentous fungi differ from yeast in morphology and growth[100]. Where yeast divide forming a unicellular population, filamentous yeast form into long, branching filamentous multi-cellular structures through hyphal growth[100]. However, in submerged cultures, the morphology of filamentous microorganisms usually varies between clumped and dispersed (unicellular) forms[95,101]. Due to a high secretion capacity and the native expression of industrially relevant enzymes, such as cellulase, filamentous fungi have become useful recombinant production hosts[102,103]. *Trichoderma reesei* and *Aspergillus niger* are the most commonly used filamentous fungi expression hosts and as such many useful genetic tools exist for recombinant expression[95,]. Despite this availability, filamentous fungi are not as easy to genetically manipulate as other expression hosts, often requiring a physiological change before genetic transformation[104],[105]. Similarly, to other fungal species, filamentous fungi have a mannosylated glycosylation profile but do not have the tendency to hypermannosylate on the scale of *S. cerevisiae*, with the glycoprofile resembling more of mammalian high-mannose glycan type ($Man_{5-9}GlcNAc_2$)[106].

1.10 Heterologous expression from multi-cellular eukaryotic species

Although fungal expression systems can overcome some of the limitations observed in prokaryotic systems, the hypermannosylation of fungal glycan chains can often limit their use. This is particularly a problem when expressing a glycoprotein that requires a more varied monosaccharide composition for its function. Therefore, other non-fungal eukaryotic expression systems from plants, vertebrates or invertebrates have been developed.

These expression systems are sourced from multicellular organisms and although it is possible to express a recombinant protein within a tissue or whole organism, unicellular expression systems offer many benefits such as secretion into the surrounding liquid culture and easier genetic manipulation[31]. A single cell expression host is established by isolating a tissue cell, which is then provided with suitable media containing growth factors or hormones

to enable cell proliferation[107]. The number of cell divisions would normally be limited, however the genetic manipulation or spontaneous mutations have created a number of immortalised cell lines capable of high numbers of cell divisions[108].

Expression systems from multicellular sources, although often they produce lower protein yields, can synthesise more diverse oligosaccharides than single-cellular sources[31]. Although they can be more challenging to genetically manipulate and slower in growth, their benefits have not prevented them becoming attractive expression hosts.

1.11 Heterologous expression from plants

The genetic engineering of plants is normally associated with improving plant fitness and pest resistance, however transgenic plants have been developed to produce commercially interesting recombinant proteins including vaccines and antibodies[109]. Although, similar to any plant, limitations can be observed due to the seasonal growing period, the long development times, and pests. Moreover, the protein is not secreted so a further extraction step is required[110,111]. There is also the need to minimise any possibility of contamination with agrochemicals or fertilizers[111]. Nevertheless, the benefits of the plant cell have motivated researchers to develop propagated plant-cell cultures, examples of which can be found from several different plant species including *Arabidopsis thaliana* and the important domestic crops of tobacco and soybean[110]. The plant suspension cells are isolated from undifferentiated callus tissues and grown in the presence of plant hormones to maintain the undifferentiated state[110].

Although plant cell cultures have primarily been used for generation of secondary metabolites, such as the chemotherapy drug Taxol, they have also been demonstrated to produce heterologous proteins. A recombinant human glucocerebrosidase Elelyso® produced in cultured carrot root cells became the FDA first approved biopharmaceutical from a plant host in 2012[112]. Similarly, to other eukaryotes, plant cells benefit from protein secretion and the ability to perform PTMs. Although the N-linked glycosylation is more varied than in fungal species, plants add a unique sugar xylose not found in any other eukaryotic oligosaccharides[113]. As this glycan residue is not present in humans it is immunogenic, thus limiting plant cell culture as a host for biopharmaceuticals[114]. Another disadvantage of plant cells is in the slower growth rates when compared with microbial cells, and challenges associated with gene drift and gene silencing[110].

1.12 Heterologous expression from insect cells

A number of unicellular insect expression systems have been developed mostly around a baculovirus expression systems. The baculoviridae family forms a wide group of viruses known to infect over 600 different types of invertebrates[115]. Baculovirus is an arthropod-specific virus, which favours lepidopteron (moth and butterfly) species. The virus is characterised by a large rod-shape and double stranded DNA[115]. Baculovirus systems accumulate large amounts of viral proteins at late stage insect infection, this feature led to the development of recombinant protein expression[115]. By replacing non-essential viral genes with recombinant genes, recombinant proteins can be translated from transduced insect cell lines. The most frequently used insect cell lines for infection are sourced from *Trichoplusia ni* and *Spodoptera frugiperda*[116]*.* The versatility of the baculovirus expression system has made insect cell lines a popular choice[116]. Insects also have the advantage of performing PTMs including a truncated version of mammalian N-linked glycosylation chain[117]. Insects have a greater capability to translate what are deemed more difficult proteins, including the assembly of multimeric protein complexes[118].

Drosophila cell lines have also been in use for over a decade as heterologous expression hosts[119]. *Drosophila* cell lines are derived from late stage embryo cells, which are maintained in formulated media[119]. The recombinant gene can either be integrated into the genome or expressed from a self-replicating plasmid[119]. The recombinant gene is often placed downstream of an inducible promoter to control expression in the *Drosophila* cell line; an inducible system does not exist in baculovirus expression systems. Similarly, to baculovirus expression system, *Drosophila* cell lines have the ability to perform PTMs and assemble multimeric proteins. It however requires a more complex media than the baculovirus expression system[120].

1.13 Heterologous glycoprotein production from mammalian cells

Although insect and plant expression systems contain glycosylation machinery to produce complex and/or hybrid glycan structures the end product still differs from that of vertebrate species[121,122]. The glycan structures are truncated in comparison to mammalian structures and can contain immunogenic monosaccharides (Further details of these differences will be discussed below Section 1.22). Consequently, the majority of biopharmaceuticals bearing a glycosylation are produced in mammalian hosts. Mammalian cell lines have been derived from a number of organisms with Chinese Hamster Ovary (CHO) cells being the most dominant host[123]. This is because of early industrial adoption due to the existence of auxotrophic mutant strains enabling stable integration of heterologous genes[124]. CHO cells are also capable of gene amplification which increases protein yields and are able to grow in serum-free

suspension conditions[125]. The first recombinant proteins produced in CHO cells were interferons and tissue-type plasminogen activator (tPA)[126]. In 1987, a tPA molecule Activase® became the first FDA-approved biopharmaceutical produced from CHO cells[127].

Since then the number of biopharmaceuticals produced from CHO cells has rapidly increased, from 14 products (23%) in 2000[60] to 75 products (35.5%) in 2014[7]. This is in contrast to yeast and *E. coli* hosts, which have seen an overall decline in production usage. The dominance of CHO cells is due in part to an increase in the approval of mAb-based therapeutics[128]. Seen as the magic bullet of therapeutic medicine due to its high specificity, monoclonal antibodies (mAb) were first produced within hybridoma cells, of a primary mouse lymphocyte fused with an immortal tumour cell[129]. Although useful mAb producers, the problems of hybridoma instability and human immunogenicity against mouse mAb limited their commercial use, leading to the generation of mammalian cell cultures[130]. The first FDA approved mAb rituximab, used for the treatment of low grade B-cell lymphoma, was produced from CHO cells in 1997[131]. Since then CHO cells and to a lesser extent murine myeloma cell lines (NS0 and SP2/0) have been responsible for the production of 29 mAbs with accumulative annual sales of over 100 billion US dollars[132].

Despite being viewed as the workhorse of mammalian hosts; CHO cells still have a number of limitations. For instance, phenotypic drift (i.e. random characteristic changes by the host) is not uncommon[133]. Although the main advantage of mammalian hosts is in the production of mammalian glycoprofiles, the glycoprofile can still differ from humans, particularly in the addition of non-human residues. For instance, NS0 and SP2/0 are known to add the sialic acid residue N-glycolylneuraminic acid while human systems add the sialic acid residue N-acetylneuraminic acid. This difference can cause an immunological reaction within humans. CHO cells are also known to add these non-human sialic acid residues. One solution to ensure a human-like glycoprofile is to use human cell lines as a recombinant expression host, such as human embryonic kidney 293 cells (HEK293 cells). The HEK293 cells have also been utilised to perform the PTM γ-carboxylation which CHO cells cannot produce with adequate efficiency[134].

1.14 Heterologous glycoprotein production overview in Transgenic Animals

Similarly, to GM plants, genetically modified animals have also been developed for the production of recombinant proteins. One such example is in the production and secretion of heterologous proteins into the milk of transgenic ruminants[135,136], where in essence the mammary gland becomes a bioreactor. The application is appealing due high volumetric productivity, continual production and PTM capabilities including mammalian N-glycosylation chains. In 2009, the FDA approved Atryn® a recombinant antithrombin alpha, produced in the

milk of goats[137]. However due to the ethical minefield of GM animals the commercialisation of this technology remains challenging.

1.15 Considerations for selecting a host

When selecting an expression host, a number of considerations must be made to ensure suitability of the host for the protein product. These include: the intended use of the protein product, the requirements of the protein for any modifications, the timescale needed, the yield, and the cost[138]. The above sections discussed the key attributes of a number of different expression platforms, these can be found summarised in Table 1.2. Further details of these host cell attributes are discussed below (Sections 1.16 to 1.20).

Table 1.2 Summarising the key features of expression hosts

	Expression system					
	Prokaryotic		**Eukaryotic**			
	E. coli	***B. subtilis***	**Yeast**	**Insect**	**Plant**	**Mammal**
Cell growth	**Rapid**	**Rapid**	**Rapid**	**Slow**	**Slow**	**Slow**
Media complexity	**Low**	**Low**	**Low**	**High**	**Medium**	**High**
Production Cost	**Low**	**Low**	**Low**	**Medium**	**Medium**	**High**
Expression levels	**High**	**High**	**High**	**Moderate to High**	**Low to Moderate**	**Low to Moderate**
Secretion to media	**No**	**Yes**	**Yes**	**No**	**Yes**	**Yes**
Post translational modifications						
Disulphide bonds	**Yes ***	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Acylation	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Acetylation	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Phosphoryl-ation	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
N-linked glycosylation	**No**	**No**	**Yes; High-man type**	**Yes; simple Complex type**	**Yes; trunc. Complex type**	**Yes; Complex type**
O-linked glycosylation	**No**	**No**	**Yes**	**Yes**	**Yes**	**Yes**
Carboxylation	**No**	**No**	**No**	**No**	**Yes**	**Yes**

*Genetically engineered strains have demonstrated the formation of disulphide bonds

1.16 Timescale and cell growth

When considering timescale, one needs to keep in mind not just the population growth rate (i.e. the speed at which the cell divides) but also the amount of time it takes to establish a recombinant expression host and the downstream processing steps. The establishment of the

host consists of introducing the recombinant DNA followed by optimising of growth and expression conditions[120]. Recombinant DNA can either be integrated into the host genome or carried on a plasmid DNA. Ease of genetic manipulation of a host is seen as a major advantage[120]. A general rule a thumb is slower dividing cells are more difficult to genetically manipulate. Brondyk *et al.* evaluated the process of establishing a host to take as little as one week in *E. coli,* over a month in yeast, while in mammalian cells the development of a stable cell line a minimum of three months is estimated[139]. Therefore, if a fast expression schedule is essential then bacterial expression hosts are more favourable.

1.17 Secretion and protein purification

Once a protein is expressed a number of downstream processing steps are normally required to allow the protein product to be recovered and removed from process-related impurities, including host cell proteins (HCPs), DNA/RNA and lipids, as well as product related impurities such as aggregates[140]. The timescale of the downstream processing steps is dependent on the cell type and the required purity needed. In cases where a protein is expressed intracellularly, a further step is required which will disrupt the cell, releasing the recombinant product. This can be done by chemical, enzymatic, mechanical or physical methods[141]. The secretion of a recombinant protein is often seen as an advantage as fewer HCPs are present therefore purification is simpler. The amount of HCP which are secreted vary between species, for instance *P. pastoris* secretes far fewer proteins than other yeast species[142]. The removal HCPs can be essential, especially when producing a biopharmaceutical where there is potential adverse immunological effect to HCPs[143]. Recently two clinical trials were suspended due to patients showing antibodies against CHO cell HCPs[143,144].

1.18 Heterologous protein production yields

When choosing a host if the aim is to produce large protein titres then considering Table 1.2, unicellular species would make the best choice. For instance, *P. pastoris* has been reported to produce 1 g L^{-1} of the simple polypeptide elastin[145]. However, the general rule that unicellular species are high producers is not always the case, with unicellular species often struggling to express to high yields of more complex multimeric proteins. For instance, the pharmacologically relevant recombinant protein IgG has been produced in a number of expression hosts (Table 1.3). The multimeric protein IgG titres are much higher from mammalian cell cultures than the unicellular expression hosts. The reason for the difference is that unicellular species, particularly prokaryotic species, cannot perform PTMs. Consequently, these hosts find it difficult to assembly the IgG protein, which impacts the protein titres. A further reason is that cell cultures for mammalian hosts have been optimised to increase the amount of the valuable IgG product. These improvements have seen CHO

cells routinely produce IgG titres of between 1 and 5 g L^{-1}, with some companies attaining as much as 10–13 g L^{-1} for extended culture durations[129,146]. It must be noted that a true comparison between expression hosts is complicated by the differences in bioprocessing techniques; for instance, culture volumes and duration.

Table 1.3 Production of recombinant monoclonal antibody IgG by expression host
Table has been modified from Frenzel *at al.*[147]

Expression Host	Culture Volume (Litres)	Culture method	Production yield mg L^{-1}	Reference
E. coli	10	fermenter	130–150	148
Native *P. pastoris*	3	Bioreactor	227	149
Humanized *P. pastoris* ($Man_5GlcNAc_2$)	3	Bioreactor	148	149
***Drosophila,* S2**	Not stated	Shake flasks	5–35	150
***Trichoplusia n*i BTI-TN5B1-4 "High Five**	0.1	Shake flasks	12	151
Nicotiana tabacum	0.05	Shake flasks	1	152
HEK-293F cells	1	Shake flasks	300–400	153
NS0 cells	5	bioreactor	800	154

1.19 Media and production

The culture medium requirements which are essential to support an expression host is another very important consideration when evaluating host choices. The cell culture media components can be highly variable between expression hosts. In mammalian cells a more complex medium is required, with components such as vitamins, metal ions, hormones and growth factors needed[107]. The lack of availability of certain nutrients and/or components can affect the protein product. For instance, the lack of donor nucleotide sugars, the substrates for glycosyltransferase reactions, can alter the end oligosaccharide chain[155]. Bacterial expression hosts do not require as complex media as mammalian hosts.

As well as medium components, the environment conditions of the cell culture are also important for maintaining expression. The control of pH, temperature and dissolvable gases (mainly CO2 and oxygen), are important for the quality and quantity of the protein produced. For instance, in a *P. pastoris* fermentation reaction, maintained at an environment of above pH 5.0 slow degrades a protein product, cellulose, by cell-associated proteolysis[156]. By reducing and maintaining a pH of 4.0 in bioreactors an increase in the fraction of full-length protein product from 40 to 90% was observed[156].

Although medium cost is often cited as a disadvantage to producing recombinant protein in CHO cells, calculated at about €1 per litre, it is worth noting that the average cost for sterile manufacturing is equivalent between CHO cell or non-CHO cell host[157]. Also, the use of defined media, as in the case of mammalian cell cultures, is much easier to optimise enabling an increase in quality and quantity.

1.20 Post translation modifications

The suitability of a host becomes important when the activity of recombinant protein requires a certain specific modification; this is especially apparent when the protein has a medical application. When considering a heterologous recombinant protein expression system, the functionality requirements of protein will often dictate the choice of host. Throughout this chapter the differences in the PTM abilities between expression hosts, in particular glycosylation, became the defining attribute of many expression hosts. A PTM is classified as the alteration of a nascent or fully folded protein through proteolytic cleavage, or the addition of chemical functional group(s) and/or cofactors[15]. The different expression hosts used for biopharmaceutical synthesis have different properties, with some unable to produce certain PTMs or synthesising unsuitable PTMs. For instance, an IgG based drug may require a N-linked glycan chain for it pharmacokinetics, approximately 70% of all approved therapeutic proteins contain at least one N-linked oligosaccharide[158]. Although as previously discussed a range of organisms have the capability of performing N-linked glycosylation, this does not however mean they all produce a suitable glycan motif. An unsuitable oligosaccharide can cause a biopharmaceutical to be ineffective, lower its efficacy, or at worse cause a negative effect. The hyper-mannosylated glycan motif produced by many yeast species is capable of eliciting an immune response[159,160]. and as such no biopharmaceutical pertaining a N-linked glycosylation chain has been licensed for production in a native yeast species. Similarly, an expression hosts which is unable to glycosylate a protein would be an unwelcome choice. This therefore limits the choice of expression host to mammalian cells.

1.21 Glycosylation

Glycosylation is the enzymatic synthesis of oligosaccharides and the subsequent attachment to a biomolecule[1,161]. Glycoproteins are a class of proteins which have a carbohydrate chain attached to the polypeptide chain. Protein glycosylation involves the attachment of an oligosaccharide, via a glycosidic linkage, to an asparagine residue (N-linked) or serine/threonine residue (O-linked) on a peptide backbone. The glycosylation of a protein is an example of a post-translational modification (PTM). Glycosylation is one of the most abundant and ubiquitous PTMs, with over two-thirds of deposited UniProt protein sequences identified using bioinformatics containing at least one glycosylation site[162]. There are up to 350

glycosidase and glycosyltransferase related genes in the human genome[163] and these create a diverse array of oligosaccharide structures. Further to this, a mass spectrometry study of the mouse proteome mapped 2352 proteins containing a total of 6367 N-glycosylation sites collectively[162]. This abundance allows for an increase in the functional diversity of the proteome, producing a number of protein isoforms from a single gene.

The evolutionary conservation of the eukaryotic ER process demonstrates the importance of N-linked glycosylation in promoting protein folding[164]. In addition to this primary function, mature glycan chains serve to increase the functionality of a protein by forming interactions, typically with other proteins, lipids or carbohydrate chains. In humans N-linked glycosylation plays a pivotal function in health, as disruption can lead to a range of congenital diseases, covering neurological disorders, congenital muscular dystrophies, connective tissue disorders and immune deficiencies[165]. To date nearly 70 genetic disorders associated with incomplete glycosylation have been found in humans[166]. Moreover the interruption of the tetradecasaccharide synthesis is lethal in mice[167].

The diversity in the Golgi process produces a wide-range of N-linked oligosaccharides structures. There are many examples where an oligosaccharide, although attached to a common polypeptide sequence, can differ in composition and structure, these differences are known as glycoforms. While naturally occurring, glycoforms of one protein can also occur due to (1) the cell being at different stage of differentiation, (2) translation is from a different tissue type or (3) translation is from a diseased tissue[168]. Although several hundred glycoforms can be observed on one protein, usually the different glycoforms cause no or little change to the overall function[161]. This is because the glycan structure has no specific function or is able to serve its function through a common or conserved part of the glycan[161]. However, there are instances where differences in a glycan structure can alter the function of a protein. One such example is where a glycoform impacts the tertiary structure of a protein, an alteration of a protein structure can affect the binding affinity and therefore impact the activity. This has been observed in Immunoglobulin G (IgG) which has been reported with several hundred glycoforms[155]. In the immunological process of ADCC, the IgG binds to an antigen and an antibody receptor on the surface of hosts natural killer cells or neutrophils. The IgG binds through the Fc region to a Fcγ-receptor. This binding triggers the release of cytokines which attack the target cell causing apoptosis. The Fc region contains the N-glycosylation tripeptide site on Asparagine 297, and as IgG is a homodimer it contains two glycosylation sites. The glycan structure is known to affect the tertiary structure of the IgG through steric hindrance affecting the binding affinity of the mAb to the antibody receptors (Fcγ-receptor). For example, the absence of a core fucose monosaccharide on N-glycan motif alters the mAb structure to increase ADCC up to 50-fold[169].

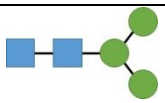
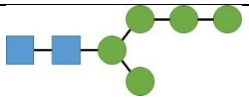
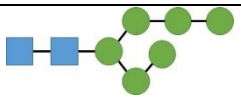
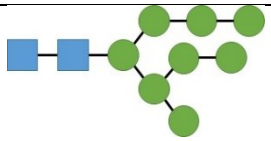
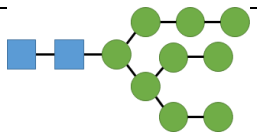
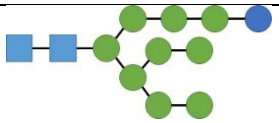
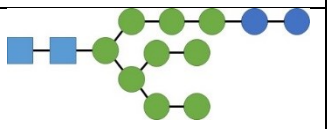
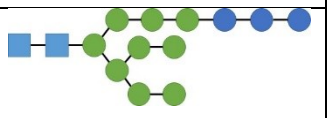
1.22 N-linked glycosylation overview in Eukaryotes

N-linked glycosylation is a ubiquitous post-translational modification in eukaryotic systems. A matured N-linked oligosaccharide contains varying compositions and/or numbers of monosaccharides[1]. However, despite this diversity, the early steps of the glycosylation pathway are highly conserved. This includes the initial oligosaccharide construction, its 'en bloc' transfer to a nascent polypeptide and the ER-based enzymatic oligosaccharide remodelling steps[1]. The highly conserved nature of the ER process is related to its importance in promoting protein folding and it serving as quality control check for destruction of misfolded proteins[1,170].

The construction of an oligosaccharide chain must begin with the synthesis of a polyisoprenoid lipid dolichol. The lipid chain serves as a carrier and anchor for oligosaccharide synthesis. Dolichol (Dol) is synthesised on the cytoplasmic face of the ER through the mevalonate pathway[171,172]. The intermediate farnesyl pyrophosphate undergoes the head-to-tail cis addition of C5 isoprenoid units[172]. This is completed by the actions of cis-prenyltransferase[171,172]. The number of isoprenoid units has been shown to vary between cell-type and species[173,174]. For instance, yeast typically have between 14 and 17 and mammals have between 18 and 21[173]. Once assembled the Dol is then phosphorylated by Dol-kinase and the resulting Dol-P acts as a lipid-precursor for oligosaccharide construction[175].

The first monosaccharide assembled onto the Dol-P is N-acetylglucosamine (GlcNAc), which is attached through a pyrophosphate linkage by the enzyme GlcNAc-1-phosphotransferase (*ALG7*), giving rise to GlcNAc-PP-Dol[176,177]. The ALG7 enzyme is the only enzyme in the glycosylation process which catalyses a sugar to a phosphate linkage. ALG7 uses a nucleotide-activated sugar donor uridine diphosphate GlcNAc (UDP-GlcNAc) as its substrate[178]. The nucleotide-activated sugar donors are used as substrates throughout the cytosolic construction of oligosaccharides. Once the first GlcNAc sugar is added a further GlcNAc sugar and 5 mannose residues are sequentially attached. The assembly of the lipid-linked oligosaccharide (LLO) is catalysed by a number of oligosaccharyl transferases, which were first identified by mutation studies in the yeast model organism *S. cerevisiae*[61,179]. Over a dozen different mutants were identified which presented a truncated oligosaccharide and designated as alg mutants[179,167]. The corresponding genes were named *ALG* genes, where *ALG* can either mean *a*sparagine-*l*inked *g*lycosylation or *al*tered in *g*lycosylation, dependent on the literature. Table 1.1 shows the different *ALG* enzymes which have been identified with the linkage which they catalyse.

Table 1.4 Genes for enzymes of LLO synthesis
A list of genes encoding ER-based oligosaccharyl transferases which function to build the LLO. The Human sequences marked by asterisks were identified or their specific function has been elucidated on the basis of known yeast sequences.

S. cerevisiae gene	GenBank Annotation number	Human gene	End glycosylation product
ALG7	AAH00325	DPAGT1	
ALG13	NP_060936	hALG13*	
ALG14	NP_659425	hALG14	
ALG1	AA67521	MAT-1*	
ALG2	AAH31095	hALG2*	
ALG11	CAA95916.1	ALG11	
ALG3	AAH02839	NOT56*	
ALG12	Q9BV10	hALG12*	
ALG9	AF395532	hALG9*/DIBD1	
ALG6	Q9Y672	hALG6*	
ALG8	NP_076984	hALG8*	
ALG10	AAH0347	hALG10*	

Once the LLO intermediate of two GlcNAc and five Mannose sugars is synthesised it is then translocated (flipped) across the membrane into the lumen of the ER. The high-energy requirement of the flipping suggests it is assisted by the actions of a protein, which has been designated as a flippase, although currently the protein identity remains elusive in eukaryotes[180]. Once within the ER lumen the LLO intermediate is elongated through the attachment of 4 mannose residues and 3 glucose residues. Where before monosaccharides were supplied through nucleotide-activated donor sugars, within the ER lumen, sugar donors are attached to Dol-P (Dol-P-Man and Dol-P-Glc)[171,176]. This is important as mammals do not have GDP-Man transporter, so require Dol-P-Man as a donor which can be transported into the ER[181]. The Dol-P-Man and Dol-P-Glc are synthesised from nucleotide-activated donor sugars, GDP-Man and UDP-Glucose, respectively[178]. The attachment of monosaccharides allows the substrates to be located within close proximity to the oligosaccharyl transferase enzymes[178]. The mature LLO ($Glc_3Man_9GlcNAc_2$-PP-Dol) is now ready for transferral to a protein.

A hetero-multimeric protein complex, known as oligosaccharyltransferase (OST), catalyses the *en bloc* transfer of the tetradecasaccharide to the amide nitrogen of an asparagine residue on a nascent protein. Although there is preference for $Glc_3Man_9GlcNAc_2$ as a substrate, mutation studies have also demonstrated the transfer of truncated oligosaccharides[182]. However the OST binding affinity is strongest for the tetradecasaccharide, with a strong preference observed for the α1,2-linked glucose residue at the non-reducing end[178,183]. It has been proposed that the OST also interacts with the ribosome so only targets translocating peptide chains[184]. The OST targets the asparagine in the consensus tripeptide sequence asparagine-X-serine/threonine (N-X-S/T), where X is any amino acid residue except for proline[185]. It is generally accepted that proline decreases the flexibility of the amino acid chain preventing correct peptide positioning needed for the transfer[186].

The eukaryotic OST complex is made up of a number of protein subunits. The heptameric human OST shares a minimum of 90% sequence homology with mouse, bovine, canine, rat, and golden hamster OSTs[182]. The model yeast *S. cerevisiae* OST also shares structural homology to the human OST, therefore functional studies can be performed using yeast[187]. These studies have identified five essential subunits for yeast viability and identified the subunit *Stt3* (Staurosporine and temperature sensitivity 3) as the catalytic subunit[182]. The other subunits of the OST function to define the substrate specificities[182].

The Stt3 is the largest subunit, with a high structural and sequence interspecies homology; the human OSTs also contain two Stt3 isotypes[188]. The OST complex binds to the membrane-anchored Dol-P-P-oligosaccharide and transfers the glycan to nascent protein by cleavage of

the high-energy GlcNAc-P bond, releasing Dol-P-P in the process, which can be recycled in glycosylation pathway[178].

Following attachment, the maturation of the oligosaccharide chain begins. The first enzymatic processing steps occur within the ER, involving the removal of three glucose residues and a mannose residue[189,170]. This processing starts immediately after the transfer of the oligosaccharide, beginning with the removal of glucose residues by two glucosidases: α-glucosidases I and α-glucosidases II. The glucosidases both function within the ER lumen to sequentially remove glucose, with α-glucosidase I acting first on the terminal α1,2 glucose and α-glucosidase II acting to remove the two inner α1,3 glucose residues[190]. The cleavage of the first α1,3 glucose by α-glucosidase II occurs at a faster rate than the second cleavage[189]. This is to allow the monoglucosylated chain to be identified by a chaperone complex which catalyses protein folding[191]. The chaperone complex consists of lectin-like chaperones (either calnexin or calreticulin), a protein disulphide isomerase protein (ERp571), a peptidyl–prolyl isomerase protein (CypB), and the glycoprotein containing a $Glc_1Man_9GlcNAc_2$ oligosaccharide[192]. The removal of the final glucose residues by α-glucosidase II can indicate a properly folded protein and dissociation of the chaperone complex[193]. However, if an aglucosylated glycoprotein remains in a unfolded or misfolded state the glycoprotein can be recognised through its exposed hydrophobic regions by UDP-glucose:glycoprotein glucosyltransferase (UGGT)[194,195]. The UGGT enzyme then re-glucosylates the $Man_9GlcNAc_2$ oligosaccharide allowing for the reassociation of the chaperone complex[196]. The chaperone complex can either assist in protein maturation or promote the retro-translocation of misfolded proteins into the cytoplasm for degradation by ER associated degradation (ERAD). The reglycosylation and deglycosylation cycle by UGGT and α-glucosidase II can occur a number of times until complete protein maturation or ERAD occurs[192,197]. This reglycosylation and deglycosylation cycling does not occur in *S. cerevisiae* due to the lack of a UGGT or UGGT homolog[198]. However, it is still possible reglycosylation will not occur on a misfolded protein and therefore, the cleavage of the mannose residue within the ER lumen also serves as an important step to promote protein folding[195,199].

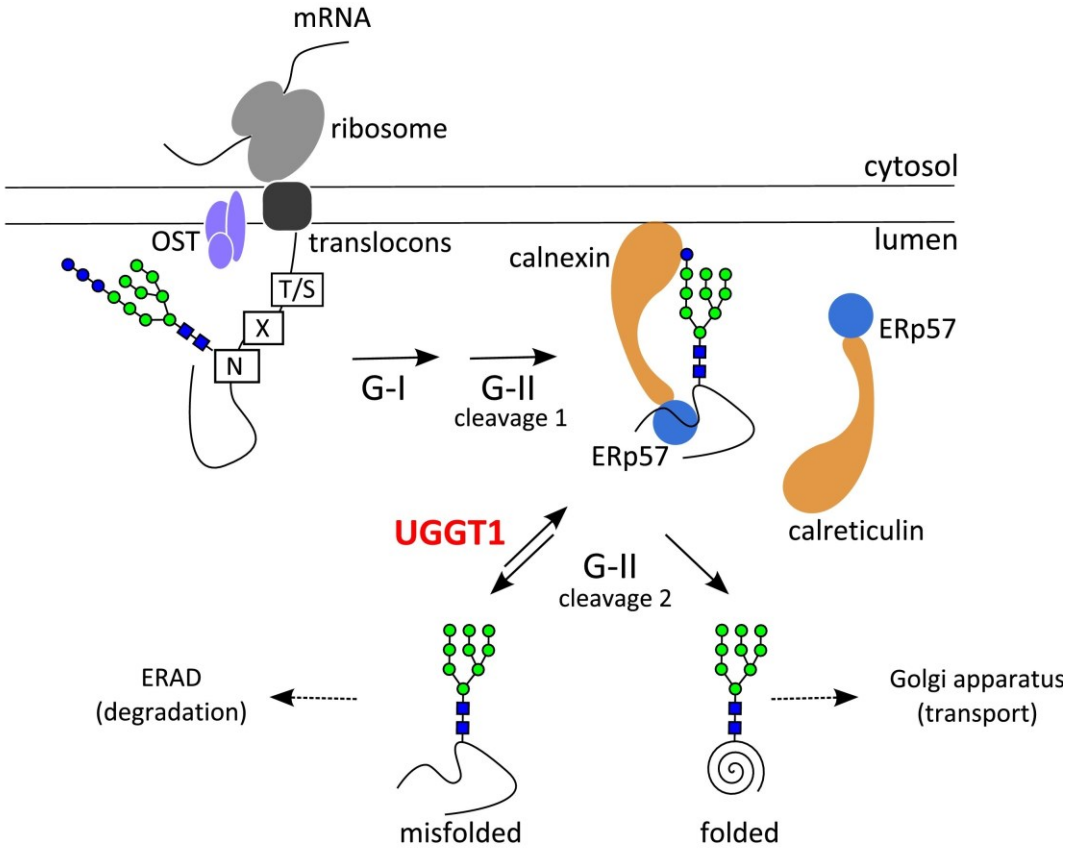


Figure 1.1 Overview of glycoprotein folding and quality control process in the ER. In the ER of most eukaryotes, the OST complex transfers a tetradecasaccharide (Glc3Man9GlcNAc2) to a nascent protein. Upon transfer the tetradecasaccharide is trimmed into smaller glycans by various glycosidases in the ER. This highly conserved cleavage process is important for protein folding and QC. With glycan intermediates forming interacts with distinct lectins or enzymes to aid protein folding and QC. This figure has been adapted from Ito *et al.*[189] (copyright license can be found in Appendix Section 11)

The mannose residue on the central branch of the oligosaccharide structure is cleaved by a class I α-1,2-mannosidase residing in the ER. The class 1 α-1,2 mannosidases are classified by their inhibition by pyranose analogues and are members of glycosyl hydrolase family 47[200]. Similar to other enzymes involved in N-linked glycosylation they are a type II membrane protein. The α-1,2 mannosidase are divided into three subfamilies based upon substrate specificity, ER Mannosidases, Golgi Mannosidases and ER degradation-enhancing mannosidase-like (EDEM) proteins.

The ER mannosidase I cleaves a single mannose residue from the central chain of a $Man_9GlcNAc_2$ to generate the $Man_8GlcNAc_2$[201]. The ER mannosidase has been demonstrated to cleave to a $Man_5GlcNAc_2$ when overexpressed or *in vitro*[201]. However, the generation of $Man_5GlcNAc_2$ is normally performed by the Golgi mannosidase I subfamily. The ER mannosidase I and Golgi 1,2-mannosidase subgroups have complementary cleavage action

to each other, that is, the central mannose residue preferentially removed from $Man_9GlcNAc_2$ by the ER enzyme is the last mannose residue removed by the Golgi enzymes[201]. The cleavage of this central mannose residue by the ER 1,2-mannosidases is at least 10-fold slower than the cleavage of other 1,2 linkages by the Golgi 1,2-mannosidases[202]. This rate-limiting step is thought to be important for protein quality control for identifying misfolded proteins by the EDEM subfamily[164].

After these processing steps in the ER a properly folded protein is anterograde trafficked to the Golgi network. The glycoprotein is packaged into Coat Protein complex II (COPII) coated vesicles[203,204]. These vesicles bud from the ER membrane and ultimately fuse with the *cis*-Golgi[205]. The packaging of the vesicle can be through a non-selective process, known as bulk flow, or by a selective receptor mediated process via ER lectins[203]. ER lectins (VIP36, ERGIC-53 and VIPL), located at ER exit sites (ERES), act as cargo receptors by binding high mannose oligosaccharides ($Man_8GlcNAc_2$) for packaging into the COPII-coated transport vesicles[204]. The ERGIC-53 lectin has been noted to have a broad binding specificity against high mannose N-glycans, including glucosylated moieties, but showed little to no affinity to low mannose structures ($Man_{5-6}GlcNAc_2$)[206]. After budding from the ER membrane a COPII coated vesicle fuses with the cis-Golgi releasing the cargo of oligosaccharide linked proteins[205].

Once a fully-folded N-linked glycoprotein enters the cis-Golgi compartment the high-mannose structure is further processed by the action of a number of different resident glycosyltransferases and glycosidases[181]. The glycosyltransferases (GT) constitute a very large family of enzymes each specific for the transfer of one monosaccharide type to a specified branch or branches on the glycan chain[181]. The glycosyltransferases and glycosidases from N-linked glycosylation pathway are all type-II transmembrane proteins, each classified by a hydrophobic transmembrane domain, within the ER or Golgi membrane, a short cytosolic amino terminus tail and a luminal carboxyl terminus[181]. The carboxyl region consists of a stem region to which a globular catalytic domain is attached. Many glycosyltransferases and glycosidases require a metal ion to optimize catalysis. The ER Mannosidase I for instance requires a calcium ion to function; fortunately the ER lumen is the main cellular reservoir of calcium[207]. Eukaryotic species contain different types of glycosyltransferases and glycosidases. This results in divergent processing of the oligosaccharide substrate, allowing oligosaccharides with differences in monosaccharide compositions, numbers and differences in chain length and branching.

In multicellular eukaryotic species, the cleavage of α1–2 linked mannose residues continues. A set of Golgi Mannosidases trim the $Man_{9-8}GlcNAc_2$ substrate to a $Man_5GlcNAc_2$ oligosaccharide[201]. In humans, the Golgi α1,2 mannosidase subfamily consists of three

isozymes, mannosidases IA, IA2, and IC which process the $Man_{9-8}GlcNAc_2$ substrate[1]. The three isoforms show high amino acid sequence homology, for instance mannosidases IA and IB are about 65% identical in amino acid sequence[208]. Despite the sequence similarity each isozymes differs in their substrate preference[208]. The enzymes collectively cleave mannose sugars from a $Man_{9-8}GlcNAc_2$, preferring the $Man_8GlcNAc_2$ processed by the ER mannosidase I. As was noted before, the Golgi and ER mannosidase I have complementary cleavage specificity, this is due to significant differences in the positions of individual sugars in the binding pockets of the individual active sites[201]. A highly conserved Arginine residue in the ER mannosidase I forms the largest number of contacts with the oligosaccharide[209]. This interaction stabilises the α1,6-arm of N-glycan carbohydrate chain and thus may dictate the required position of the oligosaccharide for cleavage of the glycosyl linkage of the central mannose. The mutation of this conserved Arginine residue to a Leucine residue in the *S. cerevisiae* ER mannosidase I altered the activity to produce a $Man_5GlcNAc_2$[209].

The low mannose oligosaccharide, produced in multicellular eukaryotic species, is further processed in the medial-Golgi, but ultimately the maturation of the oligosaccharide is completed in the trans-Golgi. As noted previously, the Golgi contains an array of glycosyltransferases that assemble different monosaccharide moieties into linear and branched glycan chains. The $Man_5GlcNAc_2$ serves as a key intermediate structure for the biosynthesis of a hybrid or complex type glycan. Hybrid structures are identified by having mannose residues exclusively attached to the Man α1,6 arm of the core, whereas complex structures have a variety of monosaccharide on the Man α1,6 arm[1].

Construction of the diverging complex and hybrid structures begins within the medial-Golgi with the enzyme N-Acetylglucosaminyltransferase-I (GlcNAcT-I), which adds a GlcNAc residue to the α1,3 mannose in the $Man_5GlcNAc_2$ glycan structure[210,211]. In the majority of cases, a medial-Golgi type II α-mannosidase (Man-II) enzyme removes the terminal α1,3 mannose and α1,6 mannose sugar residues on the Man α1,6 arm of the core[212]. This allows for the addition of a second GlcNAc residue to the terminal mannose by the actions of the GlcNAcT-II enzyme, thus giving way to a complex type glycan structure with biantennary branching; $GlcNAc_2Man_3GlcNAc_2$[213,214]. If the medial-Golgi Man-II does not act or GlcNAcT-III does not bisect the $GlcNAc_1Man_3GlcNAc_2$ structure[215], hybrid glycans are then generated. Hybrid structures are typically limited to the linear elongation of one branch whereas complex oligosaccharides structures are often multifaceted and multi-branching[1]. The monosaccharide composition and length of the complex oligosaccharide structures differ between organisms due to species-specific glycotransferases.

For vertebrates the GlcNAc residues are typically extended through a galactose followed by the addition of a sialic acid residue, this terminal sugar is often known as a capping structure[161]. The galactose residue is added via a β1,4 linkage where the sialic acid can be added either by a α2,3 or α2,6 linkage. There are also a number of glycosyltransferases, which bind monosaccharide residues directly to the core of the oligosaccharide. This is observed in some vertebrates through addition of core fucose residues to the first GlcNAc sugar, via either α1,3 or α1,6 linkages[161]. Vertebrate species typically have a higher diversity in the monosaccharide composition and branching in the final oligosaccharide compared to other eukaryotes[216].

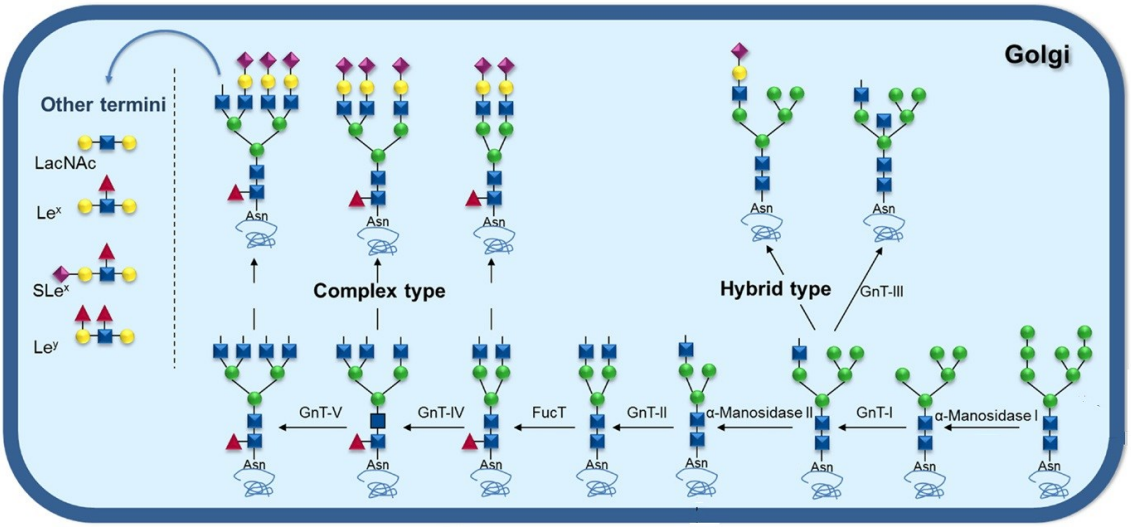


Figure 1.2 Schematic representation of N-linked oligosaccharides processing in the Golgi of a vertebrate species
After transfer of a glycoprotein to the Golgi, the attached oligosaccharide is demannosylated by the Golgi α-mannosidase I forming the $Man_5GlcNAc_2$ structure. This is a substrate for the GlcNAcT-I, which adds the first GlcNAc residue. This generates the $GlcNAcMan_5GlcNAc_2$, a substrate for GlcNAcT-III or Man-II. The Man-II trims two Mannose residues from the intermediate to form the core GlcNAcMan3GlcNAc2 precursor for complex N-glycans. Conversely, the action of GlcNAcT-III on the GlcNAcT-I product before α-mannosidase II directs the pathway to hybrid structures (110). As the pathway progresses through the Golgi, the core GlcNAcMan3GlcNAc2 can be further modified by a series of glycotransferases. This figure has been adapted from Vasconcelos-Dos-Santos *et al*.[217] (copyright license can be found in Appendix Section 11)

Plants, in contrast to vertebrates, produce an oligosaccharide of reduced complexity and diversity, which often lack decorating and capping reactions. Despite this, plants oligosaccharides contain a unique xylose sugar not found in other eukaryotes and also show differences in linkages[113,114]. The last known step in the N-glycan processing pathway in plants is the generation of what are known as Lewis A-epitope. This completed through the transfer of Galactose by a β1,3-linkage and a Fucose by a α1,4-linkage to the terminal GlcNAc residues[113]. Invertebrates display the simplest composition of oligosaccharide from multicellular organisms[122]. The key intermediate structure is only acted on by GlcNAc-TI and

ManII with some invertebrates also displaying a core α1,3-fucosylated variant[122]. Some insect cells also harbour a membrane-bound N-acetylglucosaminidase which processes a $GlcNAc_1Man_3\,GlcNAc_2$ to the trimannosyl core structure ($Man_3GlcNAc_2$)[218,219].

The processing observed in the multicellular eukaryotic Golgi is in stark contrast to that seen in yeast species. Yeast contain only mannosyltransferases within the Golgi, resulting in a hypermannose glycoprofile[220]. In *S. cerevisiae* a number of different mannosyltransferases add mannose residues via α1,2, α1,3 or α1,6 linkages, synthesising a oligosaccharide in excess of 200 mannose residues[221]. However, as previously stated other yeast species produce a much smaller hypermannosylated chain (Table 1.1).

Although an array of oligosaccharide structures is produced, they can all be classified into three types based on their peripheral structures: (i) high-mannose, which contain only mannose residues, (ii) complex, which contain only antennae initiated by GlcNAc residues, and (iii) hybrid, which contain only mannose residues on the Manα1-6 arm of the core and GlcNAc initiated antennae on the Manα1,3 arm. All three types share a common core structure of three mannose sugars and two GlcNAc residues[1,161].

1.23 N-linked glycosylation overview in Prokaryotes

Glycosylation was previously thought to be exclusive to the eukaryotic domain, however the discovery of N-linked glycosylation in the prokaryote *Campylobacter jejuni,* associated with gastroenteritis in humans, has changed this notion[222,223]. Protein glycosylation has been identified in a number of prokaryotic species, with the N-glycosylation pathway of *C. jejuni* being the most extensively characterized[222]. A genetic locus termed *pgl* (for protein glycosylation) was first identified to encode the glycosylation proteins. Various *pgl* genes share high sequence similarity to other glycosyltransferases and sugar biosynthesis enzymes. In particular, the protein PglB, which shares significant similarity to the yeast OST catalytic subunit *STT3* even containing the WWDYG catalytic domain[224–226].

To analyse the *pgl* gene cluster a mutant *C. jejuni* strains was developed with various *pgl* genes deleted[227]. A *C. jejuni pgl*B mutant strain was noted to have reduced cell surface glycosylation[223]. To further test the roles of *pgl* gene cluster, different combinations of the *C. jejuni pgl* genes were introduced into the model organism *E. coli*[59]. This helped to identify the glycosyltransferases responsible for generation of the oligosaccharide chain. The pathway begins with the synthesise of 2,4-diacetamido-2,4,6-trideoxyglucopyranose, also known as bacillosamine (Bac), from UDP-GlcNAc or UDP-GalNAc. The donor sugar nucleotides are converted sequentially by a dehydratase (PglF), aminotransferase (PglE), and acetyltransferase (PglD) to yield UDP-Bac[228–230]. The UDP-Bac is transferred to the phosphate

of the lipid carrier (UndP) by the actions Bacillosamine-1-phosphorytransferase PglC an integral membrane protein to create the Und-PP-Bac[231,232]. After transfer of the Bac monosaccharide, 5 GlcNAc sugars are subsequently added. The glycosyltransferase PglA adds the first GalNAc to create the disaccharide; PglJ adds the second GalNAc to create the trisaccharide; PglH adds the remaining three GalNAc residues to create the hexasaccharide[233]; and PglI adds the branching glucose to create the conserved heptasaccharide[59,234] (Figure 1.3). The glycosylation in model organism *E. coli* was only observed on the recombinant glycoprotein AcrA from *C. jejuni*, with glycosylation not being present on native *E. coli* proteins[59].

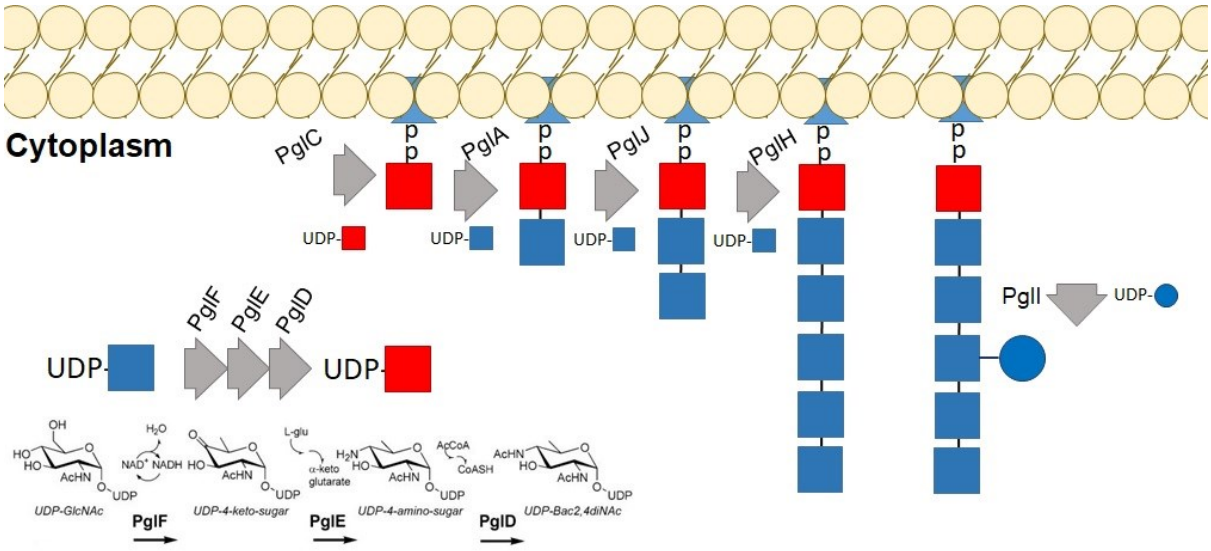


Figure 1.3 the N-linked glycosylation pathway in the prokaryote *C. jejuni*

Once the building of the heptasaccharide is complete, the phospholipid bound oligosaccharide is flipped or transported across the inner membrane into the periplasmic space. This flipping action is catalysed the flippase Wzx, a common flippase which transfers many und-PP-linked repeat units across the inner membrane[231,235]. After the heptasaccharide enters the periplasmic space it is transferred by the membrane bound OST to a fully-folded protein[236]. The OST of prokaryotes only consists of one subunit (PglB), compared with the multi-subunit eukaryotic OST. The PglB does however share significant structural similarity to the catalytic subunit *STT3*[224,237]. In *C. jejuni* the PglB transfers the oligosaccharide to an asparagine in the motif Aspartic Acid/Glutamic Acid-X_1-N-X_2-S/T (D/EX_1-N-X_2-S/T), where X is again any amino acid except proline[59,238]. With DQNAT having been defined experimentally as the optimal acceptor sequence[239]. The sequence is not a consensus acceptor sequence as other bacteria such as *Desulfovibrio desulfuricans* have different site requirements[240]. Once transferred the glycoprotein is relocated onto the surface of the bacterial cell wall. N-Linked glycans resulting

from this pathway have been identified on more than 65 *C. jejuni* proteins[241]. The oligosaccharide is not processed 'post *en bloc* transfer' as is observed in eukaryotes.

Although prokaryotic N-linked glycosylation shares the sequential process with eukaryotes; namely lipid based assembly of an oligosaccharide and *en bloc* transfer to the nitrogen residue of an asparagine, there are still many differences in the actual undertaking between the two domains. For instance, *C. jejuni* synthesises a heptasaccharide consisting of five GalNAc sugars, a glucose sugar and a bacillosamine sugar (a derivative of UDP-GlcNAc which is not present in eukaryotes). This is compared with the multi-branched tetradecasaccharide of eukaryotes. The oligosaccharide in *C. jejuni* is assembled on a undecaprenyl phosphate (Und-PP) lipid carrier, a different lipid carrier than eukaryotes. The assembly of the heptasaccharide occurs on the inner membrane within the cytoplasm, before being subsequently flipped into the periplasm where it is transferred e*n bloc* to a fully folded protein by the actions of an OST.

1.24 O-linked glycosylation overview

N-glycosylation and O-glycosylation share a number of similar building blocks (i.e. monosaccharides) but the end carbohydrates are synthesised through different enzymatic pathways[1]. This results in structural and compositional differences and therefore functional differences. O-glycans are attached via a glycosidic linkage to the hydroxyl group of an amino acid residue, usually serine or threonine, with O-glycosylation processing occurring mostly in the ER/Golgi[1,161]. The initial monosaccharides attached to the peptide chain can vary giving rise to a number of O-glycan types, namely O-N-acetylgalactosamine (O-GalNAc), O-mannose, O-N-acetylglucosamine (O-GlcNAc), O-glucose, O-galactose (O-Gal), O-fucose and O-xylose (glycosaminoglycans). GalNAc-linked glycans are known as mucin-type glycans, due to the fact that mucin glycoproteins are highly glycosylated with S/T-GalNAc[181]. The S/T-GalNAc is the simplest O-glycan structure and is known as the Tn-antigen. The Tn-antigen can be extended into a variety of different structures[71]. Mucin-type glycans function to hydrate and protect the underlying epithelial surfaces of the respiratory, gastrointestinal, and reproductive tracts[242].

1.25 A glycoengineering approach in bioprocessing

Recombinant protein expression has revolutionised pharmaceuticals creating a multibillion dollar industry[3]. As is clear from above a large number of recombinant hosts exist from a range of organisms, each offering certain advantages over another. The capacity of a host to perform N-linked glycosylation is a major consideration when choosing an expression host for biopharmaceuticals. The N-linked glycosylation can affect stability, half-life, clearance from the body and performance of a biopharmaceutical and in certain cases cause an immunogenic

response to a human patient[243–245]. Consequently, most biopharmaceuticals possessing an N-linked glycan motif are produced in a mammalian expression system. This is represented by the predominance of mammalian overexpression systems over non-mammalian based expression systems in FDA approved biopharmaceuticals with an N-linked glycosylation[7]. The preferred mammalian expression host is CHO cells[7].

As has previously been discussed, mammalian cells contain a number of glycosyltransferases which have the potential to generate 100's of different glycoforms. During the generation of therapeutic protein, it is important to tightly control the bioprocessing method to ensure a suitable and homogeneous glycan motif is produced. Parameters such as nutrient availability, dissolved oxygen, temperature, and pH can affect the quality of a product and glycoform[246]. For example, the lack of availability of the donor sugar UDP-GlcNAc can result in a high-mannose glycan profile ($Man_5GlcNAc_2$)[247]. Although the ambition is to synthesise a biopharmaceutical with a homogeneous oligosaccharide, and that future manufactured batches, of that biopharmaceutical, contain a matching homogeneous oligosaccharide, this is rarely the case. A degree of heterogeneity is expected and accepted by regulatory bodies, typically 3 or 4 glycoforms are seen as the predominant glycoform in CHO cells[248]. By repeating the bioprocess parameters each time, the drug is manufactured, this ensures consistency in reproducing the same glycoforms.

There are a number of methods which exist that allow producers to control the end glycoprofile. One such method in the use of chemical inhibitors or monosaccharide analogues which block the action of glycosylation enzymes therefore halting the elongation and processing of a glycan chain. For instance, the chemical swainsonine blocks α-mannosidase II, therefore resulting in the hybrid-type ($GlcNAc_1Man_5GlcNAc_2$) glycoprofile[249,250]. A fucose analogue (2-fluorofucose) can also act as an inhibitor[251], blocking GDP-mannose-4,6-dehydratase, a central enzyme for the synthesis of the donor nucleotide sugar UDP-fucose. Consequently the fucosylation of the glycan core is then prevented due to limitations in cellular UDP-fucose concentration[251]. Chemical inhibitors are highly specific and can be beneficial when the removal of enzyme impacts the cellular fitness or is lethal. However, cell toxicity from chemicals inhibitors still exists, as evident from the potential of swainsonine and 2-fluorofucose as chemotherapy agents. Therefore, to ensure there is no adverse reactions to a patient a costly chemical removal step would be necessary[251,252]. Also as the name suggests chemical inhibitors prevent and do not enhance glycosylation.

A second method is in the genetic engineering of the glycosylation pathway of an expression host, termed glycoengineering. Strain improvement through genetic engineering has been shown to be very effective, as demonstrated by the enhancement of disulphide bond formation

in *E. coli* (see Section 1.4). Similar principles, i.e. the addition of recombinant and/or removal of native proteins, can be exploited to alter glycosylation pathways within a host.

From the outset of mAb production in CHO cells, the tendency to add non-human sialic acid residues was noted to be a problem[253,254]. Consequently, in 1989 a CHO cell line was developed to overexpress the human α-2,6-sialyltransferase to produce a more human-like glycoprofile[255]. The need to deliver a more human glycoprofile in CHO cells has seen the undertaking of a number of glycoengineering strategies, which have been extensively reviewed by Fischer *et al*.[108]. The glycoengineering strategies includes the overexpression of native and recombinant genes, and the knockdown of native genes (Table 1.5).

Table 1.5 list of glycoengineering strategies employed on CHO cells.
CHO cells represent the most frequently applied host cell for the production of biopharmaceuticals with post-translational modifications. Despite the capacity to produce N-linked glycosylation chain, the oligosaccharide is only human-like, with certain differences negative effecting a biologics attributes. Host cell engineering technologies targeting the N-linked glycosylation pathway, comprising introduction, knock-out or post-transcriptional silencing of native genes, aim to diminish this effect. The table below has been adapted from Fischer *et al.*

CHO cell glycoengineering by using gene introduction.			
Recombinant gene name	**Recombinant gene source**	**Glycoengineering function**	**Reference**
α-2,6 sialyltransferase (ST6GAL)	Human Rat	Increase in the extent of sialylation	255–260
α-2,3 sialyltransferase (ST3GAL) and	Mouse Human	Increase in the extent of sialylation	261–264
β-1,4 N-acetylglucosaminyltransferase III (GlcNAcT-III)	Rat	Increase in biantennary glycan chains	265,266
α-1,3-D-mannoside β-1,4 N-acetylglucosaminyltransferase (GlcNAcT-IV)	Bovine	Increase in tetraantennary glycan chains	261
α-1,6 D-mannoside β-1,6 N-acetylglucosaminyltransferase (GlcNAcT-V)	Human	Increase in tetraantennary glycan chains	261

β-1,4 galactosyltransferase (GalT)	Human	Increased sialylation when co-expressed with a sialyltransferase	262,267
CMP-sialic acid transporter (CMP-SAT)	CHO Human	Increase in the extent of sialylation	263,264,268
CMP-sialic acid synthetase (CMP-SAS)	Human	Increase in the extent of sialylation	263
Mutant uridine diphosphate-N-acetyl glucosamine 2-epimerase (GNE),	CHO	Increase in CMP-sialic acid concentration leading an increase in the extent of sialylation	264
Golgi α-mannosidase II (ManII)	Human	Expression of non-fucosylated antibodies possessing N-glycans of the complex type	266
GDP-6-deoxy-D-lyxo-4-hexulose reductase (RMD)	*Pseudomonas aeruginosa*	Involved in fucose metabolism, increase fucylation	269
CHO cell glycoengineering using genomic gene knock-outs			
Native gene name	**Glycoengineering function**		**Reference**
α-1,6-fucosyltransferase (FUT8)	Production of non-fucosylated glycan chain		270–273
CMP-Sialic acid transporter	Production of glycan chains lacking sialic acid		274
GlcNAcT-I	Production of recombinant proteins with Man5 as the predominant N-linked glycosylation species		275,276
GDP-Fucose transporter	Production of non-fucosylated glycan chain		274

GlcNAcT-IVA, GlcNAcT-IVB GlcNAcT-V	Production of almost homogenous biantennary N-glycans	273
Galactosyltransferase 1 (GalT-I)	reduction in galactosylation	273
Galactosyltransferase-4/6	Production of non-sialylated glycan chain	273
CHO cell glycoengineering using genomic gene knock-down approaches		
Native gene name	**Glycoengineering function**	**Reference**
Sialidase (NEU1, NEU2 and NEU3)	Increase in the extent of sialylation	277–279
FUT8	Reduction in core fucose by 60–88% resulted in 100-fold improved ADCC of the produced IgG	280–282
GDP-fucose 4,6-dehydratase	Production of non-fucosylated glycan chain	282,283
GDP-fucose transporter (GFT)	Production of non-fucosylated glycan chain	284

The glycoengineering of CHO cells aims to produce a more human-like glycoprofile, a further benefit to glycoengineering is in the production of a homogenous glycan motif. This is especially apparent when an enzyme is removed. For instance, it is common for CHO cells to synthesise glycoforms with and without a core fucose residue, which as discussed earlier affects the pharmacokinetics of a mAb. The glycoengineering of CHO cells to produce an non-fucosylated glycan motif has been used to produce mAbs with increased ADCC activity[270]. This was achieved by removing a α-1,6 fucosyltransferase (FucT-VIII) which transfers fucose residues from GDP-fucose to the innermost GlcNAc of the tri-mannosyl core structure via the α-1,6 linkage[270]. The elimination of the core fucose in the mAb Gazyva®, which is used in lymphoma treatment, enhances the efficacy of the drug over similar mAbs which target lymphoma, but produced in CHO cells which can add core-fucose[7]. Gazyva became the FDA's first approved drug from an glycoengineered expression host in 2013[7].

1.26 Glycoengineering in yeast

Much of the early work to elucidate the pathways for glycan processing was performed in the model organism *S. cerevisiae* thus making it an ideal initial candidate for yeast glycoengineering attempts. In the 1992 the Golgi based α1,6 mannosyltransferases *och1p* and α1,3-mannosyltransferase Mnn1p were identified to be responsible for the outer chain elongation of $Man_8GlcNAc_2$ glycan chain[177]. Each enzyme responsible for the targeting and transferring mannose residue onto the incoming ER intermediate oligosaccharide. The glycoprofile of a double knockout *OCH1* and *MNN1* strain was noted to contain only the high-mannose structure of $Man_8GlcNAc_2$, which, as previously stated, acts as a substrate for α1,2 mannosidase[83]. In 1998 Chiba *et al.* demonstrated that by inserting a α1,2-D-mannosidase (EC 3.2.1.113) from *Aspergillus saitoi* into a *Δoch1:mnn1* strain of *S. cerevisiae,* a $Man_5GlcNAc_2$ glycoprofile could be produced[285]. Expression of the recombinant soluble 1,2-α-D-mannosidase enzyme was placed under the control of a constitutive promoter from Glyceraldehyde-3-phosphate dehydrogenase (GAP) and contained a C-terminal His-Asp-Glu-Leu (HDEL) tag[285]. The terminal HDEL sequence on the protein acts as a retention/retrieval signal for the ER. The conversion of $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$ was found to be incomplete, with only 27% of a native carboxypeptidase Y (CBY) containing the $Man_5GlcNAc_2$ glycoprofile, with the remaining glycoprofile comprising a mix of $Man_8GlcNAc_2$, $Man_7GlcNAc_2$ and $Man_6GlcNAc_2$ at a ratio of 29:22:22[285].

Although eliminating hypermannosylation, the removal of the *OCH1* and *MNN1* genes is severely detrimental to *S. cerevisiae* viability[177]. In *S. cerevisiae* the mannose-type glycans cover the cell surface and function to protect and maintain the cell wall. The removal or even reduction in length of the mannose chain causes temperature-sensitive growth defects, cellular aggregation, deficiencies in cytokinesis and hypersensitivity to agents that attack the cell wall, such as tunicamycin[155,286,287]. Despite the deletion of *OCH1* being unfavourable it is essential for a successful conversion of $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$. As demonstrated in 2001 by Callewaert *et al.* who expressed a HDEL-tagged *T. reesei* α1,2-D-mannosidase within a native *P. pastoris* (*OCH1*$^+$)[288]. Despite the observable cleavage of α1,2 linked mannose in the glycan chain of recombinant influenza A haemagglutinin, no $Man_5GlcNAc_2$ profile was observed[288]. Instead $Man_6GlcNAc_2$ was the predominant glycoprofile due to the transfer of an α1,6 linked mannose by *och1p* onto the α1,3-linked mannose that is attached to the β1,4-mannose of the $Man_8GlcNAc_2$[288]. The $Man_6GlcNAc_2$ glycoprofile cannot act as a substrate for biosynthesis of hybrid-type and complex-type structures, as the additional mannose residue will prevent the binding of downstream glycan processing enzymes[288]. In the same study Callewaert *et al.* also demonstrated the *in vivo* conversion of $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$

on the recombinant protein trans-sialidase from *T. cruzi*, although $Man_8GlcNAc_2$, $Man_7GlcNAc_2$ and $Man_6GlcNAc_2$ glycoprofiles were also present[288].

It is worth noting here that although the initial glycoengineering was performed within *S. cerevisiae,* effort soon shifted to *P. pastoris*. As *P. pastoris* has a mannose profile of less residues, 12-16 residues compared to over 200 in *S. cerevisiae* (Table 1.1), the removal of hyperglycosylation was considered to be less detrimental to the fitness of the host. Furthermore, *P. pastoris* had been responsible for the translation of 1000's of recombinant proteins, and has been noted to produce higher titres and secrete less host cell proteins than *S. cerevisiae.*

These initial studies demonstrated that by expressing an α-1,2-mannosidase in yeast a $Man_5GlcNAc_2$ could be achieved, although the efficiency of $Man_5GlcNAc_2$ conversion was low. Despite this Chiba *et al.* suggested that the addition of further mammalian downstream glycan processing enzymes (e.g. GlcNAcT-I) could result in the biosynthesis of hybrid-type and complex-type structures. However, before this could happen the conversion efficiency needed to be improved. Two possible reasons for the low conversion included an insufficient concentration of the enzyme or environmental conditions affecting its catalytic activity. The first point seemed unlikely, as α-1,2-mannosidase expression was under the control of the GAP promoter which provides strong constitutive expression[289] and both researchers demonstrated high expression of the α-1,2-mannosidase. However, in relation to the enzyme environment it is known that unfavourable conditions will inhibit the activity of enzyme. Aspects such as pH optimum, metal ion requirements, and temperature optima must be considered when expressing a recombinant enzyme. For instance, the different compartments within the secretory pathway have different pH conditions, the ER has a more neutral environment (pH 7.4 in S*. cerevisiae*) compared to more acidic environment in the trans-Golgi network (pH 6.4 in *S. cerevisiae*)[290,291]. The *A. saitoi* and *T. reesei* α-1,2-mannosidase both had a pH optimum of 5.0, therefore the inefficiency in conversion is likely due to the poor enzymatic activity within the pH 7.4 environment[292]. Therefore, the next logical step for researchers was to use a α-1,2-mannosidase with a neutral pH optimum.

To produce an α-1,2-mannosidases with 100% or at least near 100% cleavage to $Man_5GlcNAc_2$, within a *Δoch1* strain of *P. pastoris,* Choi *et al.* tested a number of heterologous α-1,2-mannosidases[293,294]. The enzymes were sourced from *C. elegans, D. melanogaster, Homo sapiens, Penicillium citrinum, Mus musculus* and *Aspergillus nidulans,* with each enzyme having a different optimum pH[293]. Previously, the recombinant α-1,2-mannosidases were soluble, whereas Choi *et al.* placed yeast type II membrane protein leader domain onto each of the proteins[293]. The native leader domain sequences which localise the enzyme to the

Golgi, are non-functional in yeast and therefore could not be used. The leader domains are designed to ensure effective localisation of α-1,2-mannosidase into the ER membrane, therefore preventing the secretion of α-1,2-mannosidase as had been observed in previous studies. It will be crucial to complete the conversion to $Man_5GlcNAc_2$ before glycoproteins exit the ER, as this substrate is necessary for the sequential heterologous downstream glycan processing enzymes which will be localised within the native location of Golgi, ensuring a similar pH environment. Choi *et al.* tested a number of leader domains, on a number of different truncated versions of α-1,2-mannosidases[293]. In total over 600 combinations of heterologous α-1,2-mannosidases and leader sequences were tested[293,294]. A putative *C. elegans* mannosidase was shown to have the highest activity with a 70–80% conversion to $Man_5GlcNAc_2$[293,294]. This improvement in conversion efficiency was probably due to optimal pH of the enzyme being closer to the neutral pH environment of the *P. pastoris* ER.

The increased conversion efficiency to $Man_5GlcNAc_2$ enabled Choi *et al*. to add a GlcNAcT-I to *P. pastoris och1p* knockout strain for the biosynthesis of a $GlcNAc_1Man_5GlcNAc_2$ glycoprofile[293,294]. A recombinant library of GlcNAcT-I was constructed containing over 60 different truncated versions of GlcNAcT-I from different mammalian species with yeast type II membrane protein Golgi leader domains[294]. A human GlcNAcT-I with the leader sequence *MNN9 from S. cerevisiae* was identified as the most active in producing a $GlcNAc_1Man_5GlcNAc_2$ within the *Δoch1* strain of *P. pastoris* expressing α-1,2-mannosidases[294]. To ensure efficient addition, a UDP-GlcNAc transporter from *Kluyveromyces lactis* was inserted into the genome[294]. The efficiency of the transfer is linked to the availability of UDP-GlcNAc, as a lack of UDP-GlcNAc will result in a lack of transfer. A strain deficient in a Golgi UDP-GlcNAc transporter produced significantly less $GlcNAc_1Man_5GlcNAc_2$. Interestingly, there appears to be no consensus on whether a transporter is necessary, as a second group (Vervecken *et al.*) showed that a UDP-GlcNAc transporter was not required for efficient conversion[295]. In this study, Vervecken *et al.* overexpressed a GlcNAcT-I with a *S. cerevisiae* Kre2 leader sequence for Golgi localisation within a *P. pastoris Δoch1:Man1a*$^+$ strain[295]. Although a recombinant UDP-GlcNAc transporter was not expressed, a near complete conversion of $Man_5GlcNAc_2$ to $GlcNAcMan_5GlcNAc_2$ was observed[295]. The main difference between the two studies is the reporter protein sampled for glycosylation analysis, with both studies utilising the same growth media. Choi *et al*. overexpressed a Kringle 3 domain of human plasminogen (K3) whereas Vervecken *et al.* sampled native proteins. Although titres of proteins were not reported, one can be certain that the amount of recombinant protein (i.e. K3) would be in excess of native proteins. Thus, one can assume that as protein levels increase, as observed during recombinant protein expression, the Golgi UDP-GlcNAc will quickly become depleted therefore resulting in an

incomplete transfer. Consequently, the inclusion of sugar donor nucleotide transporters provides the Golgi with excess UDP-GlcNAc enhancing GlcNAc transfer.

An increase in the Golgi UDP-GlcNAc availability was successfully achieved by introducing a transporter, however in cases were the nucleotide donor sugar is not native to the host this strategy would be ineffective and a synthesis step will be essential. Although Vervecken *et al.* successfully expressed a human β-1,4-galactosyltranferase I (GalT-I), the number of native proteins with a Galactose residue was less than 10%, as *P. pastoris* is unable to synthesise a UDP-Galactose (UDP-Gal) nucleotide donor sugar[248]. An attempt to increase conversion by supplementing the culture medium with 0.2% galactose also failed. To increase the Golgi reservoir of UDP-Gal, a UDP-galactose 4-epimerase (UGE) from *S. pombe* was introduced[296]. This enzyme has been demonstrated to convert UDP-Glucose to UDP-Gal in *P. pastoris*[297]. Bobrowicz *et al.* localised a fusion enzyme of UGE and bovine GalT to the Golgi membrane and demonstrated a conversion rate over 80% from $GlcNAc_2Man_5GlcNAc_2$ to $Gal_2GlcNAc_2Man_5GlcNAc_2$ in a recombinant protein[296].

Where glycotransferases for addition of GlcNAc and Galactose required a transporter and a single enzymatic synthesis step, respectively, the final step involving transfer of sialic acid was seen as more challenging[298]. This step will require both donor nucleotide sugar transporters and a multi-step enzymatic pathway for synthesise of CMP-sialic acid, which is not produced by *P. pastoris*. In total, six enzymes are required: UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase (GNE), N-acetylneuraminate-9-phosphate synthase (SPS), sialylate-9-P phosphatase (SPP) and CMP–sialic acid synthase, each of which are involved in the biosynthesis of CMP-sialic acid from UDP-GlcNAc[299]. The CMP-sialic acid biosynthesis enzymes were expressed as soluble enzymes for localisation to the cytoplasm[299]. A *M. musculus* CMP–sialic acid transporter was added to transport CMP-sialic acid into the Golgi, where finally a *H. sapiens* chimeric sialyltransferases, linked to yeast type-II transmembrane Golgi localization peptides (ST), would transfer sialic acid on the $Gal_2GlcNAc_2Man_3GlcNAc_2$ glycoprofile[299]. Initially the percentage of glycan conversion to $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$, was 60% on the recombinant protein Erythropoietin (EPO), however after codon optimisation of human genes for translation in *P. pastoris* this conversion rate increased to 100%[299]. In total 14 heterologous genes are required, including genes involved in donor nucleotide sugar synthesise, transporters, glycosidase and glycosyltransferases which have led to the synthesis of therapeutic relevant $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$ glycan chain.

The extensive engineering which is required to generate a humanized glycosylation pathway will exert an unnatural load onto the *P. pastoris.* This could negatively impact the overall

performance of *P. pastoris* reducing the growth rate and protein productivity. The glycoengineered *P. pastoris* strain has been noted to divide slower, with the engineered strain producing $Gal_2GlcNAc_2Man_5GlcNAc_2$ having a doubling time 4.62 hours compared to 2.40 hours from the native *P. pastoris* [99]. However, the protein yields do not appear overly affected, with a glycoengineered *P. pastoris* $Man_5GlcNAc_2$ strain producing titres >1g L^{-1} for recombinant IgG[300], which is comparable to a mammalian cell culture (Table 1.3). In addition to this, similar to strategies which seen increasing mAb titres in CHO cells, such as optimisation of media and culture condition, have been employed in glycoengineered *P. pastoris* ($Man_5GlcNAc_2$). Barnard *et al.* demonstrated high-throughput screening to select for high producers[301] and Berdichevsky *et al.* increased final mAb concentration from 1 g L^{-1} reported earlier to 1.9 g L^{-1} by limiting the oxygen content of the culture medium[302]. However, it must be noted that the proteins titres from a fully glycoengineered *P. pastoris* strain have not reported. Potentially, a fully glycoengineered *P. pastoris* strain with 14 heterologous genes may not produce the titres seen from the $Man_5GlcNAc_2$ glycoengineered *P. pastoris*.

One of the major benefits of using an engineered *P. pastoris* is only non-fucosylated glycoforms will be produced, as the machinery capable of adding fucose does not exist. This is evident from the expression of an anti-HER2 IgG, where CHO cells added a core fucose the engineered *P. pastoris* only had a-fucosylated glycan chains[300]. Although *P. pastoris* contains just one sequential glycosylation pathway compared to a multivariable pathway seen in mammalian cells, the glycoprofile displayed on *P. pastoris* are still not homogenous. For instance, the anti-HER2 IgG contains two glycoprofiles $Sia_1Gal_2GlcNAc_2Man_3GlcNAc_2$ $Gal_2GlcNAc_2Man_3GlcNAc_2$ at a ratio of 25:75, whereas CHO cells produced at a 50:50 ratio (although with core fucose)[300]. The addition of sialic acid can increase the half-life of a number of glycoproteins, as sialic acid acts as a cap that protects the galactose residue from hepatic asialoglycoprotein receptor (ASGPR)[303]. The binding of galactose to ASGPR can promote the delivery of the glycoprotein to lysosomes for catabolism[304]. The steric hindrance caused by the mAb structure normally prevents the addition of terminal sialic acid to the galactose on a glycan chain.

1.27 Insect cell-line glycoengineering

The processing pathway of the insect N-linked glycosylation pathway is truncated relative to that of the mammalian pathway[305]. Insect cells lack of elongation enzymes and presence of a N-acetylglucosaminidase (NAGLU), consequently produce a shortened trimannosyl core glycan structure[219,306]. Although, under certain conditions, a terminal sialylated glycan has been observed from several insect species[307] such as in the central nervous system of *Drosophila melanogaster*[308] and the silkworm *Bombyx mori*[309]. This potential sialylation

capacity of insect cell lines has motivated researchers to attempt to direct a glycosylation pathway, through host cell engineering technologies comprising introduction and knock-out of genes, to produce a more mammalian-like glycoprofile[306]. Much of the glycoengineering effort in insect cells has been concentrated on lepidopteran insect cell lines, specifically, Spodoptera frugiperda (Sf9) cells[306,310].

The first step in glycoengineering Sf9 was to reduce the trimannosyl glycoprofile. This was achieved by overexpressing a human GlcNAcT-I, which resulted in a fourfold increase in generation of $GlcNAc_1Man_3GlcNAc_2$ structures[311]. However due to the presence of an NAGLU, the trimannosyl core structure still remained. After the insertion of human GlcNAcT-I, the Sf9 cell line was engineered to constitutively express integrated copies of a bovine β1,4-galactosyltransferase (1998)[312] and rat α2,6-sialyltransferase (2001)[313]. An initial lectin-binding assay demonstrated the transfer of Gal and sialic acid residues, however a later detailed analysis by mass spectrometry revealed only a monoantennary structure of $Sia_1Gal_1GlcNAc_1Man_3GlcNAc_2$ to be present. This structure represented less than 10% of the glycoprofile of the recombinant protein with the predominant profile being a trimannosyl structure $Man_3GlcNAc_2$[314]. In addition, N-glycan structures also displayed an immunogenic core fucose residue due to the activity of a α1,6-fucosyltransferase within the Sf9 cell line.

The blocking of fucosylation was achieved by preventing the biosynthesis of GDP-fucose. This was achieved by incorporating a *Pseudomonas aeruginosa* GDP-4-dehydro-6-deoxy-D-mannose reductase (Rmd), which consumes the immediate precursor to GDP-L-fucose, GDP-4-keto-6-deoxy-D-mannose (Figure 1.4)[315]. The insertion of Rmd was demonstrated to eliminate the immunogenic core α1,6-fucosylation of a recombinant glycoprotein[315]. It seems an interesting choice by the authors to remove fucosylation function by inserting an enzyme to scavenge nucleotide sugar donor, rather than create a fucosyltransferase deletion strain. However, currently no reported examples of targeted gene knockouts exist in any lepidopteran insect cell line and gene knockout is further complicated by the non-availability of an annotated whole genome sequence of Sf9. Consequently, the methodology of preventing a monosaccharide addition can only be achieved by causing a competing reaction.

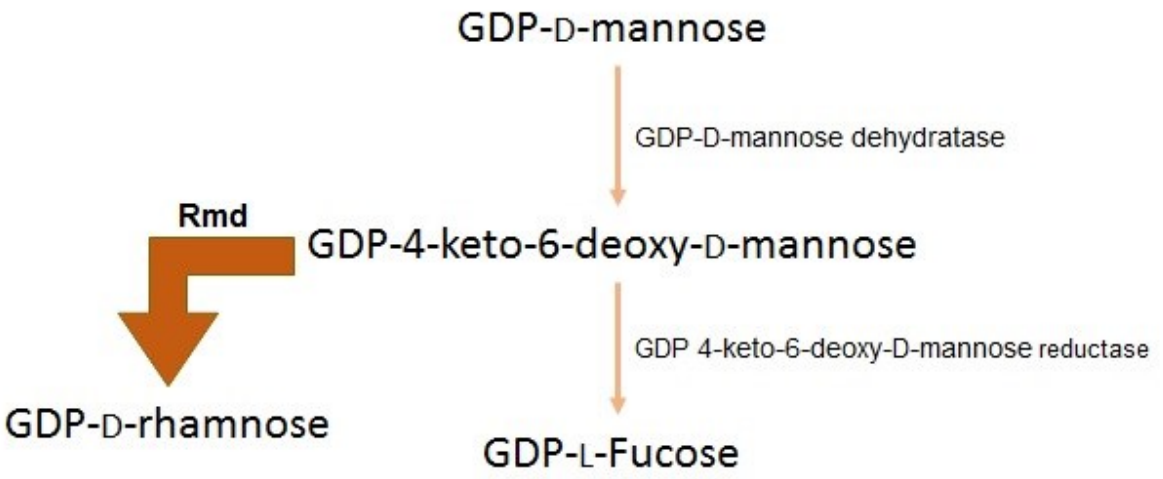


Figure 1.4 the biosynthetic pathway of GDP-L-fucose
Mabashi-Asazuma *et al.* postulated that with the addition of the bacterial Rmd, the biosynthesis of GDP-L-Fucose will be blocked. As Rmd diverts GDP-L-Fucose precursor toward synthesis of GDP-D-rhamnose. The reaction occurs within the cytoplasm of engineered insect cell-line.

The lack of elongation observed in the second branch was due to the low activity levels of endogenous GlcNAcT-II in Sf9 cell lines, which had been reported at less than 1% in comparison to mammalian GlcNAcT-II[316]. To increase the levels of biantennary structures a Sf9 strain was developed which overexpressed human GlcNAcT-II[317]. Although an increase in biantennary structures was observed, very few contained a sialylated cap[317]. As Sf9 cell lines are unable to biosynthesise CMP-sialic acid, this limits the intracellular supply leading to inefficient transfer of sialic acid. Any observable sialylation was reasoned to be due to the supplementation of media with CMP-sialic acid[317,318]. To increase the intracellular concentration of nucleotide donor sugar, and therefore increase transfer, two mouse enzymes involved in CMP-sialic acid biosynthesis, sialic acid synthase (SAS) and CMP-sialic acid synthetase (CMP-SAS), were incorporated into a Sf9 cell line[318]. As Sf9 had previously demonstrated transfer of sialic acid, a CMP-sialic acid transporter was not essential, this is in contrast to the glycoengineered *P. pastoris*. Despite this to ensure that native transporter systems are not overwhelmed by foreign nucleotide donor sugars a recombinant human CMP-sialic acid transporter (hCSAT) was also incorporated[319]. Utilising a fluorescent lectin-binding assay to measure sialylated glycoprofiles, an increase in terminal sialylation was demonstrated for the strain expressing SAS, CMP-SAS and hCSAT[318,319].

Although CMP-sialic acid biosynthesis led to increased sialylation, this increase was only viable in presence of the CMP-sialic acid precursor N-acetylmannosamine (ManNAc)[319]. Decreased levels of ManNAc resulted in a decrease in sialylation[320]. As previously mentioned a disadvantage of using insect cell lines over yeast strains for recombinant protein expression is the need for more complex and costly media to sustain growth (Table 1.2). The addition of

precursor and nucleotide sugar donors only increases the cost. Consequently, the incorporation of a *de novo* pathway for the biosynthesis of CMP-sialic acid, via ManNAc and UDP-GlcNAc would require a sufficient intracellular supply of ManNAc. This problem was addressed by incorporating an *E. coli* GlcNAc-6-phosphate-2'-epimerase (GNPE), within the glycoengineered Sf9 cell line[321]. GNPE normally functions in bacterial sialic acid degradation converting GlcNAc into ManNAc, but within a eukaryotic setting the reverse reaction occurs leading to an increase in intracellular concentration ManNAc[321]. The authors demonstrate that these improvements to the Sf9 cell line removed the need for addition of costly supplements without impacting the end glycosylation product[321].

Despite the insertion of seven constitutively expressed recombinant proteins into the glycoengineered Sf9 cell line no discernible effect was observed in the cellular fitness. The glycoengineered strain was shown to have similar growth rates to the parental cell line achieving similar peak densities. The cell morphologies were also quite consistent, except in one case after incorporation of GlcNAcT-II where a decrease in cell size was observed[317]. However, the cell size was restored after a subsequent gene insertion of GalT. This minimal effect on cellular fitness and morphology is in stark contrast to a glycoengineered *P. pastoris* strain, which displayed a decrease in growth rate and significant morphological differences[99]. This difference is probably due to the end function of glycoproteins, which in yeast function to protect and maintain the cellular wall, whereas within invertebrates, glycoproteins are not essential for cell membrane integrity.

Nevertheless, there is still the potential for metabolic burden, which could negatively impact the expression levels of a recombinant protein. Further to this, the ie1 constitutive promoter utilised to control expression of the glycosylation genes functions during the early stages of the baculovirus life cycle. Thus, there is a possibility of baculovirus-mediated shutdown of transgene expression during the later stages of the baculovirus life cycle. These potential risks led researchers to develop a strain where glycosylation genes are placed under an inducible promoter system[322,323]. Using the inducible baculovirus promoter, 39K, higher levels of transgene expression were observed over the previous constitutive system[323]. The quantitative analysis of a recombinant glycoprotein revealed a decrease in the percentage of trimannosyl structures from 93% in the native Sf9 strain, 25% in the non-inducible strain and 18% in the inducible system[323]. More strikingly, the quantitative analysis revealed a two-fold increase in the proportion of sialylated N-glycans between the inducible and non-inducible glycoengineered strain, 39% versus 18%[323].

In total nine glycosylation genes, including genes involved in donor nucleotide sugar synthesis, transporters and glycosyltransferases, were incorporated into a Sf9 cell line with the aim of

synthesising the therapeutically relevant $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$ glycan chain. However, the glycoengineered insect cells lines were not capable of converting all N-linked glycosylation chains to the biantennary terminally sialylated N-glycan structure. The conversion efficiencies ranged from 0.1%[319] to 39%[323], with a mixture of hybrid-and complex-type N-glycans with terminal GlcNAc, Gal, and/or sialic acid residues. Despite the glycoengineering efforts employed the native trimannosyl structure still persisted although at considerably lower quantities than the native cell line. The presence of an active N-acetylglucosaminidase within the glycoengineered insect cell line will continually divert some GlcNAc1Man3GlcNac2 structures to Man3GlcNac2, and until an effective knockout strategy is developed for baculoviral insect cell lines this will continue to be a problem. Recently an N-acetylglucosaminidase knockout strain of *D. melanogaster* was generated using a CRISPR-cas9 genomic engineering strategy[324]. It is anticipated that this technology could be extended to Sf9 in the near future to create a N-acetylglucosaminidase deficient cell line.

Despite the authors demonstrating the glycoengineered insect cell line's ability to synthesise the therapeutically relevant $Sia_2Gal_2GlcNAc_2Man_3GlcNAc_2$ glycan chain, no volumetric measurements of recombinant proteins were reported. Therefore, it is unclear if the introduction of seven heterologous genes led to a metabolic overload affecting protein titre.

1.28 Glycoengineering in E. coli

The previous examples of glycoengineering succeeded by altering the native eukaryotic glycosylation pathway through the introduction and/or removal of genes, this enabled the synthesis of more human-like glycosylation chain. However, glycoengineering is not limited to species which already possess N-linked glycosylation machinery. The successful functional transfer of *C. jejuni* N-linked glycosylation into *E. coli,* although conceived to aid identification, opened the possibility of using this established expression host as a platform for the translation of N-linked glycoproteins[325]. This was important as the utilisation of *C. jejuni* as a recombinant expression host would be most challenging, due to the extreme pathogenicity, the difficulty in maintaining a microaerobic atmosphere for viability (85% N_2, 10% CO_2, 5% O_2) and the complex media requirements (such as horse blood)[225,326]. Moreover, the developed *C. jejuni* chromosomal integration vectors were specifically designed to study the underlying genetics associated with virulence and not explicitly for recombinant protein expression[327].

E. coli, on the other hand, is an established recombinant expression host with a huge wealth of knowledge and tools available to enable effective protein translation. As *E. coli* does not perform N-linked glycosylation, and therefore lacks any N-linked glycosylation machinery, a complete *de novo* synthetic pathway was required. The absence of a glycosylation pathway can be seen as beneficial as flow of metabolites and sugar donor nucleotides is directed

toward the engineered biological pathway rather than competing with native glycosylation, as is the case in semi-synthetic pathway of glycoengineered eukaryotic strains. Furthermore, the synthetic pathway can be tailored to remove undesirable *C. jejuni* traits. For instance, the exclusion of genes leading to the biosynthesis of immunogenic prokaryote-specific sugars will exclude Bac from the end glycoform[328].

Although *E. coli* does not perform N-linked glycosylation, *E. coli* can synthesise carbohydrate polymers in the form of O antigens[329]. As such, four Und-P lipid carriers which act as anchors for the biosynthesis of O antigens have been identified within *E. coli,* located on the plasma membrane[330]. The O antigens are assembled on the cytosolic side of the plasma membrane before the Und-PP-carbohydrate chain is flipped into the periplasmic space by the actions of the flippase *Wzx*[331]. The completed O antigens chains are released and covalently linked to lipid A by actions of the ligase *WaaL*[329,330]. The O antigens are subsequently exported to the bacterial cell envelope to form a component of the endotoxin lipopolysaccharides (LPS)[329,330]. In *E. coli*, an Und-P-GlcNAc-1-phosphatetransferase known as *WecA* transfers GlcNAc-1-phosphate to Und-P lipid carrier[332]. This is in contrast to *C. jejuni* where *PglC* adds a Bac substrate to the lipid carrier[333]. By omitting the PglC glycosyltransferase and utilising the endogenous WecA, a more eukaryotic-like Und-PP-GlcNAc substrate can be produced in *E. coli*[59,325]. The Und-PP-GlcNAc can then act as an anchor for the assembly of a eukaryotic N-linked glycosylation chain within *E. coli*.

The elongation of Und-PP-GlcNAc can be achieved by inserting the different eukaryotic *ALG* enzymes involved in the assembly of LLO (Table 1.4). The *S. cerevisiae* genes *alg13, alg14, alg1, and alg2* were incorporated into the *E. coli* to produce a $Man_3GlcNAc_2$, a common core structure shared in all eukaryotes[334]. The *S. cerevisiae* genes were selected due to the previous soluble expression in *E. coli*. The four *ALG* enzymes were demonstrated to localise to the plasma membrane of the glycoengineered *E. coli*[334]. To demonstrate the activity of the oligosaccharyl transferase the cell surface of the glycoengineered *E. coli* was tested for the presence of non-native mannose[334]. Due to the presence of the native lipid ligase *WaaL* the LPS was confirmed by a lectin-binding assay to contain mannosylated oligosaccharides[334]. The native *E. coli* showed no activity with the lectin-binding assay. The presence of mannose sugars in the LPS inferred that the endogenous flippase *Wzx* must have relocated the $Man_3GlcNAc_2$ into the periplasmic space for transfer by *WaaL*[334].

A recombinant PglB when expressed in *E. coli* can utilise O antigens as a substrate, but only if the initial saccharide residue contains an acetamido group at the C-2 position[335]. A PglB successfully transferred a 16 chain O antigen, with GlcNAc at the first position, to the acceptor protein AcrA[335]. The *pglB* was incorporated into the *E. coli* strain producing $GlcNAc_2Man_3$,

alongside an acceptor glycoprotein[334]. Various acceptor proteins were utilised including the prokaryotic maltose binding protein (MBP) and the therapeutically relevant eukaryotic proteins of bovine RNaseA, human growth hormone (hGHv) and the fragment crystallisable (Fc) domain of human IgG1[334]. The acceptor proteins were targeted to the periplasmic space and contained a modified bacterial glycan acceptor motif (AspGlnAsnAlaThr)[334]. A fluorescent lectin-binding assay demonstrated the addition a mannose glycan chain to the acceptor protein[334]. A later detailed analysis by mass spectrometry revealed that trimannosyl structure $Man_3GlcNAc_2$ was the dominant glycoprofile, however other mannosyl structures ($Man_{1-4}GlcNAc_2$) were also revealed[334]. The presence of a $Man_4GlcNAc_2$ is notable because elongation of $Man_3GlcNAc_2$ is attributed to Alg11[336], and not Alg1 or Alg2. However, both $Man_3GlcNAc_2$ and $Man_4GlcNAc_2$ have been observed in an *S. cerevisiae ΔALG11* mutant strain, suggesting that Alg1 or Alg2 may catalyse $Man_4GlcNAc_2$-PP-Dol production *in vivo*[337]. However, a later study which replicated the initial glycoengineering strategy of expressing the four *alg13, alg14, alg1, and alg2* enzymes with *PglB,* within *E. coli,* only observed $Man_3GlcNAc_2$ and $Man_1GlcNAc_2$ glycoprofiles, and did not detect the $Man_4GlcNAc_2$ profile[338]. This later strategy utilised an Alg1 enzyme with the transmembrane domain deleted and thioredoxin-Alg2 fusion protein[338]. It is possible that these genetic differences between the strains allowed the enzymes to have an increase activity, however the authors did not delve deeper into the mechanisms underlying this phenotype.

Although researchers demonstrated the N-linked glycosylation, the site-occupancy within the recombinant acceptor glycoproteins was relatively low, with only ~1% of proteins displaying an oligosaccharide, with the later strategy observing an increase in site occupancy of 3.2%[334,338]. This increase was suggested to be due to codon optimization of *S. cerevisiae* and *C. jejuni* proteins for expression in *E. coli.* Previously a *PglB* codon optimised for translation in *E. coli,* showed an overall increase in glycosylation efficiency of 77 ± 27% over the native *PglB*[339]*.* In addition to the low site occupancy, protein titre levels of glycosylated proteins reached up to ~50 μg L^{-1}, which considerable amount lower than >5g L^{-1} of glycoprotein titres obtained from highly optimized CHO cells[334].

Despite the OST of *C. jejuni* demonstrating a relaxed specificity toward the oligosaccharide substrate, the *PglB* will only recognise the asparagine in the motif D/EX_1-N-X_2-S/T (X≠P)[238]. This is contrast to eukaryotic OST which transfers the oligosaccharide to an asparagine in N-X-S/T (X≠P). Furthermore, the *PglB* requires an acidic residue in -2 position of the sequon[326], a feature which is unfavourable for a eukaryotic OST[340]. Consequently, eukaryotic proteins will require residue changes around glycosylation acceptor sites in order to achieve efficient glycosylation. This could be problematic with changes in amino acid features possibly effecting the protein attributes or structure. Recently however, 24 different bacterial OSTs were

functionally screened in *E. coli* to examine the oligosaccharide transfer abilities[341]. A *PglB* from *Desulfovibrio gigas* was identified to transfer a fully assembled *C. jejuni* heptasaccharide $GalNAc_5(Glc)Bac$ to a eukaryotic acceptor site N-X-T with an absent acidic residue in -2 position[341]. A future glycoengineering effort in *E. coli* could utilise the *PglB* for a more-eukaryotic sequon oligosaccharide transfer.

Although the absence of an N-linked glycosylation pathway can be seen as beneficial over semi-synthetic pathway of glycoengineered eukaryotic strains, many challenges still remain in the glycoengineering of *E. coli*. Including (i) increasing the glycosylation site occupancy (ii) improving the protein titres and (iii) the need to expand the pathway to produce a therapeutically relevant glycoform. Similarly, to other glycoengineering strategies researchers are employing techniques to enhance glycosylation. For instance, the availability of the GDP-man was increased by deleting a GDP-mannose dehydratase (GMD), which similarly to *Alg1* and *Alg2* uses GDP-man as substrate[334]. Furthermore, a mutant *E. coli* strain with *ΔwaaL* genotype was used to ensure the ligase did not compete with *pglB*[334]*.* While recognisable that generating proteins with eukaryotic N-linked oligosaccharide is currently inadequate, the use of a glycoengineered *E. coli* strain for the synthesis of glycoconjugate vaccines has seen more promising[326,342]. Utilising the functional *Pgl* operon in *E. coli,* engineered glycoproteins with a bacterial glycan can be applied for vaccines and diagnosis. A further interesting application for glycoengineered *E. coli* is of a cell-free system for the production of N-linked glycoproteins[343].

1.29 Cell Free Protein Synthesis

The inherent focus of a cell is in its survival and fitness; therefore, cellular processes are directed to ensure cell viability. The introduction of a heterologous expression system exerts an addition burden to the cell, which typically leads to slower growth, cellular complications, proteolytic degradation and ultimately lower protein yields[344]. Therefore, by eliminating the need of a living-system all resources can be focused onto the orthogonal expression system. This is the aim of *in vitro* expression through Cell Free Protein Synthesis (CFPS) systems.

The use of CFPS systems has been around for over half century[345], initially utilised for the seminal research to decipher the codon usage for amino acid incorporation[346]. It has been especially utilised in protein structural studies for the translation of what are deemed difficult to express proteins. For instance, due to the hydrophobic nature of membrane proteins it is challenging to obtain yields to enable structural studies. The use of CFPS systems, where environmental conditions can be controlled, has facilitated the expression of adequate yields of membrane proteins for structural analysis[347,348].

As CFPS facilitates alterations in hydrophobic environment, this enables changes for optimal protein folding[349]. As CFPS negates a cell culture it has many benefits over traditional *in vivo* protein synthesis, mostly associated with the high-throughput translation of recombinant proteins. CFPS also offers advantages in translation of proteins that would otherwise be toxic for *in vivo* systems, preventing the inhibition of expression, misfolding or formation of inclusion bodies[350]. The incorporation of unnatural/noncanonical amino acids and modified amino acids is also made simpler using a CFPS[351].

As the name implies, CFPS is a recombinant protein expression system that occurs outside of a cellular chassis. Instead a CFPS is made up of a cellular crude extract (or cell lysate) supplemented with components to aid and enhance gene expression from a recombinant genetic template (DNA or mRNA)[352]. The cell lysate is obtained by bursting open a cell normally through mechanical or physical disruption. The essential translational machinery such as ribosomes, tRNAs, elongation and termination factors is then separated from cellular debris, to form the cell lysate. To increase the efficiency of translation the lysate is supplemented with amino acids, cofactors, essential metal ions and energy source such as ATP. The CFPS is the combination of the cell lysate with supplementary components.

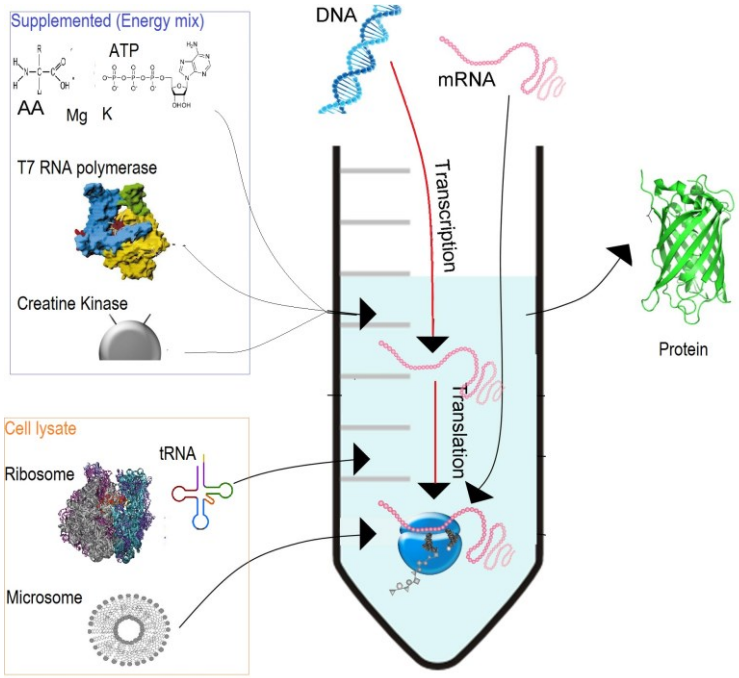


Figure 1.5 . Schematic overview of a Cell-free protein synthesis system.
The cell lysate of a species (shown in the orange box) containing translational machinery and microsomes, supplemented with amino acids, cofactors, essential metal ions and energy source such as ATP (shown in the purple box) to create a CFPS. Translation commences upon the addition of DNA (PCR product or plasmid) in a coupled system (indicated by the red lined arrow) or by adding separately transcribed mRNA in a linked reaction (indicated by the black lined arrow). This figure has been adapted from Rosenblum *et al.*[353] (copyright license can be found in Appendix Section 11)

Protein expression from a CFPS can be carried out in either 'linked' or 'coupled' reaction. In the linked reaction, a DNA template is first transcribed in a separate reaction followed by the addition of the transcribed mRNA to CFPS mixture for protein translation[352]. In a coupled reaction both transcription and translation steps take place within CFPS, i.e. within the same tube, this is known as TX-TL. CFPS allows for the rapid production of a protein as it removes the need for the cloning steps and expression optimisation required for *in vivo* production[353].

CFPS systems have been reported from a number of different cellular crude extracts, but predominantly from *E. coli,* wheat germ extract (WGE) and rabbit reticulocytes (RRL). There has recently been a surge in new CFPS systems from different sources such as Insect cells (*Sf21*)[354], CHO cells[355], *Saccharomyces cerevisiae*[356], Xenopus eggs[357], *Trypanosoma brucei*[358], and many others. This increase in the number of available CFPS systems is possibly linked to emergence of the Synthetic Biology field and shows the growing importance of this translational system[350].

The field of Synthetic Biology aims to introduce synthetic gene networks in order to re-engineer the cell for a specific task[359]. The discipline has led to engineering of unicellular organisms to produce chemicals, biofuels and in the creation of whole-cell biosensors[359]. The use of whole-cell biosensors in a real world situation, has been restricted by concerns of biosafety and by bioethical issues, due to release of genetically modified organisms into the environment. The introduction of a synthetic pathway in a cell-free system aims to eliminate these concerns[360]. A foundation of the synthetic biology discipline is in the prototyping of DNA elements, otherwise known as DNA parts. The use of CFPS, which aims to mirror the *in vivo* host, allows for the rapid and simple characterisation of DNA parts for rapid prototyping. This prototyping can be further sped up by using PCR products, therefore completely negating any cellular cloning[361].

Despite their advantages over *in vivo* systems, the high cost associated with energy sources and supplements and the lower titre levels of CFPS often limits its use[362]. However, the continual optimisation of the CFPS reaction aims to address these issues and make them more economically viable. Protein yields can be increased by prolonging the translation period in CFPS; this can be achieved by ensuring a readily available energy source and limiting by-products that inhibit translation[363]. For instance, the use of continuous synthesis reaction, where energy molecules such as ATP are regenerated in the system, and performing a CFPS reaction in dialysis chambers aims to extend the translation reaction and produce more protein[363]. For example, a CFPS system based on *S. cerevisiae* produced 0.38 ± 0.0022 μg mL^{-1} [356] of recombinant luciferase when first developed in 2013 but with improvements in cell

lysate preparation[356], the energy availability[364,362] and increasing genetic template stability[365] the CFPS can now yields 7 ± 2 μg mL^{-1} of recombinant luciferase[365].

Although *E. coli* CFPS remains a popular choice, similarly to an *in vivo E. coli* expression system it is unable to produce certain PTMs, such as N-linked glycosylation. Although many eukaryotic based CFPS systems such as *S. cerevisiae* do not perform N-linked glycosylation, the synthesis of a glycoprotein has been demonstrated in the lumen of microsomes which are within the extract[366,367]. Microsomes are artefactual vesicles formed from the endoplasmic reticulum when cells are disrupted. The microsomes can either be supplemented into the CFPS system, such as in the case of the commercially available RRL supplemented with microsomal fractions prepared from canine pancreas[368], or the CFPS systems can also be created with the microsomes retained[354]. Recently a CFPS based on CHO cells combined the two by supplementing microsomes to a CFPS with retained microsomes[355].

Although N-linked glycosylation can occur in a CFPS, site-occupancy can be problem. In a CFPS based on CHO less than 10% of recombinant glycoprotein EPO contained an oligosaccharide, although translocation of targeted secretory proteins was very efficient[355]. Also as glycosylation occurs within the lumen of a microsome, which is typically ER derived, the glycan chain is likely to be of high-mannose structure. Although this is difficult to determine as of yet no glycan analysis has been performed by researchers on a glycoprotein produced in a CFPS.

1.30 Biopharmaceutical production in CFPS systems

Although currently no biopharmaceutical has been licensed from a CFPS, interest is gaining in using this as an expression host. CFPS based on prokaryotic *E. coli* cell lysate has been demonstrated to translate and assemble an IgG, which as earlier discussed is an important class of biopharmaceuticals[369,370]. An IgG protein is a multi-domain Y-shaped protein consisting of two heavy (HC) and two light chains (LC). The HC and LC are bound together through disulphide bonds, a chemical bond which *E. coli* finds difficult to form. Engineering the CFPS environment to increase the redox potential can increase IgG formation. Also, the use of an engineered *E. coli* strain with improved disulphide bond formation, similar to the strains mention above, can increase the yield of IgG[371]. For the *in vivo* synthesis of IgG, the ratio of LC and HC must be correct to ensure efficient assembly, too much of one chain can lead to misfolding and a decrease in yield. In a CFPS control over chain ratio is not needed, allowing for efficient IgG formation.

E. coli CFPS has also been used to create Antibody drug-conjugates (ADCs). An ADC is where a cytotoxin, normally a small molecule, is attached to a mAb. The binding specificity of

a mAb is then utilised to deliver a toxic payload to a targeted location, such as a cancerous cell. The cytotoxin, or warhead as it also known, is normally attached to lysine or reduced thiol residues on a mAb[372]. However, as mAbs contain a number of these residues there is potential for heterogeneity in warhead numbers. For example, the FDA approved ADC Kadcyla®, is comprised of the conjugation of an average of 3.4 warheads per antibody[372].

Directing the attachment of drugs at defined sites aims to reduce site occupancy heterogeneity. One method is the use of engineered cysteines, as native cysteines in a mAb are already paired the introduction of new cysteines via site-directed mutagenesis which can be modified and used as conjugation sites[373]. However engineered cysteines may interfere with the mAb structure (and therefore binding), disrupt correct formation of the native disulphide bonds, or form disulphide bonds with free cysteines in other proteins[372].

An alternative method to engineered residues is in the use of site-specific incorporation of unnatural amino acids (uAAs)[351,353]. An uAA is a non-proteinogenic amino acid that either occur naturally or is chemically synthesized[374]. There are limitless possibilities in chemistry or structure of uAAs, therefore uAAs can extensively expand the proteome[375]. A uAA can be used as a site for drug conjugation, and as they can be designed to have a neutral effect they do not suffer from the same disadvantages of engineered cysteines. One major drawback in the use of uAAs is in the site specific incorporation into a polypeptide during *in vivo* production[376]. A cell-free protein synthesis (CFPS) system has been used successfully to overcome this problem, and an ADC has been generated.

1.31 Summary and Project Overview

When selecting an expression host for synthesis of a N-linked glycoprotein it is especially important to consider how the protein's attributes affects the function, and if the expression host has the capacity to execute the desired post-translational modifications that affect the product quality. As has been apparent through this chapter, N-linked glycosylation can be necessary for efficient functioning and drug efficacy. Although N-linked glycosylation may not always be required, for instance an a-glycosylated biopharmaceutical ADC functions effectively and hence is manufactured in a bacterial CFPS system. However, in most cases N-linked glycosylation is strongly associated with efficacy and hence, the reason why many mAbs have traditionally been synthesised in mammalian expression systems.

Differences in glycan structures can cause alterations in the drug efficacy, mode of action and half-life of biopharmaceutical, and although one glycoform can have a positive effect for one biopharmaceutical it maybe undesirable for another. For instance, as discussed earlier in the chapter some glycoforms can enhance half-life of a mAb. This attribute is beneficial in some

therapeutic application as it requires a smaller dose to achieve a therapeutically relevant serum concentration. However, when using mAb for medical imaging, the mAb can be radiolabelled, therefore a rapid clearance i.e. shorter half-life is a more desirable attribute[377,378].

To aid in generation of a specific glycoform, manufacturers tightly regulate the bioprocessing parameters or in recent years have begun using glycoengineered expressions hosts. Whereas tight regulation aims to generate specific glycoforms, glycoengineered expressions hosts are designed in an aim to synthesis one specific glycoform. Although this strategy limits the possible glycoforms it does not mean the most therapeutically relevant glycoform is synthesised. Differences in glycoform can reclassify a biopharmaceutical as a biobetter adding a financial incentive for a manufacturer to understand how different glycoforms affect drug function. Currently, to change a glycoform this would require a change in the bioprocess, by either cloning into a new expression host or altering expression conditions. These changes can be time consuming, do not offer certainty in the end glycoform, and do not make for certain the glycoform is the therapeutic optimum.

With this in mind, this project aims to develop a proof-of-principle system which will allow for the generation of a user defined oligosaccharide. The project aims to take the principles of glycoengineering and CFPS to develop this system

1.32 Project aims

The aim of this work was to tackle some of the challenges in bioprocessing to produce homogenous therapeutically relevant oligosaccharide. A CFPS allows for the quick translation of proteins and presence of microsomes will allow for the transfer of a N-linked glycan chain. As discussed earlier in the chapter only few expression systems can produce therapeutically relevant glycoforms. These expression hosts are often mammalian, which are often more costly than other expression hosts. So, disadvantages associated with mammalian expression systems would be replicated in a CFPS system based on these hosts. Secondly, as glycosylation occurs in the microsomes of a CFPS and no Golgi is present (or translocation does not occur) only the therapeutically irrelevant $Man_{8-9}GlcNAc_2$ would be produced.

A glycoengineering approach is opening the door to use of unconventional expression hosts for biopharmaceutical production, with the humanized *Pichia pastoris* being the most advanced example. It has been demonstrated that removal of homologous α1.6 mannosyltransferase *OCH1* and addition of heterologous mammalian Golgi α1,2 mannosidases allows the synthesis of $Man_5GlcNAc_2$. By localising the α1,2 mannosidase to ER this ensured that glycoproteins exiting the ER have a $Man_5GlcNAc_2$ profile. Noting this, a

CFPS based on a glycoengineered *Pichia pastoris* with retained/supplemented microsomes should produce glycoproteins with a $Man_5GlcNAc_2$.

The high-mannose $Man_5GlcNAc_2$ glycoprofile is not therapeutically relevant and would require further processing by Golgi-based glycotransferases and glycosidases. The development of an *in vitro* metabolic pathway which contains these enzymes would allow for the completion of therapeutic relevant glycan chain. To address the issue of user defined glycoprofile, the order and multivariate metabolic pathway could be created to produce many different glycoforms.

To summarise the project contains three main aims;

Generation of glycoengineered *Pichia pastoris* which produces $Man_5GlcNAc_2$ glycoprofile.

Generation of a CFPS based on glycoengineered *Pichia pastoris*.

Generation of an *in vitro* metabolic pathway which can process a $Man_5GlcNAc_2$ glycoprofile.

1.33 Thesis overview

The generation of glycoengineered *P. pastoris* in Chapter 3 involved the cloning of a number of plasmids, initial specifications for the plasmids were drawn up from previous studies. The plasmids generated contain a mammalian enzyme and a selective marker between genetic branches of complementary yeast genome sequence. After generation of knock-out strain experiments proceeded to define the glycoprofile of the species.

The generation of *in vitro* metabolic pathway in Chapter 4 involved the heterologous expression of 3 enzymes involved in oligosaccharide processing. The chapter proceeded with optimisation of expression protocols. After successful expression the activity of the enzymes was tested.

The generation of a cell-free protein synthesis (CFPS) system based on *Pichia pastoris* cell lysate was attempted in Chapter 5. The chapter proceeded with the cloning of plasmids with different DNA elements. The DNA elements are known to increase RNA stability and/or enhance translation. The *P. pastoris* CFPS was generated and the activity was tested with the earlier generated plasmids.

Chapter 2. Materials and Methods

2.1 Strains and plasmids

Table 2.1 *E. coli* strains utilised in this project

Strain	Description (Genotype)	Source
E coli One Shot® TOP10	F- *mcr*A Δ*(mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Life Technologies, CA, USA
E.coli DH5α™	F- Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A*sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-	Life Technologies, CA, USA
E.coli BL21 (DE3)	*fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5*	NEB, MA, USA
E. coli SHuffle® T7	*F´ lac, pro, lacI^q / Δ(ara-leu)7697 araD139 fhuA2 lacZ::T7 gene1 Δ(phoA)PvuII phoR ahpC* galE (or U) galK λatt::pNEB3-r1-cDsbC (Spec^R, lacI^q) ΔtrxB rpsL150(Str^R) Δgor Δ(malF)3*	NEB, MA, USA

Table 2.2 Yeast strains utilised in this project

Strain	Description (Genotype)	Source
Pichia pastoris X33	Wild type	Life Technologies, CA, USA
Pichia pastoris GS115	*his4*	Life Technologies, CA, USA
Pichia pastoris Δku70	*his4 Δku70*	
Saccharomyces cerevisiae BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	ATCC®, VA, USA

Table 2.3 Plasmids utilised in this project

Plasmid Name	Description	Source
pGEM®-T Easy Vector System	pGEM®-T Easy Vector is linearized vector with a single 3´-terminal thymidine at both ends providing a compatible overhang for PCR products. T7 and SP6 RNA polymerase promoters; a phage f1 region origin (ori); lac operon sequence	Promega, WI, USA

pET Expression System 28 (a)- Novagen	Expression vector based on the N-terminal His•Tag®/thrombin/T7•Tag®configuration plus an optional C-terminal His•Tag sequence. T7 RNA polymerase promoter and terminator; a phage f1 region ori	Merck Millipore, Germany
pPICzα A	Expression vector based on AOX1 promoter α mating factor; MCS; AOX1 Terminator; Sh ble; pUC ori;	Life Technologies, CA, USA
pJAN–s1	Expression vector based on AOX1 promoter; α mating factor; MCS; α mating factor; pMB ori;	BioGrammatics, Inc. CA, USA

2.2 Media

2.2.1 Miller Lysogeny broth (LB) media

LB Media (Merck Millipore, Germany) [1% (w/v) Peptone au casein 0.5% (w/v) Yeast Extract, 1% (w/v) NaCl]. For agar plates 1.5% (w/v) agar (Merck Millipore, Germany) included.

2.2.2 2XYT media

2XYT Media (Sigma-Aldrich Corp. MO, USA) [1.6% (w/v) Peptone au casein 1% (w/v) Yeast Extract, 0.5% (w/v) NaCl]

2.2.3 Terrific broth (TB) media

2.36% (w/v) Yeast Extract (Merck Millipore, Germany), 1.18% (w/v) Tryptone (Merck Millipore, Germany), 9.4g/L K_2HPO_4(Sigma-Aldrich Corp. MO, USA), 2.2g/L KH_2PO_4(Sigma-Aldrich Corp. MO, USA), 0.004% (v/v) Glycerol (Merck Millipore, Germany)

2.2.4 Yeast extract peptone medium (YPD)

1% (w/v) Yeast Extract (Merck Millipore, Germany), 2% (w/v) Peptone au casein (Merck Millipore, Germany), 2% (w/v) Glucose (Merck Millipore, Germany). For agar plates 2% (w/v) agar (Merck Millipore, Germany) included.

2.2.5 Buffered glycerol-complex medium (BMGY) or buffered methanol-complex medium (BMMY)

1% (w/v) Yeast Extract (Merck Millipore, Germany), 2% (w/v) Peptone au casein (Merck Millipore, Germany), 100mM potassium phosphate buffer, pH6.0, 1.34% (w/v) Yeast Nitrogen Base (YNB) (Sigma-Aldrich Corp. MO, USA), 0.02% (w/v) d-Biotin (Sigma-Aldrich Corp. MO,

USA), 1% (v/v) glycerol (Merck Millipore, Germany) or 0.5% (v/v) methanol (Merck Millipore, Germany).

2.2.3.1 Potassium phosphate buffer, pH6.0

132 mL of 1 M K_2HPO_4 (Sigma-Aldrich Corp. MO, USA), 868 mL of 1 M KH_2PO_4 (Sigma-Aldrich Corp. MO, USA) and confirm that the pH = 6.0 ± 0.1 (if the pH needs to be adjusted, use phosphoric acid or KOH).

2.3 Molecular biology methods

2.3.1 Nucleic acid agarose gel electrophoresis

Agarose gel electrophoresis was used to separate DNA and RNA fragments based on the molecular size. Typically, a 1% (w/v) agarose was used, with agarose dissolved in 1x TAE buffer (50X TAE Buffer; 242 g L^{-1} Tris base, 54.1 mL L^{-1} glacial acetic acid, 100mL L^{-1} 0.5 M EDTA, pH 8.0).The nucleic acid was stained with either SYBR® Safe (Life Technologies, CA, USA) or GelRed™ (Biotium, CA, USA). The dyes were added to the agarose gel at the manufacturer's recommended concentration. DNA samples were diluted in a 5X Loading Buffer (30% [v/v] glycerol, 0.25% [w/v] bromophenol blue) and loaded onto the gel. The molecular weight marker Hyperladder™ 1kb (Bioline, MA, USA) or Hyperladder™ 100bp (Bioline, MA, USA) were added as a standard(s). Agarose gel electrophoresis was typically performed in Bio-Rad Wide Mini-Sub® Cell GT Tanks using power supplied from a Bio-Rad PowerPac™ Basic set at 120V (Bio-Rad, CA, USA). The resolved nucleic acid fragments were visualised using a UV transilluminator (Bio-Rad, CA, USA).

2.3.2 Nucleic Acid isolation methods

2.3.2.1 DNA agarose gel and PCR isolation

The Wizard® SV Gel and PCR Clean-Up System (Promega, WI, USA) was utilised for the extraction of DNA fragments agarose gels in TAE. PCR products were purified directly from an amplification reaction using Wizard® SV Gel and PCR Clean-Up System (Promega, WI, USA) following the manufacturer's instructions were followed for the DNA clean-up reactions.

2.3.2.2 Plasmid DNA isolation

Plasmid DNA was isolated from *E. coli* using peqGOLD Plasmid Miniprep Kit I (PEQlab, Erlangen, Germany) according to the manufacturer's protocols. If large concentrations of DNA were required a QIAGEN Plasmid Midi Kit (QIAGEN, Limburg, Netherlands) was utilised.

2.3.2.3 Isolation of yeast genomic DNA

Genomic DNA was isolated from a 5 mL overnight liquid YPD culture of *Pichia pastoris,* supplemented with antibiotics if required. The isolation method used was developed by Hoffman *et al.*[379]

2.3.2.4 Isolation of RNA

Total RNA was isolated from *P. pastoris* using RiboPure™ RNA Purification Kit, yeast (Thermo Fisher, MA, USA) according to the manufacturer's instructions.

2.3.2.5 cDNA preparation

cDNA was prepared using the High-Capacity cDNA Archive Kit (Life Technologies, CA, USA) according to the manufacturer's instructions. 1 µg RNA was used in a total reaction volume of 20 µL. The cDNA can be used as a template for PCR (2.3.5), the preparation of cDNA and use in a PCR reaction is known as reverse transcriptase PCR (RT-PCR).

2.3.3 Restriction enzymes reaction

Restriction enzymes were purchased from the same supplier (NEB, MA, USA) and digestions were performed according to the manufacturer's instruction manual, typically at 37°C for 1-3 hours. Digested products were purified on agarose gels using a Wizard® SV Gel and PCR Clean-Up System (Promega, WI, USA).

2.3.4 Ligation

The Ligation of DNA is performed using T4 DNA Ligase (NEB, MA, USA) according to the manufacturer's instructions. The molar ratio of plasmid DNA to DNA due to be inserted was typically 1:3, respectively, for cohesive end ligation. The ligation mixture was incubated for 1 hour at room temperature or overnight at 16°C.

2.3.5 Polymerase Chain Reaction

For high fidelity Polymerase Chain Reaction (PCR), Phusion Hot Start II (Thermo Fisher Scientific, MA, USA) DNA polymerase was utilised and reactions were set-up according to manufacturer's instructions. PCR was assembled in sterile 0.2 ml PCR tubes (Thermo Fisher Scientific, MA, USA) and performed in an Eppendorf Mastercycle Gradient (Eppendorf, Hamburg, Germany) machine. Oligonucleotides were synthesised by Sigma Aldrich (Sigma Aldrich, MI, USA) and a full list of primers used can be seen in the Appendix Section 1. Annealing temperatures were determined by using an online calculator (http://tmcalculator.neb.com/#!/). Reactions were normally performed for 35 cycles. For

analytical PCR such as colony PCR, REDTaq ReadyMix (Sigma Aldrich, MI, USA) was utilised and reactions were set-up according to manufacturer's instructions.

2.3.6 DNA assembly methods

Construction of vectors used the Gibson assembly method developed by Gibson *et al.*[380]. Briefly, 5x isothermal reaction buffer was prepared containing 25% PEG-8000, 500 mM Tris-HCL pH7.5, 50 mM MgCl2, 50 mM DTT, 1mM of each dNTPs and 5 mM NAD. 320 µL of the 5x isothermal reaction buffer was used to create the assembly master mix, with the addition of 0.64 µL 10U μL^{-1} of Phusion DNA polymerase (NEB, MA, USA) and 160 µL of 40U μL^{-1} 2U μL^{-1} Taq DNA Ligase (NEB, MA, USA) the final volume was adjusted to 1.2 mL with H_2O. Primers were designed with a 30-50 bp complementary sequence to the adjoining DNA fragment. Once all the relevant fragments had been amplified with the added linker regions, equimolar concentrations of up to 100 ng of DNA from each fragment were added to 15 µL master mix and incubated for one hour at 50°C before being transformed into chemically competent *E. coli.*

2.4 Microbiology techniques

2.4.1 Measuring optical density

The optical density of a microbiology sample measured at a wavelength of 600 nm (OD_{600}), using a Digital Spectrophotometer. Before measuring samples, the Digital Spectrophotometer is calibrated to zero ('blanked') with a sterile liquid containing no biomass.

2.4.2 Transformation and preparation of chemically competent *E. coli*

Chemically TOP10 Competent *E. coli* cells were either purchased (Life Technologies, CA, USA) or manufactured in house using protocol developed by Dr Martin J. Blaser's research group (http://www.med.nyu.edu/medicine/labs/blaserlab/Protocols). The manufacturer's transformation protocol or Blaser transformation protocol was utilised depending on the source of the chemically-competent *E. coli* cells. After transformation 1 mL of S.O.C media (Life Technologies, CA, USA) was added to the microcentrifuge tube and cells were recovered at 37°C with shaking at 250 rpm for 1 hour. Typically, 100 µL of transformation mixture was plated on LB agar plates with suitable antibiotics. The remaining volume of cells within the microcentrifuge was centrifuged, supernatant removed and pellet resuspended in 100 µL LB, which was spread on appropriate LB agar containing suitable antibiotic. Petri dishes containing transformed cells were incubated overnight at 37°C.

The SHuffle® T7 Competent *E. coli* (NEB, MA, USA) and BL21(DE3) Competen*t E. coli* (NEB, MA, USA) were used for recombinant protein expression. The transformation of heterologous plasmid DNA was performed according to the manufacturer's instructions.

2.4.3 Approximating the ATP concentration

The BacTiter-Glo™ Microbial Cell Viability Assay provides a method for quantitation of the ATP present in a cell population. The cell population was normalised to an OD_{600} of 1.0 and 25 µL was added to a well within an opaque 96-well plate. The same volume of BacTiter-Glo™ Reagent was added, and sample was place on orbital shaker for 1 minute at 100 rpm. The plate was then placed in a POLARstar Omega plate reader (BMG labtech, Ortenberg, Germany) luminescence was measured after plate had been incubated for a total of 5 minutes. The test samples were done in biological triplicates and a standard curve was used to determine ATP concentration.

2.4.4 Flow Cytometry Analysis

Pichia pastoris were grown for 48 hours in 500 mL of YPD media in 2 L baffled shake flasks as previously described. At regular intervals, a 1 mL sample of culture was collected and placed in microcentrifuge tube. The sample was pelleted by centrifugation at 5000g for 5 minutes and the pellet was washed in 1x TBS (0.605% Tris, 0.876% NaCl, pH adjusted with 1 M HCl). Samples were washed for a second time before being resuspended in a final volume of 10 mL of 1x TBS buffer to create a 1 in 10 dilution. Samples were stained with LIVE/DEAD® FungalLight™ yeast viability kit (Life Technologies, CA, USA) according to manufacturer's instructions. Cells were measured using a FACscan flow cytometer (Becton Dickinson, Oxford, UK) with settings Excitation/Emission 480/500 nm (SYTO® 9 stain) 490/635 nm (propidium iodide). Data was acquired using CellQuest software (Becton Dickinson, Oxford, UK) with samples measured on a high flow rate for 30 seconds. Data was analysed using Cylogic software (CyFlo Ltd, Turku, Finland).

2.4.5 Transformation and preparation of *P. pastoris* for electroporation

2.4.5.1 Preparation of DNA for Transformation

To achieve site directed homologous recombination in *P. pastoris* the vector was linearised within the region of homology. After confirmation that the plasmid had been digested, by analysis on agarose gel, *P. pastoris* was transformed with 5-10 µg of the linearised DNA.

2.4.5.2 Electroporation Protocol

The electroporation protocol was modified from the Pichia manual (Life Technologies, CA, USA). A 5 mL culture of *Pichia pastoris* in YPD was grown overnight at 30°C with shaking at 250 rpm. The 0.5 mL of overnight culture was used to inoculate 200 mL YPD. The culture was incubated until OD_{600} 1.3-1.5 was achieved. Cells were centrifuged at 1500x g for 5 minutes at 4°C, and washed with 200 mL ice-cold autoclaved deionised water, and centrifuged again with the same settings. Cells were washed with 100 mL water and finally with 10 mL of ice-cold 1 M sorbitol (Sigma Aldrich MI, USA), before being resuspended in 200 µL of 1 M sorbitol. An 80 µL aliquot of cells was incubated with 5-10 µL of linearised DNA on ice for 5 minutes in a 0.2 cm electroporation cuvette (Bio-Rad Laboratories Inc., CA, USA). The cells were pulsed at 2000V, 25 µF, 200 Ω, for approximately 5 milliseconds, using the GenePulser electroporator. After the pulse 1 mL of ice-cold 1M sorbitol was immediately added to the electroporation cuvette and the contents transferred to a sterile 15 mL conical tube. The transformation was incubated at 30°C without shaking for 1-2 hours. After suitable incubation the transformation was spread onto a YPD agar supplemented with an appropriate selective pressure (e.g. antibiotics). The plates were incubated at 30°C for 3-5 days or until clones appeared.

2.5 Heterologous Protein Expression

2.5.1 Heterologous protein expression from recombinant *E. coli* strains

The SHuffle® T7 Competent *E. coli* (NEB, MA, USA) and BL21(DE3) Competent *E. coli* (NEB, MA, USA) were utilised as heterologous protein expression hosts. After successful transformation clones was selected, and resuspend in 5 mL liquid LB culture with antibiotic and grown overnight at 30°C for SHuffle® T7 Competent *E. coli* or 37°C for BL21(DE3) Competent *E. coli*. Following day 1 mL of overnight culture was used to inoculate a liquid culture contain with a baffled flask and incubated at 30°C for SHuffle® T7 Competent *E. coli* or 37°C for BL21(DE3) Competent *E. coli* with shaking at 250 rpm until OD_{600} reaches 0.4–0.8. Expression was then induced by the addition of the chemical inducer IPTG (Sigma Aldrich, MI, USA). The final concentration of inducer, temperature of expression, media used and length induction (translation) time, varies and actual conditions used will be documented in results. All expression was performed with shaking in a temperature controlled environment.

2.5.2 Heterologous protein expression from recombinant *P. pastoris*

The expression protocol was modified from the Pichia manual (Life Technologies, CA, USA).

A single colony was selected and grown overnight in 5 mL of YPD at 30°C with shaking at 250 rpm. 1 mL of the overnight culture was used to inoculate 25 mL BMGY medium in a 250 mL baffled flask. The flask was placed at 30°C with shaking at 250 rpm, and left overnight. The following day cells were harvested by centrifuging at 1,500x g for 5 minutes at room temperature. The supernatant was removed and the cell pellet was resuspended in 25 mL of BMMY medium to induce expression. The flask was placed at 16°C or 30°C with shaking at 250 rpm, and incubated for up to 5 days at 250 rpm, with 100% methanol added to a final concentration of 0.5% (v/v) every 24 hours.

2.6 Protein Preparation from *E. coli* lysates

2.6.1 Protein preparation buffers

Lysis buffer; 50 mM NaH_2PO_4, 300 mM NaCl and 10 mM imidazole. The pH was adjusted to 8.0 using NaOH.

Wash buffer; 50 mM NaH_2PO_4, 300 mM NaCl and 20 mM imidazole. The pH was adjusted to 8.0 using NaOH.

Elution buffer; 50 mM NaH_2PO_4, 300 mM NaCl and 250 mM imidazole. The pH was adjusted to 8.0 using NaOH.

All buffers were prepared with filter sterilized deionised water and chilled for use.

2.6.2 *E. coli* cell disruption

Once the duration of expression had finished, *E. coli* cells were pelleted by centrifuging at 4000x g for 5 minutes at 4°C. Cell pellets were either frozen at this stage at -80°C or the cell disruption procedure was initiated. Cell pellets are placed on ice and the pellet was resuspended in lysis buffer (approximately 10 mL). *E. coli* cells were disrupted by sonication, with a microtip and six 20 second bursts at 200–300 W with a 20 second cooling period between each burst. After sonication cell debris was removed by centrifugation at 16,000x g for 30 minutes and at 4°C, after which supernatant was decanted into a sterile conical tube.

2.6.3 Protein purification of 6xHis-tagged proteins

A Ni-NTA Slurry (QIAGEN, Venlo, Netherlands) was used for purification of 6xHis-tagged proteins. 1 mL of Ni-NTA slurry was added to every 4 mL of lysed *E. coli* protein sample from Section 2.6.2. After addition, the sample was placed at 4°C for 1 hour with gentle agitation. The lysate–Ni-NTA mixture was loaded into a prepared column, and lysate was allowed to flow through the column and the flow-through was collected. The Ni-NTA was washed with 10

mL of chilled wash buffer and flow-through was collected. The protein was eluted from the Ni-NTA slurry with 4x 0.5 mL elution buffer, each eluate fraction was collected in a separate tube. Each collected flow-through samples and eluted samples were analysed by SDS-PAGE.

2.7 Protein analytical methods

2.7.1 Protein Gel Electrophoresis

2.7.1.1 Gel preparation

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to separate proteins based on molecular size. Typically, a 12% (w/v) SDS-PAGE gel was set between two glass plates with a 50 mm gap with a total volume of 5 mL of SDS-PAGE mixture containing 2 mL acrylamide mix (30% v/v acrylamide, 0.8% v/v bis-acrylamide), 1.3 mL Tris-HCl (1.5 M, pH8.8), 50 µL 10% (w/v) SDS, 50 µL 10% (w/v) Ammonium persulphate, 2 µL Tetramethylethylenediamine (TEMED) and 1.5 mL distilled water. The SDS-PAGE mixture was poured and after polymerization of the gel (approximately 20 minutes), a 1 mL stacking gel was added. This consisted of 170 µL acrylamide mix (30% v/v acrylamide, 0.8% v/v bisacrylamide), 130 µL Tris-HCl (1 M, pH 6.8), 10 µL 10% (w/v) SDS, 10 µL 10% (w/v) Ammonium persulphate, 1 µL TEMED and 680 µL distilled water. A comb was inserted between the glass plates and the stacking gel was allowed to polymerise before the gel could be used.

2.7.1.2 Precipitation of protein sample

Trichloroacetic acid (TCA) protein precipitation was performed on some samples for protein analysis. 1 volume of 100% TCA solution was added to 4 volumes of protein sample and incubated on ice for 10 minutes. The sample was centrifuged at 16,000x g for 5 mins and supernatant removed. Protein pellet was resuspended in 500 µL of ice cold acetone and samples were centrifuged at 16,000x g for 5 minutes, this step was repeated a further 2 times. The pellet was placed at 95°C to remove any excess acetone, before samples were prepared for gel electrophoresis.

2.7.1.3 Protein sample preparation

Protein samples were boiled in a heating block set at 99.9°C with 1x SDS-PAGE loading buffer for 10 minutes. A 5X SDS-PAGE loading buffer contained 2.5% (w/v) 2M Tris-HCl pH 6.8, 0.2% (w/v) SDS, 0.6% (v/v) Glycerol, 0.05% (v/v) β-mercaptoethanol and 0.416% (w/v) bromophenol blue. After boiling samples were briefly spun in a microcentrifuge to remove precipitate.

2.7.1.4 Protein gel electrophoresis

SDS-PAGE gels were placed within a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (Bio-Rad, CA, USA) containing 1x Running Buffer. A 1X Running Buffer consisted of 3%Tris-base, 14.4% Glycine, 1% SDS. Protein samples were loaded on the SDS-page alongside a protein molecular marker, typically PageRuler™ Prestained Protein ladder (Thermo Fisher Scientific, MA, USA). Electrophoresis was run at 35 mA for approximately one hour.

2.7.2 Protein Gel Visualisation

After completion of gel electrophoresis, protein samples were visualised by staining SDS-PAGE gels with SimplyBlue™ SafeStain (Thermo Fisher Scientific MA, USA) following the manufacturer's instructions. The resolved protein fragments were visualised using a UV transilluminator (Bio-Rad, CA, USA) with white light setting switched on.

2.7.3 Protein immunoblotting

2.7.3.1 Protein transfer by dot-blotting

Protein samples where pipetted onto a polyvinylidene difluoride (PVDF) membrane and left to dry at 37°C until excess liquid evaporated away. Alternatively, a Bio-Dot® and Bio-Dot SF Microfiltration Apparatus (Bio-Rad, CA, USA) was utilised. This functions by creating a vacuum drawing protein sample onto the PVDF membrane.

2.7.3.2 Protein transfer by electroblotting

Protein samples which were immobilised on an SDS-PAGE gel were transferred to PVDF membrane (Bio-Rad, CA, USA), using a Trans-Blot® SD semi-dry transfer cell (Bio-Rad, CA, USA) following the manufacturer's instructions. Transfer typically took 30 minutes at 25 volts.

2.7.3.3 Western Blotting

A WesternBreeze® Chromogenic kit (Thermo Fisher Scientific, MA, USA) was utilised to detect proteins that have been immobilized on PVDF membranes, following the manufacturer's instructions. A list of primary antibodies used is shown in

Table 2.4 Table of utilised primary antibodies, the table also indicated what secondary antibody was required, therefore which WesternBreeze® Chromogenic kit was utilised.

Table 2.4 Table of utilised primary antibodies

Primary Antibody target	Catalogue Number	Manufacturer	Secondary Antibody
Man-II	MAN2C1 sc-377132	Santa Cruz Biotechnology, TX, USA	mouse
6X His tag (c-terminus)	46-0693	Thermo Fisher Scientific, MA, USA	mouse
6X His tag	652502	Biolegend, CA, USA	mouse
Flag-tag	T503-2	Signalway Antibody, MD, USA	rabbit

2.7.3.4 Measuring fluorescent proteins

Samples were placed into wells of 96-well black walled plate with clear bottoms (Greiner Bio-One, Kremsmünster, Austria). Fluorescence was measured using Synergy HT Microplate Reader (Biotek Instruments, Inc, VT, USA). Data was acquired using Gen5 Data Analysis Software (Biotek Instruments, Inc, VT, USA).

2.8 Carbohydrate analytical methods

2.8.1 Carbohydrate isolation

A N-linked oligosaccharide chain can be removed from a peptide backbone using the enzyme PNGaseF which cleaves between the innermost GlcNAc and asparagine residue. PNGaseF was purchased (NEB, MA, USA) and the manufacturer's protocol was followed. Where needed, after the cleavage reaction the carbohydrate chain was isolated from the polypeptide using a porous graphite column (PGC), Hypercarb™ SPE Cartridges (Thermo Fisher Scientific, MA, USA) following the protocol as outlined by Chen *et al.*[381]

2.8.2 Carbohydrate Electrophoresis

Fluorophore assisted carbohydrate electrophoresis (FACE) can be used for the detection of N-linked oligosaccharides. Similarly, to other electrophoresis techniques, oligosaccharides are resolved depending on the molecular weight. An oligosaccharide cleaved from peptide by PNGaseF (Section 2.8.1) is labelled with a fluorescent chemical which allows for detection on a gel electrophoresis.

2.8.2.1 Staining derivatisation of oligosaccharide chains

A 150 mM ANTS (8-aminonapthalene-1,3,6-trisulphonic acid) (Sigma Aldrich, MO, USA) was prepared in 15% Acetic acid (Sigma Aldrich, MO, USA). A 1 M solution of sodium cyanoborohydride ($NaCNBH_3$) (Sigma Aldrich, MO, USA) was prepared in dimethyl sulfoxide

(DMSO) (Sigma Aldrich, MO, USA). Oligosaccharide samples whether test samples or standard samples (Prozyme, CA, USA) were lyophilized by a vacuum concentrator and resuspended in 10 µL of staining solution, which consisted of 5 µL of freshly made ANTs solution and 5 µL of 1 M $NaCNBH_3$ solution. The staining sample was briefly centrifuged and incubate in the dark at 37°C for 16 to 18 hours. The sample was then lyophilised by a vacuum concentrator and resuspended in 20 µL of loading buffer (20% Glycerol and 0.01% Thorin (Sigma Aldrich, MO, USA)).

2.8.2.2 Gel electrophoresis preparation

A gel-casting apparatus was assembled and the resolving gel solution was prepared, containing; 4 mL of resolving gel stock solution (38% (w/v) acrylamide and 2% (w/v) N,N - methylenebisacrylamide). 1 mL 1.5M Tris–HCl, pH 8.9, 3 mL of deionised water (3ml); and 30 µL of 10% ammonium persulfate (30ul).Polymerization was initiated by addition of 10 µL of TEMED and the solution was poured into the casting apparatus to a height of 0.5 cm below the bottom of the teeth of the comb. Immediately the gel solution was overlaid with 1 cm of 100% ethanol. After polymerization occurred a stacking gel solution was prepared, containing; 2 mL stacking gel stock solution (10% (w/v) acrylamide and 2.5% (w/v) N,N-methylenebisacrylamide); 0.5 mL of 1.0 M Tris–Cl, pH 6.8, 1.5 mL of deionised water (1.5ml); and 20 µL of 10% ammonium persulfate. Polymerization was initiated by adding 5 µL of TEMED. The ethanol was poured off and remaining space was filled with stacking gel solution and a comb was inserted.

2.8.2.3 Fluorophore Assisted Carbohydrate Electrophoresis (FACE)

Carbohydrate gels were placed within a Mini-PROTEAN® Tetra Vertical Electrophoresis Cell (Bio-Rad, CA, USA) containing 1x FACE running buffer. A 10X stock consists of 1.92M glycine, 0.25M Tris base, pH 8.3. A 1X working concentration was prepared with chilled deionised water. Between 10 and 20 µL of stained carbohydrate samples from (2.8.2.1 Staining derivatisation of oligosaccharide chains) were loaded onto Carbohydrate gels. The power supplied was from a Bio-Rad PowerPac™ Basic (Bio-Rad, CA, USA) set at 15 mAmps and ran until loading dye had exited the bottom of the gel, typically for 1 hour. The resolved oligonucleotides were visualised using a UV transilluminator (Bio-Rad, CA, USA).

2.8.3 Measuring of nucleotides by HPLC

The method was adapted from Jimenez del Val *et al.*[382] An Alliance HPLC system (Waters, UK), which is composed of a Series 2695 separations module coupled to a Model 2998 photodiode array detector was used for the study. Data was acquired using Empower 2 software (Waters, UK). Separation was carried out with a CarboPac PA-1 column with a PA-

1 guard column (Thermo Fisher Scientific, MA, USA). The mobile phases used were 3 mM NaOH (E1) and 1.5 M sodium acetate in 3 mM NaOH (E2). The eluents were prepared on the day of analysis and filtered using 0.2-µm filter units (Thermo Fisher Scientific, MA, USA) and degassed. Samples injected into HPLC system were filtered using 0.2-µm syringe filters with PTFE membrane (Sigma Aldrich, MI, USA)

The column temperature was maintained at 30 °C throughout separation. Elution of the sample was carried out using the following gradient (t=time): $t_{0\ min}$, 20% E2; $t_{6\ min}$, 66% E2; $t_{18\ min}$, 66% E2; $t_{23\ min}$, 20% E2; $t_{30\ min}$, 20% E2. Detection was performed at 262.1 and 271.6 nm simultaneously to ensure maximum absorbance of all species present in the mixtures.

2.9 Extract preparation and reaction conditions for yeast cell-free protein synthesis

2.9.1 Yeast cell culture

The yeast strains *Pichia pastoris* GS155 or *Saccharomyces cerevisiae* BY4741 were used throughout the study. A single colony was used to make a starter culture, grown at 30 °C overnight in 5 mL of YPD in a conical tube shaking at 220 rpm. The 1 mL of starter culture was used to inoculate 500ml of YPD medium in 2 L baffled flasks. Cells were grown at 30°C with shaking at 250 rpm to the mid to late exponential phase.

2.9.2 Preparing spheroplasts

Cells were harvested by centrifuging at 1500x g at room temperature for 5 minutes. Cells were washed with autoclaved chilled deionised water (1 mL per 0.5 grams wet weight cell), and cells were pelleted again. The cells were resuspended with100 mM Tris.SO4 pH 9.4, 10mM DTT (1 mL per 0.5 grams wet weight cell) and incubated for 10 minutes at 30°C. Cells were pelleted by centrifugation at 1500x g at room temperature. The pellet was resuspended in 1.2 M sorbitol and 20mM Potassium phosphate buffer pH4.4 (1 mL per 0.15 grams wet weight cell). The cells were spheroplasted using Zymolyase 20 T (Sigma Aldrich, MI, USA) at 5 µL per 1g of wet weight cell. The spheroplasting reaction was carried out at 30°C with gentle shaking for 60 minutes. Spheroplast were harvested by centrifuging at 700x g for 10 minutes at RT, and pellet was resuspended in pre-chilled 1M sorbitol.

2.9.3 Cell lysis by homogenisation

Cells either spheroplasts were harvested by centrifuging at 1500x g for 5 minutes at room temperature, or 700x g for 10 minutes at room temperature, respectively. The cell pellet was resuspended in lysis buffer (30mM HEPES-KOH, pH 4.4, 100mM potassium acetate, 2mM

magnesium acetate, 2mM dithiothreitol, 8.5%(w/v) mannitol, 0.5mM PMSF) at 0.5 mL per 1 gram of wet weight. Resuspended cells were homogenised using a Dounce homogenizer with tight fitting piston. The Dounce homogeniser was placed on ice and cells were lysed by 10 to 15 strokes of the piston.

After cell lysis residual mRNA was digested by micrococcal nuclease (S7) treatment. In this respect, 10 U/mL S7 nuclease (Roche, Mannheim, Germany) and 1mM CaCl2 (final concentration) were added to the lysis and the reaction mixture was incubated for 10 min at room temperature. The reaction was inactivated by the addition of EGTA at final concentration of 6.7 mM.

The lysate subsequently underwent buffer exchange through dialysis, The extract was dialyzed against four exchanges of 200-volumes of Lysis Buffer without the addition of mannitol for 30 min each at 4°C using Slide-A-Lyzer Dialysis Cassettes (2,000 Da MWCO; Thermo Fisher Scientific, MA, USA). The translational active lysates were immediately placed in liquid nitrogen and stored at -80°C in order to preserve maximum activity.

2.9.3.1 Staining of ER and microsomes

ER-Tracker™ Blue-white DPX (BODIPY™ FL Glibenclamide), for live-cell imaging was purchased from Thermo Fisher Scientific, MA, USA. Cells were stained as per manufacturer's instructions. Prior to staining of cell lysate, the cell lysate was diluted 100-fold in Phosphate buffered saline (PBS) (Sigma Aldrich, MI, USA). Diluted cell lysate was stained with 1 µL of ER-Tracker™ Green and incubated for 1 hour at 37°C.

2.9.4 Preparation of energy and co-factor supplements

An 5x energy mixture was produced: 100mM Adenosine triphosphate and 50mM Guanosine triphosphate, 0.135mg mL^{-1} of creatine kinase, 22mM HEPES-KOH (pH 4.4). A stock solution of all 20 amino acids to concentration of 1mM amino acid was prepared. A stock solution containing of each 100 mM deoxyribonucleotide triphosphates (dUTP, dTTP, dGTP, and dCTP) dissolved in water and the pH is adjusted to 4.0 by the addition of sodium hydroxide. Solutions of 500 mM creatine phosphate, a 500 mM spermidine, a 1 M magnesium acetate, a 2 M potassium glutamate, and 100 mM DTT were all prepared in deionised water. Aliquots of each reagent were stored at -20°C

2.9.5 Protein translation reaction

The components required for a CFPS reaction and the amount required are shown in table 2.4. The DNA plasmid(s) are required to have a T7 RNA polymerase promoter as heterologous T7 RNA polymerase will be used to drive transcription.

Table 2.5 Reagents required in CFPS system

Component	Volume/Amount
Yeast cell lysate	50 %
DNA Plasmid	250 ng
T7 RNA polymerase	200 u/mL
dNTPs	1 mM
amino acid mix	0.04 mM
5X Energy mix	1x
Magnesium acetate	10mM
Potassium glutamate	120mM
Spermidine	2mM
DDT	2.7mM
RNase Inhibitor	26.7UmL
H_2O	

2.9.5.1 T7 RNA polymerase

T7 RNA polymerase (NEB. MA, USA) was added directly to the CFPS mix at a final concentration of 200 units/mL. When the T7 RNA polymerase was used independent of a CFPS the manufacturer's instructions were followed.

2.10 Enzymatic activity assays

2.10.1 N-Acetylglucosaminyltransferase I (GlcNAcT-I) activity assays (A)

The method for determining the activity of N-Acetylglucosaminyltransferase I (GlcNAcT-I) activity was adapted from Saribas *et al.*[383]; the activity was determined by measuring the release of UDP from the donor sugar UDP-GlcNAc. The recombinant GlcNAcT-I was

incubated with UDP-GlcNAc (Sigma Aldrich, MI, USA) and the substrate OM3 which is a trimannosyl core with an octyl tail ((n-octyl 3,6-Di-O- (α-mannopyranosyl) β-d-mannopyranoside (OM3) (abcam®, Cambridge, UK)). The reaction was performed in 20 µL of 20 mM HEPES pH4.0, 3 mM UDP-GlcNAc, 0.5 mM OM3, 20 mM $MnCl_2$ and 10 µL of GlcNAcT-I at 37 °C for 4 hours. The reaction was quenched with water and UDP and analysed using the RP-HPLC assay described in section 2.8.3.

2.10.2 N-Acetylglucosaminyltransferase I (GlcNAcT-I) activity assays (B)

The method for determining the activity of N-Acetylglucosaminyltransferase I (GlcNAcT-I) was adapted from Chen *et al.*[384]; activity was determined by measuring the transfer of GlcNAc to a high mannose oligosaccharide $Man_5GlcNAc_2$ (Prozyme, CA, USA). The reaction was performed in 20 µL of 20 mM HEPES, pH 4.0, 3 mM UDP-GlcNAc, 0.5 mM $Man_5GlcNAc_2$, 20 mM $MnCl_2$ and 10 µL of GlcNAcT-I at 37 °C for 4 hours. The transfer reaction and therefore enzyme activity was determined using FACE.

2.10.3 Type II α-mannosidase (Man-II) activity assays

The method for determining the activity of Type II α-mannosidase (Man-II) was adapted from Suzuki *et al.*[385]; activity was determined by measuring the release of *p*-nitrophenyl from the substrate 4-Nitrophenyl α-D-mannopyranoside. The reaction was performed in 50 µL reaction with 25 µL McIlvaine citrate-phosphate buffer, pH 6.0, 10 µL of Man-II 100 mM *p*-nitrophenyl-mannopyranoside. A positive control was also set-up with enzymatic activity produced by α-Mannosidase from *Canavalia ensiformis* (Jack bean) (Sigma Aldrich, MI, USA). The release of *p*-nitrophenol was determined spectrophotometrically. Reactions took place in wells of 96 opaque plate (Greiner Bio-One, Kremsmünster, Austria). Absorbance at 410 nm was measured using Synergy HT Microplate Reader (Biotek Instruments, Inc, VT, USA). Data was acquired using Gen5 Data Analysis Software (Biotek Instruments, Inc, VT, USA).

2.11 Inverted microscope image capture

Images were taken through 60x CPI60 objectives mounted on a Nikon Eclipse Ti inverted microscope with live cells and microsomes imaged on plain slides. To visualise the samples a Phase filter 1 was used to enhance contrast. For fluorescence imaging excitation, emission and exposures were set respectively at 375 nm, 430 nm, 1000 milliseconds. NIS-Elements Microscope Imaging Software (Nikon) was used for capturing and ImageJ (National Institutes of Health) is used for image analysis and presentation.

Chapter 3. A glycoengineering approach in *P. pastoris* for the synthesis of a recombinant glycoprotein with a core pentamannosyl glycoprofile

3.1 Introduction

The experimental aim of this chapter is to bring about a production system for the synthesis of an N-linked glycoprotein comprising a core glycoprofile. However, before the experimental procedure is delineated one must first expand upon the definition of the core glycoprofile. As has been conveyed within the introduction of this thesis, the N-linked oligosaccharide exists in many forms. These glycoforms range from those of elaborate and extended structures, with intricate sugar monomers and chemical linkages, to that of a simpler diminutive structure. However, despite the various oligosaccharide chains one witnesses between specific species, a consensus structure is common to all eukaryotes; known as the core trimannosyl glycoprofile. As one can deduce from Figure 3.1, the trimannosyl structure of $Man_3GlcNAc_2$ is universal to every carbohydrate chain. The existence of a core trimannosyl glycoprofile is owed to the highly conserved nature of the ER associated oligosaccharide construction, by *ALG* oligosaccharyl transferases.

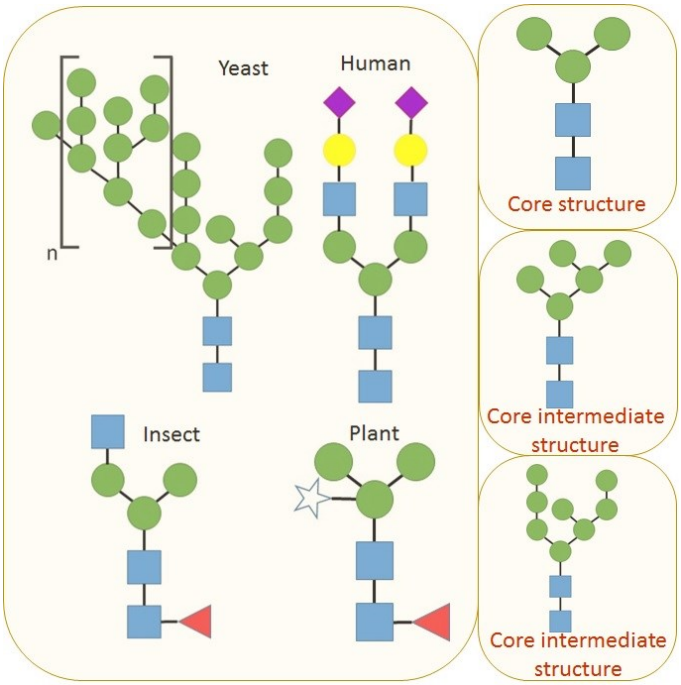


Figure 3.1 typical N-linked oligosaccharide structures from selected number of eukaryotic species.
On the right of diagram, a number of N-linked oligosaccharide structures from various species. Each of these structures, although distinct from one another, contains a consensus core structure of $Man_3GlcNAc_2$ as represented on top right of the diagram. Also, common to each of these species

is a core intermediate structure, which are a result of shared processing steps of oligosaccharides between species, as was described in Section 1.22. The $Man_8GlcNAc_2$ is a core intermediate structure resulting from the actions of an ER resident mannosidase which is present within all eukaryotes. The $Man_5GlcNAc_2$ is a second core intermediate structure resulting from the actions of a Golgi resident mannosidase, this enzyme is not present within many fungal species, therefore this core intermediate structure is not shared by all eukaryotic species.

Together with the core trimannosyl glycoprofile structure, a number of core intermediate oligosaccharide structures also exist within the N-linked glycosylation processing pathway. This includes the high mannose structure of $Man_8GlcNAc_2$ and the pentamannosyl structure of $Man_5GlcNAc_2$ (Figure 3.1). The high mannose structure is an intermediary interspecies structure, which acts as a substrate for either extension or truncation depending on the species type. For instance, whilst lower eukaryotes transfer additional mannose sugars, yielding a hypermannosylated structure, higher eukaryotes possess mannosidase activity removing mannose sugars to generate the second intermediate oligosaccharide structure of $Man_5GlcNAc_2$[1,161]. This $Man_5GlcNAc_2$ glycoform subsequently serves as a key intermediate structure for the biosynthesis of a hybrid or complex type glycan in higher eukaryotes[161].

Now that one has defined the three oligosaccharide core structures, it now must be determined to which core structure is most valuable for this project. As previously stated, the aim of this chapter is to generate a glycoprotein, which will be the substrate for the biosynthesis of a hybrid or complex type glycan. Therefore, the utilisation of the high mannose core structure would not be beneficial for the project. Consequently, the trimannosyl and pentamannosyl core structures both offer a better foundation for the synthesis of a mammalian N-linked glycoprofile. The utilisation of trimannosyl and pentamannosyl cores stood as the basis for glycoengineering efforts in insects and yeast, respectively (Sections 1.26-1.27).

As described in Section 1.22 the pentamannosyl core structure serves as a substrate for the addition of GlcNAc residue to the α1,3 mannose by GlcNAcT-I[210,211]. However, the core trimannosyl oligosaccharide structure can serve as an alternative substrate for GlcNAcT-I[383], and subsequently GlcNAcT-II, negating the need for Man-II. Man-II being the subsequent enzyme after GlcNAcT-I, which removes the terminal mannose sugars yielding a substrate for GlcNAcT-II[212]. However, the activity of GlcNAcT-I for $Man_3GlcNAc_2$ is lower than its normal substrate of $Man_5GlcNAc_2$[383], and although the $Man_3GlcNAc_2$ can occur biologically it is not easily achieved.

The $Man_3GlcNAc_2$ can occur *in vivo* when an alternate glycosylation editing pathway is induced, such as during glucose starvation[386]. However, glucose starvation yields a number of oligosaccharides (namely $Man_{0-5}GlcNAc_2$) and site occupancy is negatively affected[386]. An alternative to achieve a $Man_3GlcNAc_2$ is via the removal of enzyme(s) involved in the biosynthesis of the lipid linked oligosaccharide (LLO) (Table 1.4). The ER bound cytosolic

mannosyltransferases *ALG2* and *ALG11* are involved in the biosynthesis of the LLO. The *ALG2* is involved in the addition of two mannose monosaccharaides, producing the $Man_3GlcNAc_2$ with bisecting mannose sugars. The *ALG11* protein adds two further mannose sugars with α1,2 linkages, therefore producing the $Man_5GlcNAc_2$[336,387]. This structure differs from the core pentamannosyl consensus structure with two mannoses added onto the same oligosaccharide branch. The oligosaccharide produced by *ALG2* has the potential to be utilised as substrate for GlcNAcT-I. However the *ALG11* of *S. cerevisiae* is essential for normal growth at room temperature as the *Δalg11* knock-out has slow growth and is heat sensitive[337]. Furthermore, a glycan chain produced from a *S. cerevisiae Δalg11* strain can still be processed by downstream mannosyltransferases extending the oligosaccharide with further mannose subunits[337]. The knock-out of both *ALG11* and *ALG3*, *ALG3* being the subsequent mannosyltransferase in the LLO biosynthesis, is lethal in *S. cerevisiae*[6]. Consequently, due to difficulties in acquiring a trimannosyl structure biologically, the project focused on obtaining a system which can synthesise the core $Man_5GlcNAc_2$ glycoprofile.

As previously mentioned, the attainment of a pentamannosyl structure has been the basis of glycoengineering efforts in yeast. To achieve a $Man_5GlcNAc_2$ glycoprofile in yeast the formation of hypermannosylation must firstly be prevented. The hypermannosylation is initiated with the addition of an α1,6-linked mannose residue by *Och1p* onto the high mannose core structure of $Man_8GlcNAc_2$. The functional elimination of *Och1p* has been accomplished in a number of yeast species such as *K. lactis, H. polymorpha, Y. lipolytica, S. pombe, P. pastoris* and *S. cerevisiae*[84,287,388,389]. The exclusion of the initial α1,6-linked mannose residue in each of these species prevents the hypermannosylation, yielding glycoproteins with a $Man_8GlcNAc_2$ glycan. Although in yeast species which contain an α1,3 mannosyltransferases activity, the $Man_8GlcNAc_2$ glycan can be extended to an $Man_9GlcNAc_2$ product. The high mannose core structure acts as an intermediate structure for the biosynthesis of the pentamannosyl chain. In the presence of an α1,2 mannosidase enzyme, the $Man_8GlcNAc_2$ can be converted to $Man_5GlcNAc_2$, through the cleavage of α1,2 linked mannose residues[201]. The incorporation of the α1,2 mannosidase within an *Δoch1 P. pastoris* strain led to the generation of glycoproteins containing a $Man_5GlcNAc_2$ glycoprofile. Presently, *P. pastoris* is only the yeast species which has been genetically engineered to synthesise the pentamannosyl core structure. This strain, although demonstrated within the scientific literature, was not available to obtain commercially at the outset of the project.

As discussed in Section 1.8, *P. pastoris* is an established protein expression host known for high product titres and protein secretion. A wealth of commercially available expression vectors and kits exist for *P. pastoris* allowing for simple, scalable, and robust protein expression. Although it could be feasible to replicate the *P. pastoris* strategy of producing a

$Man_5GlcNAc_2$ glycoprofile in other yeast species, the fact that *P. pastoris* has a history of successful glycoengineering and producing biopharmaceuticals makes it illogical to choose an alternative expression host species. Therefore, the project proceeded to generate a $Man_5GlcNAc_2$ glycoprofile in *P. pastoris*.

3.2 Creating a *P. pastoris OCH1* mannosyl transferase knock-out through homologous recombination with a mammalian mannosidase

In order to achieve a $Man_5GlcNAc_2$ glycoprofile in *P. pastoris* one must first eliminate the activity of the 1,6 mannosyltransferase and replace this with activity from an α1,2 mannosidase. Previously published strategies obtained the pentamannosyl structure through two steps; firstly, by removing the *OCH1* activity, then secondly utilising the generated knockout strain to insert an α1,2 mannosidase. This two-step strategy was useful at the time as the optimal α1,2 mannosidase for conversion in *P. pastoris* had not been discovered. Therefore, utilising a *Δoch1* strain allowed for investigation to discover the best α1,2 mannosidase through combinatorial libraries of different signal sequences and different species specific α1,2 mannosidase[293,294]. As the optimal α1,2 mannosidase is known, it is decided to perform this knockout and replacement within a single reaction, as was attempted here.

The removal of *OCH1* could be achieved by disrupting the open reading frame (ORF) resulting in the knockdown of the gene activity. This can be achieved by an *in vivo* homologous recombination (HR) event, where an exogenous DNA cassette targets and replaces the gene of interest. The inclusion of a region homologous to the targeted locus will trigger the HR event resulting in genome manipulation.

The *P. pastoris* strain GS115 (the selected strain for the study) genome was published in 2009 and consists of 9.4 megabases over 4 chromosomes with 5,313 predicted protein-coding genes[85]. The *OCH1* gene is located on chromosome 1 (PAS_chr1-3_0251), with the coding region beginning at base 466,221. The coding region spans 1,215 base pairs (bp) to produce a 404 amino acid residue product. The initial design for generating the *Δoch1* deletion, targeted the genome sequence upstream and downstream of the *OCH1* coding sequence, with chromosome locations indicated on Figure 3.2. The disruption of the *OCH1* is made through a double cross-over event, between linear exogenous DNA and the *P. pastoris* chromosome. Two DNA fragments corresponding to the ends of the *OCH1* gene, of approximately 1000 bp, were obtained through PCR cloning. The 1000 bp presents a suitable length being of equivalence to the length of recombination locus of many commercially available expression vectors which have been designed for genome integration[87]. For

instance, the expression vector pPICz (Invitrogen, MA, USA) contains 942 bp of homologous DNA to the Alcohol oxidase 1 gene on chromosome 4. The actual length of the 5' end of the cassette is 1,317 bp and the 3' end contained 1,257 bp. However, once the cassette was linearised for transformation the actual sequence homology is less at 1,091 bp and 965 bp for the 5' end and 3' end, respectively.

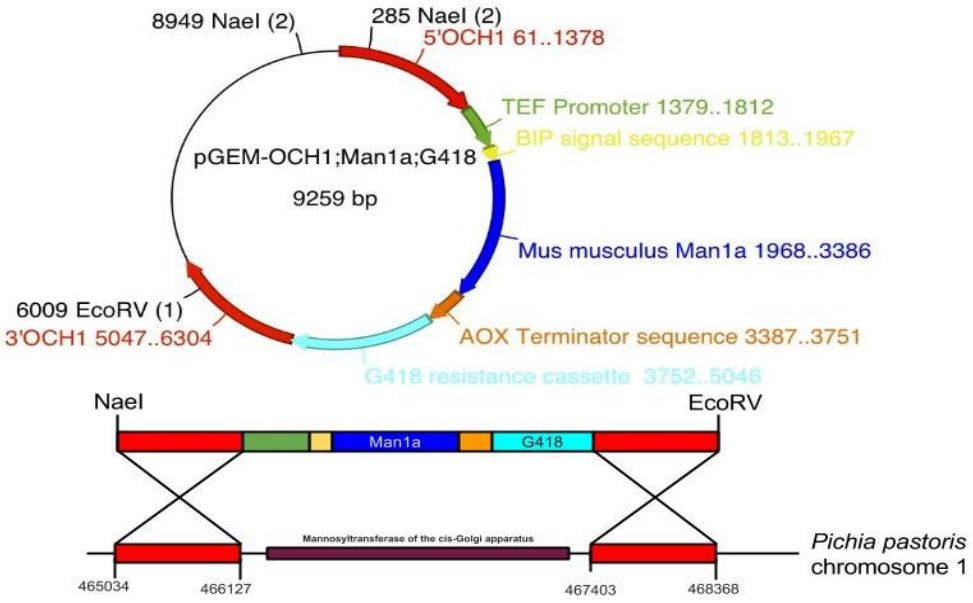


Figure 3.2 A schematic of the plasmid pGEM-OCH1:Man1a:G418.
The exogenous DNA cassette consisted of an ER localisation signal sequence from the BIP gene (M31006.1) upstream of the Man1a gene, and an ER retrieval sequence HDEL downstream. This ensures that mannose cleavage occurs in the ER. The Man1a is transcribed by a strong constitutive promoter TEF, and downstream of the Man1a gene is a terminator sequence from AOX1 to ensure the termination of transcription. To aid cell line selection a genetic cassette containing the bacterial gene Tn903kanr is used for selection against the antibiotic Geneticin® (G418)[390].

To aid in the selection of a pentamannosyl producing *P. pastoris* strain a selectable marker was inserted between the 5' and the 3' flanking sequences of the *OCH1* recombination locus. The utilisation of a selectable marker is a classic strategy which has enabled effortless direct gene deletion in many organisms. It was the basis of the highly successful *Saccharomyces* genome deletion project, which saw each individual CDS of *S. cerevisiae* deleted[391]. The project succeeded by using the kan^r gene of *E. coli* transposon *Tn903* fused to TEF gene promoter of the filamentous fungus *Ashbya gossypii*[392,393]. It has been known since the 1980s that the kan^r gene confers *S. cerevisiae* resistance to the aminoglycoside agent geneticin (G418)[394]. By inserting a kan^r gene between homologous regions of a *S. cerevisiae* gene, transformants which grew in the presence of G418 were demonstrated to contain the kan^r in replacement of the *S. cerevisiae* gene. In 2008, the kan^r gene under control of a native *GAP* gene promoter gave *P. pastoris* resistance to G418[390]. The same gene construct was initially

utilised in this study, inserted between the 5' and the 3' flanking sequences of the *OCH1* recombination locus (Figure 3.2).

As repeatedly mentioned, the $Man_5GlcNAc_2$ glycoprofile is achieved in higher eukaryotes through the actions of Golgi based α1,2 mannosidases (Section 1.22). In humans three enzymes collectively cleave mannose sugars from a $Man_{9\text{-}8}GlcNAc_2$, preferring the $Man_8GlcNAc_2$ processed by the ER mannosidase I. However, despite the substrate preferences observed between the different Golgi α1,2 mannosidases isotypes, each enzyme is individually capable of producing the $Man_5GlcNAc_2$ glycoprofile[200,201]. Hence only one Golgi α1,2 mannosidase was included within the published glycoengineered *P. pastoris* strains[288].

The activity of a number of α1,2 mannosidases has previously been tested in a *Δoch1* deletion strain of *P. pastoris*[293,294]. The mannosidase from *C. elegans* was shown to have the highest activity, closely followed by mannosidases from Drosophila, mouse and human[293]. All the α1,2 mannosidases from these species demonstrated a greater than 75% conversion rate of $Man_8GlcNAc_2$ to $Man_5GlcNAc_2$[294]. *C. elegans* was noted to have a significant level of α1,2 mannosidases leakage into the culture medium. Consequently, the hydrolysis reaction predominantly occurred extracellularly[294]. This is compared to mouse which had the least leakage of α1,2 mannosidase from the cell, so conversion was principally intracellularly[294]. This is worth noting as although *C. elegans* mannosidases exhibited higher conversion rates, the reaction was likely due to the longer duration of enzyme with substrate. Therefore, the mouse mannosidase, which spent less time in the presence of the substrate, could be deemed to have a quicker rate of conversion and is therefore more active in *P. pastoris*. To this end the *M. musculus* mannosidase 1A (Man1a) (XM_002489551.1) was utilised in this study to achieve the core structure. The Man1a gene, similarly to the published glycoengineered strains, was placed under the control of the strong constitutive native gene promoter. A promoter from Translation elongation factor 1-α (*TEF*), which has demonstrated high levels expression of recombinant proteins in *P. pastoris*[395], was utilised for Man1a transcription.

As previously mentioned, the glycosyltransferases and glycosidases from N-linked glycosylation pathway are all type II transmembrane proteins classified by a hydrophobic transmembrane domain (TMD), within the ER or Golgi membrane. For yeast glycoengineering projects, mammalian TMD cannot be used for translocation of recombinant proteins to organelle membranes and therefore need to be removed. The deleted TMD can either be replaced with a yeast TMD or the protein is left soluble i.e. no TMD. For instance, Choi *et al.* exchanged the native TMD with a *S. cerevisiae* TMD from an ER located transmembrane protein[293]. The *S. cerevisiae* TMD ensured the Man1a was targeted to the ER membrane of

P. pastoris. Whereas, Callewaert *et al.* utilised a soluble *T. reesei* mannosidase, which was targeted to the lumen of ER by adding an N-terminal ER signal sequence (SS)[292].

The Man1a selected for use in this chapter was truncated at 561 bp (or 187 amino acids) which removed the cytosolic, transmembrane domain and stem region. This position was selected as it was identical to Choi *et al.* previously published truncated Man1a[293]. However, where Choi *et al.* replaced the TMD, for this project the Man1a was left soluble. An N-terminal ER signal sequence (SS) was added to ensure that the oligosaccharide substrate was targeted directly after the *en bloc* transfer by the OST. A *S. cerevisiae* SS was selected from an ER lumen residing molecular chaperone Kar2 which is the mammalian homologue of *BiP* (Binding immunoglobulin Protein). The BiP protein is expressed and will translocate to the ER during both ER homeostasis and ER stress[396]. Therefore, a BiP-SS tagged protein will translocate even through times of cellular stress. Which is important as metabolic burden can be placed onto the cell during the overexpression of recombinant proteins, therefore by utilising the BiP-SS one is ensuring that the mannosidase will be present within the ER during all stages of cellular activity enabling oligosaccharide conversion. Previously, the SS of a human BiP had demonstrated effective ER translocation of recombinant proteins in *S. cerevisiae* and *P. pastoris*[397]. Therefore, it was reasoned that a *S. cerevisiae* BiP would trigger ER translocation and residency of a tagged recombinant protein in *P. pastoris*. In fact, a member of the Polizzi lab successfully implemented the translocation of a GFP protein to the lumen of the *P. pastoris* ER with a *S. cerevisiae* BiP SS (personal communication with Rebecca Burgoyne). To further enhance ER residency of Man1a, an HDEL retrieval sequence was added to the C-terminus of the protein. The HDEL-tag is recognized by the HDEL receptor in the Golgi apparatus and retrograde transports proteins to the ER in COP I-coated vesicles, leading to their retention in the ER instead of extracellular secretion

The BiP SS HDEL tagged Man1a gene was inserted between the 5' region of the *OCH1* recombination locus and the kan^r resistance gene, as illustrated in Figure 3.2. Once the gene elements of the exogenous DNA cassette were assembled by overlap extension PCR (OE-PCR), the cassette was cloned in pGEM®-T easy vector (Promega, WI, USA) and sequenced verified (See Appendix Section 2 for full sequence). The vector was entitled pGEM OCH1:Man1a:G418.

3.3 The knock-out of a recombinant glycoprotein producing *P. pastoris* strain with Golgi Mannosidase I construct

The pGEM OCH1:Man1a:G418 plasmid was linearised with the restriction enzymes NaeI and EcoRV, which cut at positions 227 base pairs and 5952 base pairs, respectively (Figure 3.2). The DNA cassette was isolated from the plasmid backbone by gel electrophoresis. A DNA band of size corresponding to the cassette was excised from a 1% agarose gel and purified using the Wizard SV Gel and PCR purification kit (Promega, WI, USA) according to manufacturer's instructions.

It has previously been suggested that a *P. pastoris* strain already optimised for the expression of a glycoprotein should be utilised for glycoengineering method[99]. A strain of *P. pastoris* (GS115) that expresses the N-linked glycoprotein Kringle 3 domain (K3) (PDB# 2l0sA), was kindly gifted by Xavier Ribes Calvo. The K3 of human plasminogen makes an ideal model glycoprotein as it contains a single N-linked glycosylation site which is at the surface of protein structure[398]. This is useful as buried oligosaccharide chains can be difficult to reach by enzymes and therefore be modified[399]. The K3 producing *P. pastoris* strain will be known as GS115-K3. The K3 was under the control of the *AOX1* promoter, and included an α-mating factor signal sequence for secretion into the culture medium.

An aliquot of an overnight culture of GS115-K3 was used to inoculate a 500 mL culture for preparation for electroporation as described in the *Pichia* Expression Kit manual (Life Technologies, CA, USA) and transformed with the linearised pGEM-OCH1:Man1a:G418 construct. The transformant culture was plated on YPD containing 350 μg mL^{-1} of G418. The plates were incubated at 30°C for 3-5 days. Individual colonies were selected at random and a colony PCR was performed to determine successful knockout and integration.

3.4 Genomic analysis of the *P. pastoris* strain transformed with Golgi Mannosidase I construct

The genome of the selected *P. pastoris* colonies were tested to (1) verify for the presence of Man1a construct, (2) confirm the location of the Man1a construct and (3) confirm the removal of native *OCH1* locus. The genomic DNA of the selected *P. pastoris* colonies was isolated; successful genomic extraction was confirmed by PCR using primers specific for an internal control, in this case the native *TEF* promoter. Over 50 *P. pastoris* colonies were screened, but to facilitate the presentation only a representative subset of 8 samples will be discussed here. From the 8 samples, a 435 bp product was observed in all colonies indicating a successful DNA extraction (Figure 3.4).

To confirm the presence of the Man1a construct, a PCR was performed using a forward primer specific for the start of the *TEF* promoter DNA part and a reverse primer specific for the 3' end of Man1a DNA part. This produced a 2015 bp PCR product in six of the eight transformed colonies tested (Figure 3.4). Therefore, although the 8 colonies presented the ability to grow in the presence of G418, only 75% of those colonies contained the Man1a gene and *TEF* promoter.

To test for the location and therefore ensure that the Man1a construct had integrated into the *P. pastoris* at the *OCH1* locus, a PCR was performed using primers upstream or downstream of the integration site paired with a primer within the Man1a DNA cassette (as illustrate in Figure 3.3). A PCR with a forward primer designed to amplify upstream of integration site and a reverse primer specific for the 3' end of Man1a gene should produce an expected PCR product of 3333 bp. A further PCR with a forward primer specific for the 5' end of the AOX1 terminator and a reverse primer specific for a downstream region of the integration site will produce an expected product of 2918 bp. If no PCR product is observed from either reaction this could indicate that the Man1a construct did not localise to the *OCH1* locus.

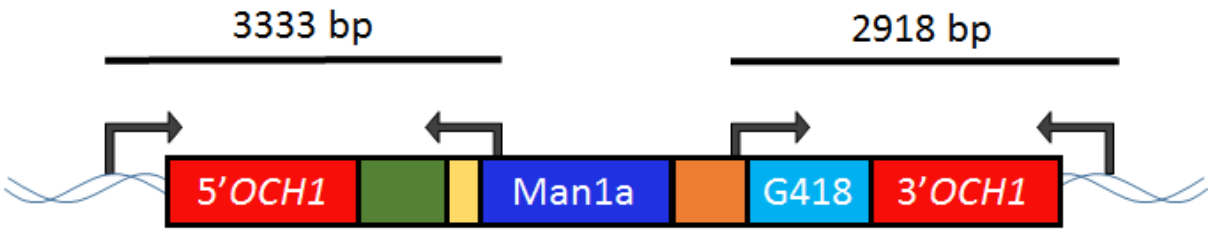


Figure 3.3 visualisation of the colony PCR which was performed on P. pastoris strains which were transformed with the exogenous DNA pGEM-OCH1:Man1a:G418.
The primers which are represented with arrows were designed to amplify a region within the exogenous DNA of the Man1a construct and a region of the native chromosome at a location upstream or downstream of the integration site. The sequences of each primer can be found within the Appendix Section 1 of this document. The expected size of each of the PCR products is shown on the figure, and was determined using gel electrophoresis.

When testing for the location of the Man1a construct it was not surprising that two colonies failed to produce a PCR product (namely 7 and 8) as both previously indicated an absence of a Man1a gene (Figure 3.4). However, a further colony, namely colony 4, also failed to produce a PCR product, suggesting that the Man1a had not integrated in the *OCH1* locus of this strain. The other five colonies all produced PCR products, with four of the eight transformed colonies returning a PCR product when utilising primers for the 3' flanking region and three returning a PCR product when testing the 5' flanking region. Only two colonies, namely 1 and 3 producing a PCR product from both reactions.

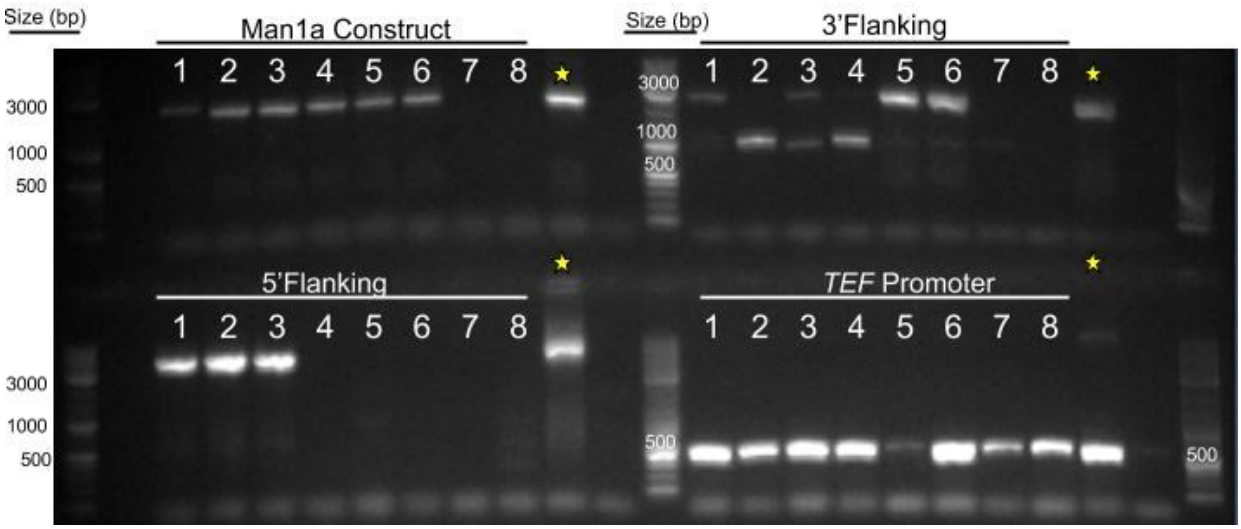


Figure 3.4 PCR analysis of the *P. pastoris* strain transformed with pGEM-OCH1:Man1a:G418.

A representative subset of 8 samples of transformants were tested for the presence and location of Man1a construct. The genomic DNA of transformant *P. pastoris* was isolated by the method indicated in Section 2.3.2.3. To verify a successful extraction a PCR was performed with primers specific for endogenous DNA. In this case the TEF promoter was amplified. All 8 colonies demonstrated the presence of native DNA, shown by the presence of a 435 bp product on the bottom right of agarose gel, in the section labelled TEF promoter. To test for the presence of the exogenous Man1a construct, a PCR reaction was performed amplifying a region within the Man1a construct. The results indicated in the top left of agarose gel that the PCR successful produced a 2015 bp PCR product in 6 of 8 tested colonies (namely colonies labelled 1 to 6). Finally, two PCR reactions were performed to verify the location of the Man1a construct, a depiction of the PCR reaction is shown in Figure 3.3. The top right of the agarose gel shows the results from the PCR reaction against the 3' region of Man1a construct. Only 3 colonies showed a 2918 bp product namely colonies labelled 1, 5 and 6. The bottom left of the agarose gel shows the results from the PCR reaction against the 5' region of Man1a construct. Only 3 colonies showed a 3333 bp product namely colonies labelled 1, 2 and 3. All PCR reactions were performed alongside a positive control indicated by the yellow star. The positive control utilises an uncut pGEM-OCH1:Man1a:G418 plasmid as template. Therefore, PCR products of the positive control will of identical size to the PCR products of transformants with Man1a constructs.

Therefore, whereas three colonies above only produced a product from one flanking region, two colonies, namely 1 and 3, produced results for both flanking regions. This could indicate that a double crossover event occurred, and the *OCH1* had successfully been removed and replaced with the Man1a construct. To confirm this a final PCR reaction was performed on the *OCH1* locus to detect if the *OCH1* gene had been deleted by utilising primers that bind within the *OCH1* CDS. As expected, colonies with prior results which indicated that the Man1a had not integrated into the *OCH1* CDS returned results indicative of an intact *OCH1* CDS (Figure 3.5). However, PCR products were also observed in the other two colonies (namely 1 and 3) suggesting that all strains contained an intact *OCH1* gene (Figure 3.5).

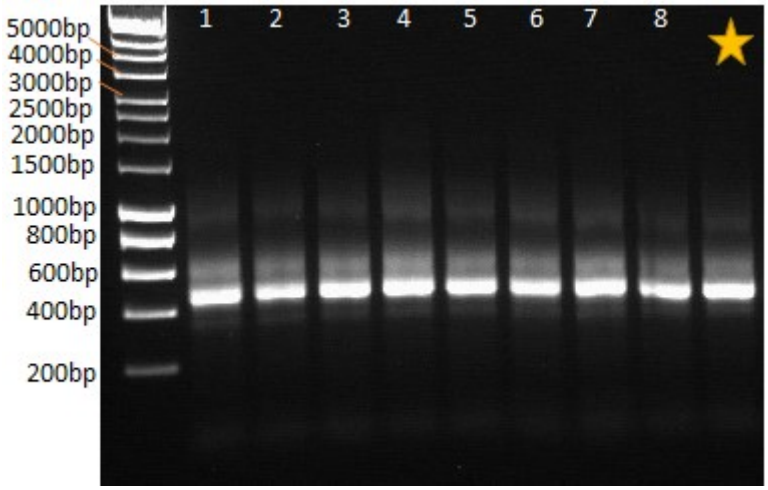


Figure 3.5 PCR analysis to detect for an intact *OCH1* gene in *P. pastoris* strains transformed with pGEM-OCH1:Man1a:G418.
A colony PCR was performed against eight selected *P. pastoris* strains. The PCR was designed to amplify a 505 bp product within the *OCH1* gene, the *OCH1* gene is 1215 bp in total. The primer sequences can be found in the Appendix Section 1. If the Man1a construct had successful integration at the target site in the *P. pastoris* genome then no PCR product would be expected, as the OCH1 gene would be deleted. However, all eight colonies displayed a PCR product of the expected size. This indicated that an intact *OCH1* CDS was present within the genome of the *P. pastoris* transformed with the Man1a construct. A PCR control reaction was performed (indicated by the yellow star) on a non-transformed *P. pastoris* GS115 strain, this resulted in a PCR product of similar size to the test samples.

The results from Figure 3.4 demonstrated that colonies numbered as 2, 5 and 6 failed one of the PCR reactions. Each demonstrated presence of the *OCH1* CDS (Figure 3.5) and that of the Man1a construct (Figure 3.4). The Man1a construct was shown to be either 5' or 3' of the integration site, however the corresponding flanking region PCR reaction failed. For instance, the colony numbered 2 produced a product for the 5' flanking region but not for the 3' flanking region (Figure 3.4). This result could be a consequence of homologous recombination by a single cross-over event, therefore not triggering a gene replacement event of Man1a with the *OCH1* CDS (As illustrated in Figure 3.6). It is generally accepted that double crossover events are much rarer than a single crossover events in *P. pastoris*[400]. As for colonies 1 and 3, the initial PCR in Figure 3.4 appeared to demonstrate the occurrence of a double crossover event. However, both colonies contained an *OCH1* CDS (Figure 3.5), indicating that a double crossover event did not occur. This could be a result of multiple integrations of the linearised pGEM-OCH1:Man1a:G418, possibly by two or more single crossover events (as illustrated in Figure 3.7). The phenomenon of multiple tandem integration in *P. pastoris* was first proposed in the 1991[401], and verified through chromosomal and whole genome sequencing of transformed *P. pastoris* species. The multiple tandem integration of heterologous genes can be beneficial as an increased gene copy number and has been shown to increase yields of the recombinant protein[402]. Although useful for recombinant expression protein expression the same cannot be said for creation of a knockout strain.

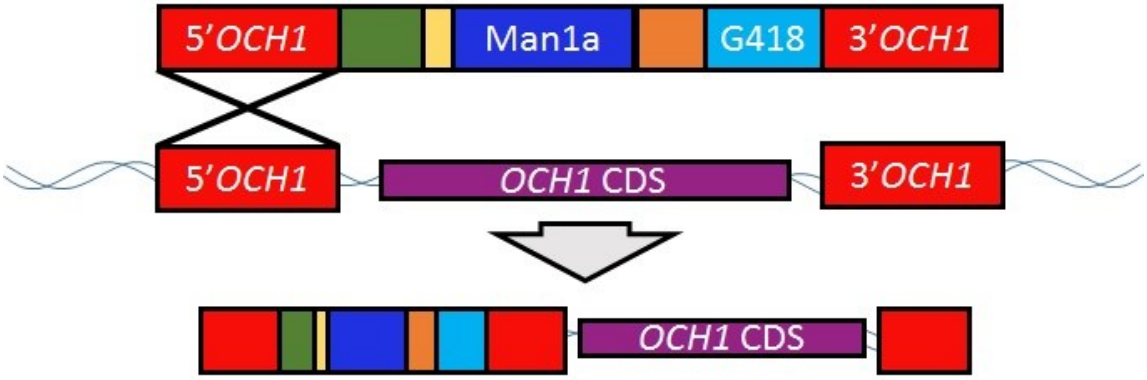


Figure 3.6 a proposed integration event by a single crossover event.
If a single crossover event occurs than only the 5' flanking region of the linearised pGEM-OCH1:Man1a:G418 recombines with the *P. pastoris* genome. Therefore, an intact OCH1 CDS is still present.



Figure 3.7 a proposed multiple tandem integration event of the Man1a construct at the OCH1 locus.
If a multiple tandem integration event has occurred the *P. pastoris* chromosome 1 could contain multiple copies of the Man1a construct. The figure above visualises the integration of three Man1a constructs into the *P. pastoris* chromosome 1. If these integration events have occurred by a single crossover an intact *OCH1* CDS will remain present within the genome.

The results show that although a large portion of screened *P. pastoris* transformants contain a Man1a construct, an intact *OCH1* CDS is still present in the chromosome. The incorrect integration of the Man1a construct into the genome will result in the *OCH1* CDS remaining intact. To move forward to achieve a *Δoch1* strain a further strategy was required as described below.

3.5 Yeast humanization in *P. pastoris Δku70* strain

The initial constructed plasmid appeared to fail in targeting the exogenous Man1a to the *OCH1* locus. It is well known that non-conventional yeast such as *P. pastoris* (other examples include *H. polymorpha, Y. lipolytica, Pichia stipitis*) have a much lower efficiency when it comes to the targeting of genes into the host genome compared with conventional yeast such as *S. cerevisiae*[403]. The integration of entire expression cassettes into a genome is achieved either by HR or non-homologous end joining (NHEJ). HR is highly accurate for integration into the correct homologous sequence, whereas NHEJ is not. Gene replacement events by HR in *P. pastoris* occur at a very low frequency, estimated at 0.1% when the total length of targeting fragments is less than 500 bp or with increased frequency of between 10% to 30% when ≥1000 bp of homology is present[404]. This is compared to *S. cerevisiae,* which has been described to be extremely efficient at HR, where gene replacement events by HR can occur

with a frequency of 70% even with targeting fragments at less than 50bp[405]. The difficulty in targeting isogenic genomic DNA into *P. pastoris* was exemplified by Chen *et al.*[406] who again aimed to knockout the *OCH1* gene. In this study of the 1000 colonies screened none returned a *Δoch1* genotype. Unfortunately, the study did not discuss what genetic differences these non-knockout transformed strains had compared to the native type making it difficult to understand if the results in Section 3.4 matched. Eventually to improve genome targeting Chen *et al.* postulated that cellular fitness is an important factor that limits the efficiency of gene targeting in non-conventional yeasts[406]. Chen *et al.* improved gene targeting by utilising a helper expression plasmid encoding *och1p* before transforming with recombinant DNA for the *OCH1* gene removal[406].

This technique could be of use here, however *P. pastoris* (GS115) preferentially integrates DNA with NHEJ, which as mentioned is a far less specific process, which can result in inaccurate integration of DNA anywhere in the genome. The protein heterodimer *ku70p/ku80p* is involved in NHEJ-integration pathway, if the *ku70p* is removed, DNA integrates mainly via the HR-pathway. During this study, a *P. pastoris Δku70* strain became available. The *Δku70* strain has been demonstrated to have an increased percentage of homologous recombination (HR) events[406]. Gene replacement events by HR in a *P. pastoris Δku70* strain have been shown to occur at a frequency of 100% with as little as 650 bp sequence homology[404]. The use of the *P. pastoris Δku70* strain should facilitate in the generation of *OCH1* gene knockout, rather than integration of Man1a cassette randomly in the genome.

As the *P. pastoris Δku70* strain had only recently become available, efforts to glycoengineer this strain have never previously been attempted. As this strain offers a better alternative for gene replacement, and therefore targeted integration of recombinant genes, it was decided to firstly generate a *Δoch1* strain, before proceeding with generating a Man1a^{+} strain. The *P. pastoris Δku70:och1* strain would be first of this genotype and could be a useful host to others for expression of recombinant proteins.

3.6 Redesign and construction of a plasmid for homologous recombination of *OCH1* gene with Nat1 resistance gene in *P. pastoris (Δku70)*

Although the utilisation of a *P. pastoris Δku70* will improve the gene targeting, to further assist in generation of a *Δoch1* knockout a second plasmid was constructed. The drawback to the first plasmid pGEM-OCH1:Man1a:G418 was that the *OCH1* homology was designed upstream and downstream of the *OCH1* coding sequence, with chromosome locations indicated on Figure 3.2. This meant that if a single crossover event occurred both the knock-

out construct and intact *OCH1* gene would be present in the genome. This possibly what is observed from screened clones in Section 3.4.

In an aim to improve the likelihood of obtaining a *Δoch1* strain a second plasmid was designed with homologous DNA targeting the *OCH1* CDS. Two DNA fragments, of approximately 650 bp, which target the *OCH1* CDS were obtained through PCR cloning with primers containing complementary sequence to the *P. pastoris* expression vector pJAN-S1. The disruption of the *OCH1* was designed to occur through a double crossover event, between linear exogenous DNA and the *P. pastoris* chromosome. However, if a single crossover event does occur this will still cause disruption of the *OCH1* ORF preventing *och1p* translation. The two OCH1 DNA fragments were introduced into the pJAN-S1 plasmid using Gibson Assembly. The pJAN-S1 contains a nourseothricin resistance gene (nat1), which will facilitate in selection of successful transformants. The constructed plasmid, was sequenced verified (Appendix Section 2) and will be known as pJAN OCH1:nat1 (Figure 3.8).

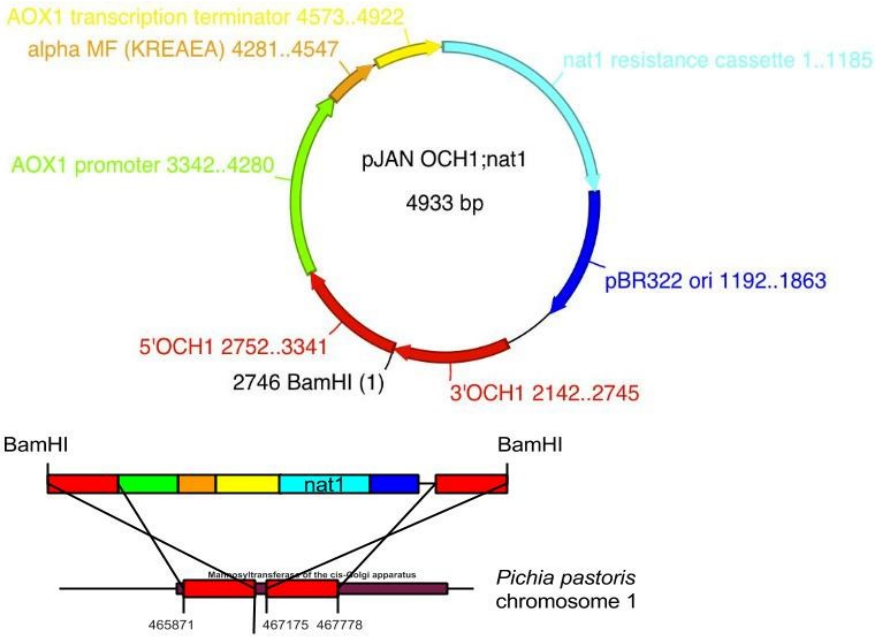


Figure 3.8 Schematic of the pJAN OCH1:nat1 plasmid.
The exogenous DNA cassette consisted of a DNA complementary to the *P. pastoris* chromosome 1. The complementary DNA will target the exogenous DNA to the *OCH1* CDS. To aid cell line selection a genetic cassette containing a Nourseothricin resistance gene is used for selection against the antibiotic Nourseothricin. Highlighted are the different DNA parts needed for homologous recombination into the *P. pastoris* genome, with positions outlined. Shown are the location of restriction sites used to linearize the plasmid.

3.7 *P. pastoris (Δku70)* knock-out of OCH1 gene with pJAN OCH1:nat1 plasmid

The pJAN OCH1:nat1 plasmid was linearised with the restriction enzyme BamHI which cuts between the *OCH1* parts in the pJAN OCH1;nat1 plasmid. (Figure 3.8). A DNA band of size

corresponding to the cassette was excised and purified by the method indicated above. A 500 mL culture of *P. pastoris (Δku70)* was prepared for electroporation as described in the *Pichia* Expression Kit manual and transformed with linearised pJAN OCH1:nat1 construct. The transformant culture was plated on YPD containing 100 µg mL^{-1} Nourseothricin. The plates were incubated at 30°C for 3-5 days.

The genomic DNA of selected *P. pastoris* colonies was isolated and a successful extraction of genomic DNA was confirmed by PCR using primers specific for the native *TEF* promoter. A PCR was performed on the *OCH1* locus to detect if the *OCH1* CDS had been disrupted with exogenous DNA, by utilising primers that amplify *P. pastoris* chromosome 1 upstream and downstream of the integration site. The observed PCR product of ~5000bp indicated a successful gene replacement (Figure 3.9). The PCR product of one of the colonies was sequence verified and showed that the DNA cassette had successfully integrated (see Appendix Section 2 for sequencing information). The generated *P. pastoris* strain will be known as *P. pastoris (Δku70:och1).*

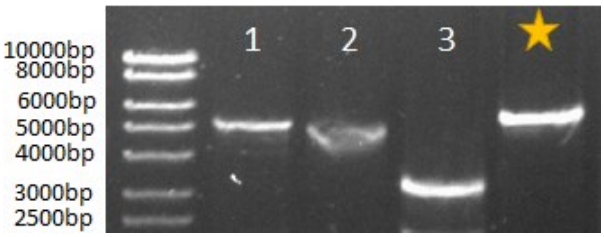


Figure 3.9 PCR analysis of the *P. pastoris* (*Δku70*) transformed with pJAN OCH1:nat1 plasmid
A Colony PCR was performed against three *P. pastoris* strains which had be transformed with pJAN OCH1:nat1. The PCR was designed to amplify a ~5000 bp product within the *OCH1* gene, the *OCH1* gene is 1215 bp in total. The primer sequences can be found in the Appendix Section 1. A positive control (indicated by the yellow star) consisting of a linearized pJAN OCH1:nat1 plasmid was included, being of a suitable molecular weight comparison. Two clones exhibited the expected PCR pattern, matching a similar size to the positive control. This indicated that the *OCH1* CDS was removed from the genome of the *P. pastoris* transformed with the pJAN OCH1:nat1. A third clone displayed a PCR product below the predicted molecular weight and was not further used in the study.

3.8 Phenotypical characterisation of the *P. pastoris Δku70:och1*

The removal of *och1p* has previously been shown to influence the growth characteristics of *P. pastoris* strains. For instance, while the native GS115 *P. pastoris* strain has a doubling time of 2 hours, the *Δoch1* strain has previously been reported to have a doubling time of 2.36 hours[99]. Although previously reports of *och1p* deficient strains of *P. pastoris* and other yeast species had documented an impediment in growth characteristics, during this study an observable difference in the cell morphology was also witnessed (Figure 3.10). *P. pastoris* colonies are classified as being spheroidal to ovoidal in shape, having a cream-coloured, dull

to faintly glistening and butyrous appearance[407]. However, the *P. pastoris Δku70:och1* strain was observed to have an abnormal morphology. Although being cream-coloured, the colonies were dull and crenulated in appearance. Following the creation of a *P. pastoris Δku70:och1* in this study, a further research group noted a similar morphological change within a *P. pastoris och1p* deficient strain[408]. As of yet the morphological phenotype of a *Δoch1* yeast strain producing a mannosidase has not been described.

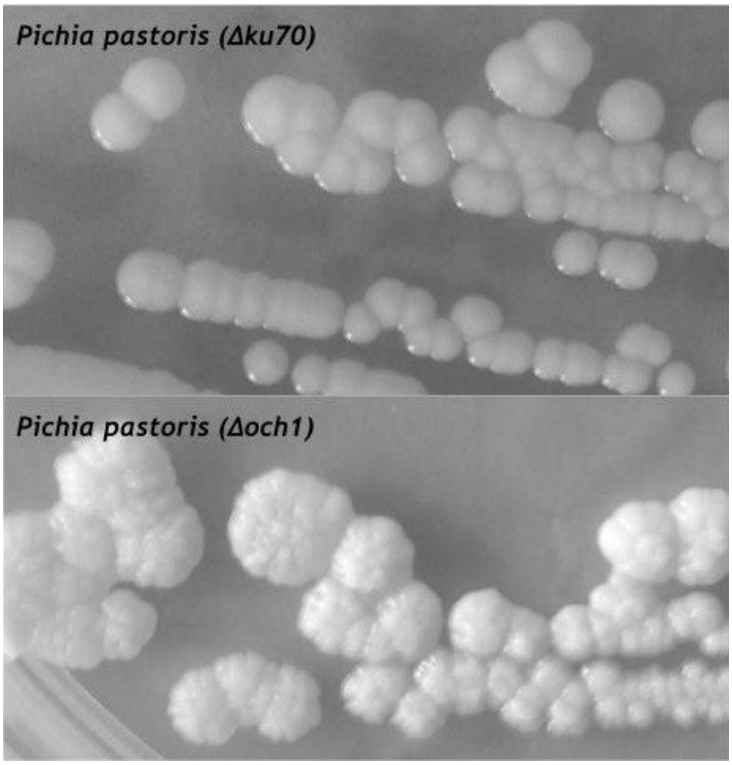


Figure 3.10 the morphological differences between *P. pastoris* (*Δku70*) and *P. pastoris* (*Δ ku70::och1*).
The cell morphology of the *P. pastoris* (*Δ ku70::och1*) was compared against the *P. pastoris* (*Δ ku70*) strain. The *P. pastoris* were plated on YPD agar without antibiotics and grown at 30°C for 3 days. The photographs displayed above were taken from a digital camera. The removal of the *och1p* appears to cause a significant morphological change, with smooth spherical colonies of native strain altering to rough, pitted and non-spherical shape.

3.9 Transformation of the *P. pastoris (Δku70:och1)* with a recombinant glycoprotein kringle-3 domain

In order to assess the glycoprofile of the glycoengineered strain *P. pastoris (Δku70:och1)* a heterologous gene encoding a glycoprotein was introduced. The K3 protein which was used initially for the *P. pastoris* transformed with the pGEM-OCH1;Man1a;G418 plasmid was selected. The pPICz K3 plasmid was linearised with the restriction enzyme SacI as recommended by supplier (Life Technologies, CA, USA). This cuts the plasmid within the 5'*AOX1* region, which promotes integration into the genome at the *AOX1* locus. A DNA band of size corresponding to the cassette was excised from a 1% agarose gel and purified using

the Wizard SV Gel and PCR purification kit (Promega, WI, USA) according to manufacturer's instructions.

A 500 mL culture of *P. pastoris* (*Δku70:och1*) was prepared for electroporation as described in the Pichia Expression Kit manual and transformed with linearised pPICz K3 construct. The transformant culture of *P. pastoris* (*Δku70:och1*) was plated on YPD containing 100 µg mL^{-1} of Zeocin™ and incubated at 30°C for 3-5 days.

3.10 The expression of recombinant glycoprotein Kringle-3 domain in *P. pastoris (Δoch1)* and assessing the N-glycosylation profile

Colonies were selected and grown overnight in 5 mL of YPD culture containing suitable antibiotic (Zeocin™) in a 50mL conical tube. Cultures were pelleted by centrifugation and supernatant removed. Cultures were resuspended in 5mL of methanol based medium (BMMY) to induce recombinant protein expression of the K3. Cultures were grown for 24 hours, before being pelleted and the supernatant collected. The protein content of the supernatant was concentrated using a TCA protein precipitation and expression of K3 was assessed by SDS-PAGE. The colony which exhibited the largest expression indicted by a denser band on the SDS-PAGE was used for further experiments.

To aid in the defining the glycoprofile of the *P. pastoris* (*Δku70:och1*:$k3^{+}$) the molecular weight of K3 was assessed against a K3 produced in a native *P. pastoris* strain. As a longer glycan chain increases the molecular weight of the protein, therefore a migration difference between the two K3s can indicate non-extension of the glycan. As a sugar subunit is 180 g mol^{-1} the size difference between K3 with wild type N-glycan chain and K3 $Man_8GlcNac_2$ is approximately 1.1 KDa.

The wild-type *P. pastoris* (K3), and *P. pastoris* (*Δku70:och1*:$K3^{+}$) strains were grown overnight at 30°C with shaking in 5 mL of YPD. A sample of culture was used to inoculate a 50 mL glycerol based medium (BMGY) and left to grow for 24 hours before being induced with methanol (50mL of BMMY) for expression of the K3 protein. The expression of recombinant protein K3 from each strain was assessed by SDS-PAGE (Figure 3.11). The wild-type K3 migrated at a slower rate than K3 produced in the knock-out strain, suggesting the wild-type K3 has a higher molecular mass. This would be due to K3 from the wild type strain having an oligosaccharide of higher molecular weight. However, one can assume that the knockout strain is producing a $Man_8GlcNAc_2$ until a more accurate assessment of the oligosaccharide chain is completed, such as utilising Mass Spectrometry (MS).

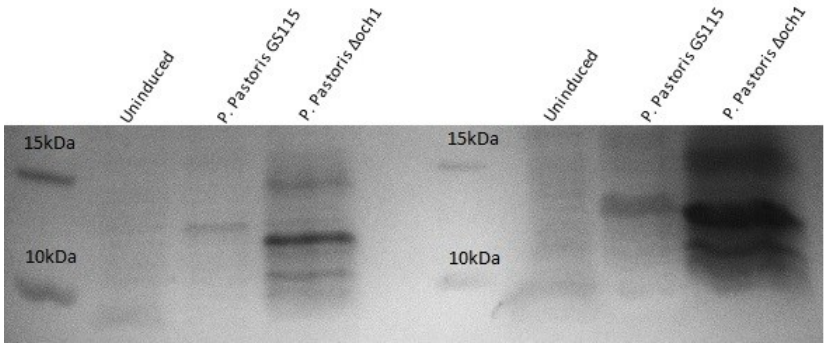


Figure 3.11 Assessing the protein expression of K3 in wild-type *P. pastoris* (GS115) and *P. pastoris* (*Δoch1*).
The *P. pastoris* GS115 strain and *P. pastoris Δoch1* strain containing the recombinant glycoprotein K3, were induced under optimal growing conditions and in the presence of 0.5% methanol. After 24 hours induction cells were harvested by centrifugation and the supernatant removed. A 50 µL aliquot of the supernatant was removed and protein content was assessed using SDS-PAGE. This result is displayed on the left of the stained SDS-PAGE above. The remaining supernatant was concentrated by TCA protein precipitation, following the method outlined in Section 2.7.1.2. The protein content of the concentrated supernatant was assessed using SDS-PAGE, with the result displayed on the right of the stained SDS-PAGE above. The supernatant of an uninduced *P. pastoris* GS115 strain containing the recombinant glycoprotein K3 was also assessed by SDS-PAGE and displayed in the lane to the right of the protein ladder (PageRuler Prestained ThermoFisher). The supernatant of the uninduced *P. pastoris* GS115 strain displayed no observable protein bands. The supernatant of *P. pastoris* GS115 strain displayed one protein band of approximately 12kDa, which is of the expected size of K3 with an attached $Man_{12}GlcNAc_2$. The *P. pastoris Δoch1* displayed a number of bands but with predominant band, which is expected to be K3. The K3 band from the *P. pastoris Δoch1* is of a lower molecular weight then the K3 expressed from *P. pastoris* GS115 strain. This molecular weight difference is assumed to be due to the K3 expressed from the *P. pastoris Δoch1* containing a truncated oligosaccharide of $Man_8GlcNAc_2$, therefore being of a lower molecular weight. Display in this figure is an enhanced close-up the original gel image can be found in the Appendix Section 3.

3.11 The transformation of *P. pastoris* (Δoch1) with a mouse Golgi-Mannosidase for production of core N-glycan structure

To create a *P. pastoris* strain which will produce a $Man_5GlcNAc_2$ glycoprofile the *P. pastoris* (*Δku70:och1)* strain required to be transformed with the *M. musculus* mannosidase 1A (Man1a). The pGEM OCH1:Man1a:G418 plasmid was utilised to introduce the Man1a into the *P. pastoris* (*Δku70:och1)* strain following the same method as described in Section 3.3. The genomic DNA of selected colonies was isolated, with a successful genomic extraction indicated by performing a PCR against the native TEF promoter as described in Section 3.4. The genomic DNA of colonies exhibiting a TEF PCR product were then screened for the presence of Man1a gene. A PCR was performed with primers complementary to Man1a CDS, a successful integration of the Man1a will be indicated by the presence of a 500 bp product (Figure 3.12). A number of colonies were observed to contain Man1a genetic material (Figure 3.12). However, as with colonies screened in Section 3.4 (Figure 3.4) a proportion of colonies which survived the selective pressure of the antibiotic lacked the heterologous gene Man1a.

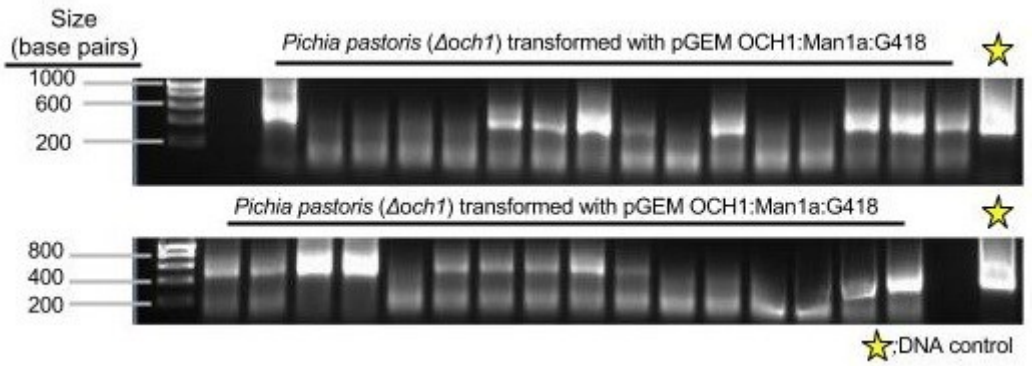


Figure 3.12 PCR analysis of the *P. pastoris Δoch1* strain transformed with pGEM-OCH1:Man1a:G418.
A colony PCR was performed on the selected colonies to check for the presence of the Man1a gene in the genome. The genomic DNA was isolated by the method outlined in Section 2.3.2.3, successful DNA extraction was assessed by performing a PCR against an endogenous target. The PCR analysis of the *P. pastoris* (*Δoch1*) transformed with pGEM OCH1;Man1a;G418 plasmid revealed that only 60% contained the Man1a construct, with the other 40% lacking any PCR product.

Seven colonies of *P. pastoris* which were identified with a Man1a gene, were then tested to ensure that the Man1a was being transcribed. The total RNA was extracted from transformed *P. pastoris* colonies containing a Man1a gene, using the Ambion® RiboPure Yeast Extraction Kit (Life Technologies, CA, USA). The presence of Man1a mRNA was determined by reverse transcription of the total RNA to cDNA using a RT-PCR kit (Life Technologies, CA, USA) following the manufacturer's instructions. The presence of the Man1a transcript was determined by PCR using primers to amplify a 331 base pair product within the Man1a CDS (Figure 3.13). None of the selected colonies which contained the Man1a gene had Man1a transcription. The transcription of housekeeping gene β-actin was also determined from each colony and used as a positive control to demonstrate a successful RT-PCR reaction; all colonies where Man1a CDS appeared showed a β-actin transcription but no Man1a transcription (Figure 3.13).

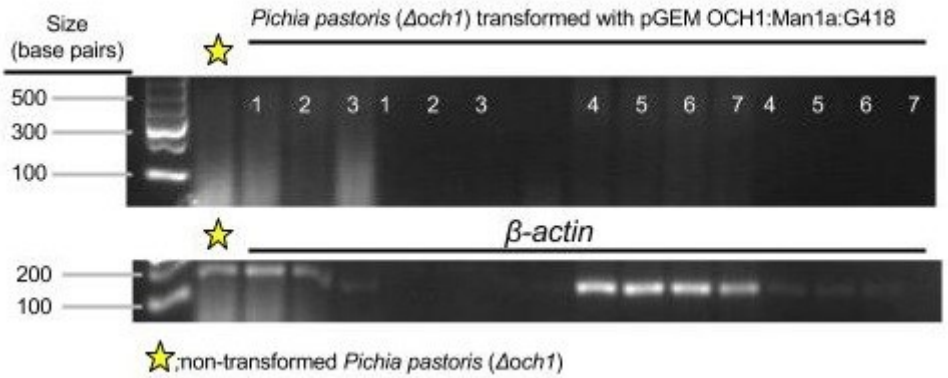


Figure 3.13 RT-PCR analysis of the Man1a transcript from a *P. pastoris* (*Δoch1*) transformed with pGEM OCH1:Man1a:G418 plasmid

The *P. pastoris* strains noted to contain the Man1a gene were analysed by RT-PCR to assess if an mRNA transcript was being produced. In total 7 colonies were examined for Man1a transcription. A 2-step RT-PCR reaction was performed, with the first step converting total RNA to cDNA after which PCR reaction was performed against Man1a and a positive control. The housekeeping gene β-actin was selected as a positive control. Two PCR conditions were utilised, with one test condition having 10 times higher concentration of cDNA. The positive control PCR reaction against β-actin demonstrated that only the reaction condition with greater DNA concentration yielded a PCR product. The PCR reaction against Man1a yielded no observable transcription product, indicating that Man1a mRNA was not being produced. The assessment of transcriptome of a non-transformed *P. pastoris Δoch1* produced no product against primers specific for Man1a.

It is difficult to find an exact reason to why the transcription of Man1a failed, one reason could be due to an off targeted integration event of the Man1a construct. Although not examined in detail here, from the previous integration attempt with the pGEM OCH1:Man1a:G418 (Section 3.4) the occurrence of an off targeted integration events was indicated. The mistargeting of a gene could possible lead to its silencing, hence no Man1a transcript. Although, one could argue that the use *P. pastoris* (*Δku70*) strain should prevent mistargeting, it has to be noted that over a third of transformants in Figure 3.12 contained no Man1a construct. The utilisation of the OCH1 *locus* for the integration site for pGEM OCH1:Man1a:G418 could be considered problematic. It would therefore be more useful to select a proven integration site for recombination, which has previously demonstrated heterologous protein expression.

3.12 Redesign of plasmid for introducing the mammalian Golgi Mannosidase to the *P. pastoris* (*Δoch1*) strain

A number of promoters have been identified and characterised as useful integration sites for recombinant protein expression. The promoter for the translation elongation factor 1α (*Tef1p*) is one of these. The *Tef1p* is one of the most abundant soluble proteins in eukaryotic cells, and functions to deliver amino acyl transfer RNAs (tRNAs) to the ribosome. The high levels of *Tef1p* signifies that the promoter is strong, and as tRNAs are essential throughout each stage of the cell cycle, that this promoter is constitutive and always strongly expressing *Tef1p*. Therefore, a further plasmid was designed which would introduce the Golgi Mannosidase into the TEF locus in the *P. pastoris* genome. This was achieved by isolating the Man1a construct from pGEM OCH1:Man1a:G418 plasmid and cloning this within the plasmid pKAN-B. The plasmid pKAN-B is an expression vector that can be used to select directly for *P. pastoris* transformants using G418 resistance conferred by a modified *Tn903kan*^{r} gene[390].

The Man1a, with *TEF* promoter, BIP signal sequence and HDEL ER retrieval tag was isolated by PCR from the pGEM OCH1:Man1a:G418 plasmid (Figure 3.2). The forward and reverse primers were designed to contain BglII and SacII sites to facilitate restriction cloning of the plasmid. A DNA band of size corresponding to the PCR product was excised from a 1%

agarose gel and purified using the Wizard SV Gel and PCR purification kit (Promega, WI, USA) according to manufacturer's instructions. The PCR product and pKAN-B vector was digested using restriction enzymes BglII and SacII. The compatible ends of the fragments were joined together using a T4 DNA ligase (NEB, MA, USA), to create the vector which will be known as pKANB TefP:Man1a (Figure 3.14) and plasmid sequence was verified (Appendix Section 2).

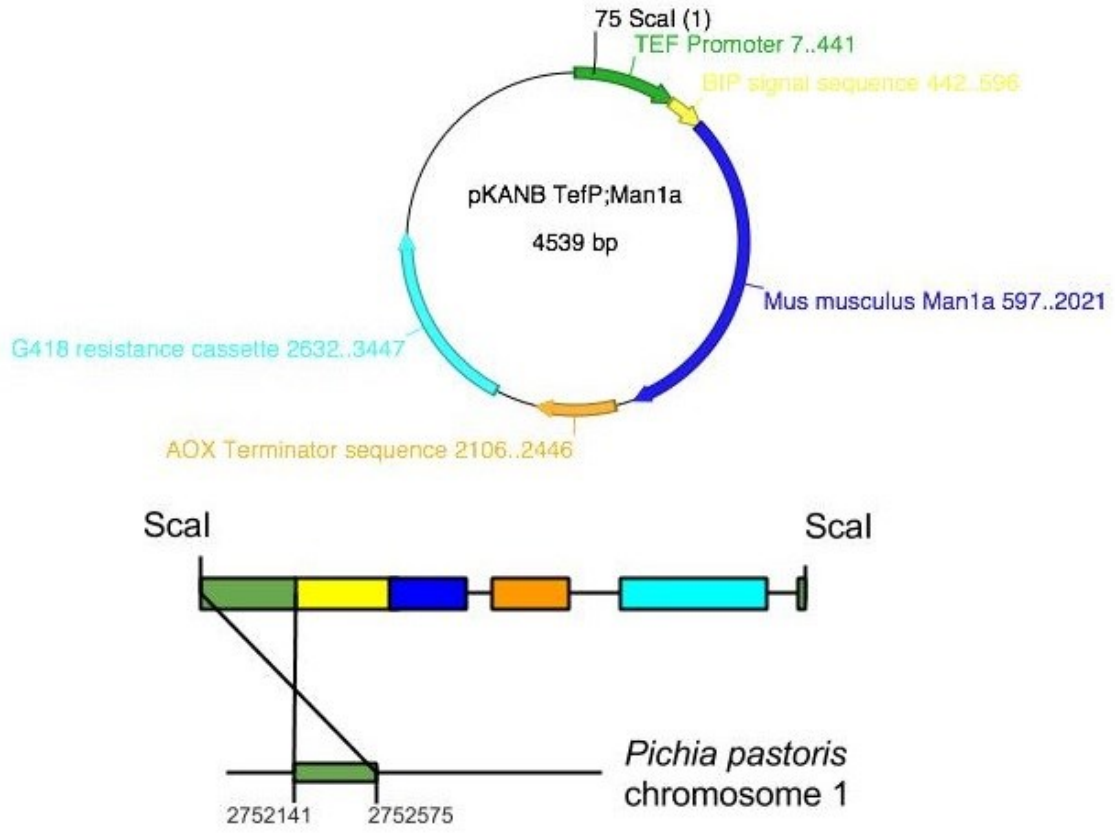


Figure 3.14 Schematic of the pKANB TefP:Man1a plasmid.
The exogenous DNA cassette consisted of an ER localisation signal sequence from the BIP gene (M31006.1) upstream of the Man1a gene, and an ER retrieval sequence HDEL downstream. This ensures that mannose cleavage occurs in the ER. Downstream of the Man1a gene is a terminator sequence from AOX1 to ensure the termination of transcription. To aid cell line selection a genetic cassette containing the bacterial gene Tn903kanr is used for selection against the antibiotic Geneticin® (G418)[390].The Man1a is transcribed by a strong constitutive promoter TEF. An enzyme restriction site within the TEF promoter (ScaI) will be utilised to linearize the plasmid, enabling the homologous recombination event between endogenous DNA of the TEF promoter and P. pastoris genome. The location of genomic integration site is shown in the figure above.

The design of the pKANB TefP:Man1a plasmid will enable a gene insertion event at the *TEF* locus to arise from a single crossover event between the locus and the *TEF* promoter on a vector. Previous efforts to introduce the Man1a to *P. pastoris* genome by HR have failed to target the Man1a cassette to the *OCH1* locus. The introduction of Man1a cassette was designed for a double crossover event. As single crossover events (insertions) are much more

likely to happen than double crossover events (replacements)[87], this plasmid was designed to insert the Man1a into the *P. pastoris* genome using a single crossover event. An insertion can also result in one or more copies of the vector upstream of the *TEF* promoter. As stated multiple gene insertion events at a single locus in a cell do occur spontaneously, but these are often at a low but detectable frequency of between 1 and 10%[404]. Vector linearized within the *TEF* promoter region will integrate the Man1a gene by insertion into the host *TEF* region. This can also occur with non-linearized plasmid and plasmid that religates, although at a lower frequency[409]. Multiple gene insertions are linked to increase translation, due to an increase in the number of copies of the recombinant gene[401].

3.13 The transformation of *P. pastoris (Δoch1)* with pKAN-B Man1a plasmid

The pKAN-B Man1a plasmid was linearized using ScaI restriction site which is 75 base pairs from the beginning of the *TEF* promoter region. A 500 mL culture of *P. pastoris* (*Δku70:och1*) was prepared for electroporation as described in the *Pichia* Expression kit manual and transformed with linearised pKAN-B TefP;Man1a plasmid. The transformant culture of *P. pastoris* (*Δoch1*) was plated on YPD containing 350 μg mL^{-1} of G418 and incubated at 30°C for 3-5 days.

Colonies were randomly selected and grown overnight in 5 mL of YPD culture containing 350 μg mL^{-1} of the antibiotic G418 in a 50mL conical tube. A PCR was performed on the selected colonies to check for the presence of the Man1a gene in the genome (Figure 3.15). The PCR reaction demonstrated that all colonies contained a copy of the Man1a gene. This 100% success rate of transformant is in stark contrast to the previous attempt with the pGEM OCH1;Man1a;G418 where only 60% contained the Man1a gene (Figure 3.12). This highlights the difficulty in utilising the homologous regions of the *OCH1* for integration, possibly due to the design for double crossover event, and that this region of the chromosome is troublesome for integration.

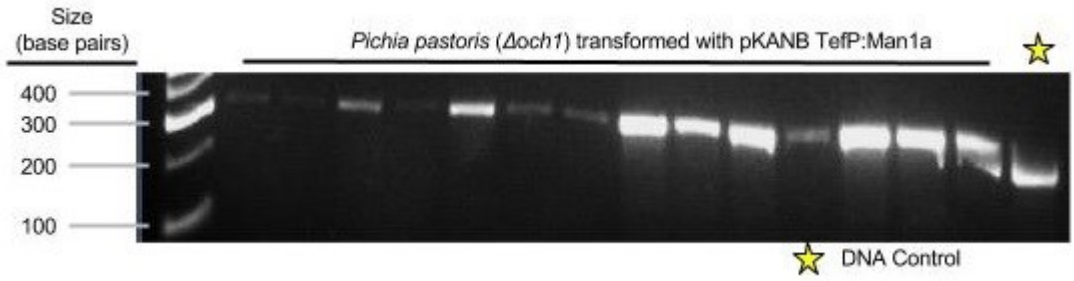


Figure 3.15 PCR analysis of the *P. pastoris* (Δoch1) transformed with pKAN-B Man1a plasmid

A colony PCR was performed on the selected colonies to check for the presence of the Man1a gene in the genome. The genomic DNA was isolated by the method outlined in Section 2.3.2.3, successful DNA extraction was assessed by performing a PCR against an endogenous target. The PCR analysis of the P. pastoris (Δoch1) transformed with pKAN-B Man1a plasmid revealed that all transformants contained the Man1a construct.

As before the transcription of the Man1a mRNA was determined using RT-PCR from selected colonies with a Man1a gene, following the same method as Section 3.11. The presence of the Man1a transcript was determined by PCR using primers to amplify a 331 bp product within the Man1a CDS (Figure 3.16). The transcription of a constitutively expressed gene (β-actin) was determined as a positive control to ensure RNA extraction and RT-PCR was successful only 3 of the 4 colonies returned a product for β-actin. From the 3 successful RT-PCR reaction no Man1a transcription were found to be present (Figure 3.16)

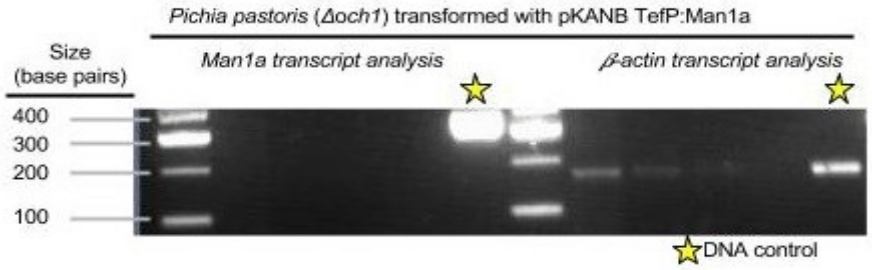


Figure 3.16 RT-PCR of the Man1a transcript from a *P. pastoris* (Δoch1) transformed with pKAN-B Man1a plasmid
The *P. pastoris* strains noted to contain the Man1a gene were analysed by RT-PCR to assess if an mRNA transcript was being produced. In total 4 colonies were examined for Man1a transcription. A 2-step RT-PCR reaction was performed, with the first step converting total RNA to cDNA after which PCR reaction was performed against Man1a and a positive control. The housekeeping gene β-actin was selected as a positive control. The positive control PCR reaction against β-actin demonstrated that only three of the four yielded a PCR product. The PCR reaction against Man1a yielded no observable transcription product, indicating that Man1a mRNA was not being produced. A positive control (indicated by the yellow star) of a plasmid containing a Man1a gene was utilised for size comparison.

3.14 Detecting the *TEF1* transcript in *P. pastoris* (*Δku70*) and *P. pastoris* (*Δoch)* strains

There are two possibilities of why there is no PCR product observed when detecting the Man1a transcript by RT-PCR; either that no transcription of the Man1a is occurring or it is at an undetectable level. As stated the *TEF* promoter is a strong constitutive promoter controlling the expression *Tef1p*, so if functional should produce a detectable level of transcript. However, to conclusively show this, the transcription of *Tef1p* in *P. pastoris* (*Δku70*) and *P. pastoris* (*Δku70:och1)* strains was determined by RT-PCR. If *Tef1* transcript is detected then it would suggest that Man1a should also be detectable.

A colony of *P. pastoris* (*Δku70*) was grown overnight in 5 mL of YPD culture. The total RNA was extracted from *P. pastoris* as described in Section 2.3.24. The presence of the *Tef1* transcript was determined in the *P. pastoris* (*Δku70*) and *P. pastoris* (*Δku70*:*Δoch)* created above by PCR using primers to amplify a 204 base pair product within the *Tef1* CDS (Figure 3.7). The *Tef1* transcript was present within both strains *P. pastoris* (*Δku70*) and *P. pastoris* (*Δku70*:*Δoch1)* (Figure 3.17). This shows that Man1a is not being transcribed in the *P. pastoris* (*Δoch1).*

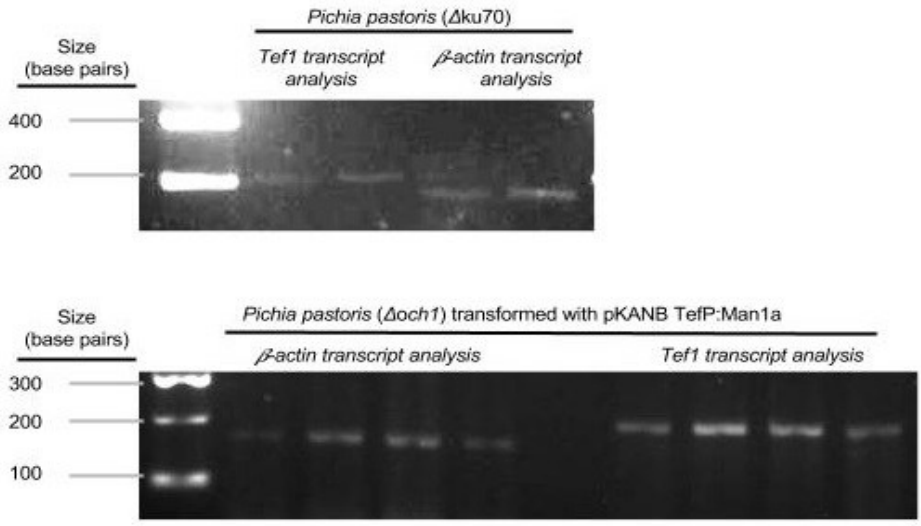


Figure 3.17 RT-PCR analysis of the Tef1 transcript from a *P. pastoris* (Δku70) and *P. pastoris* (Δoch1)
A 2-step RT-PCR reaction was performed, with the first step converting total RNA to cDNA after which PCR reaction was performed against TEF transcript. All tested clones exhibited a PCR product of expected size, indicating that transcription from the TEF promoter is observable using RT-PCR techniques. The housekeeping gene β-actin was selected as a positive control to indicate a successful RT-PCR reaction. All test clones contained a PCR product against β-actin.

3.15 Results Summary

The generation of a *P. pastoris* strain which produces a recombinant glycoprotein with the core consensus glycoprofile ($Man_5GlcNAc_2$) was attempted within this chapter. This involved the generation of several DNA plasmids for genetic engineering of a *P. pastoris* strain. The removal of the host mannosyltransferase *och1p* and addition of a mammalian mannosidases has previously been shown to synthesise an N-linked glycoproteins with a core $Man_5GlcNAc_2$ glycoprofile[285,288].

The first plasmid developed in the study aimed to remove *OCH1* and introduce the Man1a in the same homologous recombination (HR) event. This strategy differed from the other published strategies which first disrupted the *OCH1* gene before integrating an α1,2 mannosidase (Man1a) gene. Although screening of *P. pastoris* colonies demonstrated the presence of the Man1a gene, an intact *OCH1* gene was also present. It is generally accepted

that double crossover events in *P. pastoris* are much rarer than a single crossover event[87]. Vervecken *et al.* was unable to obtain a *P. pastoris Δoch1* strain by a double crossover event even with homologous regions that contained up to 3 kb of isogenic genomic DNA[295]. To put that in perspective, *S. cerevisiae* can perform HR events with homologous regions containing as little as 50 bp[410]. Vervecken *et al.* screened several hundred clones, as well as using number of selective markers and were still unable to generate a *P. pastoris Δoch1* strain via a double crossover event[295]. The Vervecken *et al.* group eventually achieved a knockout strain by single crossover event[295,99]. Although a research group from Dartmouth College led by Dr Gerngross did achieve a *Δoch1* strain using a double HR strategy[411,293]. Similar to the initial developed plasmid of this study, the *OCH1* homologous region on the Gerngross *et al.* plasmid was located upstream and downstream of the *OCH1* ORF. The Vervecken *et al.* isogenic DNA was between the *OCH1* promoter and *OCH1* ORF. Gerngross *et al.* screened 1000 colonies, to achieve one clone that exhibited the expected $Man_8GlcNAc_2$ glycoprofile[293]. It is difficult to know why Gerngross *et al.* were able to achieve a *Δoch1* strain using a double homologous recombination over others who tried, including the attempt in section 3.3. One could postulate that it was mere luck that one clone of the 1000 screened had the desired phenotype. Whatever the reason, the 1 in 1000 success rate, demonstrating an efficiency rate of 0.1%, is clearly a poor accomplishment rate. To improve upon this success rate two procedural changes were implemented. The first being the utilisation of a *P. pastoris Δku70* strain[404] to improve the efficiency of HR event and the second designing a plasmid with *OCH1* homologous sequence within the ORF.

The homologous regions to *OCH1* in the second plasmid were designed for a double-crossover event, however, if a single crossover occurred this would still cause disruption within the *OCH1* ORF preventing *och1p* translation. By utilising this strategy of homologous regions within the *OCH1* ORF and the *Δku70* strain 100% of tested colonies demonstrated the knockout genotype, this is compared the 0% from the strategy of homologous regions outside the *OCH1* ORF and *GS115* strain. The colonies of *Δoch1* knockout strain were also observed to have a morphological change similar to that described by Krainer *et al.*[408] This phenotypic change in the colony appearance was a clear indication of successful removal of the *OCH1* gene. PCR and Sanger sequencing of the *OCH1* locus confirmed a *Δoch1* genotype. The molecular weight of a glycoprotein produced from the *Δoch1* knockout strain also shifted, highlighting that mass from oligosaccharide was removed to possibly denoting a $Man_8GlcNAc_2$ glycoprofile. A future mass spectrometry analysis of the oligosaccharide will be required to verify the $Man_8GlcNAc_2$ glycoprofile.

Once the *Δoch1* knock-out strain was genetically confirmed the introduction of the Man1a gene was attempted. Two plasmids were designed for the introduction of the Man1a gene into

the *Δoch1 P. pastoris* strain, one which allowed for a double homologous recombination and the other for a single homologous recombination. Although the Man1a gene was observed by PCR in colonies from both recombination strategies, no tested clone produced a visible Man1a transcript by RT-PCR. Again, it is difficult to understand why the transcription of Man1a gene did not occur, for this is not the first time that a *M. musculus* Mannosidase 1A has been used to generate a $Man_5GlcNAc_2$ in *P. pastoris* (*Δoch1*). Yet there are some notable differences between the strategy here and that of successfully published study. One is with the architecture of the Man1a protein, the published study contained a TMD, where this study utilised a soluble enzyme. Whether this structural difference is a problem is difficult to know as the problem does not lie with translation or lack of enzyme activity but with absence of gene transcription. Therefore, one must understand the genetic differences between the two exogenous genes.

One difference is the promoter driving transcription, in this study this is *TEF* promoter, whereas previous published study utilised the *GAP* promoter. The *GAP* promoter controls the expression of Glyceraldehyde-3-phosphate dehydrogenase (GAPDH), a key enzyme in glycolysis[412]. The *GAP* promoter also provides strong constitutive expression of recombinant genes, although previous studies have demonstrated that the *TEF* promoter is as strong if not stronger than the *GAP* promoter, in one case producing double the amount of recombinant lipase[395]. However, Northern blots revealed a growth associated expression pattern of the *TEF* promoter, resulting in high TEF mRNA levels during exponential growth and reduced levels in the stationary phase[395]. Therefore, this will result in varying transcriptional levels of the TEF controlled heterologous gene during different stages of cultivation. In this study, the transcription levels of Man1a were determined during the stationary phase of growth, therefore one hypothesis could be that the lack of Man1a mRNA was due to determining transcription at this stage, rather than during exponential phase. However, previous studies did not demonstrate a complete halt in mRNA production during stationary phases, and TEF mRNA was also detected in this study during stationary phase (Figure 3.17). Nevertheless, it may be wise to use a promoter which does not demonstrate cell cycle related activation. An extensive list of promoters have been identified and characterized for heterologous expression in *P. pastoris*[413], it is possible that one of these promoters could increase the likelihood of Man1a transcription/translation.

Chapter 4. Towards the development of an *in vitro* metabolic pathway for the processing of a pentamannosyl glycoprofile

4.1 Introduction

In chapter 3, attempts were made to engineer a strain of *P. pastoris* for the synthesis of a $Man_5GlcNAc_2$ core glycoprofile. In higher eukaryotes, this oligosaccharide serves as the substrate for the biosynthesis of a hybrid or complex type glycan. The $Man_5GlcNAc_2$ is processed in the Golgi by an array of glycosyltransferases that assemble different monosaccharide moieties into linear and branched glycan chains. Although the N-linked glycosylation pathway is often diagrammatised as a sequential pathway, as exemplified in Figure 1.2, this is rarely the case. Often the final end glycoform is accomplished through a large reaction network, as illustrated in Figure 4.1. The branched series of reactions exists due to the differences observed in enzyme processing kinetics which affects oligosaccharide processing and therefore end glycoforms. For instance, the reaction scheme for the generation of a biantennary terminally sialylated N-glycan structure with core fucose and bisecting GlcNAc, has been modelled to consist of 95 reactions when starting from a high mannose structure ($Man_9GlcNAc_2$). The reaction network proceeds through eight different enzymes and a possible 77 glycoforms[414]. If the reaction network commences through the pentamannosyl glycoform the reaction scheme reduces to 78 reactions with 61 glycoforms[415].

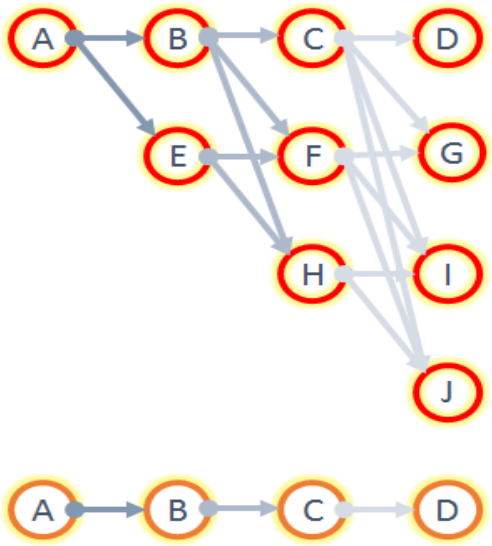


Figure 4.1 Diagram of an enzymatic reaction network.
At the top of the diagram represents a non-sequential reaction network while the bottom demonstrates a sequential pathway. The circles indicate substrates and/or products, and the arrowed nodes represent a reaction which is catalysed by an enzyme.

The non-sequential processing pathway is a consequence of the asymmetric distribution of Golgi glycosylation enzymes. Although it has been shown that glycosylation enzymes are distributed through the Golgi in the approximate order in which they process the N-glycan, variations and overlap in location is also often witnessed[416]. This dynamic enzyme distribution leads to competition over substrates. Furthermore, a resulting oligosaccharide from monosaccharide transfer reaction can inhibit downstream processing enzymes. For instance, β-1,4-mannosyl-glycoprotein 4-β-Nacetylglucosaminyltransferase (GlcNAcT-III) adds a bisecting GlcNAc to N-linked oligosaccharide, the presence of which inhibits the activity of many other enzymes (Man-II, GlcNAcT-I, GlcNAcT-IV, GlcNAcT-V and Fuc-T), therefore preventing further processing of oligosaccharide[417]. These factors can result in partially processed oligosaccharides, and therefore glycoform heterogeneity.

As previously mentioned glycoform heterogeneity can be problematic when synthesising therapeutic glycoproteins (see Section 1.21 for some examples). Tight control over the manufacturing process for biopharmaceuticals, aims to alleviate glycan heterogeneity and produce a glycan chain(s) of a defined isotype. Parameters such as nutrient availability, dissolved oxygen, temperature, and pH can affect the quality of a product and glycoform[418]. The regulation of parameters within a given range ensures end-product quality and near identical glycan chain heterogeneity. A limitation to producing a glycoprotein by these means is that the glycan chain produced may not be the optimal isotype for a desirable therapeutic effect.

The removal of the editing of a glycan chain by *in vivo* processes and into an *in vitro* system would allow for the implementation of a spatially sequential reaction, thus enabling control over the end oligosaccharide composition. A sequential pathway is what one witnesses through the glycoengineering efforts of non-mammalian expression hosts. That is, the engineered *de novo* glycosylation pathway is directed solely to the generation of a defined glycosylation chain. However, the achievement of full conversion of substrate to product does not always occur (see section 1.27 for examples). An *in vitro* reaction, where one can control reaction conditions will enable a near 100% conversion each time.

The concept of *in vitro* enzyme reaction was the basis of an artificial Golgi apparatus prototype, where by using a digital microfluidics researchers modified the glycan chain on the glycosaminoglycan, heparan sulphate (HS)[419]. By embedding recombinant HS-modifying enzymes into a microfluidic device, researchers could specifically modify the HS glycan chain to a defined chemical structure. The concept of the HS artificial Golgi can be applied to an *in vitro* metabolic pathway for the controllable synthesis of N-linked glycosylation chains.

The use of an artificial Golgi for the modification of N-linked glycosylation chains has not previously been attempted. Although previously individual N-linked glycosylation processing enzymes have been utilised for a single *in vitro* reaction, the undertaking of a metabolic pathway for glycan remodelling has never been tried. In the previous chapter, attempts were made to generate a system for the production of the key intermediate structure; $Man_5GlcNAc_2$. This core structure can be modified through an *in vitro* metabolic pathway to generate a complex or hybrid glycan structure. The use of an *in vitro* metabolic pathway will enable a sequential N-linked glycosylation pathway, permitting a user-defined glycoprofile. This chapter will work towards the development of an *in vitro* metabolic pathway which would process the $Man_5GlcNAc_2$ to a complex structure.

To assess the merits of an *in vitro* metabolic pathway a proof of concept system was devised with the first three enzymes involved in synthesis of the complex structure N-acetylglucosaminyltransferase-I (GlcNAcT-I) (NP_002397), N-acetylglucosaminyltransferase-II (GlcNAcT-II) (NP_002399) and type II α-mannosidase (Man-II) (NP_002363). The GlcNAcT-I is medial-Golgi resident which transfers a GlcNAc residue to the α1–3 mannose in the $Man_5GlcNAc_2$ glycan structure[420]. In the majority of cases, a medial-Golgi based Man-II enzyme removes the terminal α1-3 mannose and α1-6 mannose sugar residues[421]. This allows for the addition of a second GlcNAc residue to the terminal mannose by the actions of the GlcNAcT-II enzyme, giving way to complex type glycan structure the biantennary $GlcNAc_2Man_3GlcNAc_2$ (Figure 4.2).

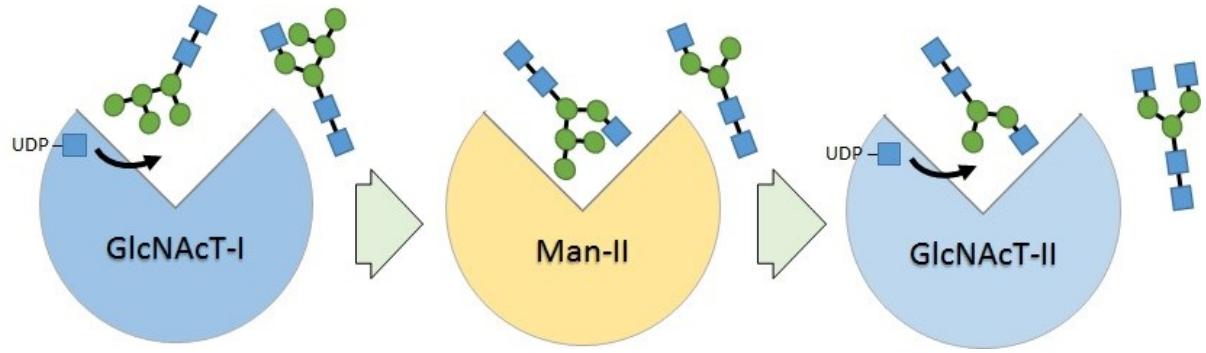


Figure 4.2 A sequential pathway for the generation of $GlcNAc_2Man_3GlcNAc_2$ glycoprofile
The figure displays an *in vitro* metabolic pathway containing the first three enzymes involved in the establishment of a complex N-glycan profile. The recombinant enzymes will be His-tagged therefore have potential to be immobilized in individual environments, to enable a sequential pathway. A glycoprotein produced with pentamannosyl structure of $Man_5GlcNAc_2$ can pass through *in vitro* metabolic pathway. The sequential addition/removal of monosaccharides which will render a glycoprotein with a $GlcNac_2Man_5GlcNac_2$ profile.

The human GlcNAcT-I had previously been expressed as a maltose binding protein (MBP) fusion protein in inclusion bodies (IBs) in *E. coli* (JM109) and refolded using an oxido-shuffling

method to achieve an active enzyme[383]. However, there are risks associated with this strategy such as misfolding of the protein into a non-natural configuration. This strategy increases the tendency for protein aggregation and generally results in a lower protein titre when compared to producing a soluble protein. The expression of human GlcNAcT-II has previously been achieved in the baculovirus/Sf9 insect cell system[422]. In this case the recombinant GlcNAcT-II gene was cloned from cDNA of HL-60, a human promyelocytic leukaemia cell line[423]. As discussed in the introduction (Section 1.15), insect cells are slow growing, requiring expensive media, and produce lower titres of protein in comparison to other expression hosts. Therefore, the expression of human GlcNAcT-II was attempted in *E. coli* within the study. A literature search showed that the expression of a recombinant human Man-II had only been performed from mammalian cells (COS cells[424], HEK293 cells[425]). Similarly, to insect cell lines, mammalian cells are recognised as economical expensive to maintain, grow slowly and produce low titres of recombinant protein. Therefore, this study offered the first that a human Man-II would be recombinantly expressed in a non-mammalian host.

4.2 The synthesis of N-acetylglucosaminyltransferase-I, N-acetylglucosaminyltransferase-II and type II α-Mannosidase genes

The genes for expression of human recombinant proteins GlcNAcT-I, Man-II and GlcNAcT-II, were all synthesised by DNA2.0. (CA, USA). The genetic material was codon optimised by DNA2.0 for expression in *E. coli* and cloned into an *E. coli* expression vector featuring an Isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter and a Kanamycin selection marker.

As previously mentioned, the enzymes of the N-linked glycosylation pathway are classified as type-II transmembrane proteins[200]. Therefore, the three enzymes share the same structural features, of a short amino-terminal cytoplasmic tail, followed by a transmembrane domain (TMD), a short stem region and finally a large carboxy-terminal catalytic domain in the lumen of the Golgi[1]. The amino acid residues of TMDs are characterised with having a hydrophobic nature; this feature helps stabilise and embed transmembrane proteins into the lipid membrane bilayer. However, the hydrophobicity can also create problems during recombinant expression, such as protein aggregation and misfolding, which often leads to the formation of inclusion bodies (IB). Several studies have demonstrated the low, functional yields of mammalian membrane proteins from *E. coli*[426,427]. However, as only the soluble carboxyl catalytic domain is required for glycan remodelling, the hydrophobic regions of each enzyme can be removed, this should help promote successful soluble expression in *E. coli*.

The genes were truncated at sites specified in the paper of Nett *et al.*, this removed the cytosolic, TMD and stem regions (Table 4.1)[294]. Nett *et al.* required the removal of these regions in order to replace them with same structure features from yeast to ensure localisation to the *P. pastoris* Golgi membrane[294]. In this case, the three recombinant genes were designed with the addition of an N-terminal polyhistidine-tag (His-tag) sequence. The His-tag will aid in separation of recombinant proteins from host cell proteins (HCP) through affinity chromatography. This is achieved by using a chelating metal cation of Nickel-nitrilotriacetic acid (Ni^{2+}-NTA), which strongly binds histidine residues[428,429]. The purification of His-tagged fused proteins is accomplished using strong concentrations of imidazole which displaces the protein from Ni^{2+}-NTA.

Table 4.1 truncation details of GlcNAcT-I, GlcNAcT-I and Man-II.
The truncated forms of the first three Golgi based enzymes involved in the production of a complex oligosaccharide for N-linked glycosylation, were synthesised by DNA2.0. Proteins were designed to have cytosolic, TMD and stem regions, with the truncated site denoted below, this site was selected based on research by Nett *et al.*[294]

Enzyme Name	Theoretical weights of proteins (kDa)	Length of polypeptide chain	Site at truncation	Theoretical weights of trunc. proteins in kDa
GlcNAcT-I	50 kDa	445 amino acids (aa)	38 aa	46 kDa
GlcNAcT-II	51 kDa	447 aa	84 aa	42 kDa
Man-II	131 kDa	1144 aa	99 aa	122 kDa

The binding properties of His-tag for Ni^{2+}-NTA can facilitate the immobilisation of fused-proteins, in similar scenario to that of the HS-modifying artificial Golgi described above (Section 4.1). The N-terminal His-tag essential acts as a TMD, immobilising the enzyme onto the Ni^{2+}-NTA surface. An immobilized enzyme can be supported within a customised environment which will have the optimal reaction conditions for activity. The glycoprotein substrate can then flow through the system, similar to a plug flow reactor, enabling the sequential addition/removal of monosaccharides, to render a glycoprotein with a defined glycoprofile.

Upon arrival of the synthesised genes, the plasmids were transformed into chemically competent *E. coli* (Life Technologies, CA, USA) by heat shock transformation. The *E. coli* transformants were streaked onto LB agar containing 30 μg mL^{-1} of Kanamycin (Sigma-Aldrich Corp. MO, USA) and incubated at 37°C overnight. Individual colonies were selected at random

and grown in 5 mL of LB media containing 30 µg mL^{-1} of Kanamycin (Sigma-Aldrich Corp. MO, USA) and incubated at 37°C overnight with shaking. The expression vectors were isolated from *E. coli* using a plasmid preparation kit following the manufacturer's instructions (PEQLAB GmbH, Germany).

4.3 Preliminary expression of N-acetylglucosaminyltransferase-I, N-acetylglucosaminyltransferase-II and type II α-Mannosidase in *E. coli* strain BL21 (DE3)

For the expression of the recombinant proteins, the plasmids were transformed into *E. coli* strain BL21 (DE3) (NEB, MA, USA) by heat shock transformation. The transformants were streaked onto a LB agar plates containing 30 µg mL^{-1} of Kanamycin (Sigma-Aldrich Corp. MO, USA) and incubated overnight at 37°C. A single colony was selected and resuspended in 5 mL of LB culture with 30 µg mL^{-1} of Kanamycin and incubated overnight at 37°C with shaking to produce a starter culture.

The starter culture was used to inoculate at a 1:100 dilution into 100 mL of LB media containing the 30 µg mL^{-1} of Kanamycin in a conical flask. The culture was incubated at 37°C with shaking until OD_{600} reached 0.4–0.8 and the culture was then induced with 500 µM IPTG (Sigma-Aldrich Corp. MO, USA). After induction the culture was incubated with shaking for 3 hours at 37°C, 5 hours at 30°C or overnight at 16°C. The expression was checked by SDS-PAGE and Western blot in both the total cell extract (soluble + insoluble) and the soluble fraction alone. An antibody with specificity for the His-tag (BioLegend, CA, USA) was used for the Western blot. No visible expression was observed for GlcNAcT-I and Man-II for 500 µM IPTG (Appendix Section 4). Expression of GlcNAcT-II was observed under all conditions (Figure 4.3).

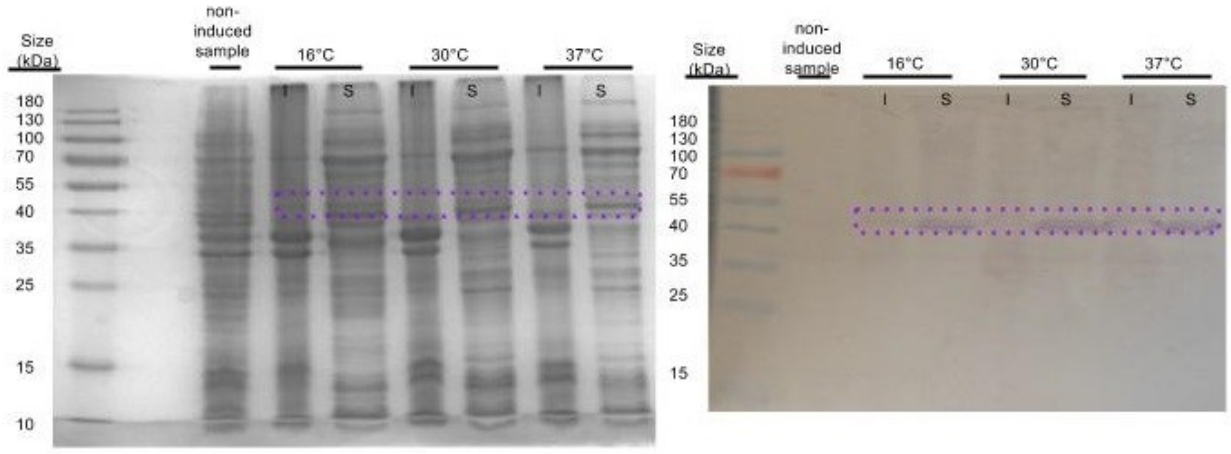


Figure 4.3 Expression of GlcNAcT-II.
A Coomassie stained SDS-PAGE and Western Blot showing [500µM] IPTG induced expression from *E. coli* strain BL21 containing human GlcNAcT-II recombinant gene in LB media at three temperatures (16°C, 30°C and 37°C). Both the Cell fraction (soluble/insoluble) denoted with the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. A protein

band of expected size 42 kDa was observed in all induced samples on the SDS-PAGE. A western blot confirmed the presence of His-tagged protein at 42 kDa. No band was present within a non-induced sample at 42 kDa. From the SDS-PAGE the optimal conditions for expression of GlcNAcT-II appeared to [500μM] IPTG, LB media at 37°C.

An optimisation procedure for protein expression was performed on the BL21 strains expressing GlcNAcT-I and Man-II, including the use of lower concentration (50 μM) of chemical inducer IPTG, lower expression temperatures of 16°C and 30°C, and expression in more nutritious media (2xYT media or TB media). Despite altering numerous expression conditions no visible expression was observed for GlcNAcT-I and Man-II (data not shown).

4.4 The attempted expression of N-acetylglucosaminyltransferase-I and type II α-Mannosidase in *E. coli* strain SHuffle®

The GlcNAcT-II was successfully expressed within *E. coli* strain BL21 (DE3), this strain is routinely used for recombinant protein expression and was seen as a suitable candidate for protein expression. However, with this strain GlcNAcT-I or Man-II could not be expressed. One of the major differences between GlcNAcT-II and the two unexpressed proteins is in the number of disulphide bonds. The GlcNAcT-I and Man-II contain 2 and 3 disulphide bonds respectively[383,421], whereas GlcNAcT-II is predicted from database searches to contain no disulphide bonds. The *E. coli* BL21 (DE3) strain can have difficulty in forming disulphide bridges, hence why researchers have engineered strains such as the *E. coli* SHuffle® T7, which as discussed in Section 1.4 has been modified to promote disulphide bond formation. The *E. coli* SHuffle® T7 strain was successfully utilised for the translation of GlcNAcT-I and Man-II.

Similar to above the GlcNAcT-I and Man-II expression plasmids were transformed into *E. coli* SHuffle® T7 strain. The transformed strains were grown with (50 μM or 500 μM) of chemical inducer IPTG, expression temperatures of 16°C and 30°C, and expression either LB media, 2xYT media or TB media). The expression was checked by Coomassie stained protein gel and Western blot in both the total cell extract (soluble + insoluble) and the soluble fraction alone. The GlcNAcT-I and Man-II were successfully expressed under all conditions. Figure 4.4 and Figure 4.5 represent the optimal expression conditions for GlcNAcT-I and Man-II, example results for other expression conditions can be found in the Appendix Section 4.

In addition to the recombinant proteins of GlcNAcT-I and Man-II, a number of nonspecific bands were observed from the analysis by Western blot. These contaminating bands could be nonspecific endogenous proteins, where if native protein contains a histidine-repeat sequence the primary antibody will form an interaction. However, these contaminating bands could also be result of the recombinant protein being cleaved by endogenous proteases of *E.*

coli, with these bands being truncated forms of GlcNAcT-I and/or Man-II. The latter appears more probably due to the lack of unspecific bands from GlcNAcT-II expression (Figure 4.3) and within some lanes of the GlcNAcT-I lanes (Figure 4.4); as if nonspecific bands were from endogenous proteins the presence of these bands would be within all lanes. Equally the nonspecific bands are all of size below the expected size of the recombinant protein, which would be result of a cleaved recombinant protein.

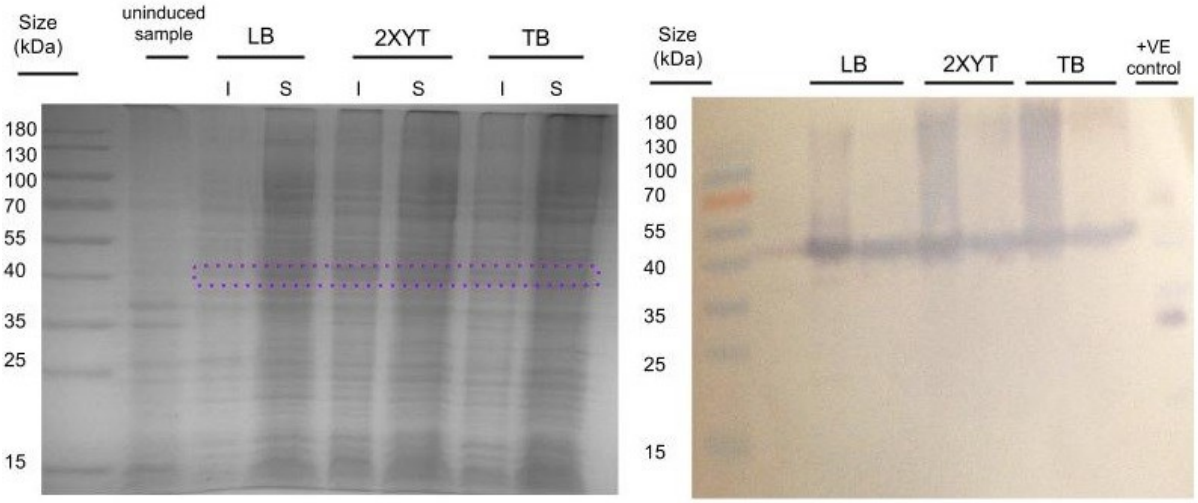


Figure 4.4 expression of GlcNAcT-I.
A Coomassie stained SDS-PAGE and Western Blot showing [500μM] IPTG induced expression from *E. coli* strain Shuffle® containing human GlcNAcT-I recombinant gene at 30°C, in three different media conditions (LB, 2XYT and TB media). Both the Cell fraction (soluble/insoluble) denoted with the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. A protein band of expected size 46 kDa was observed in all induced samples on the SDS-PAGE. A western blot confirmed the presence of His-tagged protein at 46 kDa. No band was present within a non-induced sample at 46 kDa. From the western blot the optimal conditions for expression of GlcNAcT-I appeared to [500μM] IPTG, 2xYT media, at 30°C. The SDS-PAGE and Western blot for expression at 16°C can be found in the Appendix Section 4.

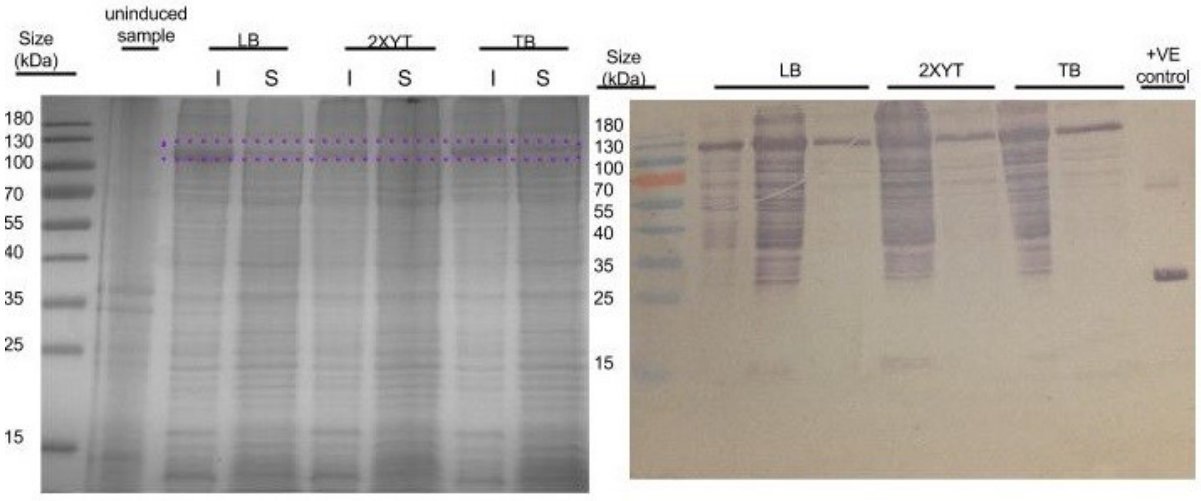


Figure 4.5 expression of Man-II.
A Coomassie stained SDS-PAGE and Western Blot showing [500μM] IPTG induced expression from *E. coli* strain Shuffle® containing human Man-II recombinant gene at 16°C, in three different media conditions (LB, 2XYT and TB media). Both the Cell fraction (soluble/insoluble) denoted with

the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. A protein band of expected size 122 kDa was observed in all induced samples on the SDS-PAGE. A western blot confirmed the presence of His-tagged protein at 122 kDa. No band was present within a non-induced sample at 122 kDa. From the western blot the optimal conditions for expression of GlcNAcT-I appeared to LB media at 16°C. The SDS-PAGE and Western blot for expression at 30°C can be found in the Appendix Section 4.

4.5 The expression, purification and characterisation of N-acetylglucosaminyltransferase-I

The optimal expression conditions for recombinant GlcNAcT-I were found to be within an *E. coli* SHuffle® T7 strain at 30°C in 2XYT media (Figure 4.4). These expression conditions were repeated for a large-scale expression, in 500 mL of media. After induction *E. coli* cells were harvested by centrifugation, the supernatant was removed and the cells were resuspended within a lysis buffer at 0.5 mL buffer per 1 gram of wet weight of cell. The cells were disrupted using sonication, and cell debris was removed by harvesting the cell lysate mixture by centrifugation. The recombinant GlcNAcT-I was purified from the supernatant by gravity-flow affinity chromatography utilising the polyhistidine tag. The supernatant containing GlcNAcT-I was added to a Ni-NTA agarose (QIAGEN, Germany) and the recombinant protein was purified from HCP following the manufacturer's instructions. The chromatography samples were analysed by SDS-PAGE (Figure 4.6).

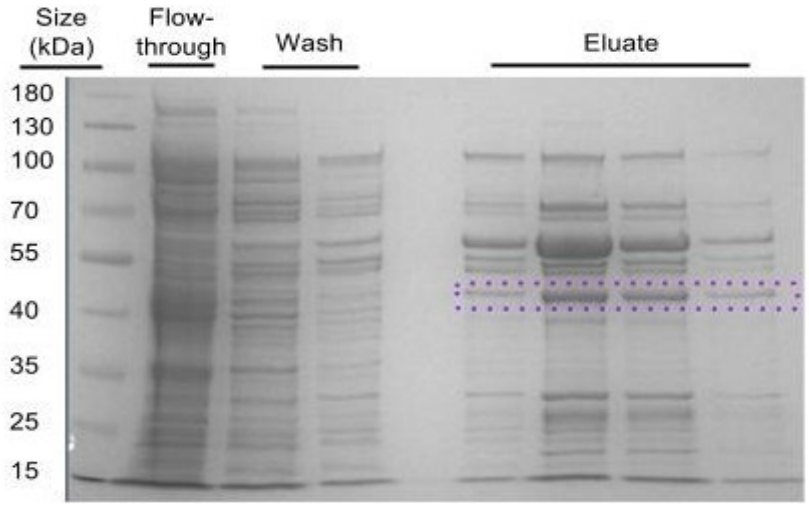


Figure 4.6 SDS-PAGE of a His-tag purification of GlcNAcT-I.
The heterologous expression of human GlcNAcT-I was performed in SHuffle® T7 strain at 30°C in 500 mL of 2xYT media and induced with [500μM] of IPTG. After 4 hours of induction cells were harvested and lysed by sonication. The GlcNAcT-I was expressed with a N-terminus His-tag, which enabled purification by affinity chromatography. The recombinant GlcNAcT-I was subjected to a purification using Ni-NTA Agarose (QIAGEN, Germany). After washing, His-tagged proteins are eluted in buffer under native conditions. Chemical composition of buffers was per the manufacturer's instructions.

A protein band of approximately 46 kDa was present within the eluate fraction after the His-tag purification, but was not observed within the flow-through or wash fractions (Figure 4.6). Therefore, this band is expected to be the recombinant GlcNAcT-I. Although an affinity chromatography was performed several contaminating HCP were observed to co-eluted with purified recombinant GlcNAcT-I. This could be as a result of inadequate removal of HCP during the wash conditions. As demonstrated by the presence of identical size protein bands within the flow-through and wash fractions, and within the eluate fraction. Presently, a fixed concentration of imidazole is utilised during wash steps, the use of an imidazole step-gradient elution, such as that used during Fast protein liquid chromatography (FPLC), may remove more contaminates from the purified recombinant GlcNAcT-I. Nevertheless, this result demonstrates that if one is to immobilise a His-tagged fused protein in the future, for use in an *in vitro* metabolic pathway, it will be firstly important to ensure all contaminants are removed. That is, one cannot expect to create an immobilised enzyme system straight after cellular lysis, and one should seek an efficient purification step before immobilising a protein.

4.6 Assessing the activity of N-acetylglucosaminyltransferase-I

The production of human GlcNAcT-I in *E. coli* SHuffle® T7 strain, offers the first time that a soluble recombinant mammalian GlcNAcT-I protein was expressed in a prokaryotic expression host. The previously published strategy utilised *E. coli* (JM109) strain, this strain is not generally used for recombinant protein expression. Therefore, by using the more suitable expression strain of *E. coli* SHuffle® T7, a soluble GlcNAcT-I was expressed, preventing the need for a refolding strategy.

The enzymatic activity of purified GlcNAcT-I was determined against the synthetic sugar acceptor, OM3 (n-octyl 3,6-Di-O-(α-mannopyranosyl) β-d-mannopyranoside) (CAS number 140147-36-0). This sugar analogue had previously been utilised to assess the activity of a recombinant GlcNacT-I[383]. The GlcNAcT-I was incubated with OM3 and the nucleotide donor sugar (NDS) UDP-GlcNAc; a successful reaction would see the transfer of GlcNAc to OM3 and the release of nucleotide, uridine diphosphate (UDP). A Reverse Phase-High Pressure Liquid Chromatography (RP-HPLC) method had recently been established in the lab for the detection of sugar nucleotides and nucleotides[382]. The RP-HPLC method was employed to monitor the amounts of UDP and UDP-GlcNAc, when OM3 was incubated with GlcNAcT-I. If GlcNAcT-I is active GlcNAc will transfer of to OM3 releasing UDP. This will be observed in a HPLC trace through a decrease in UDP-GlcNAc and an increase in UDP. The donor nucleotides and nucleotides are detected using a photodiode array detector set at a wavelength at 262.1 nm. Jimenez del Val *et al.* have previously demonstrated the nucleotide UDP-GlcNAc to have a retention time of 8.224 seconds and UDP of 8.591 seconds[382].

However, in this study the UDP-GlcNAc and UDP standards, which were purchased from Sigma Aldrich, were found to have a retention time of 8.122 seconds and 8.441 seconds, respectively (Table 4.2).

Table 4.2 the chromatogram retention times of the donor sugar nucleotide UDP-GlcNAc and the nucleotide diphosphate UDP.

The chemical standards of UDP-GlcNAc (CAS number 91183-98-1) and UDP (CAS Number 27821-45-0) were purchased from Sigma Aldrich. Each analyte was dissolve in deionised water to a final concentration of 0.08 mM to match the concentration stated in Jimenez del Val *et al.*[382]. The chromatogram was produced by injecting 30 µL of an analyte. Mobile phases used were 3 mM NaOH (E1) and 3 mM NaOH containing 1.5 M sodium acetate (E2). Separation was performed at 30°C on a CarboPac PA-1 column (Dionex, ThermoFisher USA), with the following gradient profile (T=time); $T_{0\ min}$, 20% E2; $T_{6\ min}$, 66% E2; $T_{18\ min}$, 66% E2; $T_{23\ min}$, 20% E2; $T_{30\ min}$, 20% E2. Detection was done by absorbance at 262.1 nm. An example of a full chromatogram can be found in the Appendix Section 5. The table denotes the peak times of UDP-GlcNAc and UDP standards as acquired using Empower 2 software (Waters, UK).

	Retention time in seconds	
	UDP-GlcNAc	**UDP**
	8.082	8.44
	8.123	8.427
	8.149	8.448
	8.132	8.448
Average	**8.1215**	**8.44075**
Standard Deviation	**0.024642**	**0.008584**

Two reactions, prepared in triplicates, were performed, each reaction contained an OM3 substrate and donor nucleotide sugar UDP-GlcNAc at concentrations specified in the Materials and Methods (Section 2.10.1). One reaction contained the recombinant GlcNAcT-I which was His-tag purified in Section 4.5, where the other contained no enzyme and can be considered a negative control. The peak resolution between the two chemicals is difficult to witness on RP-HPLC trace due to the millisecond difference between retention times. Therefore, the chromatograph data acquired using Empower 2 software (Waters, UK) is presented in bar-graphs (Figure 4.7, Figure 4.8). The original RP-HPLC trace can be found within the Appendix Section 5.

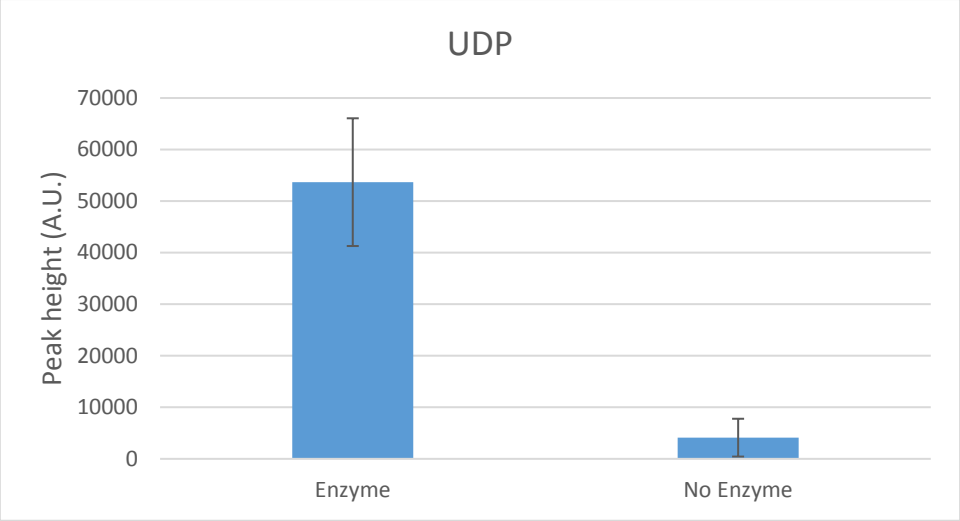


Figure 4.7 The RP-HPLC analysis of UDP concentration, from an enzyme reaction containing GlcNAcT-I and a reaction with no enzyme.
The graph represents the peak height at 8.12 seconds from a chromatograph acquired using Empower 2 software (Waters, UK). The 8.12 seconds represented the retention time of UDP (Table 4.2). As is shown from the graph, upon addition of recombinant GlcNAcT-I an increase in UDP is observed. It is assumed that this is increase is due to release of UDP from UDP-GlcNAc, due to enzymatic cleavage by GlcNAcT-I. The chromatograph of the RP-HPLC can found in the Appendix Section 5, details of RP-HPLC can be found in Section 2.8.3 or the caption to Table 4.2. Experiments were conducted in biological triplicate, the data presented is the average and error bars represent standard deviation.

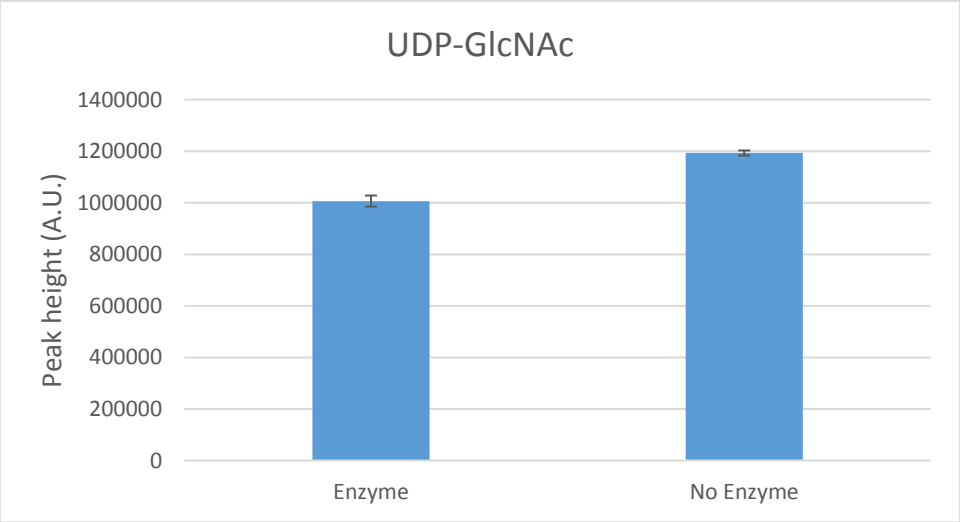


Figure 4.8 The RP-HPLC analysis of UDP-GlcNAc concentration, from an enzyme reaction containing GlcNAcT-I and a reaction with no enzyme.
The graph represents the peak height at 8.44 seconds from a chromatograph acquired using Empower 2 software (Waters, UK). The 8.44 seconds represents the retention time of UDP-GlcNAc (Table 4.2). As is shown from the graph, upon addition of recombinant GlcNAcT-I a decrease in UDP-GlcNAc is observed. It is assumed that this is decrease is due to release of UDP from UDP-GlcNAc, due to enzymatic cleavage by GlcNAcT-I. GlcNAc cannot be detected using this RP-HPLC method. The chromatograph of the RP-HPLC can found in the Appendix Section 5, details of RP-HPLC can be found in Section 2.8.3 or the caption to Table 4.2. Experiments were conducted in biological triplicate, the data presented is the average and error bars represent standard deviation

The concentration of UDP-GlcNAc was demonstrated to drop by ~1.2 fold, while the concentration of UDP had a >5-fold increase with the addition of recombinant GlcNAcT-I, compared to the reaction control where the enzyme was absent. The increase in UDP and decrease in UDP-GlcNAc demonstrated release of UDP from the UDP-GlcNAc, indicating GlcNAcT-I cleaved a bond in the UDP-GlcNAc substrate. It must be noted that that the measurements were calculated from the peak height of the HPLC trace which is measured in arbitrary units, and that the actual concentration of nucleotide were undetermined. Therefore, one could propose that the fold differences (i.e. 5-fold and 1.2-fold) matches the actual amounts of UDP released from UDP-GlcNAc. However, until the relationship between the peak area and nucleotide concentration is calculated this can only remain an assumption.

The above RP-HPLC analysis of the reaction can only monitor the release of UDP from UDP-GlcNAc, and cannot determine if GlcNAc is transferred to the OM3 substrate. The RP-HPLC method cannot detect GlcNAc, as the method only detects nucleotides (including nucleotides attached to sugars) and cannot detect monosaccharides. Therefore, a secondary analytical method was required to verify that the GlcNAc monosaccharide is being transferred to an oligosaccharide substrate by recombinant GlcNAcT-I. This was done by incubating the recombinant human GlcNAcT-I with UDP-GlcNAc and $Man_5GlcNAc_2$ (Prozyme, CA, USA) for 4 hours, after which an assessment of the oligosaccharide structures was performed using Fluorophore-Assisted Carbohydrate Electrophoresis (FACE). FACE was utilised in this study, as the technique offered a readily available and economically viable carbohydrate analysis tool. FACE involves the attachment of a fluorophore through a derivatisation step to the reducing termini of the $Man_5GlcNAc_2$ to form a Schiff base, this is then stabilized by a reductive reaction with sodium cyanoborohydride[430,431]. A single molecule of fluorescent label can be incorporated to oligosaccharide, thus allowing molar quantities to be determined[430,432]. Various labels have been used for the reductive amination of glycans, an extensive list of which can be found in the review by Anumula[433]. The fluorophore 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTs) was utilised in this study. This fluorophore was selected as it has the maximum net charge of any fluorophore at -3 in acidic conditions, whereas other common glycan labelling fluorophore, such as 2-Aminobenzamide (2-AB), have a net charge of 0 in acid conditions. Therefore, the higher negative charge will enable greater migration during gel electrophoresis and consequently better resolution of labelled-glycans. The fluorescently labelled glycan can be visualised using UV light from a UV transilluminator.

The FACE demonstrated that the reaction with enzyme and donor sugar UDP-GlcNAc contained a band of a higher molecular mass than the $Man_5GlcNAc_2$ where the enzyme was absent. This relative increased shift in mobility of the $Man_5GlcNAc_2$ demonstrated that GlcNAc

monosaccharide had transferred to yield a $GlcNAc_1Man_5GlcNAc_2$ moiety (Figure 4.9). Therefore, the recombinant GlcNAcT-I produced in Section 4.5 actively transfers GlcNAc.

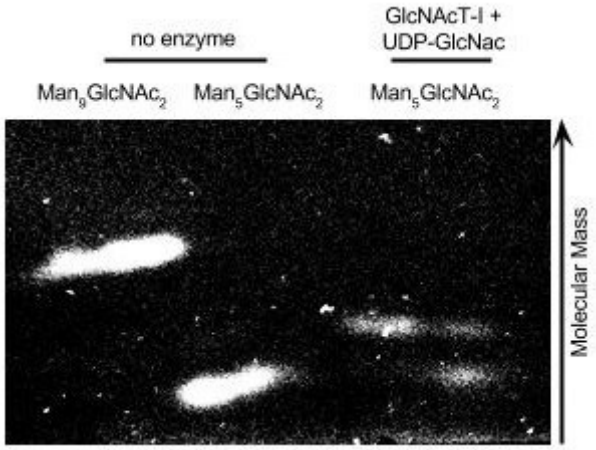


Figure 4.9 FACE analysis of $Man_5GlcNAc_2$ with enzyme GlcNAcT-I and donor sugar UDP-GlcNAc.
The image was taken by a camera attached to a UV transilluminator. The contrast of the image has been increased to enable better visualisation of the fluorescent label oligosaccharides. The original image can be found in the Appendix Section 5 of this document. The first lane of the gel image demonstrates $Man_9GlcNAc_2$ and the second lane demonstrates a $Man_5GlcNAc_2$ both of which have been labelled with ANTs. The third lane represents a labelled $Man_5GlcNAc_2$ which before labelling had been incubated with GlcNAcT-I which was obtained in Section 4.5. The shift in mobility of the $Man_5GlcNAc_2$ in the third lane toward the $Man_9GlcNAc_2$ infers an increase in molecular weight. This is indicative of addition of the monosaccharide GlcNAc to yield $GlcNAcMan_9GlcNAc_2$. Hence demonstrating the transfer activity of recombinant GlcNAcT-I.

4.7 The expression, purification and characterisation of type II α-mannosidase

The optimal expression conditions for recombinant Man-II were found to be within an *E. coli* SHuffle® T7 strain at 30°C in LB media with 500 µM IPTG (Figure 4.5). This condition was replicated for large scale expression in 500 mL of media. After induction cells were harvested by the above method for GlcNAcT-I. The recombinant Man-II was purified from the supernatant by gravity-flow affinity chromatography utilising Ni-NTA agarose (QIAGEN, Germany). The chromatography samples were analysed by Coomassie stained SDS-PAGE (Figure 4.10). A protein band of approximately 122 kDa was present within the eluate fraction after the His-tag purification, but was not observed within the flow-through or wash fractions (Figure 4.10). Therefore, this band is expected to be the recombinant Man-II.

Similarly, to GlcNAcT-I, HCP were found to co-eluate with Man-II, with both demonstrating a large contaminant protein at 55kDa. Again, this could be due to insufficient removal of contamination bands during the washing steps. Similarly, to the reason described for the contaminating bands in the Western blots from Section 4.2, a native protein may contain a histidine-repeat sequence. If this is the case the native protein will co-elute with recombinant

His-tagged proteins, therefore a further purification strategy might be required. These could include size-exclusion chromatography or ion exchange chromatography.

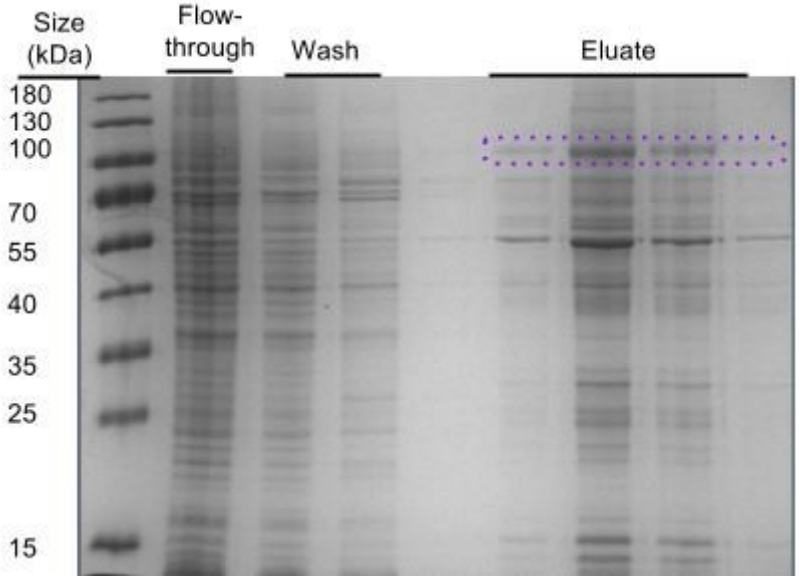


Figure 4.10 SDS-PAGE of a His-tag purification of Man-II
The heterologous expression of human Man-II was performed in *E. coli* SHuffle® T7 strain at 30°C in 500 mL of LB media and induced with [500μM] of IPTG. After 4 hours of induction cells were harvested and lysed by sonication. The Man-II was expressed with a N-terminus His-tag, which enabled purification by affinity chromatography. The recombinant Man-II was subjected to a purification using Ni-NTA Agarose (QIAGEN, Germany). After washing, His-tagged proteins are eluted in buffer under native conditions. Chemical composition of buffers were per the manufacturer's instructions

4.8 Assessing the activity of type II α-mannosidase

The translation of human Man-II in *E. coli* SHuffle® T7 strain, offers the first time that a soluble recombinant mammalian Man-II protein was expressed in a prokaryotic expression host. The enzymatic activity of the purified Man-II was determined against *p*-nitrophenyl-α-mannopyranoside as a substrate, using the same conditions as previously reported[385]. The amount of liberated *p*-nitrophenol was determined spectrophotometrically at 410 nm, an increase in absorbance would indicate an active Man-II. No increase in *p*-nitrophenol was observed, therefore the purified Man-II had no activity against the *p*-nitrophenyl-α-mannopyranoside (for data see Appendix Section 6).

The metal ions Zinc (Zn^{2+}) and Cobalt (Co^{2+})[385] have demonstrated to enhance activity of class II mannosidase[421]. A reaction was repeated on the purified Man-II with and *p*-nitrophenyl-α-mannopyranoside in the presence of Zn^{2+} or Co^{2+}. Again, no increase in absorbance was observed at 410 nm, indicating that the purified Man-II was inactive (for data see Appendix Section 6).

4.9 The attempted expression of active type II α-Mannosidase in *Pichia pastoris*

An active Man-II could not be achieved using *E. coli* as an expression host. Several researchers had previously utilised *P. pastoris* as an expression for active type II mannosidase production. For instance, the glycoengineered *P. pastoris* strain developed by Gerngross *et al.*[298] featured a Golgi based Man-II gene from *D. melanogaster*[294]. This was inserted into the genome under the control of a strong constitutive promoter (GAP promoter or PpPMA1 promoter). The Man-II was active and was able to produce the intermediate glycan structure[158]. However, the expression strategy in this study generated a soluble Man-II which will secreted from the cell, whereas Gerngross *et al.* generated an intracellular Man-II which was targeted to the Golgi membrane. *P. pastoris* is often used as a host organism for the production of recombinant proteins[97], and therefore was seen as a suitable candidate for the expression of an active Man-II. Before a recombinant protein can be expressed in *P. pastoris* the gene must be cloned into a suitable expression vector before inserting the genetic material into *P. pastoris* genome.

4.10 The generation of an expression vector for type II α-Mannosidase in *Pichia pastoris*

The Man-II gene was introduced into the *P. pastoris* expression vector pPICz-α A (Life technologies, CA, USA). The Man-II gene was fused with the α-factor secretion signal in the pPICz-α A vector and the N-terminal polyhistidine-tag was retained. The Man-II gene was isolated from the *E. coli* plasmid using PCR. The forward primer was designed to contain an EcoRI enzyme restriction site and the reverse had an XbaI enzyme restriction site. A DNA band of size corresponding to the PCR product was excised from a 1% agarose gel and purified using the Wizard SV Gel and PCR purification kit (Promega, WI, USA) according to manufacturer's instructions. The PCR product and pPICz-α A vector were cleaved using restriction enzymes EcoRI and XbaI. The compatible ends of the fragments were joined together using a T4 DNA ligase (NEB, MA, USA), to create the vector which will be known as pPICz-α Man-II and sequence was verified (Appendix Section 2).

The pPICz-α Man-II plasmid was linearized by digesting with the restriction enzyme SacI. This restriction site is located 209 base pairs within the 5' AOX1 region. A 500 mL culture of *P. pastoris* (GS115) was prepared for electroporation as described in the *Pichia* Expression kit manual and transformed with linearised pPICz-α Man-II plasmid. The transformant culture of *P. pastoris* (GS115) was plated on YPD containing 100 μg mL^{-1} of Zeocin™ and incubated at 30°C for 3-5 days.

Colonies were randomly selected and grown overnight in 5 mL of YPD culture containing 100 μg mL^{-1} of the antibiotic Zeocin™ in a 50mL conical tube. A colony PCR demonstrated that all selected colonies contained the Man-II gene in the genome. The overnight cultures in 5 mL of YPD media were pelleted by centrifugation, the supernatant was removed and pellet was resuspended in 5mL of methanol based medium (BMMY) to induce recombinant protein expression of Man-II. Cultures were grown for 24 hours, 2 days and 4 days, with 0.5% methanol added at 24 hour intervals. Cells were pelleted and the supernatant collected. The protein content of the supernatant was assessed by dot blot with an antibody with specificity for the His-tag (BioLegend, CA, USA) (Figure 4.11).

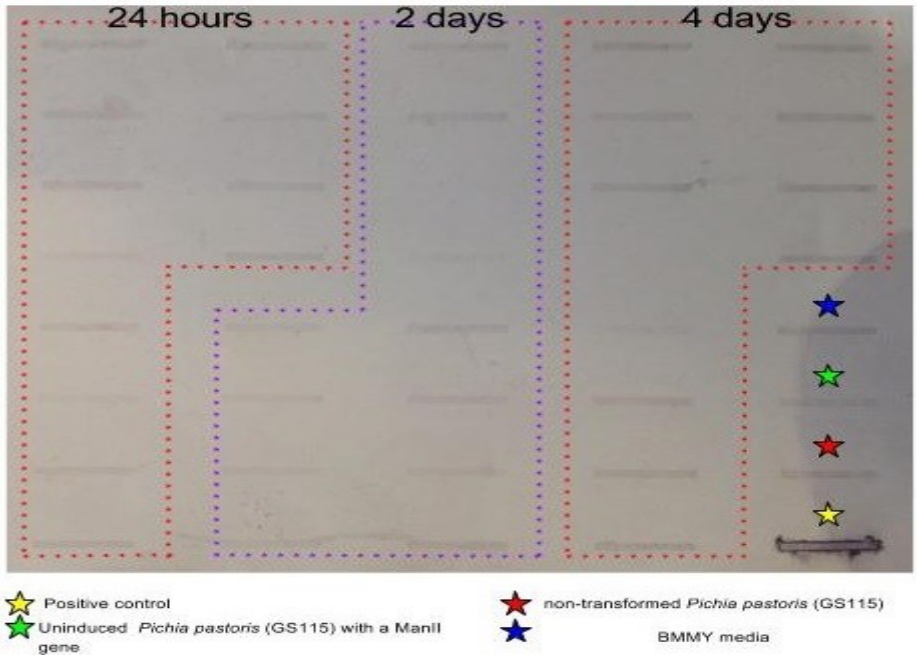


Figure 4.11 Dot Blot assessing Man-II expression from *Pichia pastoris* (GS115)
The expression of the Man-II strains was compared to expression from a non-induced strain containing the Man-II gene and from a non-transformed *P. pastoris* strain which was grown in the presence of BMMY media for 24 hours. The expression profile of the uninduced sample is indicated by the green star. The expression of the native P. pastoris is indicated by the red star. The dot blot revealed a slight purple band from these two strains which matched bands from the induced *P. pastoris* strains which had demonstrated the presence of Man-II gene. A positive control of a purified His-tagged protein (in this case a recombinant His-tagged GFP kindly donated by Dr. James Chappell) was placed on the dot blot. The expression profile of the positive control is indicated by the yellow star. The expression profile of a His-tagged protein produced a dark purple band. Therefore, as *P. pastoris* strains do not produce a dark purple band similar to the positive control but instead had a light coloured band comparable to the negative, it was concluded that Man-II was not being expressed.

As no clear protein expression was observed from the dot blot, a number of colonies were selected for an up-scaled expression, in baffled shake flasks. As aeration is important for protein expression in *P. pastoris*, baffled shake flasks were used to increase oxygen transfer compared to expressing in a conical flask[87]. Colonies were grown overnight in 5 mL of YPD culture containing 100 μg mL^{-1} of the antibiotic Zeocin™ in a 50mL conical tube. A sample of

culture was used to inoculate 200 mL of a glycerol based media (BMGY) and left to grow for 24 hours before being induced with 200 mL methanol based media (BMMY) for expression of the Man-II protein, 0.5% of methanol was added daily. Culture supernatants were collected on the second, third, fourth, and fifth days after commencement of induction. The expression of Man-II was checked by Coomassie stained SDS-PAGE and Western blot in both the total cell extract and the supernatant, again no visible expression of Man-II was observed. (Appendix Section 4).

4.11 The expression of a type II α-Mannosidase from *Caenorhabditis elegans* in *P. pastoris*

In the previous successful expression of human Man-II in *P. pastoris*, the protein was designed with a transmembrane domain[293], and as such was not tagged for secretion as what was attempted above. The soluble expression of a number of type II α-Mannosidases from *Caenorhabditis elegans* has previously been achieved in *P. pastoris*[434]. This included a type II α-mannosidase II (ceMan-II) enzyme which cleaves the terminal α1-3 mannose and α1-6 mannose sugar residues from a $GlcNAc_1Man_5GlcNAc_2$ motif. The ceMan-II from this study was truncated to have the cytosolic and TMD removed and contained an α-mating factor for secretion from the cell[434]. The ceMan-II was truncated at the 123rd amino acid residue of 1143 predicted amino acid length, resulting a protein of approximately 112 kDa.

The *P. pastoris* expression vector pPICzα- C containing the *C. elegans* gene for the synthesis of the recombinant ceMan-II was gifted by Dr Iain B. H. Wilson (plasmid will be known as pPICzα-ceMan-II). The pPICzα-ceMan-II was linearised by digesting the plasmid with the restriction enzyme PmeI, the restriction site is located 414 base pairs within the 5' *AOX1* region. A 500 mL culture of *P. pastoris* (GS115) was prepared for electroporation as described in the *Pichia* Expression kit manual and transformed with linearised pPICzα-ceMan-II plasmid. The transformant culture of *P. pastoris* (GS115) was plated on YPD containing 100 μg mL^{-1} of Zeocin™ and incubated at 30°C for 3-5 days.

Colonies were randomly selected and grown overnight in 5 mL of YPD culture containing 100 μg mL^{-1} of the antibiotic Zeocin™ in a 50 mL conical tube. The overnight cultures in 5 mL YPD were pelleted by centrifugation, the supernatant was removed and pellet was resuspended in 5mL of methanol based medium (BMMY) to induce recombinant protein expression of Man-II. As ceMan-II has been designed to be secreted from the cell, the culture media will contain recombinant protein. Cultures were grown at 30°C for 24 hours, 2 days and 4 days, with 0.5% methanol added at 24 hour intervals. An assessment of expression of ceMan-II by SDS-PAGE was not undertaken in this study due to the previous demonstration of expression and

secretion by a postdoctoral researcher, Dr Kate Royle, within the laboratory. The SDS-PAGE of successful expression and secretion of ceMan-II from *P. pastoris* X33 strain can be found within the Appendix Section 4. The expression conditions utilised in this study matched those of Dr Kate Royle. After the stated expression interval times, cells were pelleted and the supernatant collected and screened against *p*-nitrophenyl-α-mannopyranoside using the same conditions as previously reported[385]. Despite the previously reported ceMan-II demonstrating activity against *p*-nitrophenyl-α-mannopyranoside, no supernatant in this study gave a response against the substrate (Appendix Section 6)[434].

4.12 Transcriptional analysis of the type II α-Mannosidase from *Homo sapiens* and *Caenorhabditis elegans* in *Pichia pastoris*

The production assessment of human Man-II in *P. pastoris* indicated that no translation was occurring. As previous results for the expression of ManIa within *P. pastoris* had demonstrated lack of transcription (Figure 3.16), the transcription of the human Man-II mRNA was determined using RT-PCR. As the *P. pastoris* had demonstrated the translation of *C. elegans* Man-II, and consequently transcription, the ceMan-II offered a suitable positive control to assess human Man-II transcription.

The *P. pastoris* strains expressing Man-II from Human and *C. elegans* were grown overnight in 5 mL of YPD media supplemented with 100 μg mL^{-1} of Zeocin™ and incubated at 30°C with shaking. The cultures were pelleted by centrifugation and supernatant removed. Pellets were resuspended in 5 mL of methanol media (BMMY) to induce transcription of the Man-II gene. After 3 hours of induction the total RNA was extracted from *P. pastoris* strains using the RiboPure Yeast Extraction Kit (Life Technologies, Carlsbad, CA, USA), the RNA was reverse transcribed to cDNA using a RT-PCR kit (Life Technologies, CA, USA) following the manufacturer's instruction. The presence of the Man-II transcript was determined by PCR using primers to amplify a 180 bp product within the human Man-II CDS and 280bp product within the *C. elegans* Man-II CDS (Figure 4.12). The transcriptome of the colonies selected all contained a Man-II transcript. However, the band representing the transcript from the human Man-II appears much fainter than the ceMan-II counterpart. This could indicate a weaker transcription of the Human Man-II, which could explain the lack of expression. Although, it must be noted that the RT-PCR method employed here is not quantitative, thus making it difficult to support this claim. Nevertheless, transcription of human Man-II is occurring within *P. pastoris* host, however this did not lead to Man-II translation.

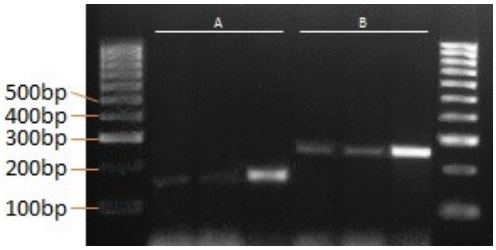


Figure 4.12 transcript analysis of *Homo sapiens* Man-II (A) and C. elegans Man-II (B) synthesised in *P. pastoris*.
The total RNA from *P. pastoris* strains with Man-II constructs were isolated. A 2-step RT-PCR was performed, first step involving convert RNA to DNA and the second step involving amplifying a specified transcript, in this case the transcript of Man-II. The Man-II gene from Human and *C. elegans* both demonstrated transcription in the *P. pastoris* host. Display in this figure is an enhanced close-up the original gel image can be found in the Appendix Section 7.

4.13 Summary

The recombinant expression of the first three Golgi-based enzymes involved in the formation of a complex-type N-linked glycosylation chain was attempted in the unicellular expression hosts of *E. coli* and *P. pastoris*. The expression of soluble protein of the human variants of these enzymes had not previously been recorded in these hosts. Although the soluble expression of GlcNAcT-I, ManII and GlcNAcT-II was observed, only GlcNAcT-I was shown to be active, with Man-II demonstrating enzyme inactivity and GlcNAcT-II activity being undetermined.

The *E. coli* BL21 (DE3) strain, which is routinely used for recombinant protein expression, was first selected as a host for the expression of the Golgi enzymes. The *E. coli* BL21 (DE3) strain successfully expressed a soluble version of GlcNAcT-II, but no visible expression of GlcNAcT-I or Man-II was observed. Although expression of GlcNAcT-II was observed the activity of the enzyme could not be determined as an appropriate substrate was not available at the time. The GlcNAcT-II adds a GlcNAc monosaccharide onto a terminal mannose on the $GlcNAc_1Man_3GlcNAc_2$ oligosaccharide. As the expression of an active Man-II could not be achieved, the $GlcNAc_1Man_3GlcNAc_2$ motif could not be generated. Unlike the $Man_5GlcNAc_2$ oligosaccharide, the $GlcNAc_1Man_3GlcNAc_2$ cannot be purchased and a synthetic analogue (such as OM3) does not exist. The GlcNAcT-I action must precede GlcNAcT-II, with previous enzyme kinetic studies with GlcNAcT-II having used a glycan motif generated by upstream processes (i.e. the actions of GlcNAcT-I or with actions of GlcNAcT-I and Man-II).

The difficulties in the translation of GlcNAcT-I and Man-II were attributed to these proteins containing a number of disulphide bonds (see Section 4.3). Therefore, the *E. coli* SHuffle® T7 strain which has been engineered to promote disulphide bond formation was selected. As discussed in Section 1.4, this strain has been genetic engineered to overexpress chaperones to assist in folding and has a higher oxidising cytoplasmic environment, thus facilitating in the

production of soluble recombinant protein with disulphide bridges. Using the *E. coli* SHuffle® T7 strain a soluble human GlcNAcT-I and Man-II were achieved in this study. However, upon testing the activity of Man-II, no catalytic cleavage of the substrate was observed (Section 4.8).

The expression of an active human Man-II enzyme was not achieved in this study. A soluble Man-II was expressed using *E. coli* SHuffle® T7 strain as a host, but the resulting enzyme appeared to be inactive. The cell pellet of the *E. coli* SHuffle® T7 expressing Man-II, which contains inclusion bodies (IBs) appears to have a large concentration of Man-II (Figure 4.5). Many methods exist in the literature which have demonstrated an improvement in the solubility of a recombinant protein translated in *E. coli*. For instance, the addition of a fusion tag, such as MBP, often improves the solubility of the fusion protein and could enhance production[435]. The large amount of the human Man-II observed within the IBs of the *E coli* SHuffle® strain, could be recovered using a refolding strategy similar to the previous study with human MBP fused GlcNAcT-I[383] and could be used to enable production of an active Man-II.

A number of studies had demonstrated the successful heterologous translation of an active Man-II in the recombinant expression host of *P. pastoris*. This included the expression of recombinant *C. elegans* Man-II[434], recombinant *D. melanogaster* Man-II[436], and a recombinant *Capra hircus* Man-II[437]. Hence why the expression of the human Man-II was attempted within *P. pastoris.* The *C. elegans* Man-II construct was also obtained and despite previously been shown to actively cleave *p*-nitrophenyl-α-mannopyranoside, the *C. elegans* Man-II disappointedly showed no activity against the *p*-nitrophenyl-α-mannopyranoside substrate. The translation of human variant of Man-II failed and this was despite the presence of a Man-II transcript.

One theory proposed on why the Man-II failed to translate is due to the large size of the recombinant protein. Larger proteins such as Man-II often have difficulties expressing, removing domains of a target protein could improve expression levels and solubility[35]. Although the Man-II construct lacks the cytosolic and transmembrane domains, further truncation of the Man-II is possible. The active recombinant Man-II proteins from *C. elegans* and *Drosophila* both had larger truncations from the N-terminal domain than the recombinant human Man-II gene in this study[434,436].

Although the production of an active Man-II failed, the *E. coli* SHuffle® T7 strain did produce an active recombinant human GlcNAcT-I. The activity of GlcNAcT-I was tested against the substrate analogue OM3 and the substrate $Man_5GlcNAc_2$. As a $Man_5GlcNAc_2$ oligosaccharide covalently attached to a protein was not available, a $Man_5GlcNAc_2$ oligosaccharide was purchased (ProZyme, CA, USA). The non-covalently bound $Man_5GlcNAc_2$ oligosaccharide

represented the most physiological acceptor available for the donor sugar UDP-GlcNAc and GlcNAcT-I. It has previously been shown that the GlcNAcT-I enzyme can use a non-covalently bound $Man_5GlcNAc_2$ as substrate[438]. This was again demonstrated in this study, with the recombinant human GlcNAcT-I transferring GlcNAc onto the $Man_5GlcNAc_2$ substrate. This transfer was observed using the electrophoresis technique, FACE, the use of this method makes it difficult to accurately quantify and therefore determine the enzyme kinetics. The release of UDP from donor sugar UDP-GlcNAc was observed when the recombinant human GlcNAcT-I was incubated with the substrate analogue OM3. The concentration of UDP can be detected using an RP-HPLC technique, allowing for an easier quantification of enzyme activity. The release of UDP and transfer of GlcNAc to OM3, demonstrated that the soluble GlcNAcT-I was active, although the relative activity compared to previous recombinant GlcNAcT-I was not determined. The concentration of UDP could be used to determine the GlcNAcT-I activity, however it has been previously demonstrated that the GlcNAcT-I has a stronger activity (higher V_{max}/K_m ratio) for $Man_5GlcNAc_2$ over Man_3 synthetic analogues, such as OM3[438]. Therefore, when determining the enzyme activity, it will be important to use $Man_5GlcNAc_2$ as the substrate over OM3. A method has previously been developed by Flynn *et al.* where fluorescence-labelled oligosaccharides, such as $Man_5GlcNAc_2$ and $GlcNAc_1Man_5GlcNAc_2$ are separated by molecular weight by RP-HPLC, will allow for quantification of UDP-GlcNAc transfer and more accurate determination of GlcNAcT-I activity[381].

To sum up, the generation of the complex-type oligosaccharide structure $GlcNAc_2Man_3GlcNAc_2$ was not achieved in the study due to the lack of an active Man-II. This therefore limited the development of an *in vitro* metabolic pathway for the processing of a core glycoprofile $Man_5GlcNAc_2$. Although there is potential to negate the action of Man-II, by expressing a glycoprotein with a $Man_3GlcNAc_2$ glycoprofile or by synthesising alternative mannosidases to Man-II, these two strategies both have their biological difficulties. Namely the difficulty in achieving a $Man_3GlcNAc_2$ profile and the enzymatic control Man-II offers over alternative mannosidases. Future effort will be placed in producing an active Man-II, to enable the synthesis of a complex-type N-linked glycosylation chain.

Chapter 5. The development of a cell-free protein synthesis system based on *P. pastoris* cell lysate

5.1 Introduction

Chapter 4 endeavoured to create an *in vitro* metabolic pathway for the remodelling of an oligosaccharide utilising the core $Man_5GlcNAc_2$ glycan as the substrate; this being the initial substrate for Golgi-based glycosyltransferases. As previously discussed, the removal of glycan processing from the cell has many merits, mainly related to improved enzymatic conversion and controlling of the end glycoform. The glycoprotein with pentamannosyl glycoform was intended to be produced from a glycoengineered *P. pastoris.* That is the glycoprotein would be produced from a cellular, *in vivo* source. As the project aims to work toward the cell-free expression of a glycoprotein, the next step was to generate a complete cell-free transcription-translation system. Cell-Free Protein Synthesis (CFPS) systems, which is the *in vitro* translation of a recombinant protein, offers a platform to enable this to occur.

As CFPS negates a cell culture during expression it has many benefits over traditional *in vivo* protein synthesis, mostly associated with the high-throughput synthesis of recombinant proteins[353,439]. In CFPS cellular complications such as proteolytic degradation and product toxicity are eliminated. The incorporation of unnatural amino acids and modified amino acids is also made simpler using a CFPS[350,351].

In its simplest form, a CFPS system consists of a crude cellular extract containing the necessary translational machinery, tRNAs, and accessory enzymes to aid translation. Over time researchers have supplemented crude cellular extract with additional catalytic components to enhance CFPS productivity[440,353]. For instance, as with *in vivo* translation, Adenosine triphosphate (ATP) is an essential component for translation required for the aminoacylation of tRNAs and peptide bond formation. The diminishment of ATP will cause a termination of translation. Compounds such as creatine phosphate and the enzyme creatine kinase can regenerate ATP (Figure 5.1) helping to maintain an ATP pool[441]. The addition of these two components to a CFPS system has shown to increase the duration of translation reaction; therefore increasing protein yields[442].

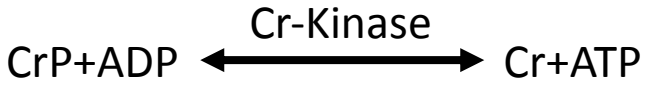


Figure 5.1 the enzyme reaction for Creatine-Kinase.

Creatine kinase (Cr-Kinase) (EC 2.7.3.2), catalyses the conversion of phosphocreatine (CrP) to creatine generating ATP. The enzyme Cr-Kinase is utilised in CFPS systems to help replenish ATP stocks, ensuring a readily available supply of ATP for translation of recombinant proteins.

Beyond a CFPS based on crude extracts is the *p*rotein synthesis *u*sing *r*ecombinant *e*lements (PURE) CFPS system[443]. This system has been developed to consist of individual purified *E. coli* translation components, omitting the need for a crude extract[443]. However, the PURE system is not as productive as a CFPS system prepared with a crude extract. Hillebrecht *et al.* demonstrated a 70% increase in the amount of Firefly luciferase being produced from an *E. coli*-based CFPS compared to a PURE system[444]. Nevertheless, a major advantage of the PURE system is that it contains neither nucleases nor proteases, both of which hinder reaction productivity.

The protein expression in CFPS can be carried out in either a 'linked' or 'coupled' reaction. In the linked reaction, a DNA template is first transcribed in a separate reaction followed by the addition of the transcribed mRNA to CFPS mixture for protein translation[352]. In a coupled reaction both transcription and translation steps take place within CFPS, i.e. within the same tube[352].

The cellular extract from *E. coli* is by far the most popular source for CFPS, due to a simple and inexpensive extraction preparation, and its scalability (from 15 µL to 100 litres)[445]. *E. coli* CFPS has been demonstrated to produce 5 mg mL^{-1} of a recombinant protein[446], which is over double amount of recombinant protein produced from the next productive CPFS, a purchasable wheat germ extract (WGE) CFPS (2.19 mg mL^{-1} of ζ-crystallin quinone reductase)[447]. However, in parallels to *in vivo E. coli* recombinant expression, the major drawback to utilising an *E. coli* CFPS is in the lack of ability to perform certain PTMs, in particular glycosylation[448]. Similarly, to *in vivo* recombinant expression, a solution to enable incorporation of PTMs is the utilisation of eukaryotic sources of crude extracts in CFPS. For instance, the N-linked glycosylation of a protein can be achieved when microsomes (ER) are retained or supplemented within a CFPS system[448]. One of the first examples of glycosylation being detected within a CFPS was in 1983, when a WGE CFPS containing canine pancreatic microsomal membranes successfully glycosylated a vesicular stomatitis virus (VSV) G protein[449]. Although supplementation of WGE CFPS with canine microsomes was presented earlier in 1977 and utilised for peptide signal cleavage, the glycosylation of the protein was not indicated[366]. As well as WGE, early examples of N-linked glycosylation by a CFPS can also be found from Rabbit Reticulocyte Lysates (RRL) supplemented with canine pancreatic microsomal membranes[450,451]. Since these early examples a number of other eukaryotic CFPS have been developed which display N-linked glycosylation abilities. These CFPS systems differ from WGE and RRL as they are often harvested with retained 'native' microsomes rather

than supplemented with an interspecies microsome; examples can be found in CHO cell CFPS[355] and insect CFPS (based on sf21 cell line)[354,363]. Although glycosylation was detected from each of these eukaryotic CFPS the actual oligosaccharide structure was never specified. As microsomes are artefacts of the ER reformed after a cell disruption processes, one could postulate that any glycosylation which occurs should only be that which inherently occurs within the ER. Therefore, microsomal membranes can be deemed to lack the ability to produce more humanised glycosylation patterns, rendering them unsuitable for biopharmaceutical production.

Eukaryotic CFPS systems are often seen as not being as cost-effective as their prokaryotic counterparts, with time consuming and relatively expensive extraction process. Eukaryotic CFPS systems are unable to match the volumetric scalability described above and therefore return lower overall yields of recombinant protein. In fact, in a review published close to the beginning of this project in 2012, identified current challenges of CFPS as being (1) the lack of a cost-effective and scalable eukaryotic CFPS platform, and (2) the inability to carry out humanized glycosylation patterns[452]. Since that review, a number of researchers have aimed to solve these challenges. Indeed, the authors of that review have made progress in generating a relatively cost-effective and scalable eukaryotic CFPS platform by utilising *S. cerevisiae* as the source for CFPS. As discussed in Section 1.15, *S. cerevisiae,* along with other yeast species, are an inexpensive and fast growing expression host, making them an attractive alternative to current eukaryotic CFPS. Nevertheless, despite working towards the generation an inexpensive eukaryotic CFPS system, the overall protein yields from *S. cerevisiae* CFPS are much lower than other eukaryotic CFPS. For example, a *S. cerevisiae* CFPS produced 0.017 mg mL^{-1} of GFP[364], whereas the commercially available WGE CFPS has been demonstrate to produced titres up to 1.6 mg mL^{-1} of GFP[453]. Further to the low titres, the *S. cerevisiae* CFPS was not designed to glycosylate proteins.

In regard to the second challenge identified, the development of a CFPS with a humanized N-linked glycoprofile, many obstacles still remain. Although the number of eukaryotic CFPS with glycosylation capability is ever expanding, including the use of mammalian cells such as CHO cells[355] and HeLa cells[454], these systems are all based on the retention of microsomes. Consequently, CFPS with microsomes will only contain a high-mannose structure. Although co-translocated into microsomes, target proteins appear to fold but not exit and therefore cannot be remodelled further to a more humanized therapeutically suitable glycan chain (within a retained Golgi).

The aim of this chapter was to address the challenge of producing a CFPS with a humanized glycoprofile. As is evident, the inclusion of microsomes within a eukaryotic CFPS enables the

glycosylation of a protein product, although with difficulty in obtaining a human-like glycoprofile. The remodelling of the core glycan structure could be achieved through an *in vitro* metabolic pathway, as described in the previous chapter. Therefore, the utilisation of microsomes from a eukaryotic organism which synthesises a $Man_5GlcNAc_2$ can offer a substrate for this *in vitro* metabolic pathway.

5.2 Plasmid construction for use in *P. pastoris* CFPS

To test for the activity of the created *P. pastoris* CFPS appropriate genetic material was required. The designs of the genetic material were based on similar designs of many previous published CFPS. The genetic material for the CFPS was all contained within a self-replicating plasmid pET28-A (Merck, NJ, USA), the vector map of which can be found within the Appendix Section 8. In total 12 plasmids were constructed each containing consensus elements, such as the promoter, CDS and terminator, along with non-consensus DNA elements which will be discussed below (Figure 5.2). The genetic element differences occurred to aid in generating an optimal construct and understand how these differences could affect transcription and/or translation in a *P. pastoris* CFPS system.

The *P. pastoris* CFPS has been set-up as a coupled reaction, that is, both transcription and translation reactions happen within *P. pastoris* CFPS. Although native polymerases exist within the cell lysate[455], their numbers can be limited or inactive, therefore diminishing the productivity of the CFPS. As such, many CFPS are enhanced with an orthogonal RNA polymerase to initiate transcription[456]. A T7 RNA polymerase was utilised within the CFPS reaction with the aim of producing the mRNA template. The T7 RNA polymerase is known to strongly transcribe a gene and has been utilised in a number of previous CFPS (see Table 5.1 for examples). The pET28-A plasmid was selected based on it containing a T7 RNA polymerase bacteriophage promoter and a T7 terminator sequence, for initiation and termination of transcription.

As well as the two consensus elements of the promoter and terminator each plasmid contained a gene for the expression of Green Fluorescent Protein (GFP). GFP is a commonly used reporter protein, its expression can be observed due to the fluorescent nature of mature protein. GFP has also demonstrated to express at high levels *in vivo* within a *P. pastoris* host[457]. GFP was obtained from BioBrick Foundation Registry of Standard Biological Parts. The GFP was isolated by PCR using primers designed to introduce XhoI and BamHI sites to the 5' and 3' ends of the GFP gene, respectively; these are compatible with pET28a expression vector multiple cloning site. The pET28a expression vector and GFP gene were cleaved with restriction enzymes XhoI and BamHI (NEB, WI, USA), The compatible ends of

the fragments were joined together using a T4 DNA ligase (NEB, WI, USA), to create the vector pGFP (Figure 5.2).

The mRNA molecules produced in a CFPS are not modified to contain a 5′ cap, a necessary feature for eukaryotic translation initiation. The 7-methyl guanylate (m^7G) cap is added post-transcriptionally to the 5' ends of transcripts and is thought to facilitate the recruitment of the ribosomal subunits to the mRNA[458]. As an m^7G cap is not added, a number of published CFPS systems have utilised Internal Ribosome Entry Site (IRES), thus bypassing the need for cap-dependent translation[459,460]. For instance, the encephalomyocarditis virus (EMCV-IRES) which is a widely used and well-studied IRES element, has previously been utilised to enhance translation in CFPS systems based on RRL extracts[461]. An IRES is an RNA sequence which forms complex secondary structures in the 5’ untranslated region (UTR) of mRNA, promoting ribosome binding and therefore allowing for cap-independent translation[462]. IRES elements are utilised by viruses to ‘hijack’ translation machinery, but can also occur natively within the genome and are utilised during stress responses[463]. For instance, *P. pastoris* has been discovered to contain at least two IRES elements within its genome[464]. Despite this the actual employment of IRES elements in *P. pastoris* for recombinant protein expression is scarce, with only one IRES (at the time of this study) shown to successfully mediate *in vivo* translation; an IRES from the 5’-UTR of *GPR1* mRNA in *S. cerevisiae*, termed GPR-IRES. Only by placing the GPR-IRES sequence between two genes in a bicistronic construct within *P. pastoris* could the second gene be translated[465,466]. IRES elements have been shown to direct internal initiation of a second cistron in a bicistronic mRNA[467].

As GPR-IRES has successfully been applied for recombinant expression in *P. pastoris*, it was selected to be used in this study. However, as the GPR-IRES had not been utilised in a CPFS system before, a second IRES element which had a proven record of translation in CFPS was also selected. The IRES of the cricket paralysis virus (CrPV) (CrPV-IRES) has been utilised in a number of eukaryotic CFPS systems. For instance, the CrPV-IRES was shown to increase expression of human epidermal growth factor receptor (EGFR) 3.8 fold in a Sf9 CFPS[459]. The CrPV-IRES has also been successful utilised to translate recombinant proteins within a *S. cerevisiae* CFPS (2.33±0.11 mg mL^{-1} of luciferase)[468] and a CHO cell CFPS (50 µg mL^{-1} of luciferase)[459]. The interspecies activity of the CrPV-IRES made it an attractive option for an IRES, as many IRES elements have shown to be largely species dependent, and therefore it was selected in this study[469,470].

A further two plasmids were constructed with an IRES element upstream of the GFP. The IRES elements CrPV IGR and GPR were synthesised by DNA2.0 and fused upstream of the GFP gene by Overlap Extension PCR (OE-PCR). The individual IRES elements and GFP

gene were successfully amplified with primers containing complementary sequences to the adjacent DNA parts. Overlap Extension PCR (OE-PCR) was used to assemble the tandem DNA parts. A further PCR was performed on the IRES-GFP fused PCR products to introduce restriction enzyme sites compatible with the pET28a expression vector. The pET28a and IRES-GFP gene were cleaved with restriction enzyme XhoI and BamHI (NEB, WI, USA). The compatible ends of the fragments were joined together using a T4 DNA ligase (NEB, WI, USA), to create the vectors which will be known as pCrPV-GFP and pGPR-GFP (Figure 5.2).

Several eukaryotic CFPS have included a poly-A tail at the 3' end of the mRNA molecule[471,472]. A poly-A tail is known to increase the stability and prevent degradation of an RNA molecule[473,474]. For instance, the translation of DNA templates is >10-fold higher when a poly-A tail is present in the 3´ UTR of a CFPS based on HeLa cell lysate[475]. A poly-A tail was added downstream of the GFP in six of the twelve plasmids constructed. Two oligonucleotides were designed with complementary sequences to each other, one containing 50 adenine bases and the other with 50 thymine bases. The NotI and EcoRI were added to the 3' end of 50(T) and 5' of 50(A) oligonucleotides, respectively. The two oligonucleotides were annealed with sequence overhangs of DNA bases matching NotI and EcoRI. The different plasmids were cleaved with restriction enzymes NotI and EcoRI. The annealed oligonucleotides and digested vector were joined together using a T4 DNA ligase (NEB, USA). The assembled constructs were known as pGFP-pA, pCrPV-GFP-pA, pGPR-GFP-pA, pSS-GFP-pA, pCrPV-SS-GFP-pA, and pGPR-SS-GFP-pA (Figure 5.2).

A further three plasmids were constructed to contain an ER signal sequence upstream of the GFP coding sequence. A signal sequence triggers the translocation of nascent proteins into the ER, and has been demonstrated *in vitro* within a CHO cell CFPS to deliver nascent proteins to microsomal vesicles[355]. The α-factor signal sequence was amplified by PCR from the *P. pastoris* expression vector pPICzαA (Life Technologies, UK), with the aim of targeting the reporter protein to the microsomes in the cell lysate. The primers were designed to add complementary sequence to the pET28a and GFP, 5' and 3' of the α-factor signal sequence. The pET28a-GFP gene was cleaved with BamHI. The compatible ends of the fragments were joined together using Gibson assembly (NEB, WI, USA), to create the vector which will be known as pSS-GFP (Figure 5.2). A further PCR reaction was performed to isolate the α-factor signal sequence using primers with complementary sequence to either of the two IRES elements and the GFP. The pCrPV-GFP and pGPR-GFP were cleaved with SacI and the signal sequence was introduced by Gibson assembly (NEB, WI, USA) to yield plasmids pCrPV-SS-GFP and pGPR-SS-GFP (Figure 5.2).

The constructed expression vectors were transformed into chemically competent *E. coli* strain by heat shock transformation. The *E. coli* transformants were streaked onto an LB agar containing 35 µg mL^{-1} of Kanamycin and incubated at 37°C overnight. Individual colonies were selected at random and grown in 5 mL of LB media containing 35 µg mL^{-1} of Kanamycin and incubated at 37°C overnight with shaking. The expression vectors were isolated from *E. coli* using a plasmid preparation kit following the manufacturer's instructions (PEQLAB GmbH, Erlangen, Germany). The DNA sequence of the 12 expression vectors were verified (Appendix Section 2).

Although a total of 12 plasmids were constructed the initial testing of the CFPS was performed with pCrPV-GFP-pA and pGPR-GFP-pA. The features of these constructs, including the poly-A tail and IRES sequences, offer best chance of a translational product in a CFPS, as demonstrated by previously published yeast and CHO cells CFPS.

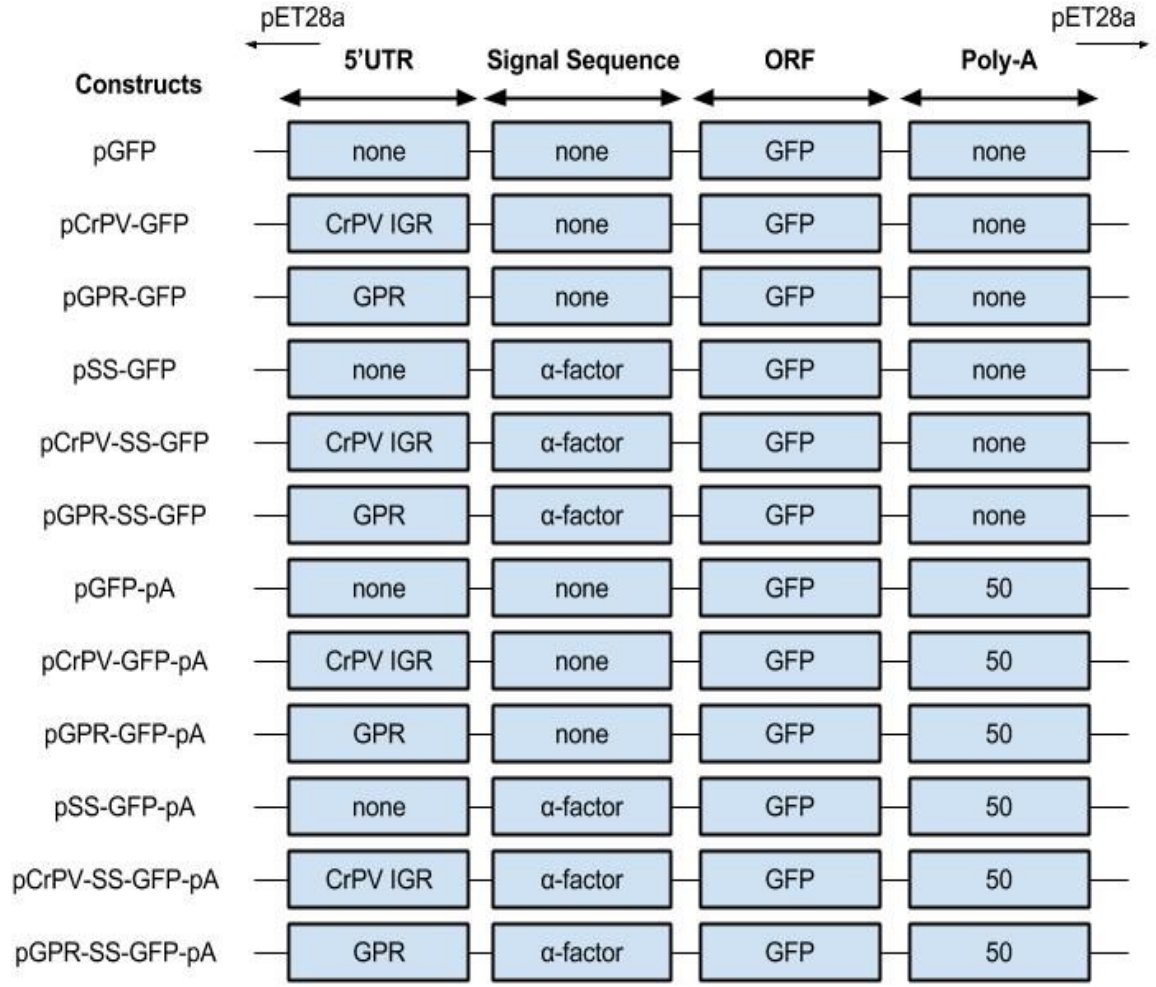


Figure 5.2 Schematic of the 12 plasmids constructed to test CFPS showing the different DNA elements.
A total of 12 genetic constructs were placed in the plasmid pET28a. The cloning of genes in pET28A enable transcription to proceed via a T7 RNA polymerase as this plasmid contains a T7 promoter and T7 terminator sequence. Each of the 12 plasmids contained a GFP (P42212). An increase in fluorescence (ex/em 485/520) over background in a CFPS reaction indicated the

successful translation of GFP. In addition to the GFP, genetic constructs were designed to contain an IRES sequence (either GPR or CrPV), an ER signal sequence (α-mating factor), and a Poly-A tail (50XAdenine). The diagram highlights which of these elements are present within which plasmid.

5.3 Developing a methodology for *Pichia pastoris* CFPS

Any organism has the potential to provide a source for crude lysate of a CFPS. However, in order to produce a pentamannosyl glycoprofile within a microsome, a eukaryotic organism which synthesises this glycoprofile in the ER is essential. Glycoengineered yeast strains have the produce a $Man_5GlcNAc_2$ within the ER[288,411]. Therefore, a glycoengineered yeast cell lysate should feasibly produce a core pentamannosyl glycoprofile ($Man_5GlcNAc_2$) within retained microsomes. Presently, only *P. pastoris* has been modified to produce a $Man_5GlcNAc_2$ within the ER. It is worth noting that a glycoengineered strain of *P. pastoris* was unavailable at this point, either commercially or through the attempted generation in Chapter 3. However, the potential methodology used to obtain a functional *P. pastoris* CFPS should be transferable to other *P. pastoris* strains. To date, no CFPS system from *P. pastoris* has been published. Therefore, this study offers the first known attempt to obtain a functional *P. pastoris* CFPS. As no previous methodology exists for the preparation of a *P. pastoris* CFPS, one first must determine suitable parameters which will enable successful transcription and translation. To achieve this, it would be ideal to evaluate some of the current methodologies for other CFPS systems (Table 5.1).

Table 5.1 a literature review of procedural steps utilised to generate an active CFPS system.
A selected few CFPS systems are represented on the table. Detailed are methods utilised to prepare and disrupt cells, the composition of chemicals supplemented to aid translation and/or transcription, and details of reaction procedures. The review of previous published active CFPS systems sourced from different species aided the development of methodologies for a *P. pastoris* CFPS for this study.

	E. coli[476,477]	***S. cerevisiae***[356,478,479]	**Insect**[363,480,481]	**CHO cells**[355,367]	**WGE**[482,483]
Cell preparation	- Growth within a medium-rich undefined media such as 2xYT media - Cells were harvested during the late exponential phase - Wash and prepare cells in S30 buffer (10 mM Tris acetate, pH 8.2, 16 mM KOAc, 14 mM $Mg(OAc)_2$)	- Growth within a medium-rich defined media such as Synthetic Complete (SC) Media - Cells were harvested once the OD_{600} reached 12.0 (late exponential phase) - Wash and prepare cells in Mannitol Buffer A (30 mM HEPES, pH 7.4 w/5 M KOH, 100 mM potassium acetate, 2 mM magnesium acetate, 2 mM DTT, 8.5% (w/v) mannitol)	- Sf21 cells were maintained in a spinner culture with highly rich defined media at 27 °C - Cells were harvested at densities of 1.5 – 4 $\times 10^6$ cells/ml. (early to mid-exponential phase) - Wash and prepare cell extraction buffer (40 mM Hepes-KOH (pH 7.95), 100 mM KOAc, 1 mM $Mg(OAc)_2$, 2 mM CaCl2, 4 mM DTT).	- CHO cells were maintained in high defined rich media - Cells were harvested at a density of approximately $2x10^6$ cells/mL. (early to mid-exponential phase) - Wash and prepare cell extraction buffer (40 mM HEPES-KOH (pH 7.5), 100 mM NaOAc, 4 mM DTT)	Grind wheat seeds using a blender, isolate embryo fractions by Sieving the ground material (710- to 850-mm mesh).
Cell disruption techniques	- Glass beads (bead beating) - High pressure homogenizer at 17,500 psi - Sonication	-high-pressure homogenizer	- Mini-Bomb cell disruption chamber	- Disruption was accomplished by syringing the harvested cell pellet through a 20-gauge needle	- Freeze embryo in liquid nitrogen and disrupt using physical force from a pestle and mortar - Sonication in nonidet P-40
Translation reaction conditions	- **Energy mix** $Mg(OAc)_2$ 22 mM creatine kinase	- **Energy mix** 120 mM potassium glutamate, 5.5 mM magnesium glutamate 25 mM creatine phosphate	- **Energy mix** 2.5 mM $Mg(OAc)_2$, 75 mM KOAc 1.75 mM ATP,	- **Energy mix** 150 mM KOAc 3.9 mM $Mg(OAc)_2$	- **Energy mix** 2.7 mM $Mg(OAc)_2$ 100 mM KOAc 0.25 Mm GTP,

	3.28 mM ATP 0.54 M potassium glutamate, **- Buffer** 155mM HEPES-KOH (pH 7.5) 4.7 mM DTT 20 mM spermidine 20 mM of each amino acid	0.4 mM cAMP 0.27 mg/mL creatine phosphokinase **- Buffer** 22 mM HEPES-KOH (pH 7.4) 1.7 mM DTT 0.5 mM spermidine 2 mM putrescine, 11% glycerol 0.08 mM of each amino acid	0.3 mM GTP 10 mM creatine phosphate **- Buffer** 30 mM HEPES-KOH (pH 7.6), 2.5 mM DTT 0.25 mM spermidine 5 mM of each amino acid 0.32 µl of biotin-lys-tRNAs	20 mM Creatine phosphate 100mg/mL Creatine kinase **- Buffer** 30 mM HEPES KOH, pH 7.6 2.5 mM DTT 0.25 mM spermidine 5 mM of each amino acid	16 mM creatine phosphate. 1.2 mM ATP, **- Buffer** - 30mM HEPES-KOH (pH 7.6) 2.5 mM DTT 0.4 mM spermidine 0.3 mM of each amino acid
Transcription reaction conditions	- T7 RNA Polymerase - 2.33 mM of each nucleoside triphosphate	- 1.3 U/µL T7 RNA Polymerase - 1.5 mM of each nucleoside triphosphate	- Capped mRNA	- 1 U/mL T7 RNA-polymerase - 0.3 mM of each nucleoside triphosphates expect ATP at 1.75 mM	- SP6 RNA polymerase
Microsomes?	- n/a	No	Yes	Yes	Yes
Temperature	30- 37°C	30°C	27°C	33°C	25–30°C
Duration of the reaction	2-4 hours	4 hours	Not displayed	Not displayed	Not displayed

5.4 *P. pastoris* cell preparation for generation of a cell lysate

Table 5.1 demonstrates that a variety of methods have been developed for the preparation of a crude cell lysate. The cell lysates have always been prepared during the exponential growth phase of a cell culture (with the exception of cell lysates sourced from tissues e.g. WGE). The cell lysate from unicellular sources are generally prepared during the late-exponential growth phase, whereas those cell lines propagated from multicellular sources appear to be harvested during the early to mid-exponential phase. As *P. pastoris* is a unicellular eukaryote the harvesting was carried out during the late-exponential growth.

The growth curve for *P. pastoris* X33 strain was determined by measuring the optical density at 600 nm (OD_{600}) over a continuous time (Figure 5.3). Similar to other strategies a nutritiously rich liquid medium, in this case YPD, was selected to maximise cellular growth.

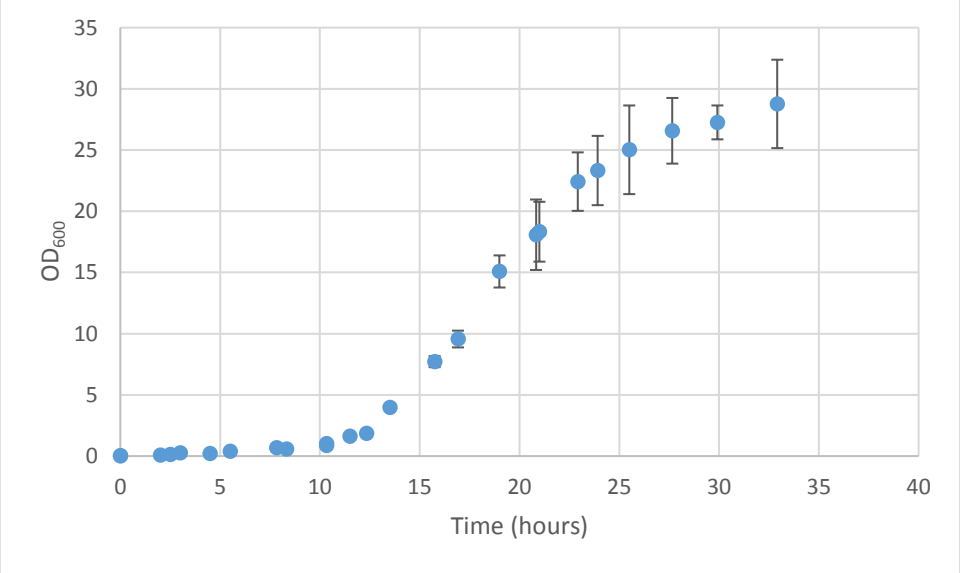


Figure 5.3 Growth curve of *Pichia pastoris* X33.
P. pastoris strain X33 was grown overnight in YPD media, diluted in 500 mL of fresh YPD in a 2.5 L baffled flask. The *P. pastoris* culture was incubated at 30°C with shaking. Samples were taken at approximately two hour intervals and the optical density was calculated. Experiments were conducted in biological triplicate; error bars represent the standard error from the average.

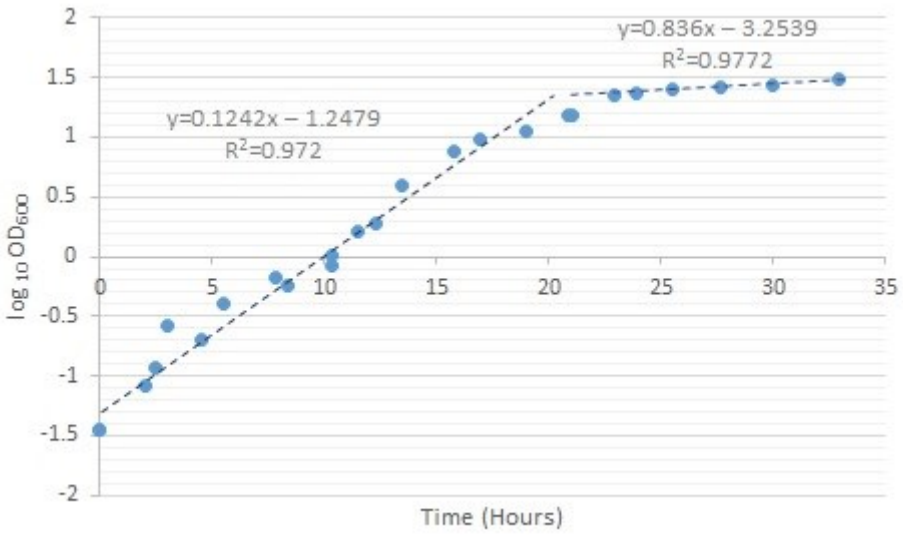


Figure 5.4 Culture profile of *P. pastoris X33.*
The *P. pastoris* strain was cultured in 500 mL of YPD media and the cell density was tracked over 33 subsequent hours. Aliquots of the culture media were taken at regular intervals and the optical density of the cell concentration was measured at wavelength of 600 nm (Figure 5.3). Experiments were conducted in biological triplicate, results were Log_{10} transformed and fitted with a linear regression. The R^2 values for *P. pastoris* is 0.972 between hours 0 and 20, and R^2 values for *P. pastoris* is 0.9772 between hours 20 and 33 hours

As previously reported *P. pastoris* can grow to extremely high optical densities compared to other yeast species, but as OD_{600} also measures dead cells it is important to ensure that when harvesting *P. pastoris* we choose a biomass with a high cell population viability. The cell population viability was measured by staining cells with LIVE/DEAD® FungalLight™ (Life Technologies, CA, USA). Stained cells were observed using flow cytometry (Figure 5.5). The percentage of live cells remained high (>98%) throughout the population growth curve (Table 5.2).

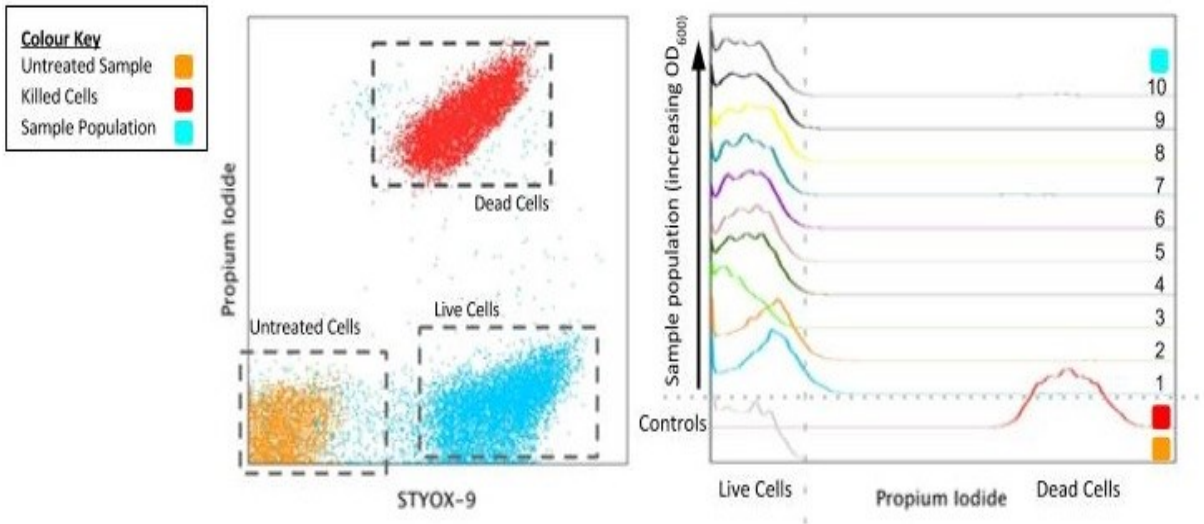


Figure 5.5 Cell viability of *P. pastoris* X33.

Cells were collect at approximately 2 hour intervals and the optical density of the cell population recorded as shown in Table 5.2. *P. pastoris* cells were stained with LIVE/DEAD® FungalLight™ yeast viability kit (Life Technologies, CA, USA). The chemical SYTO® 9 stained live cells and propidium iodide dead cells. Two controls were also added, one being a population of dead cells killed by a boiling step this control was stained, and a second control of untreated (unstained) cells. Cells were measured using a FACscan flow cytometer (Becton Dickinson, Oxford, UK) with settings Excitation/Emission 480/500 nm (SYTO® 9 stain) 490/635 nm (propidium iodide). Data was acquired using CellQuest software (Becton Dickinson, Oxford, UK) with samples measured on a high flow rate for 30 seconds. Data was analysed using Cylogic software (CyFlo Ltd, Turku, Finland).

Table 5.2 cell viability of *P. pastoris* X33.
The flow cytometry data collect from Figure 5.5 was analysed using Cylogic software (CyFlo Ltd, Turku, Finland). This calculated the number of live/dead cells from each sample. Shown in the table is the percentage of live P. pastoris cells observed at 10 different phases of the growth profile.

Sample population number	Biomass OD_{600}	Live cells (%)
1	0.864	99.55
2	2.34	99.32
3	3.83	99.01
4	7.45	99.42
5	11.60	99.54
6	12.80	99.59
7	18.00	98.47
8	19.00	98.29
9	20.00	98.53
10	22	98.11

The growth profile of the *P. pastoris* and subsequent $\log_{10}$ transformation indicated a shift in growth rate between approximately 18-22 hours. This is intersecting point between the mid-late exponential and late exponential-stationary phases. Therefore, the growth phase considered to harvest the cells was selected to occur at approximately 18 hours or at OD_{600} of 16. This time was viewed as the initial point of the mid-late exponential phase, replicating the point published CFPS are harvested for cell lysate preparation. Although the cell viability at OD_{600} of 16 was not determined, based on OD_{600} of 18, one can assume that the cell viability will be over 98%. Therefore, the cell culture will contain a healthy population of active live cells for the subsequent cell lysate.

5.5 *P. pastoris* cell disruption method for generation of a cell lysate

A number of techniques have been utilised to disrupt cells for the generation of lysates for CFPS systems as shown in Table 5.1. When selecting a lysis method for *P. pastoris*, the aspects of cell morphology and requirement for microsomes were taken into consideration. As *P. pastoris* is a yeast species it therefore shares a similar cell structure to that of *S. cerevisiae,* specifically both species contain a chitin cell wall. For the generation of a *S. cerevisiae* cell

lysate, the cell wall was enzymatically removed by zymolyase. This enable easier cell lysis, as spheroplasts are more fragile, therefore requiring less of a disruption force.

Although, the initial step of cell wall removal was shared between the two yeast species, the actual physical lysis method differed. In the creation of the *S. cerevisiae* CFPS, a high-pressure homogenizer was utilised; however, this method destroys any internal organelle structure such as microsomes. To obtain intact subcellular structures a milder treatment is required[484]. Previous studies to obtain microsomes from *P. pastoris* spheroplasts utilised a Dounce homogenizer. This was followed by low centrifugation speeds to ensure the remove of cell debris and nuclei, but retention of microsomal vesicles in the cell lysate. The method for the generation of *P. pastoris* cell lysate is outlined in further technical detail within the Methods chapter (Section 2.9). To maintain organelle structures a mannitol buffer was required, as mannitol sugars within the buffer help maintain the osmotic potential preventing the self-lysis of the microsomes[485].

After the *P. pastoris* cell lysate was generated, a dye was added to a sample of the lysate that selectively stains the ER, therefore detecting the presence of the microsomes. The photostable probe ER-Tracker™ Blue-White DPX (Life Technologies) was added to the *P. pastoris* cell lysate following the manufacturer's instructions. The stained CFPS was observed under a fluorescence microscope (Nikon, Japan) with excitation at ~374 nm and emission was detected between 430 and 640 nm. A whole cell staining of the *P. pastoris* X33 with ER-Tracker™ was performed alongside microsome staining (Figure 5.6). The subcellular location of the stained ER in this study correlated agreeably to previous reported organelle structure studies[486]. Specifically, *P. pastoris* contained discrete ER sites localised to the cell boundary due to presence of the vacuole and nucleus structures. The ER staining of the intact *P. pastoris* X33 demonstrated the affinity the dye has for the ER over other organelle structures. Thus, the ER-tracker™ should enable the identification of microsomes in the CFPS. Microsomes, prior to translation initiation in a CFPS, are characterised as rounded vesicular structures[487]. The distribution of ER-Tracker™ dye corresponds to the circular shape, suggesting the presence of stained microsomes within the *P. pastoris* CFPS (Figure 5.6).

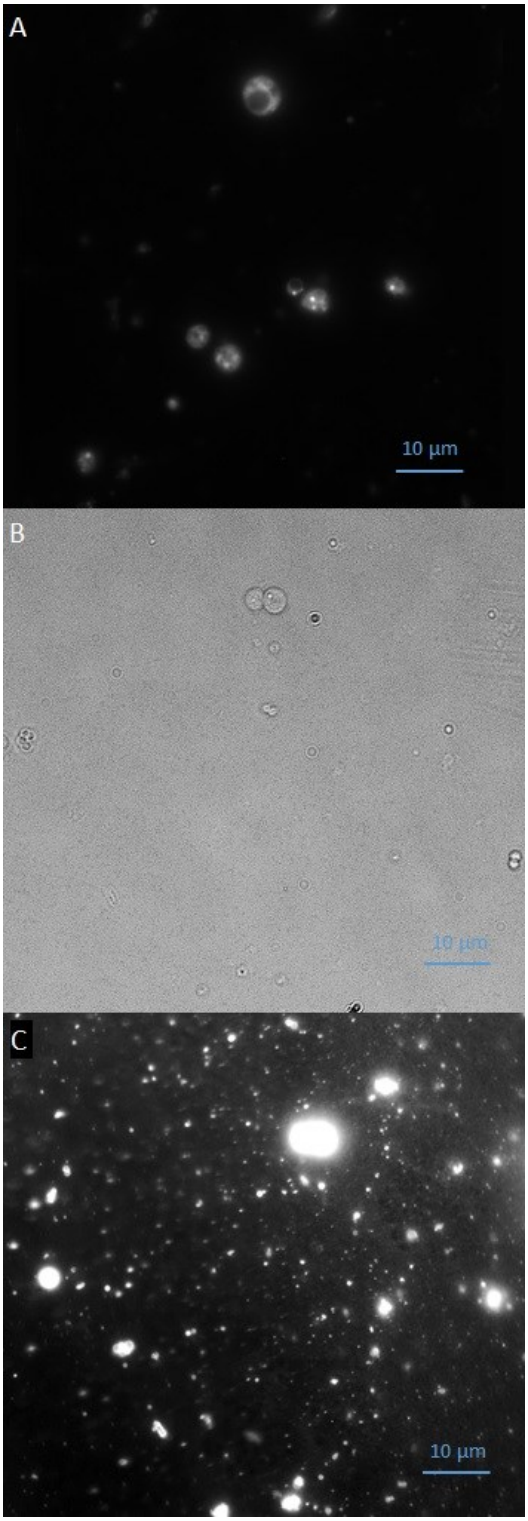


Figure 5.6 *P. pastoris* CFPS stained with ER-Tracker™ Blue-White DPX.
The photo denoted **A** represents the ER of *P. pastoris* stained with ER-Tracker™ Blue-White DPX (ThermoFisher, USA), and imaged by fluorescence microscope (Nikon, Japan). The microsome artefacts obtained after disruption of *P. pastoris* cells by a Dounce homogenizer were labelled with ER-Tracker™ (ThermoFisher, USA), and imaged by fluorescence microscope (Nikon, Japan). The photo denoted B represents the dark field and photo denoted C represents the light field.

5.6 From *P. pastoris* cell lysate to generation of a CFPS

The literature has demonstrated cell lysates to contain all the necessary machinery and factors to initiate translation, with many early examples of CFPS simply utilising the lysate to achieve recombinant protein expression. However, a number of improvements to CFPS have been made over the years to boost productivity and to reduce background (native) expression. To reduce the expression of native proteins from the CFPS, the mRNA is removed through digestion with micrococcal nuclease. Improvements to the productivity which resulted in increased product yields have been achieved through the supplementation of cell lysates with metal ions, co-factors and enzymes, as outlined in Table 5.3 within the 'Translation reaction conditions' section of that table. An energy mix and buffer was prepared for addition to the *P. pastoris* CPFS to concentrations described in Table 5.3. The concentrations of the materials were selected by amalgamating previous CFPS based on *S. cerevisiae* and CHO cells (Table 5.3).

Table 5.3 Chemical composition of the energy mix utilised in a CFPS system based on *P. pastoris* cell lysate.
The table displays different concentration of components which have previously demonstrated to enhance translation within a CFPS. As this is the first instance of a CFPS system based on *P. pastoris* the concentration of each components was based off previously published CFPS based on *S. cerevisiae* and CHO cells.

Component	Final concentration	Function
$Mg(OAc)_2$	2.0 mM	Metal ion cofactor
Kglutamate	14.4 mM	Metal ion cofactor
creatine kinase	0.27 mg	enzyme for ATP regeneration
ATP	20 mM	Energy molecule
GTP	5 mM	Energy molecule
Creatine phosphate	25 mM	high-energy phosphate bond donors
HEPES-KOH (pH 7.5)	22mM	Buffer to maintain pH conditions
Spermidine	0.2mM	Enhance AA corporation[488]
DDT	0.02 mM	Reducing agent, enhance enzyme activity
Each AA	0.08 mM	For translation

5.7 Transcriptional analysis of GFP expression vectors

The plasmids generated for the study contained a T7 promoter, which is common feature in many prokaryotic and eukaryotic CFPS systems[353]. The bacteriophage T7 RNA polymerase has high specificity for its own promoter so directs transcription exclusively to recombinant genes[489]. To ensure that a transcript is produced upon addition of T7 RNA polymerase to the CFPS, the pCrPV-GFP-pA and pGPR-GFP-pA plasmids were incubated with a T7 RNA Polymerase (NEB, USA) according to the manufacturer's instructions (Figure 5.7).

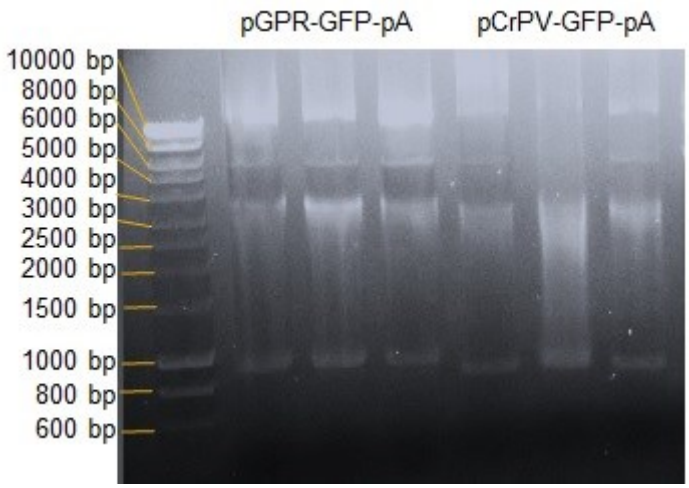


Figure 5.7 the transcription analysis of pGPR-GFP-pA and pCrPV-GFP-pA with T7 RNA Polymerase.
The *in vitro* transcription of two circularized plasmids pGPR-GFP-pA and pCrPV-GFP-pA was monitored by gel electrophoresis. The two plasmids were incubated in the presence of T7 RNA polymerase (NEB, WI,USA), transcription reactions were incubated at 37°C for either 1, 2 or 4 hours. 5 µL of each 40 µL reaction was loaded on a 1% agarose gel, and molecular weight of each transcript was compared against DNA marker Hyperladder I (Bioline, UK). Note: Transcribed ssRNA product will appear smaller than the dsDNA template.

The pCrPV-GFP-pA and pGPR-GFP-pA are expected to produce transcripts of ~983 bases of ~1191 bases. A transcript from pCrPV-GFP-pA and pGPR-GFP-pA was observed by gel electrophoresis, however the product was above the expected size. This could conceivably be due to the presence of double stranded structures, particularly from the IRES which as previously stated form secondary structures to promote ribosome binding (Section 5.2). The presence of secondary structures will alter the migration speed of the transcript, making them appear larger in size. Alternatively, as circularized plasmids were utilised, transcription potentially did not halt once the RNA polymerase met the terminator sequence, this phenomenon is known as read-through[490]. As such suppliers of T7 RNA polymerase recommend the use of linear DNA with the DNA strand concluding shortly after the terminator[491]. If read-through has occurred a larger transcript will be produced, which is what one might witness in Figure 5.7. The products of larger size (i.e. >3000bp) in Figure 5.7, could feasibly be due to read through. Nevertheless, an mRNA transcript is being produced from pCrPV-GFP-pA and pGPR-GFP-pA, which will provide a template for the translation of recombinant GFP.

5.8 Testing the expression of GFP in a *P. pastoris* CFPS

At this point, a *P. pastoris* CFPS had been produced with retained microsomes, supplemented with an energy mix outlined in Table 5.3. A successful *in vitro* transcription of an exogenous gene had been observed from plasmids generated in Section 5.2. Once the transcription had

been verified the translation of GFP was monitored from the same two plasmids. The addition of these plasmids yielded no expression of recombinant protein (data not shown), which was observed by no increase in fluorescence. The *P. pastoris* CFPS had been developed as a coupled reaction where transcription and translation takes place simultaneously in the CFPS. The transcription of a template has been observed, therefore the problem lies with initiation or incomplete translation, which in turn could be down to translation reaction conditions.

5.9 Comparison of *P. pastoris* CFPS against *S. cerevisiae* CFPS

As previously stated, a low pressure disruption method was utilised for the preparation of the cell lysate. The employment of this method for generating a CFPS is rare, with most cell lysates achieved through application of high pressure (see Table 5.1); possibly as disruption by higher pressures increases the number of full cell disruptions. An active *S. cerevisiae* CFPS which produced GFP has previously been reported from a *S. cerevisiae* cell lysate generated using a high pressure disruption method[356]. The GFP protein was under the control of T7 phage promoter, with an upstream CrPV IRES element and a downstream poly-A tail, similar to the pCrPV-GFP-pA construct used in this study. Therefore, a *S. cerevisiae* CFPS was generated in this study using an identical method for the generation of the *P. pastoris* CFPS. That is, a *S. cerevisiae* cell lysate was generated using a Dounce homogenizer (as outlined in Section 2.9.3) rather than a previously reported high pressure homogenizer. The *S. cerevisiae* CFPS offers a suitable control to test the efficiency of the disruption method due to the cell structural similarities between these two yeast species. The *S. cerevisiae* and *P. pastoris* cell lysates were both generated with a Dounce homogenizer and supplemented with energy conditions outlined in Table 5.3. The pCrPV-GFP-pA plasmid was added to each CFPS and the GFP output was tested after an eight hour incubation (Figure 5.8). A negative control containing no plasmid was utilised in the analysis to remove the impact of background fluorescence. The *S. cerevisiae* CFPS demonstrated GFP expression with the pCrPV-GFP-pA plasmid, whereas the *P. pastoris* did not (Figure 5.8).

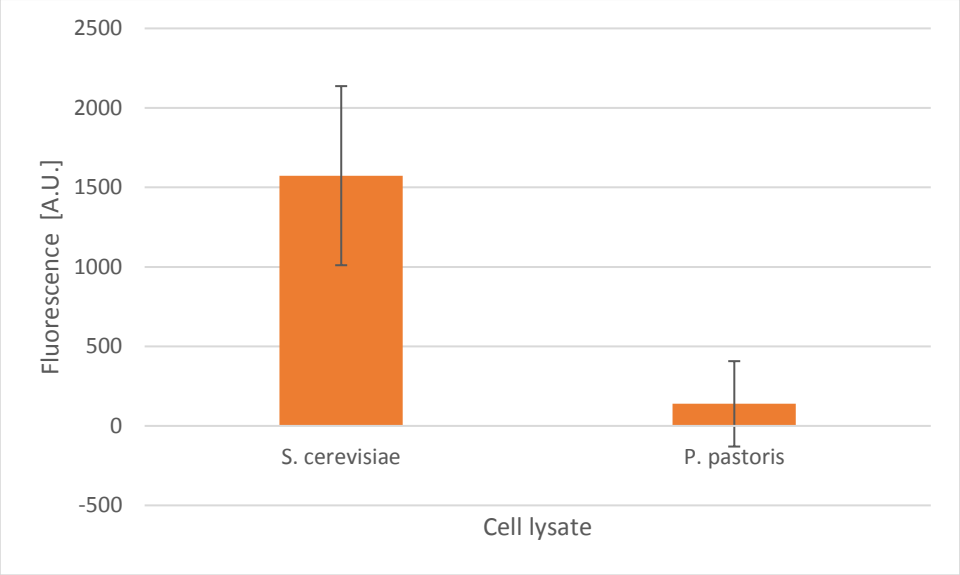


Figure 5.8 Analyses of GFP expression from the pCrPV-GFP-pA plasmid within a CFPS. The cell lysate was prepared following the procedure set out in Section 2.9, 25 µL of cell lysate was added to 25 µL of the energy mix from Table 5.3 with 250 ng of the pCrPV-GFP-pA plasmids. A negative control was set-up where no plasmid DNA was added. This was utilised in the analysis to remove the impact of background fluorescence. Experiments were conducted in biological quadruplicate; error bars represent the standard error from the average. The GFP expression was monitored over 8 hours, with fluorescence readings (excitation 485nm / emission 520nm) taken every 15 minutes using a Biotek Synergy HT plate reader, with an incubation temperature of 30°C. The bar graph represents the fluorescence of the reaction at 8 hours, the total reaction run can be found in the Appendix Section 9. The CFPS system based on S. cerevisiae lysate was demonstrated to have a fluorescent output indicative of GFP expression, whereas CFPS system based on P. pastoris had very low or possibly no fluorescence indicating no translation had occurred.

The above *in vitro* expression utilised a CrPV IRES to initiate GFP translation, this IRES has successfully been employed within another reported *S. cerevisiae* CFPS. At the time of this study the CrPV IRES had not been utilised within a *P. pastoris* expression system. As previously mentioned IRES elements can be species specific. The GPR-IRES element has previously shown cap-independent protein translation in an *in vivo P. pastoris* recombinant expression system, but never within *S. cerevisiae*. To test if the use of the CrPV IRES prevented expression this element was changed for the GPR IRES.

A plasmid containing the GPR-IRES element was tested in the *P. pastoris* CFPS and *S. cerevisiae* CFPS. The pGPR-GFP-pA was tested against the CFPS systems with energy conditions, as outlined in Table 5.3, and GFP output was tested after an eight hour incubation (Figure 5.9). A negative control containing no plasmid was utilised in the analysis to remove the impact of background fluorescence.

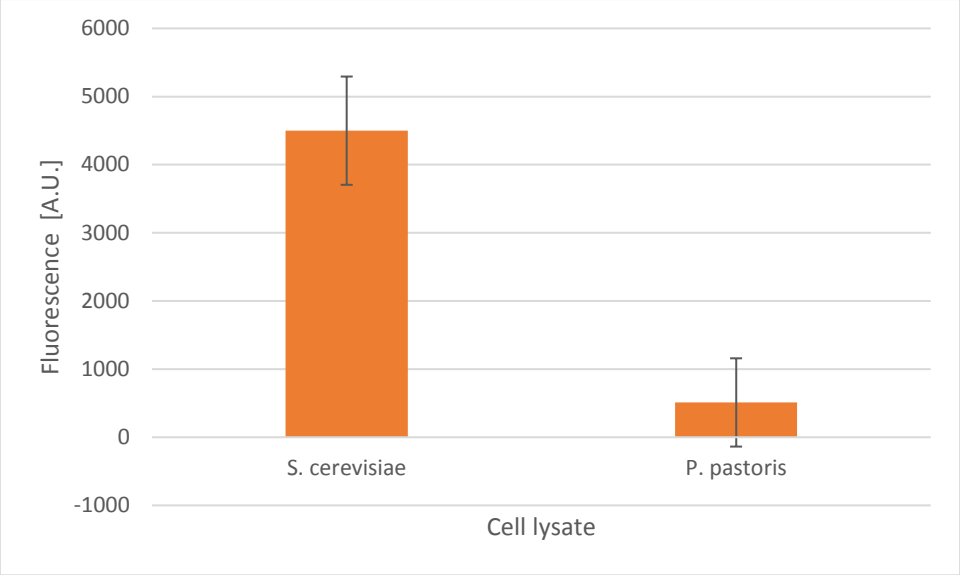


Figure 5.9 Analyses of GFP expression from the pGPR-GFP-pA plasmid within a CFPS. Where the previous CFPS reaction was with pCrPV-GFP-pA plasmids, this reaction displayed here proceeded with the pGPR-GFP-pA plasmid. The reaction set-up was identical, with the GFP expression monitored over 8 hours, with fluorescence readings (excitation 485nm / emission 520nm) taken every 15 minutes using a Biotek Synergy HT plate reader. The bar graph represents the fluorescence of the reaction at 8 hours, the total reaction run can be found in the Appendix Section 9. Experiments were conducted in biological quadruplicate; error bars represent the standard error from the average. The CFPS system based on S. cerevisiae lysate was demonstrated to have a fluorescent output indicative of GFP expression, whereas CFPS system based on P. pastoris had very low or possibly no fluorescence indicating no translation had occurred. The pGPR-GFP-pA plasmid was demonstrated to have an increased in fluorescence output over the pCrPV-GFP-pA, within the CFPS based on S. cerevisiae lysate.

The *P. pastoris* CFPS showed no detectable expression of GFP with the GPR-IRES in contrast to the *S. cerevisiae* CFPS (Figure 5.9). The GPR IRES element which has not previously been reported in a *S. cerevisiae* CFPS, demonstrated a fluorescence output three times higher than the CrPV IRES element. This is interesting as this further demonstrates that different IRES elements can affect the translational output. Therefore, it would be of interest to create a library of IRES elements not only to test if one can increase protein titres within a *S. cerevisiae* CFPS, as low protein titres in *S. cerevisiae* CFPS being a specified problem, but also to potentially discover a functional IRES for *P. pastoris* CFPS[356].

5.10 Energy limitations with a CFPS system

The observable expression of GFP within the *S. cerevisiae* CFPS indicated the low pressure lysis method did not hamper the translational activity of the CFPS. Therefore, the lack of translation within the *P. pastoris* CFPS must be due to other factors. It has previously been noted that the active metabolism of a crude extract is a function of the metabolism present in the living cell at the time of harvest[479]. For instance, the ATP consumption by non-protein producing metabolic pathways has been shown to cause energy limitations in *S. cerevisiae* CFPS reactions[362]. Moreover, an assessment of *S. cerevisiae* CFPS demonstrated that CFPS

harvested at higher optical densities when ATP concentrations were higher, yielded higher titres of recombinant proteins. Just as for *in vivo* metabolism, the initiation and continuation of transcription and translation is dependent on the energy pool, i.e. the ATP availability of the CFPS[362,492].

The internal ATP concentration of cell population was determined using the ATP bioluminescence assay CellTiter-Glo® (Promega, WI, USA). The ATP concentration of *P. pastoris* was calculated from a standard curve prepared concomitant with sampling (Appendix Section 10). The internal ATP concentration of the population steadily increased until an OD_{600} of approximately 18, when internal ATP concentration plateaued before decreasing (Figure 5.10). The observed decrease in internal ATP concentration of the population is at a relatively similar time to where a reduction in growth rate is observed (Figure 5.3). The *P. pastoris* CFPS was harvested at optical density of 16, which is close to the curve apex of *in vivo* ATP concentration.

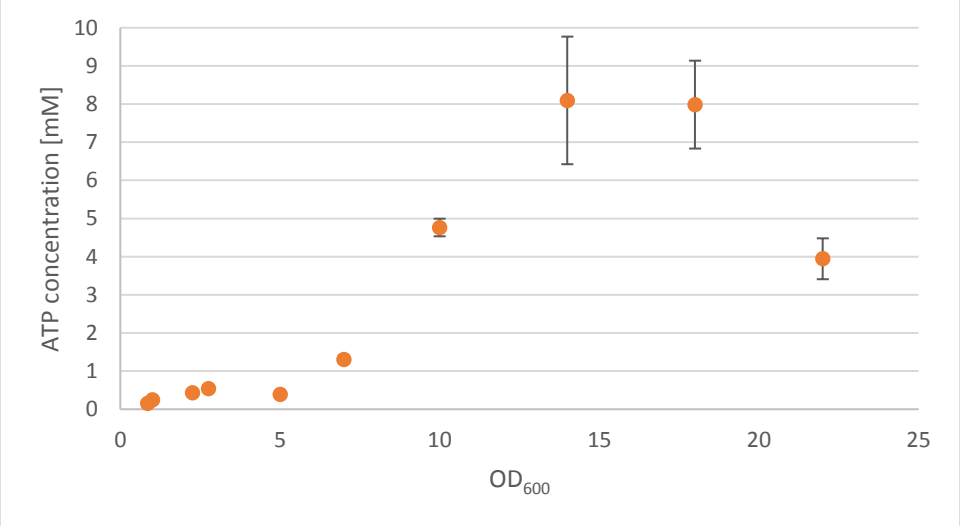


Figure 5.10 comparison of ATP concentration during *P. pastoris* growth profile
The luminescent signal produced during CellTiter-Glo® Assay and subsequent comparison to a standard curve (Appendix Section 10) enabled the calculation of ATP concentration at specified OD during the *P. pastoris* culture. Experiments were conducted in biological triplicate, the data presented is the average and error bars represent standard deviation.

Although harvesting of the *P. pastoris* fraction was executed at the highest ATP concentration, the actual concentration of ATP in the CFPS of *P. pastoris* could still be a limiting factor, with a low concentration of ATP in the *P. pastoris* CFPS preventing the translation. As *S. cerevisiae* has demonstrated translation, the ATP concentration of this CFPS is at sufficient levels. Therefore, by comparing the amount of ATP between these two systems one can elucidate if ATP is at an inadequate concentration within the *P. pastoris* CFPS. The concentration of ATP was determined in each CFPS using ATP bioluminescence assay CellTiter-Glo® (Promega,

WI, USA) (**Figure 5.11**). *P. pastoris* was observed to have a greater concentration of ATP in the cell lysate, approximately three times more than *S. cerevisiae* (Figure 5.11). The *in vivo* ATP concentration of *S. cerevisiae* was also determined (Figure 5.12). The *S. cerevisiae* was harvested at an OD_{600} of 14. The ATP concentration is lower in the *S. cerevisiae* cell population at harvest than that of the *P.* pastoris (Figure 5.12).

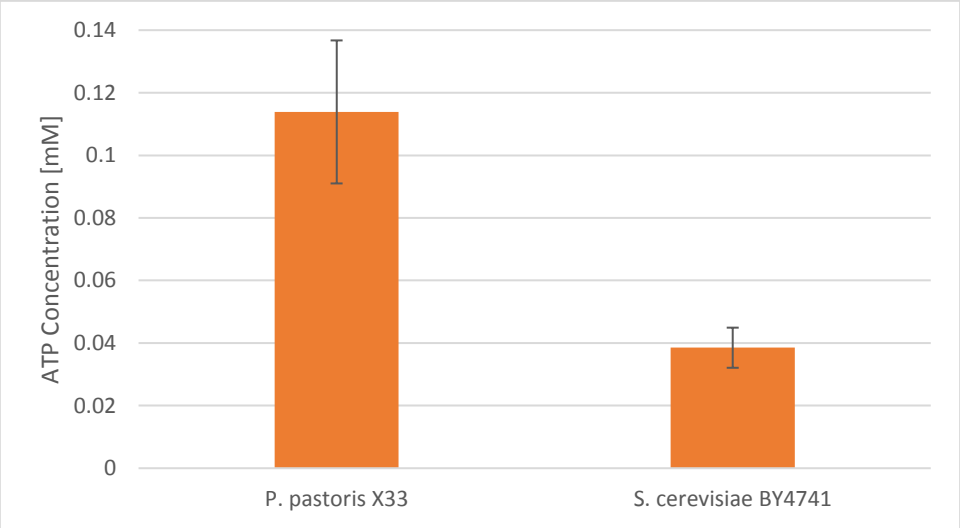


Figure 5.11 comparison of ATP concentration between *P. pastoris* and *S. cerevisiae* cell lysate.
The luminescent signal produced during CellTiter-Glo® Assay and subsequent comparison to a standard curve (Appendix Section 10) enabled the calculation of ATP concentration between the two cell lysates. The P. pastoris cell lysate was calculated to have an ATP concentration of 0.114 mM, while S. cerevisiae cell lysate had an ATP concentration 0.039 Mm. Experiments were conducted in biological triplicate, the data presented is the average and error bars represent standard deviation.

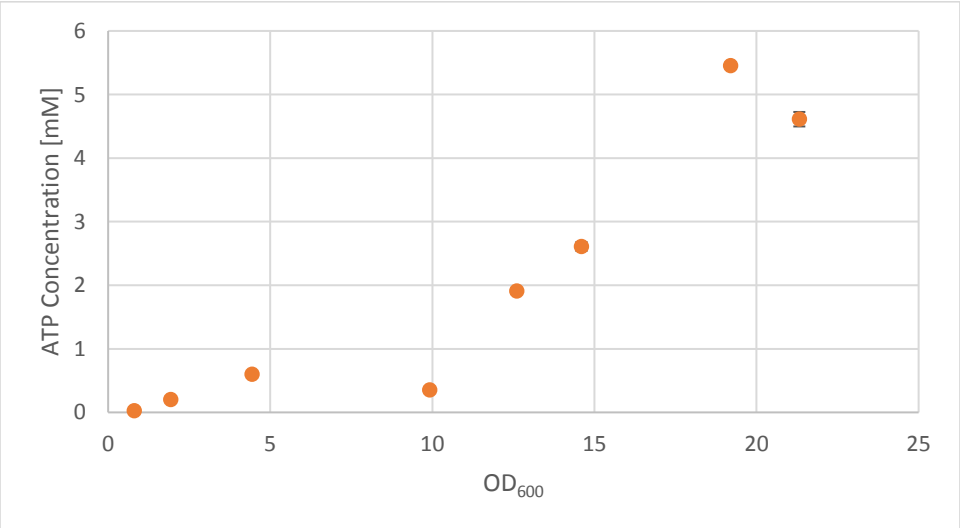


Figure 5.12 comparison of ATP concentration during *S. cerevisiae* growth profile.
The luminescent signal produced during CellTiter-Glo® Assay and subsequent comparison to a standard curve (Appendix Section 10) enabled the calculation of ATP concentration at specified OD during the *S. cerevisiae* culture. Experiments were conducted in biological triplicate, the data presented is the average and error bars represent standard deviation

The translation of an mRNA molecule to polypeptide chain is an energy rich reaction, with ATP essential for processes such as addition of amino acids to tRNAs and the ribosomal binding to the start codon. Hence, the hypothesis that a low ATP pool could inhibit translation in the *P. pastoris* CFPS. The comparison of *P. pastoris* and *S. cerevisiae,* interestingly showed that the total ATP concentration was ~2.75 times higher in *P. pastoris* CPFS. Therefore, it was incorrect to assume that energy limitations prevented translation in *P. pastoris.*

5.11 Summary

There has recently been a surge in the generation of cell free protein synthesis (CFPS) systems from an array of species. However, many of these CFPS cannot perform glycosylation or add non-therapeutically relevant glycosylation chains. To work toward solving this the development of a CFPS based on *P. pastoris* was attempted. The aim was to generate an active CFPS by following previous CFPS generation methodologies and use the *P. pastoris* CFPS to translate a recombinant protein. The generation of a yeast CFPS based on *S. cerevisiae* has previously been shown to synthesise recombinant proteins. A *S. cerevisiae* CFPS was also generated within the study and demonstrated activity. The generation of an active *P. pastoris* was not successful, indicated by the absence of reporter protein production. The absence of GFP was due to failings in the *P. pastoris* coupled reaction, in particular, the translation of mRNA to the recombinant polypeptide. The transcription of the recombinant gene occurred, indicated by a bacteriophage T7 RNA polymerase generating a transcript. A number of investigative steps were taken to understand why the *P. pastoris* CFPS was inactive. These steps included (1) checking the viability of the cell population before harvesting, (2) examining the ATP concentration and (3) the testing different IRES elements.

An obvious reason for translational inactivity in the *P. pastoris* CFPS system would be an insufficient chemical environment for ribosomal initiation and elongation; in particular a lack of the energy molecule ATP. The hypothesis was that if *P. pastoris* had a much lower ATP concentration than the active S. *cerevisiae* CFPS this could account for the lack of GFP expression. However, the *P. pastoris* was determined to contain more ATP than its *S. cerevisiae* CFPS counterpart. Therefore, the energy supply for the *P. pastoris* CFPS should be amenable to aid translation.

As previously stated the transcripts synthesised by a bacteriophage T7 RNA polymerase are not produced with a 5′ cap, which is necessary in eukaryotic cells (and therefore in a eukaryotic CFPS) for translation initiation. In order to initiate translation of uncapped mRNA transcripts, an IRES sequence was placed directly upstream of the recombinant reporter gene GFP. Two IRES elements were utilised in this study, and possibly mediated translation within a *S. cerevisiae* CFPS. IRES elements have been utilised in a number of CFPS systems to enhance

the recombinant protein translation. However, the mediation of translation by IRES elements has shown to be species dependent. For instance, the EMCV-IRES has shown to initiate translation in RRL CFPS[493], HeLa Cell CFPS[475] and CHO cell CFPS[355]. However, EMCV-IRES had previously failed to translate a recombinant protein *in vivo* within *P. pastoris* and the *S. cerevisiae* CFPS did not considered the EMCV-IRES as it requires more eukaryotic initiation factors (eIF) than other IRES type elements[468].

In this study the IRES elements of GPR and CrPV were used, and were shown to initiate GFP expression in a *S. cerevisiae* CFPS. The GPR-IRES was selected for use in the study on the basis that it was functional *in vivo* within *P. pastoris*. GPR-IRES had previously been utilised for *in vivo* translations in *P. pastoris,* with the IRES functioning to separate two genes in bicistronic construct[465,466]. Only by placing the IRES sequence between the two genes could the second gene be translated. However, the GPR-IRES has not been utilised in *P. pastoris* as a 5'UTR on a monocistronic construct with a T7 promoter, i.e. as has been designed in this study. Therefore, it could be that GPR-IRES within this context is unable to initiate a cap-independent translation. Although the GPR and CrPV IRES elements failed to mediate *in vitro* translation in *P. pastoris*, it should be noted that the *S. cerevisiae* CFPS showed a fluorescence output with these IRES elements indicating they can be functional.

It is important to select an IRES which will initiate cap-independent translation in *P. pastoris* CFPS. This may require the screening of a number of IRES elements to ensure a suitable candidate. As previously mentioned, *P. pastoris* does contain a number of native IRES elements, which could be ideal candidates[464]. The translation of a number of polypeptides has been achieved in RRL CFPS[494] and *S. cerevisiae* CFPS[468] by including capped mRNAs. The addition of capping structure to mRNA can be achieved through an *in vitro* enzymatic process E.coli RNA polymerase primed with an m^7G structure[495]. The capped mRNA in *S. cerevisiae* CFPS was shown to have higher translational activity than some IRES-dependent transcripts, which included a transcript with CrPV-IRES. Although addition of capped mRNA would contradict the coupled reaction and add an additional economic cost, it could be useful approach to help understand if IRES elements are solely responsible for lack of translation, and also could be used for IRES screening comparison.

To sum up, the generation of a *P. pastoris* CFPS was not successful. In normal circumstances the addition of IRES elements and ATP are used to enhance expression within a CFPS and are not necessary used to initiate translation within an inactive CFPS. There are many published examples of crude cellular lysate being utilised for expression, before protein titres are improved through addition of IRES elements, polyA tails, chemical and enzymatic optimisation. Currently no translation is being observed within the *P. pastoris* CFPS with what

one could deem enhanced environmental conditions for a coupled reaction. Although examining the ATP concentration helped to rule out energy limitations as a cause, the true reason of the inactivity still remains elusive. Although *P. pastoris* is seen as a highly productive expression host, it is in fact categorised as having a relatively low specific productivity at the individual cell level. The high protein titres are obtained due to the high cell densities thereby achieving high volumetric productivity[496]. This low specific productivity could explain the lack of productivity observed within CFPS. The reasons for the low cell productivity in *P. pastoris* is still being actively investigated. In the future with a better understanding of *P. pastoris* as a host it might be possible to generate an active CFPS.

Chapter 6. Discussion

6.1 Summary of the intended work

The aim of this project was to work towards the development of a cell-free based system for the synthesis of recombinant N-linked glycoproteins. The proof of concept design aimed to demonstrate the feasibility of synthesising a glycoprotein using a cell-free based system. The development of a CFPS system linked to an *in vitro* glycan processing pathway, should enable the production of user-defined homogenous glycoform.

The project began with applying a glycoengineering approach to the yeast species *P. pastoris*, to develop a strain which would synthesise a core pentamannosyl oligosaccharide on a glycoprotein. A glycoprotein synthesised from a glycoengineered strain containing the core pentamannosyl oligosaccharide can be modified by an *in vitro* metabolic pathway mimicking the actions of a mammalian Golgi apparatus. The project aimed to express a glycosidase and two glycotransferases, for the creation of an 'artificial Golgi' to edit the core oligosaccharide toward a complex type glycan structure. The feasibility of producing a cell-free protein synthesis (CFPS) system from *P. pastoris* cell lysate was also tested. A functioning CFPS system would allow for the development of a *P. pastoris* cell lysate from the glycoengineered strain. Again, an oligosaccharide synthesised from a CFPS developed from a glycoengineered strain can be modified by an 'artificial Golgi'.

This project is laying down the foundations for the generation of a CFPS and artificial Golgi which could lead to the generation of a glycoprotein with oligosaccharide consisting of a user-defined structure. Through the results presented here a number of key conclusions have been made which will be outlined below. These findings are aiding in the future experimental planning and direction of this project entitled 'Towards the cell-free expression of glycoproteins'.

6.2 Improvements and Future Work

The main objective of the project was to gain a fully functional cell free protein synthesis which can covalently attach an oligosaccharide chain of $Man_5GlcNAc_2$ to an asparagine residue in a consensus sequence, and that this oligosaccharide can be rearranged by an *in vitro* enzymatic pathway to allow for the synthesis of specific N-glycan structures, which in this study is $GlcNAc_2Man_3GlcNAc_2$. There are many further lines of work which are required to complete the main objective of the project and a number of new objectives which can be added to enhance this work. These are discussed below.

6.2.1 A glycoengineering approach in *P. pastoris* for the generation of a core pentamannosyl glycoprofile

This project began with an exerted effort to generate a *P. pastoris* strain producing the $Man_5GlcNAc_2$ glycoprofile. The glycoengineering of *P. pastoris* strain had been accomplished a number of times prior to this project, as reviewed in Section 1.26. As such, it was assumed that a replication of methodology from the literature would provide a *P. pastoris (Δoch1:*Man1a+) strain. However, this did not transpire, with no analysed *P. pastoris* transcribing the recombinant Man1a gene.

The initial complementary DNA sequence utilised for the homologous recombination event replicated the design from previously published work[497]. The subsequent attempted HR event with this design yielded no detectable strains with the correct genotype. This failure was reasoned to be due to the design requirements for a double crossover event occurring upstream and downstream of the intended gene targeted for knockout. Subsequently, any future knockout strategies in *P. pastoris* are now designed to ensure the knockout target site is within the CDS of intended target gene and that only *P. pastoris Δku70* is utilised. By implementing these design rules, a *P. pastoris Δoch1* strain was produced in this study.

Despite the utilisation of a number of different cloning strategies the generation of a *P. pastoris* with $Man_5GlcNAc_2$ failed. Consequently, a further methodology is required to help generate or identify a *P. pastoris* with $Man_5GlcNAc_2$. For instance, an innovative redesign of the Man1a genetic construct could generate a *P. pastoris* with $Man_5GlcNAc_2$. The insertion of the Man1a gene was accompanied with the insertion of an antibiotic resistance gene, to aid with selection. Although the transcription of Man1a was undetectable within the clones tested, the transcription and translation of the selective marker occurred due to the survival of clones on selective media. A strategy of redesigning a Man1a construct to utilise the essential need for antibiotic resistance gene could generate a positive clone. Constructing a synthetic operon with the Man1a and selective marker under the control of one constitutive promoter, would mean the transcription of selective marker gene and subsequent translation can only occur with transcription of Man1a gene. The utilisation of different promoters to the TEF promoter could also aid in transcription of the Man1a gene.

Although the redesigning of Man1a construct could generate a positive clone, a real drawback within the study has been the methodology utilised to test clones. The current screening method is based on genotypic and transcriptomic testing. This method can be slow and laborious. It would be preferable to have a high-throughput screening method based on phenotype. A phenotypic analysis of the glycoprofile offers an easier route to discovering the

successful translation of Man1a. Krainer *et al.* noted that inactivation of the Och1p activity shifted the interaction capacity of a mannose specific lectin[408]. The *Galanthus nivalis* lectin (GNL) and *Lens culinaris* lectin (LCA) have stronger binding specificity for high mannose and low mannose structures, respectively. The use of these fluorescent lectin-based assay would enable a high-throughput phenotypic assay for determination of successful Man1a activity. However, a lectin with binding affinity for $Man_5GlcNAc_2$ would first be required. Kanagawa *et al.* have identified a lectin which binds specifically for $Man_5GlcNAc_2$[498], but this lectin has not been modified to contain a fluorescence output, like GNL or LCA. Thus, it could not be utilised in a similar high-throughput screening method as described by Krainer[408] without first modifying it.

It must be noted that since commencing of the project a number of research groups have obtained a *P. pastoris* with a $Man_5GlcNAc_2$, and as such it is now available commercially; *Pichia pastoris* SuperMan5 strain (BioGrammatics, Inc. CA, USA). Therefore, to progress in the project this strain could be obtained and utilised.

6.2.2 Towards the development of an *in vitro* metabolic pathway for the processing of a core glycoprofile

The project aimed to develop an *in vitro* enzymatic pathway for the modification of an oligosaccharide. The expression of an active GlcNAcT-I has gone some way to beginning this effort. However, as shown in Table 6.1 a further 5 (or 6 with fucosylation) recombinant proteins are required to achieve a complex-type oligosaccharide. The project is currently being hindered without the expression of Man-II.

If the failure to produce a soluble active Man-II protein continues, one solution is the use an alternative mannosidase or mannosidases for the removal of the two mannose sugars. A large array of class II α-mannosidase exists, distributed throughout animals, plants and microorganisms. Although many class II α-mannosidases exist, the cleavage specificity is very different from a mammalian Golgi based class II α-mannosidase. For instance, the *Sulfolobus solfataricus* class II α-mannosidase can cleave an α-1,4 linkage between mannose sugars[499]. There are a number of class II α-mannosidase which exist in humans from the Golgi based Man-II from this study, to lysosomal and cytosolic based Man-II's involved in catabolism of glycoproteins with the ability to cleave α-1,2 α-1,3 and α-1,6 linkages[424,500]. The unusual aspects of the Golgi Man-II is in its dual specificity for α-1,2, α-1,6 linkages and its prerequisite for the GlcNAc residue added by GlcNAcT-I[421]. The small pocket for its active site and different substrate anchor sites in the Golgi Man-II offers greater specificity for the oligosaccharide substrate, which is not matched by other class II α-mannosidase[421,212]. Therefore, the strategy

of removing the two mannose sugars to achieve the $GlcNAc_1Man_3GlcNAc_2$ with the Golgi Man-II enzyme is the more desirable option. The use of different species Golgi Man-II could yet yield an active enzyme.

In this study the translation of first three Golgi based N-linked glycosylation enzymes was attempted. There is potential to further process the $GlcNAc_2Man_3GlcNAc_2$ oligosaccharide to a more therapeutic relevant oligosaccharide. This will involve the translation of number of different recombinant glycosyltransferase. Table 6.1 represents a selection of glycosyltransferase which could be utilised to generate a complex glycan structure. In this study the overexpression of recombinant proteins was attempted in *E. coli* and *P. pastoris.* The expression optimisation to ensure translation of an active soluble protein can be a time-consuming process. There was more confidence in successful expression of an active GlcNAcT-I due to an earlier study that demonstrated the translation of a recombinant fusion GlcNAcT-I in *E. coli.* By using a genetic engineered *E. coli* strain non-fusion enzyme was synthesised in this study. Of the six other enzymes shown in Table 6.1 only two have previously been successfully translated in *E. coli*. The expression host *P. pastoris* offers an alternative source for expression and was shown to successfully express SiaT-I (Table 6.1). Many of the required glycotransferases have been expressed in the humanized yeast, although human Man-II failed to express in *P. pastoris* in this study. However, the humanized yeast utilised glycosyltransferases from a number of mammalian sources. Therefore, translation of non-human glycosyltransferases could offer an alternative source for glycotransferases.

Table 6.1 A selection of human glycosyltransferases and glycosidases which could be utilised for the processing/maturation of N-linked glycan chain.
The information regarding recombinant expression was obtained from BRENDA (www.brenda-enzymes.org)

Enzyme	EC number	Required metal ion	pH optimum	Inhibitors	Recombinant expression in *E. coli*?	Recombinant expression in *P. pastoris*?
FucT-VIII	2.4.1.68	Mg^{2+}	4.5 - 8.5	Co^{2+} Cu^{2+} Zn^{2+} Ni^{2+} GDP	No	No
GlcNAcT-I	2.4.1.101	Co^{2+} Mn^{2+}	6.5 - 9	Ca^{2+} Zn^{2+} UDP	Yes* [383,384]	No
GlcNAcT-II	2.4.1.143	Mn^{2+}	6-7	UDP	Yes#	No

GlcNAcT-III	2.4.1.144	Mn^{2+}	6-6.5	UDP	Yes [501]	No
GalT-I	2.4.1.38	Mn^{2+}	6.5-8	UDP	Yes[502,503,504]	Yes[505]
Man-II	3.2.1.114	Co^{2+} Fe^{2+} Mn^{2+}	5.5-6.5		Yes[$]	No
SiaT-I	2.4.99.1	Mg^{2+}	6-6.5	CDP CTP Cu^{2+} UTP Mn^{2+}	No	Yes[505]

* Expressed in this study and shown to be active, # Expressed in this study but unable to test activity, $ Expressed in this study but no activity was observed.

The Golgi glycosylation enzymes are type-II membrane proteins with a transmembrane and cytosolic domain at the N-terminus. The architecture of the Golgi is such that individual enzymes are located within different compartments. The location of the enzyme is dependent on processing it performs on the oligosaccharide. For instance, GlcNAcT-I which is one of the first glycotransferases in the glycosylation processing is located within the cis-Golgi compare to GalT-I and SiaT-I which transfer sugars at the end of the glycosylation pathway and therefore located within the trans-Golgi[506]. The individual compartments of the Golgi have varying environments, as is exemplified by the pH optimums of the individual enzymes being different (Table 6.1). The different enzyme needs require different co-factors and pH for efficient activity, changes in these conditions could lower the activity or completely inhibit the enzyme.

The three Golgi enzymes translated in this study contained a polyhistidine-tag to aid in purification. The polyhistidine-tag is located on the N-terminus of the protein, the same position as the transmembrane domain. The polyhistidine-tag could be utilised to tether the protein, therefore allowing for the enzyme to be immobilised in a tailored environment for optimal activity. This can allow for the replenishing of donor sugars, and removal of waste products such as UDP which is known to inhibit enzymatic action (Table 6.1).

6.2.3 The development of a Cell-Free Protein Synthesis system based on *P. pastoris* cell lysate

The eventual aim of the *P. pastoris* CFPS system is to translate a recombinant secretory protein with N-linked glycosylation. A glycoprotein with a consensus tripeptide sequence was successfully synthesised with an oligosaccharide within a CHO CFPS which contained microsomes[355]. Using a staining method, the *P. pastoris* CFPS was shown to contain

microsomes (Figure 5.6). The eventual aim is to emulate the CHO CFPS to generate N-linked glycoproteins. The site occupancy of the protein in CHO CFPS was low and the *en bloc* transfer was likely to be $Man_{8-9}GlcNAc_2$ although this was not evaluated in the study. The site occupancy could also be a problem within a *P. pastoris* CFPS and will need to be investigated. Microsomes also enable an easier purification of proteins from CFPS[448]. As microsomes can be easily separated by gentle centrifugation, therefore the localised translate proteins can be co-purified with microsomes from CFPS; with a treatment with mild detergent releasing the protein[448].

The eventual *P. pastoris (Δoch1:Man1a$^+$)* strain is to be utilised as a cell lysate for a CFPS system. The retention of the microsomes in the CFPS and translocation of a recombinant protein to the microsomes could allow for *en bloc* transfer oligosaccharide and its processing to $Man_5GlcNAc_2$.The results demonstrated that the microsomes were present within CFPS based on *P. pastoris* cell lysates. However, the inactivity of *P. pastoris* CFPS and the inability to create *P. pastoris* Man1a$^+$ strain has meant inability to fulfil this objective.

If one is unable to create a *P. pastoris* CFPS, an alternative option would be to use the active *S. cerevisiae* CFPS supplemented with *P. pastoris* microsomes. The translocation of a protein to the ER *in vivo* is achieved by the addition of a signal secretion peptide, which is recognised by the protein-ribosomal complex SRP[507]. The SRP in *P. pastoris* recognises the α-mating factor secretion signal from *S. cerevisiae*[508]. Thus, potentially if a *S. cerevisiae* cell lysate was to be supplemented with microsomes from *P. pastoris* the secretory protein could translocate to these microsomes using the *S. cerevisiae* native SRP. This is similar idea to the commercial RRL which are supplemented with canine pancreas microsomal membranes[509]. Therefore, the fine-tuning of *P. pastoris* CFPS which could be long laborious effort may not be required to complete the objective.

.

References

1. Varki, A. *Essentials of glycobiology*. (Cold Spring Harbor Laboratory Press, 2009).

2. Ceroni, A., Dell, A. & Haslam, S. M. The GlycanBuilder: a fast, intuitive and flexible software tool for building and displaying glycan structures. *Source Code Biol. Med.* **2,** 3 (2007).

3. Demain, A. L. Small bugs, big business: The economic power of the microbe. *Biotechnol. Adv.* **18,** 499–514 (2000).

4. Junod, S. W. (FDA). Celebrating a Milestone: FDA's Approval of First Genetically-Engineered Product. *History Corner, Update Magazine, Food and Drug Law Institute* (2007). Available at: https://www.fda.gov/AboutFDA/WhatWeDo/History/ProductRegulation/SelectionsFromFDLIUpdateSeriesonFDAHistory/ucm081964.htm.

5. Goeddel, D. V *et al.* Expression in Escherichia coli of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 106–110 (1979).

6. Evens, R. P. Pharma Success in Product Development—Does Biotechnology Change the Paradigm in Product Development and Attrition. *AAPS J.* **18,** 281–285 (2016).

7. Walsh, G. Biopharmaceutical benchmarks 2014. *Nat. Biotechnol.* **32,** (2014).

8. Nicolaides, N. C., Sass, P. M. & Grasso, L. Advances in targeted therapeutic agents. *Expert Opin. Drug Discov.* **5,** 1123–40 (2010).

9. Mullard, A. Maturing antibody–drug conjugate pipeline hits 30. *Nat. Rev. Drug Discov.* **12,** 483–483 (2013).

10. Fernandez, F. J. & Vega, M. C. Technologies to keep an eye on: alternative hosts for protein production in structural biology. *Curr. Opin. Struct. Biol.* **23,** 365–373 (2013).

11. Singh, S. M. & Panda, A. K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **99,** 303–310 (2005).

12. Singh, A., Upadhyay, V., Upadhyay, A. K., Singh, S. M. & Panda, A. K. Protein recovery from inclusion bodies of Escherichia coli using mild solubilization process. *Microb Cell Fact* **14,** 41 (2015).

13. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of

Protein Interaction Networks. *PLoS Comput. Biol.* **11,** e1004049 (2015).

14. Walsh, G. & Jefferis, R. Post-translational modifications in the context of therapeutic proteins. *Nat Biotech* **24,** 1241–1252 (2006).

15. Berg, J. M. *Biochemistry*. (Palgrave Macmillan, 2011).

16. Marshall, A. GM soybeans and health safety - a controversy reexamined. *Nat Biotech* **25,** 981–987 (2007).

17. Brown, T. A. *Genomes*. (BIOS Scientific, 2002).

18. Silhavy, T. J., Kahne, D. & Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb. Perspect. Biol.* **2,** a000414 (2010).

19. Balbás, P. & Lorence, A. Recombinant Gene Expression. *Rev. Protoc.* **267,** 506 (2004).

20. Patrick Tailliez. Mini-revue : les bactéries lactiques, ces êtres vivants apparus il y a près de 3 milliards d'années. *Lait* **81,** (2001).

21. Mierau, I. & Kleerebezem, M. 10 Years of the nisin-controlled gene expression system (NICE) in Lactococcus lactis. *Appl. Microbiol. Biotechnol.* **68,** 705–717 (2005).

22. Cavanagh, D., Fitzgerald, G. F. & McAuliffe, O. From field to fermentation: The origins of Lactococcus lactis and its domestication to the dairy environment. *Food Microbiol.* **47,** 45–61 (2015).

23. Mierau, I. *et al.* Industrial-scale production and purification of a heterologous protein in Lactococcus lactis using the nisin-controlled gene expression system NICE: the case of lysostaphin. *Microb. Cell Fact.* **4,** 15 (2005).

24. Guo, S. *et al.* The recombinant Lactococcus lactis oral vaccine induces protection against C. difficile spore challenge in a mouse model. *Vaccine* **33,** 1586–1595 (2015).

25. Bahey-El-Din, M. Lactococcus lactis-based vaccines from laboratory bench to human use: An overview. *Vaccine* **30,** 685–690 (2012).

26. Pontes, D. S. *et al.* Lactococcus lactis as a live vector: Heterologous protein production and DNA delivery systems. *Protein Expr. Purif.* **79,** 165–175 (2011).

27. Sandkvist, M. & Bagdasarian, M. Secretion of recombinant proteins by Gram-negative bacteria. *Curr. Opin. Biotechnol.* **7,** 505–511 (1996).

28. Song, Y., Nikoloff, J. M. & Zhang, D. Improving protein production on the level of

regulation of both expression and secretion pathways in *Bacillus subtilis*. *J. Microbiol. Biotechnol.* **25,** 963–977 (2015).

29. Lee, J. Y., Na, Y. A., Kim, E., Lee, H. S. & Kim, P. The actinobacterium *Corynebacterium glutamicum*, an industrial workhorse. *J Microbiol Biotechnol* **26,** 807–822 (2016).

30. Anné, J., Maldonado, B., Van Impe, J., Van Mellaert, L. & Bernaerts, K. Recombinant protein production and streptomycetes. *J. Biotechnol.* **158,** 159–167 (2012).

31. Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27,** 297–306 (2009).

32. Wu, S. *et al.* Functional production and characterization of a fibrin-specific single-chain antibody fragment from *Bacillus subtilis*: Effects of molecular chaperones and a wall-bound protease on antibody fragment production. *Appl. Environ. Microbiol.* **68,** 3261–3269 (2002).

33. Harwood, C. R., Pohl, S., Smith, W. & Wipat, A. Bacillus subtilis. *Methods Microbiol.* **40,** 87–117 (2013).

34. Ferrer-Miralles, N., Domingo-Espín, J., Corchero, J. L., Vázquez, E. & Villaverde, A. Microbial factories for recombinant pharmaceuticals. *Microb. Cell Fact.* **8,** 17 (2009).

35. Rosano, G. L. & Ceccarelli, E. A. Recombinant protein expression in Escherichia coli: Advances and challenges. *Front. Microbiol.* **5,** 1–17 (2014).

36. Sezonov, G., Joseleau-Petit, D. & D'Ari, R. Escherichia coli Physiology in Luria-Bertani Broth . *J. Bacteriol.* **189,** 8746–8749 (2007).

37. Wasi, S., Tabrez, S. & Ahmad, M. Use of Pseudomonas spp. for the bioremediation of environmental pollutants: a review. *Environ. Monit. Assess.* **185,** 8147–8155 (2013).

38. Overton, T. W. Recombinant protein production in bacterial hosts. *Drug Discov. Today* **19,** 590–601 (2014).

39. Asmana Ningrum, R. Human Interferon Alpha-2b: A Therapeutic Protein for Cancer Treatment. *Scientifica (Cairo).* **2014,** 970315 (2014).

40. Walsh, G. Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* **28,** 917–24 (2010).

41. Hershfield, M. S. *et al.* Treating gout with pegloticase, a PEGylated urate oxidase, provides insight into the importance of uric acid as an antioxidant in vivo. *Proc. Natl.*

Acad. Sci. U. S. A. **107,** 14351–14356 (2010).

42. Baeshen, M. N. *et al.* Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *J. Microbiol. Biotechnol.* **25,** 953–962 (2015).

43. de Marco, A. Strategies for successful recombinant expression of disulfide bond-dependent proteins in Escherichia coli. *Microb. Cell Fact.* **8,** 26 (2009).

44. Schlegel, S. *et al.* Optimizing heterologous protein production in the periplasm of E. coli by regulating gene expression levels. *Microb. Cell Fact.* **12,** 1–12 (2013).

45. Denoncin, K. & Collet, J.-F. Disulfide Bond Formation in the Bacterial Periplasm: Major Achievements and Challenges Ahead. *Antioxid. Redox Signal.* **19,** 121002061152001 (2012).

46. Fakruddin, M., Mohammad Mazumdar, R., Bin Mannan, K. S., Chowdhury, A. & Hossain, M. N. Critical Factors Affecting the Success of Cloning, Expression, and Mass Production of Enzymes by Recombinant E. coli. *ISRN Biotechnol.* **2013,** 590587 (2013).

47. Gao, X. *et al.* Soluble cytoplasmic expression, rapid purification, and characterization of cyanovirin-N as a His-SUMO fusion. *Appl. Microbiol. Biotechnol.* **85,** 1051–1060 (2010).

48. Robinson, M.-P. *et al.* Efficient expression of full-length antibodies in the cytoplasm of engineered bacteria. *Nat. Commun.* **6,** 8072 (2015).

49. Bessette, P. H., Aslund, F., Beckwith, J. & Georgiou, G. Efficient folding of proteins with multiple disulfide bonds in the Escherichia coli cytoplasm. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 13703–13708 (1999).

50. Levy, R., Weiss, R., Chen, G., Iverson, B. L. & Georgiou, G. Production of correctly folded Fab antibody fragment in the cytoplasm of Escherichia coli trxB gor mutants via the coexpression of molecular chaperones. *Protein Expr. Purif.* **23,** 338–347 (2001).

51. Chen, J. *et al.* Chaperone activity of DsbC. *J. Biol. Chem.* **274,** 19601–19605 (1999).

52. Kamionka, M. Engineering of therapeutic proteins production in Escherichia coli. *Curr. Pharm. Biotechnol.* **12,** 268–74 (2011).

53. Harris, J. M., Martin, N. E. & Modi, M. Pegylation: a novel process for modifying pharmacokinetics. *Clin. Pharmacokinet.* **40,** 539–551 (2001).

54. Bailon, P. & Won, C.-Y. PEG-modified biopharmaceuticals. *Expert Opin. Drug Deliv.* **6,**

1–16 (2009).

55. Harris, J. M. & Chess, R. B. Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* **2,** 214–21 (2003).

56. Evens, R. & Kaitin, K. The evolution of biotechnology and its impact on health care. *Health Aff. (Millwood).* **34,** 210–219 (2015).

57. Swierczewska, M., Lee, K. C. & Lee, S. What is the future of PEGylated therapies? *Expert Opin. Emerg. Drugs* **20,** 531–536 (2015).

58. Khow, O. & Suntrarachun, S. Strategies for production of active eukaryotic proteins in bacterial expression system. *Asian Pac. J. Trop. Biomed.* **2,** 159–162 (2012).

59. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional transfer into E. coli. *Science* **298,** 1790–3 (2002).

60. Walsh, G. Biopharmaceutical benchmarks. *Nat Biotech* **18,** 831–833 (2000).

61. Bill, R. M. Recombinant Protein Production in Yeast. *Life Sci.* **531,** 588 (2009).

62. Bill, R. M. Playing catch-up with escherichia coli: Using yeast to increase success rates in recombinant protein production experiments. *Front. Microbiol.* **5,** 1–5 (2014).

63. Bergman, L. W. Growth and maintenance of yeast. *Methods Mol. Biol.* **177,** 9–14 (2001).

64. Munro, S. What can yeast tell us about N -linked glycosylation in the Golgi apparatus? *FEBS Lett.* **498,** 223–227 (2001).

65. Mattanovich, D., Sauer, M. & Gasser, B. Yeast biotechnology: teaching the old dog new tricks. *Microbial cell factories* **13,** 34 (2014).

66. Beach, D., Durkacz, B. & Nurse, P. Functionally homologous cell cycle control genes in budding and fission yeast. *Nature* **300,** 706–709 (1982).

67. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274,** 546,563-567 (1996).

68. Idiris, A., Tohda, H., Kumagai, H. & Takegawa, K. Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* **86,** 403–417 (2010).

69. Damasceno, L. M., Huang, C.-J. & Batt, C. A. Protein secretion in Pichia pastoris and advances in protein production. *Appl. Microbiol. Biotechnol.* **93,** 31–39 (2012).

70. Van den Steen, P., Rudd, P. M., Dwek, R. A. & Opdenakker, G. Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **33,** 151–208 (1998).

71. Hang, H. C. & Bertozzi, C. R. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorg. Med. Chem.* **13,** 5021–5034 (2005).

72. Tanner, W. & Lehle, L. Protein glycosylation in yeast. *Biochim. Biophys. Acta* **906,** 81–99 (1987).

73. Jungmann, J. & Munro, S. Multi-protein complexes in the cis Golgi of Saccharomyces cerevisiae with alpha-1,6-mannosyltransferase activity. *EMBO J.* **17,** 423–434 (1998).

74. Nakajima, T. & Ballou, C. E. Yeast manno-protein biosynthesis: solubilization and selective assay of four mannosyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* **72,** 3912–3916 (1975).

75. Liu, H.-H., Ji, X.-J. & Huang, H. Biotechnological applications of Yarrowia lipolytica: Past, present and future. *Biotechnol. Adv.* **33,** 1522–1546 (2015).

76. Martínez, J. L., Liu, L., Petranovic, D. & Nielsen, J. Pharmaceutical protein production by yeast: Towards production of human blood proteins by microbial fermentation. *Curr. Opin. Biotechnol.* **23,** 965–971 (2012).

77. Nielsen, J. Production of biopharmaceutical proteins by yeast: Advances through metabolic engineering. *Bioengineered* **4,** 207–211 (2013).

78. Wood, V. *et al.* The genome sequence of Schizosaccharomyces pombe. *Nature* **415,** 871–880 (2002).

79. Ohashi, T. & Takegawa, K. N- and O-linked oligosaccharides completely lack galactose residues in the gms1och1 mutant of Schizosaccharomyces pombe. *Appl. Microbiol. Biotechnol.* **86,** 263–272 (2010).

80. Ramezani-Rad, M. *et al.* The Hansenula polymorpha (strain CBS4732) genome sequencing and analysis. *FEMS Yeast Res.* **4,** 207–215 (2003).

81. Kim, M. W. *et al.* Characterization of N-linked oligosaccharides assembled on secretory recombinant glucose oxidase and cell wall mannoproteins from the methylotrophic yeast Hansenula polymorpha. *Glycobiology* **14,** 243–251 (2004).

82. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430,** 35–44 (2004).

83. Liu, B. *et al.* Disruption of the OCH1 and MNN1 genes decrease N-glycosylation on

glycoprotein expressed in Kluyveromyces lactis. *J. Biotechnol.* **143,** 95–102 (2009).

84. Song, Y. *et al.* Engineering of the yeast Yarrowia lipolytica for the production of glycoproteins lacking the outer-chain mannose residues of N-glycans. *Appl. Environ. Microbiol.* **73,** 4446–4454 (2007).

85. De Schutter, K. *et al.* Genome sequence of the recombinant protein production host Pichia pastoris. *Nat. Biotechnol.* **27,** 561–566 (2009).

86. Miele, R. G., Nilsen, S. L., Brito, T., Bretthauer, R. K. & Castellino, F. J. Glycosylationproperties of the Pichia pastoris-expressed recombinant kringle 2 domain oftissue-type plasminogen activator. *Biotechnol. Appl. Biochem.* **25,** 151–157 (1997).

87. Cregg, J. M., Barringer, K. J., Hessler, A. Y. & Madden, K. R. Pichia pastoris as a host system for transformations. *Mol. Cell. Biol.* **5,** 3376–85 (1985).

88. Cregg, J. M., Madden, K. R., Barringer, K. J., Thill, G. P. & Stillman, C. A. Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris. *Mol. Cell. Biol.* **9,** 1316–1323 (1989).

89. Couderc, R. & Baratti, J. Oxidation of Methanol by the Yeast, Pichia pastoris. Purification and Properties of the Alcohol Oxidase. *Agric. Biol. Chem.* **44,** 2279–2289 (1980).

90. Heyland, J., Fu, J., Blank, L. M. & Schmid, A. Quantitative physiology of Pichia pastoris during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnol. Bioeng.* **107,** 357–368 (2010).

91. Looser, V. *et al.* Cultivation strategies to enhance productivity of Pichia pastoris: A review. *Biotechnol. Adv.* **33,** 1177–1193 (2015).

92. Vieira, É. D., Andrietta, M. D. G. S. & Andrietta, S. R. Yeast Biomass Production: a New Approach in Glucose-limited Feeding Strategy. *Braz. J. Microbiol.* **44,** 551–8 (2013).

93. Li, E. & Mira De Orduña, R. A rapid method for the determination of microbial biomass by dry weight using a moisture analyser with an infrared heating source and an analytical balance. *Lett. Appl. Microbiol.* **50,** 283–288 (2010).

94. Dashko, S., Zhou, N., Compagno, C. & Piskur, J. Why, when, and how did yeast evolve alcoholic fermentation? *FEMS Yeast Res.* **14,** 826–832 (2014).

95. Nevalainen, K. M. H., Te'o, V. S. J. & Bergquist, P. L. Heterologous protein expression

in filamentous fungi. *Trends Biotechnol.* **23,** 468–474 (2005).

96. Werten, M. W., van den Bosch, T. J., Wind, R. D., Mooibroek, H. & de Wolf, F. A. High-yield secretion of recombinant gelatins by Pichia pastoris. *Yeast* **15,** 1087–1096 (1999).

97. Darby, R. A. J., Cartwright, S. P., Dilworth, M. V & Bill, R. M. Which yeast species shall I choose? Saccharomyces cerevisiae versus Pichia pastoris (review). *Methods Mol. Biol.* **866,** 11–23 (2012).

98. Beck, A., Cochet, O. & Wurch, T. GlycoFi's technology to control the glycosylation of recombinant therapeutic proteins. *Expert Opin. Drug Discov.* **5,** 95–111 (2010).

99. Jacobs, P. P., Geysens, S., Vervecken, W., Contreras, R. & Callewaert, N. Engineering complex-type N-glycosylation in Pichia pastoris using GlycoSwitch technology. *Nat. Protoc.* **4,** 58–70 (2009).

100. Wendland, J. Comparison of morphogenetic networks of filamentous fungi and yeast. *Fungal Genet. Biol.* **34,** 63–82 (2001).

101. Domnigues, F. C., Queiroz, J. a., Cabral, J. M. S. & Fonseca, L. P. The influence of culture conditions on mycelial structure and cellulose production by Trichoderma reeseiRut C-30. *Enzym. Microb. Technol* **26,** 394–401 (2000).

102. Schmoll, M. & Schuster, A. Biology and biotechnology of Trichoderma. *Appl. Microbiol. Biotechnol.* **87,** 787–799 (2010).

103. Gupta, V. K. *et al.* The Post-genomic Era of Trichoderma reesei: What's Next? *Trends Biotechnol.* **34,** 970–982 (2016).

104. Deane, E. E., Whipps, J. M., Lynch, J. M. & Peberdy, J. F. Transformation of Trichoderma reesei with a constitutively expressed heterologous fungal chitinase gene. *Enzyme Microb. Technol.* **24,** 419–424 (1999).

105. Nevalainen, H. & Peterson, R. Making recombinant proteins in filamentous fungi- Are we expecting too much? *Front. Microbiol.* **5,** 1–10 (2014).

106. Deshpande, N., Wilkins, M. R., Packer, N. & Nevalainen, H. Protein glycosylation pathways in filamentous fungi. *Glycobiology* **18,** 626–637 (2008).

107. Yao, T. & Asayama, Y. Animal-cell culture media: History, characteristics, and current issues. *Reprod. Med. Biol.* **16,** 99–117 (2017).

108. Fischer, S., Handrick, R. & Otte, K. The art of CHO cell engineering: A comprehensive

retrospect and future perspectives. *Biotechnol. Adv.* **33,** 1878–1896 (2015).

109. Rigano, M. M. & Walmsley, A. M. Expression systems and developments in plant-made vaccines. *Immunol. Cell Biol.* **83,** 271–277 (2005).

110. Weathers, P. J., Towler, M. J. & Xu, J. Bench to batch: Advances in plant cell culture for producing useful products. *Appl. Microbiol. Biotechnol.* **85,** 1339–1351 (2010).

111. Hellwig, S., Drossard, J., Twyman, R. M. & Fischer, R. Plant cell cultures for the production of recombinant proteins. *Nat. Biotechnol.* **22,** 1415–1422 (2004).

112. Fox, J. L. First plant-made biologic approved. *Nat Biotech* **30,** 472 (2012).

113. Strasser, R. Plant protein glycosylation. *Glycobiology* **26,** 926–939 (2016).

114. Kim, H. S., Jeon, J. H., Lee, K. J. & Ko, K. N-glycosylation modification of plant-derived virus-like particles: An application in vaccines. *Biomed Res. Int.* **2014,** (2014).

115. Galleno, M. & Sick, A. J. 12 - BACULOVIRUS EXPRESSION VECTOR SYSTEM BT - Gene Expression Systems. in 331–363 (Academic Press, 1999). doi:https://doi.org/10.1016/B978-012253840-7/50013-4

116. Hunt, I. From gene to protein: A review of new and enabling technologies for multi-parallel protein expression. *Protein Expr. Purif.* **40,** 1–22 (2005).

117. Gutternigg, M. *et al.* Biosynthesis of truncated N-linked oligosaccharides results from non-orthologous hexosaminidase-mediated mechanisms in nematodes, plants, and insects. *J. Biol. Chem.* **282,** 27825–27840 (2007).

118. Kost, T. A., Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat Biotech* **23,** 567–575 (2005).

119. Kirkpatrick, R. B. & Shatzman, A. 11 - Drosophila S2 SYSTEM FOR HETEROLOGOUS GENE EXPRESSION BT - Gene Expression Systems. in 289–330 (Academic Press, 1999). doi:https://doi.org/10.1016/B978-012253840-7/50012-2

120. Fernandez, J. M. & Hoeffler, J. P. *Gene expression systems : using nature for the art of expression*. (Academic Press, 1999).

121. Schiller, B., Hykollari, A., Yan, S., Paschinger, K. & Wilson, I. B. H. Complicated N-linked glycans in simple organisms. *Biol. Chem.* **393,** 661–673 (2012).

122. Rendic, D., Wilson, I. B. H. & Paschinger, K. The Glycosylation Capacity of Insect Cells.

Croat. Chem. Acta **81,** 7–21 (2008).

123. Noh, S. M., Sathyamurthy, M. & Lee, G. M. Development of recombinant Chinese hamster ovary cell lines for therapeutic protein production. *Curr. Opin. Chem. Eng.* **2,** 391–397 (2013).

124. De Jesus, M. & Wurm, F. M. Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors. *Eur. J. Pharm. Biopharm.* **78,** 184–188 (2011).

125. Kim, J. Y., Kim, Y. G. & Lee, G. M. CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Appl. Microbiol. Biotechnol.* **93,** 917–930 (2012).

126. Walsh, G. Biopharmaceutical benchmarks 2010. *Nat. Biotechnol.* **28,** 917–24 (2010).

127. Jayapal, Karthik P.; Wlaschin, Katie F.; Hu, Wei Shou; Yap, M. G. S. Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chem. Eng. Prog.* **103,** 40–47 (2007).

128. Zhu, J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol. Adv.* **30,** 1158–1170 (2012).

129. Kelley, B. Industrialization of mAb production technology: the bioprocessing industry at a crossroads. *MAbs* **1,** 443–452 (2009).

130. Liu, J. K. H. The history of monoclonal antibody development – Progress, remaining challenges and future innovations. *Ann. Med. Surg.* **3,** 113–116 (2014).

131. Boyiadzis, M. & Foon, K. A. Approved monoclonal antibodies for cancer therapy. *Expert Opin. Biol. Ther.* **8,** 1151–1158 (2008).

132. Aggarwal, S. What's fueling the biotech engine[mdash]2010 to 2011. *Nat Biotech* **29,** 1083–1089 (2011).

133. Jadhav, V. *et al.* CHO microRNA engineering is growing up: Recent successes and future challenges. *Biotechnol. Adv.* **31,** 1501–1513 (2013).

134. Dumont, J., Euwart, D., Mei, B., Estes, S. & Kshirsagar, R. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit. Rev. Biotechnol.* **36,** 1110–1122 (2016).

135. Maksimenko, O. G., Deykin, a V, Khodarovich, Y. M. & Georgiev, P. G. Use of transgenic animals in biotechnology: prospects and problems. *Acta Naturae* **5,** 33–46

(2013).

136. Wang, Y., Zhao, S., Bai, L., Fan, J. & Liu, E. Expression Systems and Species Used for Transgenic Animal Bioreactors. *Biomed Res. Int.* **2013,** 580463 (2013).

137. FDA approves second transgenic milk drug. *Nat Rev Drug Discov* **13,** 644 (2014).

138. Gellisen, G., Strasser, A. W. M. & Suckow, M. Key and criteria to the selection of an expression platform. *Prod. Recomb. Proteins Nov. Microb. Eukaryot. Syst.* 3–4 (2006). doi:10.1002/3527603670

139. Brondyk, W. H. Chapter 11 Selecting an Appropriate Method for Expressing a Recombinant Protein. in *Guide to Protein Purification, 2nd Edition* (eds. Burgess, R. R. & Deutscher, M. P.) **463,** 131–147 (Academic Press, 2009).

140. Hogwood, C. E. M., Bracewell, D. G. & Smales, C. M. Measurement and control of host cell proteins (HCPs) in CHO cell bioprocesses. *Curr. Opin. Biotechnol.* **30,** 153–160 (2014).

141. Chisti, Y. & Moo-Young, M. Disruption of microbial cells for intracellular products. *Enzyme Microb. Technol.* **8,** 194–204 (1986).

142. Gonçalves, A. M. *et al.* Pichia pastoris: A recombinant microfactory for antibodies and human membrane proteins. *J. Microbiol. Biotechnol.* **23,** 587–601 (2013).

143. Wang, X., Hunter, A. K. & Mozier, N. M. Host cell proteins in biologics development: Identification, quantitation and risk assessment. *Biotechnol. Bioeng.* **103,** 446–458 (2009).

144. Gutierrez, A. H., Moise, L. & De Groot, A. S. Of [Hamsters] and men: a new perspective on host cell proteins. *Hum. Vaccin. Immunother.* **8,** 1172–1174 (2012).

145. Fahnestock, S. R. & Bedzyk, L. A. Production of synthetic spider dragline silk protein in Pichia pastoris. *Appl. Microbiol. Biotechnol.* **47,** 33–39 (1997).

146. Chon, J. H. & Zarbis-Papastoitsis, G. Advances in the production and downstream processing of antibodies. *N. Biotechnol.* **28,** 458–463 (2011).

147. Frenzel, A., Hust, M. & Schirrmann, T. Expression of recombinant antibodies. *Front. Immunol.* **4,** 1–20 (2013).

148. Simmons, L. C. *et al.* Expression of full-length immunoglobulins in Escherichia coli: rapid and efficient production of aglycosylated antibodies. *J. Immunol. Methods* **263,**

133–147 (2002).

149. Chen, M.-T. *et al.* Generation of diploid Pichia pastoris strains by mating and their application for recombinant protein production. *Microb. Cell Fact.* **11,** 91 (2012).

150. Johansson, D. X. *et al.* Efficient expression of recombinant human monoclonal antibodies in Drosophila S2 cells. *J. Immunol. Methods* **318,** 37–46 (2007).

151. Palmberger, D. *et al.* Insect cells for antibody production: evaluation of an efficient alternative. *J. Biotechnol.* **153,** 160–166 (2011).

152. Sharp, J. M. & Doran, P. M. Strategies for enhancing monoclonal antibody accumulation in plant cell and organ cultures. *Biotechnol. Prog.* **17,** 979–992 (2001).

153. van Berkel, P. H. C. *et al.* Rapid production of recombinant human IgG With improved ADCC effector function in a transient expression system. *Biotechnol. Bioeng.* **105,** 350–357 (2010).

154. Spens, E. & Haggstrom, L. Defined protein and animal component-free NS0 fed-batch culture. *Biotechnol. Bioeng.* **98,** 1183–1194 (2007).

155. del Val, I. J. *et al.* Application of Quality by Design Paradigm to the Manufacture of Protein Therapeutics. *Glycosylation* 353–396 (2012). doi:10.5772/50261

156. Jahic, M., Gustavsson, M., Jansen, A.-K., Martinelle, M. & Enfors, S.-O. Analysis and control of proteolysis of a fusion protein in Pichia pastoris fed-batch processes. *J. Biotechnol.* **102,** 45–53 (2003).

157. Kunert, R., Gach, J. & Katinger, H. Expression of a Fab Fragment in CHO and Pichia pastoris: A comparative case study. *Bioprocess Int.* 34–40 (2008).

158. Wildt, S. & Gerngross, T. U. The humanization of N-glycosylation pathways in yeast. *Nat. Rev. Microbiol.* **3,** 119–128 (2005).

159. Ahrens, P. B. & Ankel, H. Natural killer cells discriminate between high mannose- and complex-type asparagine-linked oligosaccharides. *Biochimie* **70,** 1619–1625 (1988).

160. Muchmore, V., Sathyamoorthy, N., Decker, J. & Sherblom, A. Evidence That Specific High-Mannose Can Directly Inhibit Antigen-Driven Oligosaccharides T-Cell Responses. *J. Leukoc. Biol.* **464,** 457–464 (1990).

161. Taylor, M. E. *Introduction to glycobiology*. (Oxford University Press, 2011).

162. Zielinska, D. F., Gnad, F., Wiśniewski, J. R. & Mann, M. Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141,** 897–907 (2010).

163. Coutinho Pedro M., Stam Mark, B. E. and H. B. Why are there so many carbohydrate-active enzyme-related genes in plants? *Trends Plant Sci.* **8,** 561–563 (2003).

164. Aebi, M., Bernasconi, R., Clerc, S. & Molinari, M. N-glycan structures: recognition and processing in the ER. *Trends Biochem. Sci.* **35,** 74–82 (2010).

165. Freeze, H. H. & Ng, B. G. Golgi Glycosylation and human inherited diseases. *Cold Spring Harb. Perspect. Biol.* **3,** 1–22 (2011).

166. Freeze, H. H., Eklund, E. A., Ng, B. G. & Patterson, M. C. Neurology of inherited glycosylation disorders. *Lancet Neurol.* **11,** 453–466 (2012).

167. Lehle, L., Strahl, S. & Tanner, W. Protein glycosylation, conserved from yeast to man: A model organism helps elucidate congenital human diseases. *Angew. Chemie - Int. Ed.* **45,** 6802–6818 (2006).

168. Lowe, J. B. & Marth, J. D. A genetic approach to Mammalian glycan function. *Annu. Rev. Biochem.* **72,** 643–691 (2003).

169. Matsumiya, S. *et al.* Structural comparison of fucosylated and nonfucosylated Fc fragments of human immunoglobulin G1. *J. Mol. Biol.* **368,** 767–79 (2007).

170. Aebi, M. N-linked protein glycosylation in the ER. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 2430–2437 (2013).

171. Burda, P. & Aebi, M. The dolichol pathway of N-linked glycosylation. *Biochim. Biophys. Acta - Gen. Subj.* **1426,** 239–257 (1999).

172. Welti, M. Regulation of dolichol-linked glycosylation. *Glycoconj. J.* **30,** 51–56 (2013).

173. Jones, M. B., Rosenberg, J. N., Betenbaugh, M. J. & Krag, S. S. Structure and synthesis of polyisoprenoids used in N-glycosylation across the three domains of life. *Biochim. Biophys. Acta* **1790,** 485–494 (2009).

174. Schenk, B., Fernandez, F. & Waechter, C. J. The ins(ide) and out(side) of dolichyl phosphate biosynthesis and recycling in the endoplasmic reticulum. *Glycobiology* **11,** 61R–70R (2001).

175. Shridas, P. & Waechter, C. J. Human dolichol kinase, a polytopic endoplasmic reticulum

membrane protein with a cytoplasmically oriented CTP-binding site. *J. Biol. Chem.* **281,** 31696–31704 (2006).

176. Jones, J., Krag, S. S. & Betenbaugh, M. J. Controlling N-linked glycan site occupancy. *Biochim. Biophys. Acta - Gen. Subj.* **1726,** 121–137 (2005).

177. Nakayama, K., Nagasu, T., Shimma, Y., Kuromitsu, J. & Jigami, Y. OCH1 encodes a novel membrane bound mannosyltransferase: outer chain elongation of asparagine-linked oligosaccharides. *EMBO J.* **11,** 2511–2519 (1992).

178. Breitling, J. & Aebi, M. N-linked protein glycosylation in the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.* **5,** 1–15 (2013).

179. Huffaker, T. C. & Robbins, P. W. Yeast mutants deficient in protein glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **80,** 7466–7470 (1983).

180. Rush, J. S. Role of Flippases in Protein Glycosylation in the Endoplasmic Reticulum Supplementary Issue: Cellular Anatomy of Lipid Traffic. *Lipid Insights* **88,** 45–53 (2015).

181. Stanley, P. Golgi glycosylation. *Cold Spring Harb. Perspect. Biol.* **3,** 1–13 (2011).

182. Mohorko, E., Glockshuber, R. & Aebi, M. Oligosaccharyltransferase: The central enzyme of N-linked protein glycosylation. *J. Inherit. Metab. Dis.* **34,** 869–878 (2011).

183. Burda, P., Jakob, C. A., Beinhauer, J., Hegemann, J. H. & Aebi, M. Ordered assembly of the asymmetrically branched lipid-linked oligosaccharide in the endoplasmic reticulum is ensured by the substrate specificity of the individual glycosyltransferases. *Glycobiology* **9,** 617–625 (1999).

184. Harada, Y., Li, H., Li, H. & Lennarz, W. J. Oligosaccharyltransferase directly binds to ribosome at a location near the translocon-binding site. *Proc. Natl. Acad. Sci.* **106,** 6945–6949 (2009).

185. Bause, E. Structural requirements of N-glycosylation of proteins. Studies with proline peptides as conformational probes. *Biochem. J.* **209,** 331–336 (1983).

186. Ho, B. K., Coutsias, E. A., Seok, C. & Dill, K. A. The flexibility in the proline ring couples to the protein backbone. *Protein Sci.* **14,** 1011–1018 (2005).

187. Spirig, U. *et al.* The STT3 protein is a component of the yeast oligosaccharyltransferase complex. *Mol. Gen. Genet.* **256,** 628–637 (1997).

188. Kelleher, D. J., Karaoglu, D., Mandon, E. C. & Gilmore, R. Oligosaccharyltransferase

isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell* **12,** 101–111 (2003).

189. Ito, Y., Takeda, Y., Seko, A., Izumi, M. & Kajihara, Y. Functional analysis of endoplasmic reticulum glucosyltransferase (UGGT): Synthetic chemistry's initiative in glycobiology. *Semin. Cell Dev. Biol.* **41,** 90–98 (2015).

190. Satoh, T., Toshimori, T., Yan, G., Yamaguchi, T. & Kato, K. Structural basis for two-step glucose trimming by glucosidase II involved in ER glycoprotein quality control. *Sci. Rep.* **6,** 20575 (2016).

191. Rutkevich, L. A. & Williams, D. B. Participation of lectin chaperones and thiol oxidoreductases in protein folding within the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **23,** 157–166 (2011).

192. Cherepanova, N., Shrimal, S. & Gilmore, R. N-linked glycosylation and homeostasis of the endoplasmic reticulum. *Curr. Opin. Cell Biol.* **41,** 57–65 (2016).

193. D'Alessio, C., Caramelo, J. J. & Parodi, A. J. UDP-GLC:GLYCOPROTEIN GLUCOSYLTRANSFERASE-GLUCOSIDASE II, THE YING-YANG OF THE ER QUALITY CONTROL. *Semin. Cell Dev. Biol.* **21,** 491–499 (2010).

194. Ito, Y. & Takeda, Y. Deciphering the Roles of Glycan Processing in Glycoprotein Quality Control through Organic Synthesis. *Biosci. Biotechnol. Biochem.* **77,** 2331–2338 (2013).

195. Caramelo, J. J. & Parodi, A. J. A sweet code for glycoprotein folding. *FEBS Lett.* **589,** 3379–3387 (2015).

196. Tannous, A., Patel, N., Tamura, T. & Hebert, D. N. Reglucosylation by UDP-glucose:glycoprotein glucosyltransferase 1 delays glycoprotein secretion but not degradation. *Mol. Biol. Cell* **26,** 390–405 (2015).

197. Tannous, A., Pisoni, G. B., Hebert, D. N. & Molinari, M. N-linked sugar-regulated protein folding and quality control in the ER. *Semin. Cell Dev. Biol.* **41,** 79–89 (2015).

198. Benyair, R., Ogen-Shtern, N. & Lederkremer, G. Z. Glycan regulation of ER-associated degradation through compartmentalization. *Semin. Cell Dev. Biol.* **41,** 99–109 (2015).

199. Roth, J. *et al.* Protein N-glycosylation, protein folding, and protein quality control. *Mol. Cells* **30,** 497–506 (2010).

200. Mast, S. W. & Moremen, K. W. Family 47 α-Mannosidases in N-Glycan Processing. *Methods Enzymol.* **415,** 31–46 (2006).

201. Tempel, W. *et al.* Structure of mouse Golgi α-mannosidase IA reveals the molecular basis for substrate specificity among Class 1 (family 47 glycosylhydrolase) α1,2-mannosidases. *J. Biol. Chem.* **279,** 29774–29786 (2004).

202. Lal, A. *et al.* Substrate specificities of recombinant murine Golgi alpha1, 2-mannosidases IA and IB and comparison with endoplasmic reticulum and Golgi processing alpha1,2-mannosidases. *Glycobiology* **8,** 981–995 (1998).

203. Geva, Y. & Schuldiner, M. The back and forth of cargo exit from the endoplasmic reticulum. *Curr. Biol.* **24,** R130–R136 (2014).

204. D'Arcangelo, J. G., Stahmer, K. R. & Miller, E. A. Vesicle-mediated export from the ER: COPII coat function and regulation. *Biochim. Biophys. Acta - Mol. Cell Res.* **1833,** 2464–2472 (2013).

205. De Matteis, M. A. & Rega, L. R. Endoplasmic reticulum-Golgi complex membrane contact sites. *Curr. Opin. Cell Biol.* **35,** 43–50 (2015).

206. Hauri, H. P., Kappeler, F., Andersson, H. & Appenzeller, C. ERGIC-53 and traffic in the secretory pathway. *J. Cell Sci.* **113 (Pt 4,** 587–596 (2000).

207. Brostrom, M. A. & Brostrom, C. O. Calcium dynamics and endoplasmic reticular function in the regulation of protein synthesis: implications for cell growth and adaptability. *Cell Calcium* **34,** 345–363 (2003).

208. Tremblay, L. O. & Herscovics, A. Characterization of a cDNA encoding a novel human Golgi alpha 1, 2-mannosidase (IC) involved in N-glycan biosynthesis. *J. Biol. Chem.* **275,** 31655–31660 (2000).

209. Romero, P. A., Vallée, F., Howell, P. L. & Herscovics, A. Mutation of Arg273 to Leu Alters the Specificity of the Yeast N-Glycan Processing Class I α1,2-Mannosidase . *J. Biol. Chem.* **275,** 11071–11074 (2000).

210. Schachter, H. The 'yellow brick road' to branched complex N-glycans. *Glycobiology* **1,** 453–461 (1991).

211. Schachter, H. Complex N-glycans: the story of the 'yellow brick road'. *Glycoconj. J.* **31,** 1–5 (2014).

212. Rose, D. R. Structure, mechanism and inhibition of Golgi α-mannosidase II. *Curr. Opin. Struct. Biol.* **22,** 558–62 (2012).

213. Tan, J. *et al.* The human UDP-N-acetylglucosamine: alpha-6-D-mannoside-beta-1,2-N-acetylglucosaminyltransferase II gene (MGAT2). Cloning of genomic DNA, localization to chromosome 14q21, expression in insect cells and purification of the recombinant protein. *Eur. J. Biochem.* **231,** 317–328 (1995).

214. Schachter, H. The joys of HexNAc. The synthesis and function of N- and O-glycan branches. *Glycoconj. J.* **17,** 465–483 (2000).

215. Rouwendal, G. J. A. *et al.* Efficient introduction of a bisecting GlcNAc residue in tobacco N-glycans by expression of the gene encoding human N-acetylglucosaminyltransferase III. *Glycobiology* **17,** 334–344 (2007).

216. Mariño, K., Bones, J., Kattla, J. J. & Rudd, P. M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat. Chem. Biol.* **6,** 713–723 (2010).

217. Vasconcelos-dos-Santos, A. *et al.* Biosynthetic Machinery Involved in Aberrant Glycosylation: Promising Targets for Developing of Drugs Against Cancer. *Front. Oncol.* **5,** 138 (2015).

218. Sarkar, M. *et al.* Null mutations in Drosophila N-acetylglucosaminyltransferase I produce defects in locomotion and a reduced life span. *J. Biol. Chem.* **281,** 12776–12785 (2006).

219. Altmann, F., Schwihla, H., Staudacher, E., Glossl, J. & Marz, L. Insect cells contain an unusual, membrane-bound beta-N-acetylglucosaminidase probably involved in the processing of protein N-glycans. *J. Biol. Chem.* **270,** 17344–17349 (1995).

220. Fabre, E., Hurtaux, T. & Fradin, C. Mannosylation of fungal glycoconjugates in the Golgi apparatus. *Curr. Opin. Microbiol.* **20,** 103–110 (2014).

221. Munro, S. What can yeast tell us about N-linked glycosylation in the Golgi apparatus? *FEBS Lett.* **498,** 223–227 (2001).

222. Szymanski, C. M., Yao, R., Ewing, C. P., Trust, T. J. & Guerry, P. Evidence for a system of general protein glycosylation in Campylobacter jejuni. *Mol. Microbiol.* **32,** 1022–30 (1999).

223. Young, N. M. *et al.* Structure of the N-linked glycan present on multiple glycoproteins in the Gram-negative bacterium, Campylobacter jejuni. *J. Biol. Chem.* **277,** 42530–9

(2002).

224. Li, L. *et al.* Overexpression and topology of bacterial oligosaccharyltransferase PglB. *Biochem. Biophys. Res. Commun.* **394,** 1069–1074 (2010).

225. Mohawk, K. L., Poly, F., Sahl, J. W., Rasko, D. A. & Guerry, P. High frequency, spontaneous motA mutations in Campylobacter jejuni strain 81-176. *PLoS One* **9,** 1–8 (2014).

226. Yan, Q. & Lennarz, W. J. Studies on the function of oligosaccharyl transferase subunits: Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* **277,** 47692–47700 (2002).

227. Kelly, J. *et al.* Biosynthesis of the N-linked glycan in Campylobacter jejuni and addition onto protein through block transfer. *J. Bacteriol.* **188,** 2427–2434 (2006).

228. Schoenhofen, I. C. *et al.* Functional characterization of dehydratase/aminotransferase pairs from Helicobacter and Campylobacter: enzymes distinguishing the pseudaminic acid and bacillosamine biosynthetic pathways. *J. Biol. Chem.* **281,** 723–732 (2006).

229. Vijayakumar, S. *et al.* Cj1121c, a novel UDP-4-keto-6-deoxy-GlcNAc C-4 aminotransferase essential for protein glycosylation and virulence in Campylobacter jejuni. *J. Biol. Chem.* **281,** 27733–27743 (2006).

230. Olivier, N. B., Chen, M. M., Behr, J. R. & Imperiali, B. In vitro biosynthesis of UDP-N,N'-diacetylbacillosamine by enzymes of the Campylobacter jejuni general protein glycosylation system. *Biochemistry* **45,** 13659–13669 (2006).

231. Tytgat, H. L. P. & Lebeer, S. The sweet tooth of bacteria: common themes in bacterial glycoconjugates. *Microbiol. Mol. Biol. Rev.* **78,** 372–417 (2014).

232. Glover, K. J., Weerapana, E., Chen, M. M. & Imperiali, B. Direct biochemical evidence for the utilization of UDP-bacillosamine by PglC, an essential glycosyl-1-phosphate transferase in the Campylobacter jejuni N-linked glycosylation pathway. *Biochemistry* **45,** 5343–5350 (2006).

233. Troutman, J. M. & Imperiali, B. Campylobacter jejuni PglH is a single active site processive polymerase that utilizes product inhibition to limit sequential glycosyl transfer reactions. *Biochemistry* **48,** 2807–2816 (2009).

234. Glover, K. J., Weerapana, E. & Imperiali, B. In vitro assembly of the undecaprenylpyrophosphate-linked heptasaccharide for prokaryotic N-linked

glycosylation. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14255–14259 (2005).

235. Whitfield, C. & Paiment, A. Biosynthesis and assembly of Group 1 capsular polysaccharides in Escherichia coli and related extracellular polysaccharides in other bacteria. *Carbohydr. Res.* **338,** 2491–2502 (2003).

236. Kowarik, M. *et al.* N-Linked Glycosylation of Folded. *Science (80-.).* **314,** 1148–1150 (2006).

237. Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J. & Kohda, D. Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases. *J. Biol. Chem.* **285,** 4941–4950 (2010).

238. Kowarik, M. *et al.* Definition of the bacterial N-glycosylation site consensus sequence. *EMBO J.* **25,** 1957–66 (2006).

239. Chen, M. M., Glover, K. J. & Imperiali, B. From peptide to protein: comparative analysis of the substrate specificity of N-linked glycosylation in C. jejuni. *Biochemistry* **46,** 5579–5585 (2007).

240. Ielmini, M. V. & Feldman, M. F. Desulfovibrio desulfuricans PglB homolog possesses oligosaccharyltransferase activity with relaxed glycan specificity and distinct protein acceptor sequence requirements. *Glycobiology* **21,** 734–742 (2011).

241. Scott, N. E. *et al.* Simultaneous glycan-peptide characterization using hydrophilic interaction chromatography and parallel fragmentation by CID, higher energy collisional dissociation, and electron transfer dissociation MS applied to the N-linked glycoproteome of Campylobact. *Mol. Cell. Proteomics* **10,** M000031–MCP201 (2011).

242. Hattrup, C. L. & Gendler, S. J. Structure and function of the cell surface (tethered) mucins. *Annu. Rev. Physiol.* **70,** 431–457 (2008).

243. Raju, T. S. Terminal sugars of Fc glycans influence antibody effector functions of IgGs. *Curr. Opin. Immunol.* **20,** 471–478 (2008).

244. Jefferis, R. Posttranslational modifications and the immunogenicity of biotherapeutics. *J. Immunol. Res.* **2016,** (2016).

245. Wright, A. & Morrison, S. L. Effect of glycosylation on antibody function: implications for genetic engineering. *Trends Biotechnol.* **15,** 26–32 (1997).

246. Rudd, P. M. & Dwek, R. A. Glycosylation : Heterogeneity and the 3D Structure of

Proteins. *Crit. Rev. Biochem. Mol. Biol.* **32,** 1–100 (1997).

247. Fan, Y. *et al.* Amino acid and glucose metabolism in fed-batch CHO cell culture affects antibody production and glycosylation. *Biotechnol. Bioeng.* **112,** 521–35 (2015).

248. Wacker, C., Berger, C. N., Girard, P. & Meier, R. Glycosylation profiles of therapeutic antibody pharmaceuticals. *Eur. J. Pharm. Biopharm.* **79,** 503–7 (2011).

249. Crispin, M. *et al.* Disruption of alpha-mannosidase processing induces non-canonical hybrid-type glycosylation. *FEBS Lett.* **581,** 1963–8 (2007).

250. Crispin, M. *et al.* Inhibition of hybrid- and complex-type glycosylation reveals the presence of the GlcNAc transferase I-independent fucosylation pathway. *Glycobiology* **16,** 748–756 (2006).

251. Okeley, N. M. *et al.* Development of orally active inhibitors of protein and cellular fucosylation. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 5404–5409 (2013).

252. Goss, P. E., Reid, C. L., Bailey, D. & Dennis, J. W. Phase IB clinical trial of the oligosaccharide processing inhibitor swainsonine in patients with advanced malignancies. *Clin. Cancer Res.* **3,** 1077–1086 (1997).

253. Ghaderi, D., Zhang, M., Hurtado-Ziola, N. & Varki, A. Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnol. Genet. Eng. Rev.* **28,** 147–175 (2012).

254. El Mai, N., Donadio-Andrei, S., Iss, C., Calabro, V. & Ronin, C. Engineering a human-like glycosylation to produce therapeutic glycoproteins based on 6-linked sialylation in CHO cells. *Methods Mol. Biol.* **988,** 19–29 (2013).

255. Lee, E. U., Roth, J. & Paulson, J. C. Alteration of terminal glycosylation sequences on N-linked oligosaccharides of Chinese hamster ovary cells by expression of beta-galactoside alpha 2,6-sialyltransferase. *J. Biol. Chem.* **264,** 13848–13855 (1989).

256. Zhang, X., Lok, S. H. L. & Kon, O. L. Stable expression of human α-2,6-sialyltransferase in Chinese hamster ovary cells: functional consequences for human erythropoietin expression and bioactivity. *Biochim. Biophys. Acta - Gen. Subj.* **1425,** 441–452 (1998).

257. Bragonzi, A. *et al.* A new Chinese hamster ovary cell line expressing α2,6-sialyltransferase used as universal host for the production of human-like sialylated recombinant glycoproteins. *Biochim. Biophys. Acta - Gen. Subj.* **1474,** 273–282 (2000).

258. Minch, S. L., Kallio, P. T. & Bailey, J. E. Tissue Plasminogen Activator Coexpressed in Chinese Hamster Ovary Cells with α(2,6)-Sialyltransferase Contains NeuAcα(2,6)Galβ(1,4)Gl-N-AcR Linkages. *Biotechnol. Prog.* **11,** 348–351 (1995).

259. Fukuta, K., Yokomatsu, T., Abe, R., Asanagi, M. & Makino, T. Genetic engineering of CHO cells producing human interferon-γ by transfection of sialyltransferases. *Glycoconj. J.* **17,** 895–904 (2000).

260. Jassal, R. *et al.* Sialylation of Human IgG-Fc Carbohydrate by Transfected Rat α2,6-Sialyltransferase. *Biochem. Biophys. Res. Commun.* **286,** 243–249 (2001).

261. Fukuta, K. *et al.* Remodeling of sugar chain structures of human interferon-γ. *Glycobiology* **10,** 421–430 (2000).

262. Weikert, S. *et al.* Engineering Chinese hamster ovary cells to maximize sialic acid content of recombinant glycoproteins. *Nat. Biotechnol.* **17,** 1116–1121 (1999).

263. Jeong, Y. T., Choi, O., Son, Y. D., Park, S. Y. & Kim, J. H. Enhanced sialylation of recombinant erythropoietin in genetically engineered Chinese-hamster ovary cells. *Biotechnol. Appl. Biochem.* **52,** 283–291 (2009).

264. Son, Y.-D., Jeong, Y. T., Park, S.-Y. & Kim, J. H. Enhanced sialylation of recombinant human erythropoietin in Chinese hamster ovary cells by combinatorial engineering of selected genes. *Glycobiology* **21,** 1019–1028 (2011).

265. Davies, J. *et al.* Expression of GnTIII in a recombinant anti-CD20 CHO production cell line: Expression of antibodies with altered glycoforms leads to an increase in ADCC through higher affinity for FC gamma RIII. *Biotechnol. Bioeng.* **74,** 288–294 (2001).

266. Ferrara, C. *et al.* Modulation of therapeutic antibody effector functions by glycosylation engineering: Influence of Golgi enzyme localization domain and co-expression of heterologous β1, 4-N-acetylglucosaminyltransferase III and Golgi α-mannosidase II. *Biotechnol. Bioeng.* **93,** 851–861 (2006).

267. Jeong, Y. T. *et al.* Enhanced sialylation of recombinant erythropoietin in CHO cells by human glycosyltransferase expression. *J. Microbiol. Biotechnol.* **18,** 1945–1952 (2008).

268. Wong, N. S. C., Yap, M. G. S. & Wang, D. I. C. Enhancing recombinant glycoprotein sialylation through CMP-sialic acid transporter over expression in Chinese hamster ovary cells. *Biotechnol. Bioeng.* **93,** 1005–1016 (2006).

269. von Horsten, H. H. *et al.* Production of non-fucosylated antibodies by co-expression of

heterologous GDP-6-deoxy-d-lyxo-4-hexulose reductase. *Glycobiology* **20,** 1607–1618 (2010).

270. Yamane-Ohnuki, N. & Satoh, M. Production of therapeutic antibodies with controlled fucosylation. *MAbs* **1,** 230–6 (2009).

271. Malphettes, L. *et al.* Highly efficient deletion of FUT8 in CHO cell lines using zinc-finger nucleases yields cells that produce completely nonfucosylated antibodies. *Biotechnol. Bioeng.* **106,** 774–783 (2010).

272. Cristea, S. *et al.* In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnol. Bioeng.* **110,** 871–880 (2013).

273. Yang, Z. *et al.* Engineered CHO cells for production of diverse, homogeneous glycoproteins. *Nat Biotech* **33,** 842–844 (2015).

274. Zhang, P. *et al.* Identification of functional elements of the GDP-fucose transporter SLC35C1 using a novel Chinese hamster ovary mutant. *Glycobiology* **22,** 897–911 (2012).

275. Zhong, X. *et al.* Engineering novel Lec1 glycosylation mutants in CHO–DUKX cells: Molecular insights and effector modulation of N-acetylglucosaminyltransferase I. *Biotechnol. Bioeng.* **109,** 1723–1734 (2012).

276. Sealover, N. R. *et al.* Engineering Chinese Hamster Ovary (CHO) cells for producing recombinant proteins with simple glycoforms by zinc-finger nuclease (ZFN)—mediated gene knockout of mannosyl (alpha-1,3-)-glycoprotein beta-1,2-N-acetylglucosaminyltransferase (Mgat1). *J. Biotechnol.* **167,** 24–32 (2013).

277. Ferrari, J., Gunson, J., Lofgren, J., Krummen, L. & Warner, T. G. Chinese hamster ovary cells with constitutively expressed sialidase antisense RNA produce recombinant dnase in batch culture with increased sialic acid. *Biotechnol. Bioeng.* **60,** 589–595 (1998).

278. Ngantung, F. A., Miller, P. G., Brushett, F. R., Tang, G. L. & Wang, D. I. C. RNA interference of sialidase improves glycoprotein sialic acid content consistency. *Biotechnol. Bioeng.* **95,** 106–119 (2006).

279. Zhang, M., Koskie, K., Ross, J. S., Kayser, K. J. & Caple, M. V. Enhancing glycoprotein sialylation by targeted gene silencing in mammalian cells. *Biotechnol. Bioeng.* **105,** 1094–1105 (2010).

280. Mori, K. *et al.* Engineering Chinese hamster ovary cells to maximize effector function of

produced antibodies using FUT8 siRNA. *Biotechnol. Bioeng.* **88,** 901–908 (2004).

281. Beuger, V. *et al.* Short-hairpin-RNA-mediated silencing of fucosyltransferase 8 in Chinese-hamster ovary cells for the production of antibodies with enhanced antibody immune effector function. *Biotechnol. Appl. Biochem.* **53,** 31–37 (2009).

282. Imai-Nishiya, H. *et al.* Double knockdown of α1,6-fucosyltransferase (FUT8) and GDP-mannose 4,6-dehydratase (GMD) in antibody-producing cells: a new strategy for generating fully non-fucosylated therapeutic antibodies with enhanced ADCC. *BMC Biotechnol.* **7,** 84 (2007).

283. Kanda, Y. *et al.* Comparison of biological activity among nonfucosylated therapeutic IgG1 antibodies with three different N-linked Fc oligosaccharides: The high-mannose, hybrid, and complex types. *Glycobiology* **17,** 104–118 (2007).

284. Omasa, T. *et al.* Overexpression of GADD34 Enhances Production of Recombinant Human Antithrombin III in Chinese Hamster Ovary Cells. *J. Biosci. Bioeng.* **106,** 568–573 (2008).

285. Chiba, Y., Suzuki, M., Yoshida, S., Yoshida, A. & Ikenaga, H. Production of Human Compatible High Mannose-type (Man 5 GlcNAc 2) Sugar Chains in Saccharomyces cerevisiae *. *October* **273,** 26298–26304 (1998).

286. Okada, H., Ohnuki, S., Roncero, C., Konopka, J. B. & Ohya, Y. Distinct roles of cell wall biogenesis in yeast morphogenesis as revealed by multivariate analysis of high-dimensional morphometric data. *Mol. Biol. Cell* **25,** 222–233 (2014).

287. Moo, W. K. *et al.* Functional characterization of the Hansenula polymorpha HOC1, OCH1, and OCR1 genes as members of the yeast OCH1 mannosyltransferase family involved in protein glycosylation. *J. Biol. Chem.* **281,** 6261–6272 (2006).

288. Callewaert, N. *et al.* Use of HDEL-tagged Trichoderma reesei mannosyl oligosaccharide 1,2-α- D -mannosidase for N -glycan engineering in Pichia pastoris. *FEBS Lett.* **503,** 173–178 (2001).

289. Qin, X. *et al.* GAP promoter library for fine-tuning of gene expression in Pichia pastoris. *Appl. Environ. Microbiol.* **77,** 3600–3608 (2011).

290. Wu, M. M. *et al.* Organelle pH studies using targeted avidin and fluorescein-biotin. *Chem. Biol.* **7,** 197–209 (2000).

291. Paroutis, P., Touret, N. & Grinstein, S. The pH of the secretory pathway: measurement,

determinants, and regulation. *Physiology (Bethesda).* **19,** 207–15 (2004).

292. Maras, M. *et al.* Molecular cloning and enzymatic characterization of a Trichoderma reesei 1,2-α-d-mannosidase. *J. Biotechnol.* **77,** 255–263 (2000).

293. Choi, B.-K. *et al.* Use of combinatorial genetic libraries to humanize N-linked glycosylation in the yeast Pichia pastoris. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 5022–5027 (2003).

294. Nett, J. H. *et al.* A combinatorial genetic library approach to target heterologous glycosylation enzymes to the endoplasmic reticulum or the Golgi apparatus of Pichia pastoris. *Yeast* **28,** 237–252 (2011).

295. Vervecken, W. *et al.* In vivo synthesis of mammalian-like, hybrid-type N-glycans in Pichia pastoris. *Appl. Environ. Microbiol.* **70,** 2639–2646 (2004).

296. Bobrowicz, P. *et al.* Engineering of an artificial glycosylation pathway blocked in core oligosaccharide assembly in the yeast Pichia pastoris: Production of complex humanized glycoproteins with terminal galactose. *Glycobiology* **14,** 757–766 (2004).

297. Chen, X. *et al.* Changing the donor cofactor of bovine alpha 1, 3-galactosyltransferase by fusion with UDP-galactose 4-epimerase. More efficient biocatalysis for synthesis of alpha-Gal epitopes. *J. Biol. Chem.* **275,** 31594–31600 (2000).

298. Hamilton, S. R. & Gerngross, T. U. Glycosylation engineering in yeast: the advent of fully humanized yeast. *Curr. Opin. Biotechnol.* **18,** 387–92 (2007).

299. Hamilton, S. R. *et al.* Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science* **313,** 1441–1443 (2006).

300. Potgieter, T. I. *et al.* Production of monoclonal antibodies by glycoengineered Pichia pastoris. *J. Biotechnol.* **139,** 318–325 (2009).

301. Barnard, G. C. *et al.* High-throughput screening and selection of yeast cell lines expressing monoclonal antibodies. *J. Ind. Microbiol. Biotechnol.* **37,** 961–971 (2010).

302. Berdichevsky, M., d'Anjou, M., Mallem, M. R., Shaikh, S. S. & Potgieter, T. I. Improved production of monoclonal antibodies through oxygen-limited cultivation of glycoengineered yeast. *J. Biotechnol.* **155,** 217–224 (2011).

303. Morell, A. G., Gregoriadis, G., Scheinberg, H., Hickman, J. & Ashwell, G. The role of sialic acid in determining the survival of glycoproteins in the circulation. *J. Biol. Chem.*

246, 1461–1467 (1971).

304. Schwartz, A. L. The hepatic asialoglycoprotein receptor. *CRC Crit. Rev. Biochem.* **16,** 207–233 (1984).

305. Tomiya, N., Narang, S., Lee, Y. C. & Betenbaugh, M. J. Comparing N-glycan processing in mammalian cell lines to native and engineered lepidopteran insect cell lines. *Glycoconj. J.* **21,** 343–360 (2004).

306. Shi, X. & Jarvis, D. L. Protein N-glycosylation in the baculovirus-insect cell system. *Curr. Drug Targets* **8,** 1116–25 (2007).

307. Marchal, I., Jarvis, D. L., Cacan, R. & Verbert, A. Glycoproteins from insect cells: Sialylated or not? *Biol. Chem.* **382,** 151–159 (2001).

308. Koles, K. *et al.* Identification of N-glycosylated proteins from the central nervous system of Drosophila melanogaster. *Glycobiology* **17,** 1388–1403 (2007).

309. Kajiura, H., Hamaguchi, Y., Mizushima, H., Misaki, R. & Fujiyama, K. Sialylation potentials of the silkworm, Bombyx mori; B. mori possesses an active ??2,6-sialyltransferase. *Glycobiology* **25,** 1441–1453 (2015).

310. Palmberger, D., Wilson, I. B. H., Berger, I., Grabherr, R. & Rendic, D. Sweetbac: A new approach for the production of mammalianised glycoproteins in insect cells. *PLoS One* **7,** 1–8 (2012).

311. Wagner, R. *et al.* Elongation of the N-glycans of fowl plague virus hemagglutinin expressed in Spodoptera frugiperda (Sf9) cells by coexpression of human ??1,2-N-acetylglucosaminyltransferase I. *Glycobiology* **6,** 165–175 (1996).

312. Hollister, J. R., Shaper, J. H. & Jarvis, D. L. Stable expression of mammalian beta 1,4-galactosyltransferase extends the N-glycosylation pathway in insect cells. *Glycobiology* **8,** 473–480 (1998).

313. Hollister, J. R. & Jarvis, D. L. Engineering lepidopteran insect cells for sialoglycoprotein production by genetic transformation with mammalian beta 1,4-galactosyltransferase and alpha 2,6-sialyltransferase genes. *Glycobiology* **11,** 1–9 (2001).

314. Yun, E. Y. *et al.* Galatosylation and sialylation of mammalian glycoproteins produced by baculovirus-madiated gene expression in insect cells. *Biotechnol. Lett.* **27,** 1035–1039 (2005).

315. Mabashi-Asazuma, H., Kuo, C. W., Khoo, K. H. & Jarvis, D. L. A novel baculovirus vector for the production of nonfucosylated recombinant glycoproteins in insect cells. *Glycobiology* **24,** 325–340 (2014).

316. Altmann, F., Kornfeld, G., Dalik, T., Staudacher, E. & Gl??ssl, J. Processing of asparagine-linked oligosaccharides in insect cells. N-acetylglucosaminyltransferase I and II activities in cultured lepidopteran cells. *Glycobiology* **3,** 619–625 (1993).

317. Hollister, J., Grabenhorst, E., Nimtz, M., Conradt, H. & Jarvis, D. L. Engineering the protein N-glycosylation pathway in insect cells for production of biantennary, complex N-glycans. *Biochemistry* **41,** 15093–15104 (2002).

318. Aumiller, J. J., Hollister, J. R. & Jarvis, D. L. A transgenic insect cell line engineered to produce CMP-sialic acid and sialylated glycoproteins. *Glycobiology* **13,** 497–507 (2003).

319. Mabashi-Asazuma, H. *et al.* Impact of a human CMP-sialic acid transporter on recombinant glycoprotein sialylation in glycoengineered insect cells. *Glycobiology* **23,** 199–210 (2013).

320. Viswanathan, K. *et al.* Engineering Sialic Acid Synthetic Ability into Insect Cells: Identifying Metabolic Bottlenecks and Devising Strategies to Overcome Them. *Biochemistry* **42,** 15215–15225 (2003).

321. Geisler, C. & Jarvis, D. L. Innovative use of a bacterial enzyme involved in sialic acid degradation to initiate sialic acid biosynthesis in glycoengineered insect cells. *Metab. Eng.* **14,** 642–652 (2012).

322. Aumiller, J. J., Mabashi-Asazuma, H., Hillar, A., Shi, X. & Jarvis, D. L. A new glycoengineered insect cell line with an inducibly mammalianized protein N-glycosylation pathway. *Glycobiology* **22,** 417–428 (2012).

323. Toth, A. M., Kuo, C. W., Khoo, K. H. & Jarvis, D. L. A new insect cell glycoengineering approach provides baculovirus-inducible glycogene expression and increases human-type glycosylation efficiency. *J. Biotechnol.* **182–183,** 19–29 (2014).

324. Mabashi-Asazuma, H., Kuo, C. W., Khoo, K. H. & Jarvis, D. L. Modifying an Insect Cell N-Glycan Processing Pathway Using CRISPR-Cas Technology. *ACS Chem. Biol.* **10,** 2199–2208 (2015).

325. Wacker, M. *et al.* N-linked glycosylation in Campylobacter jejuni and its functional

transfer into E. coli. *Science* **298,** 1790–1793 (2002).

326. Valguarnera, E., Kinsella, R. L. & Feldman, M. F. Sugar and Spice Make Bacteria Not Nice: Protein Glycosylation and Its Influence in Pathogenesis. *J. Mol. Biol.* **428,** 3206–3220 (2016).

327. Jervis, A. J., Butler, J. A., Wren, B. W. & Linton, D. Chromosomal integration vectors allowing flexible expression of foreign genes in Campylobacter jejuni. *BMC Microbiol.* **15,** 230 (2015).

328. Morrison, M. J. & Imperiali, B. The renaissance of bacillosamine and its derivatives: Pathway characterization and implications in pathogenicity. *Biochemistry* **53,** 624–638 (2014).

329. Raetz, C. R. H. & Whitfield, C. Lipopolysaccharide endotoxins. *Annu. Rev. Biochem.* **71,** 635–700 (2002).

330. Tatar, L. D., Marolda, C. L., Polischuk, A. N., van Leeuwen, D. & Valvano, M. A. An Escherichia coli undecaprenyl-pyrophosphate phosphatase implicated in undecaprenyl phosphate recycling. *Microbiology* **153,** 2518–2529 (2007).

331. Alaimo, C. *et al.* Two distinct but interchangeable mechanisms for flipping of lipid-linked oligosaccharides. *EMBO J.* **25,** 967–976 (2006).

332. Lehrer, J., Vigeant, K. A., Tatar, L. D. & Valvano, M. A. Functional Characterization and Membrane Topology of Escherichia coli WecA, a Sugar-Phosphate Transferase Initiating the Biosynthesis of Enterobacterial Common Antigen and O-Antigen Lipopolysaccharide . *J. Bacteriol.* **189,** 2618–2628 (2007).

333. Nothaft, H. & Szymanski, C. M. Bacterial protein n-glycosylation: New perspectives and applications. *J. Biol. Chem.* **288,** 6912–6920 (2013).

334. Valderrama-Rincon, J. D. *et al.* An engineered eukaryotic protein glycosylation pathway in Escherichia coli. *Nat. Chem. Biol.* **8,** 434–436 (2012).

335. Wacker, M. *et al.* Substrate specificity of bacterial oligosaccharyltransferase suggests a common transfer mechanism for the bacterial and eukaryotic systems. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7088–93 (2006).

336. O'Reilly, M. K., Zhang, G. & Imperiali, B. In vitro evidence for the dual function of Alg2 and Alg11: essential mannosyltransferases in N-linked glycoprotein biosynthesis. *Biochemistry* **45,** 9593–9603 (2006).

337. Cipollo, J. F., Trimble, R. B., Chi, J. H., Yan, Q. & Dean, N. The Yeast ALG11 Gene Specifies Addition of the Terminal 1,2-Man to the Man5GlcNAc2-PP-dolicholN-Glycosylation Intermediate Formed on the Cytosolic Side of the Endoplasmic Reticulum. *J. Biol. Chem.* **276,** 21828–21840 (2001).

338. Srichaisupakit, A., Ohashi, T., Misaki, R. & Fujiyama, K. Production of initial-stage eukaryotic N-glycan and its protein glycosylation in Escherichia coli. *J. Biosci. Bioeng.* **119,** 399–405 (2015).

339. Pandhal, J. *et al.* Systematic metabolic engineering for improvement of glycosylation efficiency in Escherichia coli. *Biochem. Biophys. Res. Commun.* **419,** 472–476 (2012).

340. Petrescu, A.-J., Milac, A.-L., Petrescu, S. M., Dwek, R. A. & Wormald, M. R. Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. *Glycobiology* **14,** 103–114 (2004).

341. Ollis, A. A. *et al.* Substitute sweeteners: diverse bacterial oligosaccharyltransferases with unique N-glycosylation site preferences. *Sci. Rep.* **5,** 15237 (2015).

342. Baeshen, M. N. *et al.* Production of biopharmaceuticals in E. Coli: Current scenario and future perspectives. *J. Microbiol. Biotechnol.* **25,** 953–962 (2015).

343. Guarino, C. & DeLisa, M. P. A prokaryote-based cell-free translation system that efficiently synthesizes glycoproteins. *Glycobiology* **22,** 596–601 (2012).

344. Borkowski, O., Ceroni, F., Stan, G. B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* **33,** 123–130 (2016).

345. Zamecnik, P. C. & Keller, E. B. Relation between phosphate energy donors and incorporation of labeled amino acids into proteins. *J. Biol. Chem.* **209,** 337–354 (1954).

346. Matthaei, J. H., Jones, O. W., Martin, R. G. & Nirenberg, M. W. Characteristics and composition of RNA coding units. *Proc. Natl. Acad. Sci. U. S. A.* **48,** 666–77 (1962).

347. Isaksson, L., Enberg, J., Neutze, R., Göran Karlsson, B. & Pedersen, A. Expression screening of membrane proteins with cell-free protein synthesis. *Protein Expr. Purif.* **82,** 218–225 (2012).

348. Tokmakov, A. A. *et al.* Multiple post-translational modifications affect heterologous protein synthesis. *J. Biol. Chem.* **287,** 27106–27116 (2012).

349. Focke, P. J. *et al.* Combining in Vitro Folding with Cell Free Protein Synthesis for Membrane Protein Expression. *Biochemistry* **55,** 4212–4219 (2016).

350. Smith, M. T., Wilding, K. M., Hunt, J. M., Bennett, A. M. & Bundy, B. C. The emerging age of cell-free synthetic biology. *FEBS Lett.* **588,** 2755–2761 (2014).

351. Shrestha, P., Smith, M. T. & Bundy, B. C. Cell-free unnatural amino acid incorporation with alternative energy systems and linear expression templates. *N. Biotechnol.* **31,** 28–34 (2014).

352. Casteleijn, M. G., Urtti, A. & Sarkhel, S. Expression without boundaries: Cell-free protein synthesis in pharmaceutical research. *Int. J. Pharm.* **440,** 39–47 (2013).

353. Rosenblum, G. & Cooperman, B. S. Engine out of the Chassis: Cell-Free Protein Synthesis and its Uses. *FEBS Lett.* **588,** 261–268 (2014).

354. Tarui, H. *et al.* Establishment and characterization of cell-free translation/glycosylation in insect cell (Spodoptera frugiperda 21) extract prepared with high pressure treatment. *Appl. Microbiol. Biotechnol.* **55,** 446–453 (2001).

355. Brödel, A. K., Sonnabend, A. & Kubick, S. Cell-free protein expression based on extracts from CHO cells. *Biotechnol. Bioeng.* **111,** 25–36 (2014).

356. Hodgman, C. E. & Jewett, M. C. Optimized extract preparation methods and reaction conditions for improved yeast cell-free protein synthesis. *Biotechnol. Bioeng.* **110,** n/a-n/a (2013).

357. Matthews, G. & Colman, a. A highly efficient, cell-free translation/translocation system prepared from Xenopus eggs. *Nucleic Acids Res.* **19,** 6405–6412 (1991).

358. Duszenko, M., Kang, X., Böhme, U., Hömke, R. & Lehner, M. In vitro translation in a cell-free system from Trypanosoma brucei yields glycosylated and glycosylphosphatidylinositol-anchored proteins. *Eur. J. Biochem.* **266,** 789–97 (1999).

359. Baldwin, G. *Synthetic Biology - A Primer (revised Edition)*. (2015).

360. Pardee, K. *et al.* Paper-Based Synthetic Gene Networks. *Cell* **159,** 940–954 (2014).

361. Chappell, J., Jensen, K. & Freemont, P. S. Validation of an entirely in vitro approach for rapid prototyping of DNA regulatory elements for synthetic biology. *Nucleic Acids Res.* **41,** 3471–3481 (2013).

362. Anderson, M. J., Stark, J. C., Hodgman, C. E. & Jewett, M. C. Energizing eukaryotic

cell-free protein synthesis with glucose metabolism. *FEBS Lett.* **589,** 1723–1727 (2015).

363. Stech, M. *et al.* A continuous-exchange cell-free protein synthesis system based on extracts from cultured insect cells. *PLoS One* **9,** e96635 (2014).

364. Schoborg, J. a, Hodgman, C. E., Anderson, M. J. & Jewett, M. C. Substrate replenishment and byproduct removal improve yeast cell-free protein synthesis. *Biotechnol. J.* **9,** 1–11 (2013).

365. Gan, R. & Jewett, M. C. A combined cell-free transcription-translation system from Saccharomyces cerevisiae for rapid and robust protein synthesis. *Biotechnol. J.* **9,** 641–651 (2014).

366. Shields, D. & Blobel, G. Cell-free synthesis of fish preproinsulin, and processing by heterologous mammalian microsomal membranes. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 2059–2063 (1977).

367. Thoring, L. *et al.* Cell-free systems based on CHO cell lysates: Optimization strategies, synthesis of 'difficult-to-express' proteins and future perspectives. *PLoS One* **11,** 1–21 (2016).

368. Walter, P. & Blobel, G. Preparation of microsomal membranes for cotranslational protein translocation. *Methods Enzymol.* **96,** 84–93 (1983).

369. Yin, G. *et al.* Aglycosylated antibodies and antibody fragments produced in a scalable in vitro transcription-translation system. *MAbs* **4,** 217–225 (2012).

370. Frey, S., Haslbeck, M., Hainzl, O. & Buchner, J. Synthesis and characterization of a functional intact IgG in a prokaryotic cell-free expression system. *Biol. Chem.* **389,** 37–45 (2008).

371. Groff, D. *et al.* Engineering toward a bacterial 'endoplasmic reticulum' for the rapid expression of immunoglobulin proteins. *MAbs* **6,** 671–678 (2014).

372. Kline, T. *et al.* Methods to Make Homogenous Antibody Drug Conjugates. *Pharm. Res.* 1–14 (2014). doi:10.1007/s11095-014-1596-8

373. Junutula, J. R. *et al.* Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* **26,** 925–932 (2008).

374. Link, A. J., Mock, M. L. & Tirrell, D. A. Non-canonical amino acids in protein engineering.

Curr. Opin. Biotechnol. **14,** 603–609 (2003).

375. Wang, Q., Parrish, A. R. & Wang, L. Expanding the Genetic Code for Biological Studies. *Chem. Biol.* **16,** 323–336 (2009).

376. Wals, K. & Ovaa, H. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2,** 15 (2014).

377. Kataoka, K., Harada, a & Nagasaki, Y. Block copolymer micelles for drug delivery: design, characterization and biological significance. *Adv. Drug Deliv. Rev.* **47,** 113–131 (2001).

378. Olafsen, T. & Wu, A. M. Antibody Vectors for Imaging. *Semin. Nucl. Med.* **40,** 167–181 (2010).

379. Hoffman, C. S. & Winston, F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene* **57,** 267–272 (1987).

380. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6,** 343–5 (2009).

381. Chen, X. & Flynn, G. C. Gas-Phase Oligosaccharide Nonreducing End (GONE) Sequencing and Structural Analysis by Reversed Phase HPLC/Mass Spectrometry with Polarity Switching. *J. Am. Soc. Mass Spectrom.* **20,** 1821–1833 (2009).

382. Del Val, I. J., Kyriakopoulos, S., Polizzi, K. M. & Kontoravdi, C. An optimized method for extraction and quantification of nucleotides and nucleotide sugars from mammalian cells. *Anal. Biochem.* **443,** 172–180 (2013).

383. Saribas, a S., Johnson, K., Liu, L., Bezila, D. & Hakes, D. Refolding of human beta-1-2 GlcNAc transferase (GnT1) and the role of its unpaired Cys 121. *Biochemical and biophysical research communications* **362,** 381–6 (2007).

384. Chen, R., Pawlicki, M. A., Hamilton, B. S. & Tolbert, T. J. Enzyme-Catalyzed Synthesis of a Hybrid N-Linked Oligosaccharide using N-Acetylglucosaminyltransferase I. *Adv. Synth. Catal.* **350,** 1689–1695 (2008).

385. Suzuki, T. *et al.* Man2C1, an alpha-mannosidase, is involved in the trimming of free oligosaccharides in the cytosol. *Biochem. J.* **400,** 33–41 (2006).

386. Harada, Y. *et al.* Metabolically programmed quality control system for dolichol-linked

oligosaccharides. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 19366–71 (2013).

387. Cipollo, J. F., Trimble, R. B., Chi, J. H., Yan, Q. & Dean, N. The yeast ALG11 gene specifies addition of the terminal alpha 1,2-Man to the Man5GlcNAc2-PP-dolichol N-glycosylation intermediate formed on the cytosolic side of the endoplasmic reticulum. *J. Biol. Chem.* **276,** 21828–21840 (2001).

388. Uccelletti, D. *et al.* The Kluyveromyces lactis alpha1,6-mannosyltransferase KlOch1p is required for cell-wall organization and proper functioning of the secretory pathway. *FEMS Yeast Res.* **6,** 449–457 (2006).

389. Yoko-o, T. *et al.* Schizosaccharomyces pombe och1(+) encodes alpha-1,6-mannosyltransferase that is involved in outer chain elongation of N-linked oligosaccharides. *FEBS Lett.* **489,** 75–80 (2001).

390. Lin-cereghino, J. *et al.* Direct selection of Pichia pastoris expression strains using new G418 resistance vectors. *Yeast* **25,** 293–299 (2008).

391. Winzeler, E. A. *et al.* Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285,** 901–906 (1999).

392. Wach, A., Brachat, A., Pohlmann, R. & Philippsen, P. New heterologous modules for classical or PCR-based gene disruptions in Saccharomyces cerevisiae. *Yeast* **10,** 1793–1808 (1994).

393. Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. & Cullin, C. A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res.* **21,** 3329–3330 (1993).

394. Jimenez, A. & Davies, J. Expression of a transposable antibiotic resistance element in Saccharomyces. *Nature* **287,** 869–871 (1980).

395. Ahn, J. *et al.* Translation elongation factor 1-alpha gene from Pichia pastoris: molecular cloning, sequence, and use of its promoter. *Appl. Microbiol. Biotechnol.* **74,** 601–608 (2007).

396. Sou, S. N., Ilieva, K. M. & Polizzi, K. M. Binding of human BiP to the ER stress transducers IRE1 and PERK requires ATP. *Biochem. Biophys. Res. Commun.* **420,** 473–478 (2012).

397. ČIPLYS, E., SLIBINSKAS, R., SASNAUSKAS, K., Michalak, M. & GOLD, L. I. Generation of native recombinant secreted human endoplasmic reticulum chaperones

by using their native signal sequences in yeast expression systems. (2014).

398. Christen, M. T., Frank, P., Schaller, J. & Llinás, M. Human plasminogen kringle 3: solution structure, functional insights, phylogenetic landscape. *Biochemistry* **49,** 7131–50 (2010).

399. Higel, F., Seidl, A., S??rgel, F. & Friess, W. N-glycosylation heterogeneity and the influence on structure, function and pharmacokinetics of monoclonal antibodies and Fc fusion proteins. *Eur. J. Pharm. Biopharm.* **100,** 94–100 (2016).

400. Cereghino, J. L. & Cregg, J. M. Heterologous protein expression in the methylotrophic yeast Pichia pastoris. *FEMS Microbiol. Rev.* **24,** 45–66 (2000).

401. Clare, J. J., Rayment, F. B., Ballantine, S. P., Sreekrishna, K. & Romanos, M. A. High-level expression of tetanus toxin fragment C in Pichia pastoris strains containing multiple tandem integrations of the gene. *Biotechnology. (N. Y).* **9,** 455–460 (1991).

402. Menendez, J., Garcia, B. & Hidalgo, Y. New method for the selection of multicopy transformants of Pichia pastoris, using 3-amino-1,2,4 triazol. *Biotechniques* **29,** 1094–1099 (2000).

403. Klinner, U. & Schäfer, B. Genetic aspects of targeted insertion mutagenesis in yeasts. *FEMS Microbiol. Rev.* **28,** 201–223 (2004).

404. Näätsaari, L. *et al.* Deletion of the Pichia pastoris KU70 Homologue Facilitates Platform Strain Generation for Gene Expression and Synthetic Biology. *PLoS One* **7,** e39720 (2012).

405. Oldenburg, K. R., Vo, K. T., Michaelis, S. & Paddon, C. Recombination-mediated PCR-directed plasmid construction in vivo in yeast. *Nucleic Acids Res.* **25,** 451–452 (1997).

406. Chen, Z. *et al.* Enhancement of the gene targeting efficiency of non-conventional yeasts by increasing genetic redundancy. *PLoS One* **8,** e57952 (2013).

407. American Type Culture Collection. Komagataella pastoris (Guilliermond) Yamada et al. (ATCC® 28485™). Available at: https://www.lgcstandards-atcc.org/Products/All/28485.aspx#specifications.

408. Krainer, F. W. *et al.* Knockout of an endogenous mannosyltransferase increases the homogeneity of glycoproteins produced in Pichia pastoris. *Sci. Rep.* **3,** 3279 (2013).

409. Näätsaari, L. *et al.* Deletion of the Pichia pastoris KU70 Homologue Facilitates Platform

Strain Generation for Gene Expression and Synthetic Biology. *PLoS One* **7,** e39720 (2012).

410. Bulter, T. & Alcalde, M. Preparing Libraries in Saccharomyces cerevisiae BT - Directed Evolution Library Creation: Methods and Protocols. in (eds. Arnold, F. H. & Georgiou, G.) 17–22 (Humana Press, 2003). doi:10.1385/1-59259-395-X:17

411. Hamilton, S. R. *et al.* Production of complex human glycoproteins in yeast. *Science* **301,** 1244–1246 (2003).

412. Waterham, H. R., Digan, M. E., Koutz, P. J., Lair, S. V & Cregg, J. M. Isolation of the Pichia pastoris glyceraldehyde-3-phosphate dehydrogenase gene and regulation and use of its promoter. *Gene* **186,** 37–44 (1997).

413. Vogl, T. & Glieder, A. Regulation of Pichia pastoris promoters and its consequences for protein production. *N. Biotechnol.* **30,** 385–404 (2013).

414. Jimenez del Val, I., Nagy, J. M. & Kontoravdi, C. A dynamic mathematical model for monoclonal antibody N-linked glycosylation and nucleotide sugar donor transport within a maturing Golgi apparatus. *Biotechnol. Prog.* **27,** 1730–1743 (2011).

415. Klymenko, O. V, Shah, N., Kontoravdi, C., Royle, K. E. & Polizzi, K. M. Designing an Artificial Golgi Reactor to Achieve Targeted Glycosylation of Monoclonal Antibodies. *AICHE J.* **62,** 2959–2973 (2016).

416. Saint-Jore-Dupas, C. *et al.* Plant N-Glycan Processing Enzymes Employ Different Targeting Mechanisms for Their Spatial Arrangement along the Secretory Pathway. *Plant Cell* **18,** 3182–3200 (2006).

417. Kim, P.-J., Lee, D.-Y. & Jeong, H. Centralized modularity of N-linked glycosylation pathways in mammalian cells. *PLoS One* **4,** e7317 (2009).

418. del Val, I. J., Kontoravdi, C. & Nagy, J. M. Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. *Biotechnol. Prog.* **26,** 1505–27 (2010).

419. Martin, J. G. *et al.* Toward an artificial Golgi: redesigning the biological activities of heparan sulfate on a digital microfluidic chip. *J. Am. Chem. Soc.* **131,** 11041–8 (2009).

420. Gordon, R. D. *et al.* X-ray crystal structures of rabbit N-acetylglucosaminyltransferase I (GnT I) in complex with donor substrate analogues. *J. Mol. Biol.* **360,** 67–79 (2006).

421. van den Elsen, J. M., Kuntz, D. a & Rose, D. R. Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *EMBO J.* **20,** 3008–17 (2001).

422. Toki, D., Sarkar, M., Yip, B. & Reck, F. Expression of stable human O-glycan core 2 β-1, 6-N-acetylglucosaminyltransferase in Sf9 insect cells. *Biochem. J* **325,** 63–69 (1997).

423. Bierhuizen, M. F. & Fukuda, M. Expression cloning of a cDNA encoding UDP-GlcNAc:Gal beta 1-3-GalNAc-R (GlcNAc to GalNAc) beta 1-6GlcNAc transferase by gene transfer into CHO cells expressing polyoma large tumor antigen. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 9326–9330 (1992).

424. Oh-Eda, M. *et al.* Overexpression of the Golgi-localized enzyme alpha-mannosidase IIx in Chinese hamster ovary cells results in the conversion of hexamannosyl-N-acetylchitobiose to tetramannosyl-N-acetylchitobiose in the N-glycan-processing pathway. *Eur. J. Biochem.* **268,** 1280–1288 (2001).

425. Li, B. *et al.* Inhibition of Golgi mannosidase II with mannostatin A analogues: synthesis, biological evaluation, and structure-activity relationship studies. *Chembiochem* **5,** 1220–1227 (2004).

426. Tate, C. G. Overexpression of mammalian integral membrane proteins for structural studies. *FEBS Lett.* **504,** 94–98 (2001).

427. Wagner, S. *et al.* Tuning Escherichia coli for membrane protein overexpression. *Proc. Natl. Acad. Sci.* **105,** 14371–14376 (2008).

428. Franken, K. L. *et al.* Purification of his-tagged proteins by immobilized chelate affinity chromatography: the benefits from the use of organic solvent. *Protein Expr. Purif.* **18,** 95–99 (2000).

429. Schmitt, J., Hess, H. & Stunnenberg, H. G. Affinity purification of histidine-tagged proteins. *Mol. Biol. Rep.* **18,** 223–230 (1993).

430. Ruhaak, L. R. *et al.* Glycan labeling strategies and their use in identification and quantification. *Anal. Bioanal. Chem.* **397,** 3457–81 (2010).

431. Laroy, W., Contreras, R. & Callewaert, N. Glycome mapping on DNA sequencing equipment. *Nat. Protoc.* **1,** 397–405 (2006).

432. Kumar, H. P. M., Denny, P. A. & Denny, P. C. Use of Fluorophore-Assisted Carbohydrate Electrophoresis (FACE®) in the Elucidation of N-Linked Oligosaccharide

Structures. *Carbohydr. Bioetchnology Protoc.* **10,** 221–234 (1999).

433. Anumula, K. R. Advances in fluorescence derivatization methods for high-performance liquid chromatographic analysis of glycoprotein carbohydrates. *Anal. Biochem.* **350,** 1–23 (2006).

434. Paschinger, K. *et al.* A deletion in the golgi alpha-mannosidase II gene of Caenorhabditis elegans results in unexpected non-wild-type N-glycan structures. *J. Biol. Chem.* **281,** 28265–77 (2006).

435. Nallamsetty, S., Austin, B. P., Penrose, K. J. & Waugh, D. S. Gateway vectors for the production of combinatorially-tagged His 6 -MBP fusion proteins in the cytoplasm and periplasm of Escherichia coli. *Protein Sci.* **14,** 2964–2971 (2005).

436. Polakova, M. *et al.* alpha-D-mannose derivatives as models designed for selective inhibition of Golgi alpha-mannosidase II. *Eur. J. Med. Chem.* **46,** 944–952 (2011).

437. Li, J., Zhang, J., Lai, B., Zhao, Y. & Li, Q. Cloning, Expression, and Characterization of Capra hircus Golgi ??-Mannosidase II. *Appl. Biochem. Biotechnol.* **177,** 1241–1251 (2015).

438. Schachter, H., Reck, F. & Paulsen, H. Use of synthetic oligosaccharide substrate analogs to map the active sites of N-acetylglucosaminyltransferases I and II. *Methods Enzymol.* **363,** 459–75 (2003).

439. Carlson, E. D., Gan, R., Hodgman, E. C. & Jewet, M. C. Cell-Free Protein Synthesis : Applications Come of Age. **30,** 1185–1194 (2014).

440. Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnol. Adv.* **30,** 1185–94

441. Kim, D. M. & Swartz, J. R. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnol. Bioeng.* **74,** 309–316 (2001).

442. Kim, T.-W. *et al.* Prolonged cell-free protein synthesis using dual energy sources: Combined use of creatine phosphate and glucose for the efficient supply of ATP and retarded accumulation of phosphate. *Biotechnol. Bioeng.* **97,** 1510–1515 (2007).

443. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19,** 751–755 (2001).

444. Hillebrecht, J. R. & Chong, S. A comparative study of protein synthesis in in vitro

systems: from the prokaryotic reconstituted to the eukaryotic extract-based. *BMC Biotechnol.* **8,** 58 (2008).

445. Zawada, J. F. *et al.* Microscale to manufacturing scale-up of cell-free cytokine production--a new approach for shortening protein production development timelines. *Biotechnol. Bioeng.* **108,** 1570–1578 (2011).

446. Park, N., Um, S. H., Funabashi, H., Xu, J. & Luo, D. A cell-free protein-producing gel. *Nat. Mater.* **8,** 432–437 (2009).

447. Sawasaki, T., Ogasawara, T., Morishita, R. & Endo, Y. A cell-free protein synthesis system for high-throughput proteomics. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 14652–14657 (2002).

448. Zemella, A., Thoring, L., Hoffmeister, C. & Kubick, S. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem* **16,** 2420–2431 (2015).

449. Rubenstein, J. L. & Chappell, T. G. Construction of a synthetic messenger RNA encoding a membrane protein. *J. Cell Biol.* **96,** 1464–1469 (1983).

450. Hirose, S., Kim, S., Miyazaki, H., Park, Y. S. & Murakami, K. In vitro biosynthesis of human renin and identification of plasma inactive renin as an activation intermediate. *J. Biol. Chem.* **260,** 16400–16405 (1985).

451. MacDonald, M. R., McCourt, D. W. & Krause, J. E. Posttranslational processing of alpha-, beta-, and gamma-preprotachykinins. Cell-free translation and early posttranslational processing events. *J. Biol. Chem.* **263,** 15176–15183 (1988).

452. Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: Applications come of age. *Biotechnol. Adv.* **30,** 1185–1194 (2012).

453. Endo, Y. & Sawasaki, T. Cell-free expression systems for eukaryotic protein production. *Curr. Opin. Biotechnol.* **17,** 373–380 (2006).

454. Mikami, S., Kobayashi, T. & Imataka, H. Cell-free protein synthesis systems with extracts from cultured human cells. *Methods Mol. Biol.* **607,** 43–52 (2010).

455. Shin, J. & Noireaux, V. Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *J. Biol. Eng.* **4,** 8 (2010).

456. Noireaux, V., Bar-Ziv, R. & Libchaber, A. Principles of cell-free genetic circuit assembly.

Proc. Natl. Acad. Sci. U. S. A. **100,** 12672–12677 (2003).

457. Papakonstantinou, T., Harris, S. & Hearn, M. T. W. Expression of GFP using Pichia pastoris vectors with zeocin or G-418 sulphate as the primary selectable marker. *Yeast* **26,** 311–321 (2009).

458. Iizuka, N., Najita, L., Franzusoff, A. & Sarnow, P. Cap-dependent and cap-independent translation by internal initiation of mRNAs in cell extracts prepared from Saccharomyces cerevisiae. *Mol. Cell. Biol.* **14,** 7322–7330 (1994).

459. Brodel, A. K. *et al.* IRES-mediated translation of membrane proteins and glycoproteins in eukaryotic cell-free systems. *PLoS One* **8,** e82234 (2013).

460. Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J. Virol.* **62,** 2636–2643 (1988).

461. Bochkov, Y. a. & Palmenberg, A. C. Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location. *Biotechniques* **41,** 283–292 (2006).

462. Komar, A. A. & Hatzoglou, M. Cellular IRES-mediated translation: the war of ITAFs in pathophysiological states. *Cell Cycle* **10,** 229–240 (2011).

463. Fitzgerald, K. D. & Semler, B. L. Bridging IRES elements in mRNAs to the eukaryotic translation apparatus. *Biochim. Biophys. Acta* **1789,** 518–528 (2009).

464. Liang, S. *et al.* Comprehensive structural annotation of Pichia pastoris transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. *BMC Genomics* **13,** 738 (2012).

465. Chen, W.-S. *et al.* Development of a prokaryotic-like polycistronic baculovirus expression vector by the linkage of two internal ribosome entry sites. *J. Virol. Methods* **159,** 152–9 (2009).

466. Liang, S., Lin, Y., Li, C. & Ye, Y. Internal ribosome entry site mediates protein synthesis in yeast Pichia pastoris. *Biotechnol. Lett.* **34,** 957–964 (2012).

467. Jackson, R. J., Howell, M. T. & Kaminski, A. The novel mechanism of initiation of picornavirus RNA translation. *Trends Biochem. Sci.* **15,** 477–483 (1990).

468. Hodgman, C. E. & Jewett, M. C. Characterizing IGR IRES-mediated translation initiation

for use in yeast cell-free protein synthesis. *N. Biotechnol.* **31,** 499–505 (2014).

469. Woolaway, K. E., Lazaridis, K., Belsham, G. J., Carter, M. J. & Roberts, L. O. The 5' untranslated region of Rhopalosiphum padi virus contains an internal ribosome entry site which functions efficiently in mammalian, plant, and insect translation systems. *J. Virol.* **75,** 10244–10249 (2001).

470. Cevallos, R. C. & Sarnow, P. Factor-independent assembly of elongation-competent ribosomes by an internal ribosome entry site located in an RNA virus that infects penaeid shrimp. *J. Virol.* **79,** 677–683 (2005).

471. Bergamini, G., Preiss, T. & Hentze, M. W. Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA* **6,** 1781–1790 (2000).

472. Savage, D. F. Cell-free protein synthesis: Methods and protocols, edited by Alexander S. Spirin and James R. Swartz. 2008. Wiley-VCH, Weinheim, Germany. 262 pp. $110 (hardcover). *Protein Sci.* **17,** 962–963 (2008).

473. Gallie, D. R. The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.* **5,** 2108–2116 (1991).

474. Jackson, R. J. & Standart, N. Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell* **62,** 15–24 (1990).

475. Bergamini, G., Preiss, T. & Hentze, M. W. Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. *RNA* **6,** 1781–1790 (2000).

476. Kigawa, T. Cell-Free Protein Production System with the E. coli Crude Extract for Determination of Protein Folds BT - Cell-Free Protein Production: Methods and Protocols. in (eds. Endo, Y., Takai, K. & Ueda, T.) 101–111 (Humana Press, 2010). doi:10.1007/978-1-60327-331-2_10

477. Zawada, J. F. Preparation and testing of E. coli S30 in vitro transcription translation extracts. *Methods Mol. Biol.* **805,** 31–41 (2012).

478. Sullivan, C. J. *et al.* A cell-free expression and purification process for rapid production of protein biologics. *Biotechnol. J.* **11,** 238–248 (2016).

479. Choudhury, A., Hodgman, C. E., Anderson, M. J. & Jewett, M. C. Evaluating fermentation effects on cell growth and crude extract metabolic activity for improved

yeast cell-free protein synthesis. *Biochem. Eng. J.* **91,** 140–148 (2014).

480. Ezure, T. *et al.* A cell-free translocation system using extracts of cultured insect cells to yield functional membrane proteins. *PLoS One* **9,** 1–17 (2014).

481. Ezure, T., Suzuki, T., Shikata, M., Ito, M. & Ando, E. A Cell-Free Protein Synthesis System from Insect Cells BT - Cell-Free Protein Production: Methods and Protocols. in (eds. Endo, Y., Takai, K. & Ueda, T.) 31–42 (Humana Press, 2010). doi:10.1007/978-1-60327-331-2_4

482. Takai, K. & Endo, Y. The Cell-Free Protein Synthesis System from Wheat Germ BT - Cell-Free Protein Production: Methods and Protocols. in (eds. Endo, Y., Takai, K. & Ueda, T.) 23–30 (Humana Press, 2010). doi:10.1007/978-1-60327-331-2_3

483. Harbers, M. Wheat germ systems for cell-free protein expression. *FEBS Lett.* **588,** 2762–2773 (2014).

484. Kubick, S., Gerrits, M., Merk, H., Stiege, W. & Erdmann, V. A. Chapter 2 In Vitro Synthesis of Posttranslationally Modified Membrane Proteins. *Curr. Top. Membr.* **63,** 25–49 (2009).

485. Albro, P. W., Corbett, J. T. & Schroeder, J. L. Rapid isolation of microsomes for studies of lipid peroxidation. *Lipids* **22,** 751–756 (1987).

486. Rossanese, O. W. *et al.* Golgi Structure Correlates with Transitional Endoplasmic Reticulum Organization in Pichia pastoris and Saccharomyces cerevisiae. *J. Cell Biol.* **145,** 69 LP-81 (1999).

487. Teckman, J. H. & Perlmutter, D. H. Retention of mutant alpha(1)-antitrypsin Z in endoplasmic reticulum is associated with an autophagic response. *Am. J. Physiol. Gastrointest. Liver Physiol.* **279,** G961-74 (2000).

488. Hunter, A. R., Farrell, P. J., Jackson, R. J. & Hunt, T. The role of polyamines in cell-free protein synthesis in the wheat-germ system. *Eur. J. Biochem.* **75,** 149–157 (1977).

489. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–130 (1986).

490. Mairhofer, J., Wittwer, A., Cserjan-Puschmann, M. & Striedner, G. Preventing T7 RNA Polymerase Read-through Transcription—A Synthetic Termination Signal Capable of Improving Bioprocess Stability. *ACS Synth. Biol.* **4,** 265–273 (2015).

491. NEB. HiScribe™ T7 High Yield RNA Synthesis Kit. Available at: https://www.neb.com/products/e2040-hiscribe-t7-high-yield-rna-synthesis-kit.

492. Calhoun, K. A. & Swartz, J. R. Energizing cell-free protein synthesis with glucose metabolism. *Biotechnol. Bioeng.* **90,** 606–613 (2005).

493. Zeenko, V. V *et al.* An efficient in vitro translation system from mammalian cells lacking the translational inhibition caused by eIF2 phosphorylation. *RNA* **14,** 593–602 (2008).

494. Dasso, M. C. & Jackson, R. J. On the fidelity of mRNA translation in the nuclease-treated rabbit reticulocyte lysate system. *Nucleic Acids Res.* **17,** 3129–3144 (1989).

495. Contreras, R., Cheroutre, H., Degrave, W. & Fiers, W. Simple, efficient in vitro synthesis of capped RNA useful for direct expression of cloned eukaryotic genes. *Nucleic Acids Res.* **10,** 6353–6362 (1982).

496. Aw, R. & Polizzi, K. M. Can too many copies spoil the broth? *Microb. Cell Fact.* **12,** 128 (2013).

497. Vervecken, W. *et al.* In vivo synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. *Appl. Environ. Microbiol.* **70,** 2639–2646 (2004).

498. Kanagawa, M. *et al.* Structural Basis for Multiple Sugar Recognition of Jacalin-related Human ZG16p Lectin. *J. Biol. Chem.* **289,** 16954–16965 (2014).

499. Hansen, D. K. *et al.* Mutational Analysis of Divalent Metal Ion Binding in the Active Site of Class II α-Mannosidase from *Sulfolobus solfataricus*. *Biochemistry* **54,** 2032–2039 (2015).

500. Zhong, W. *et al.* Probing the substrate specificity of Golgi alpha-mannosidase II by use of synthetic oligosaccharides and a catalytic nucleophile mutant. *J. Am. Chem. Soc.* **130,** 8975–83 (2008).

501. Song, E. *et al.* A simplified procedure for the purification of human N -acetylglucosaminyltransferase-III from human hepatocellular carcinoma tissues. 201–204 (2000).

502. Bajazawa, K., Furukawa, K., Narimatsu, H. & Kobata, A. Kinetic Study of Human {beta}-1,4-Galactosyltrasgerase Expressed in E.coli. *J Biochem* **113,** 747–753 (1993).

503. Shibatani, S., Fujiyama, K., Nishiguchi, S., Seki, T. & Maekawa, Y. Production and characterization of active soluble human beta1,4-galactosyltransferase in Escherichia

coli as a useful catalyst in synthesis of the Gal beta1-->4 GlcNAc linkage. *J. Biosci. Bioeng.* **91,** 85–7 (2001).

504. Fujiyama, K. *et al.* Human N-acetylglucosaminyltransferase I. Expression in Escherichia coli as a soluble enzyme, and application as an immobilized enzyme for the chemoenzymatic synthesis of N-linked oligosaccharides. *J. Biosci. Bioeng.* **92,** 569–74 (2001).

505. Malissard, M., Zeng, S. & Berger, E. G. Expression of functional soluble forms of human beta-1, 4-galactosyltransferase I, alpha-2,6-sialyltransferase, and alpha-1, 3-fucosyltransferase VI in the methylotrophic yeast Pichia pastoris. *Biochem. Biophys. Res. Commun.* **267,** 169–173 (2000).

506. Schaub, B. E., Berger, B., Berger, E. G. & Rohrer, J. Transition of galactosyltransferase 1 from trans-Golgi cisterna to the trans-Golgi network is signal mediated. *Mol. Biol. Cell* **17,** 5153–62 (2006).

507. Akopian, D., Shen, K., Zhang, X. & Shan, S. Signal Recognition Particle: An essential protein targeting machine. *Annu. Rev. Biochem.* **82,** 693–721 (2013).

508. Lin-Cereghino, G. P. *et al.* The effect of alpha-mating factor secretion signal mutations on recombinant protein expression in Pichia pastoris. *Gene* **519,** 311–317 (2013).

509. Sachse, R. *et al.* Synthesis of membrane proteins in eukaryotic cell-free systems. *Eng. Life Sci.* **13,** 39–48 (2013).

Appendix

Section 1 Primer sequences

A list of primers utilised throughout the study, with sequence information and a brief description.

Primer Name	Description	Sequence (5'→3')
5OCH1_F	Forward primer of the OCH1 5' recombination locus	CAAAGATTGTAGGAGCGAACCC
5OCH1_TEFp_R	Reverse primer of the 5'OCH1 plus the beginning of tEF promoter	GTTATTACCTGATTGGGAAGGAGAGCGATAG
TEFp_5OCH1_F	Forward primer of the TEFp plus the end of the 5'OCH1	CTATCGCTCTCCTTCCCAATCAGGTAATAAC
TEFp_BiP_R	Reverse primer of the TEFp plus the beginning of the BiP tag	CTGTTGAAAAACATGGTGGCGAATAAC
BiP_TEFp_F	Forward primer of the BiP signal sequence and the end of GAP promoter	GTTATTCGCCACCATGTTTTTCAACAG
BiP_Man1a_R	Reverse primer of the BiP signal sequence and the beginning of Man1a CDS	GTCAGCGGGCTCTGTTGCTACTAGTG
Man1a_BiP_F	Forward primer of the Man1a plus the end of the BiP tag	CACTAGTAGCAACAGAGCCCGCTGAC
Man1a_AOXTerm_R	Reverse primer of the Man1a plus the beginning of the AOX-TT	CTAAGGCTAAAACTCACAATTCATCATG
AOXTerm_Man1a_F	Forward primer of the AOX-TT plus the end of the Man1a	CATGATGAATTGTGAGTTTTAGCCTTAG
AOXTerm_G418_R	Reverse primer of the AOX-TT plus the beginning of the GAP-KanR construct (annealing some after the end of the vector AOX-TT site)	GACGAGGACACCACATTTTGAAGCTAT
G418_AOXTT_F	Forward primer of the GAP-KanR plus the end of the AOX-TT	ATAGCTTCAAAATGTGGTGTCCTCGTC
G418_3OCH1_R	Reverse primer of the GAP-KanR plus the beginning of the 3' OCH1 recombination locus	CTTCAAATGTATGCCTTTAGAAAAACTCATC

3OCH1_G418_F	Forward primer of the 3' OCH1 recombination locus plus the end of the GAP-KanR	GATGAGTTTTTCTAAAGG CATACATTTGAAG
3OCH1_R	Reverse primer of the OCH1 3' recombination locus	GCGATGTACTGACGAAC ACCG
Man1a 5'part_(SacII)_Forward	To add the SacII restriction site to the 5’ end	CATGGCCGCGGGAATTC GAT
Man1a 5'part_(SacII)_Reverse	To add the SacII restriction site to the 5’ end	GCCCGCGGTCACAATTC ATCATG
Man1A(colony PCR)_Forward	Primer to amplify a region within the Man1a gene	GCAGTGGAACTTGGGGT AAA
Man1A(colony PCR)_Reverse	Primer to amplify a region within the Man1a gene	AGCTGTCTCCAAGTCCT CCA
Man1A(colony PCR)_5ochGenomic_Fo ward	Primer site upstream of the integration of the site	GCTGCGCCCACACATAT AAA
Man1A(colony PCR)_3’OCHGenomic_ Reverse	Primer site downstream of the integration of the site	GCCACTGTTACGTGAAG TCC
Man1a construct sequence primer#1	Sequencing primer for Man1a construct binding site at 722 bp	TGTTTGTCCATTTCATAG GC
Man1a construct sequence primer#2	Sequencing primer for Man1a construct binding site at 1459 bp	CGTAAAAACAATCCTTGT C
Man1a construct sequence primer#3	Sequencing primer for Man1a construct binding site at 2213 bp	AACATACGCTTCGTCGG T
Man1a construct sequence primer#4	Sequencing primer for Man1a construct binding site at 2974 bp	ACGAGGCAAAATGAAAA GTA
Man1a construct sequence primer#5	Sequencing primer for Man1a construct binding site at 3744 bp	TCAGGTAGCCATCTCTG AAA
Man1a construct sequence primer#6	Sequencing primer for Man1a construct binding site at 4489 bp	GGTTACTCACCACTGCG AT
Man1a construct sequence primer#7	Sequencing primer for Man1a construct binding site at 5230 bp	TATCAAAGCAAGAACCG ATG
3’OCH1_pJANS1_Forwa rd	Forward primer of the 3’OCH1 gene plus end of the pJAN-S1 plasmid for Gibson assembly	CGAAAAGTGCCACCTGG GCCTTATCGTTGGAAGA ACG
3’OCH1_5’OCH1_Rever se	Reverse primer for 3’OCH1 with complementary 5’OCH1 Sequence for Gibson assembly	GGACTCCATTGGACTAG CTAGGATCCGGTTACCA GCATTATCTACCA

5'OCH1_3'OCH1_Forward	Forward primer for 5'OCH1 with complementary 3'OCH1 Sequence for Gibson assembly	TGGTAGATAATGCTGGTAACCGGATCCTAGCTAGTCCAATGGAGTCC
5'OCH1_pJANS1_Reverse	Reverse primer of the 5'OCH1 gene plus end of the pJAN-S1 plasmid for Gibson assembly	TTTCGTCTTTGGATGTTAGATCAATGAAACTCCAATTGTCTTCG
OCH1_500_F	Primer used for amplifying 500bp product from the *OCH1* CDS gene after transformation pGEM-och1:man1a:G418	TGCTGGGTTGGTCATTGGTA
OCH1_500_R	Primer used for amplifying 500bp product from the *OCH1* CDS gene after transformation pGEM-och1:man1a:G418	ACTGAAGCTGGTGATCGGAA
beta Actin Forward	Primer to amplify β-actin gene for RT-PCR	GCT TTG TTC CAC CA TCT GT
beta Actin Reverse	Primer to amplify β-actin gene for RT-PCR	TGC ATA CGC TCA GCA ATA CC
TefP(BglII)_Forward	Forward primer to introduce BglII restriction site for construction of pKANB TefP;Man1a plasmid	GTCAGATCTCCAATCAGGTAATAACTGTCG
Man1a(SacIIx2)_Reverse	Reverse primer to introduce SacII restriction site for construction of pKANB TefP;Man1a plasmid	ACCGCGGCCCGCGGTCACAATTCATCAT
ManII(pPICz)_EcoRI_Foward	Primer to introduce a EcoRI site for cloning of ManII gene into pPICz plasmid	TGAGCAGAATTCGGAGATATACCATGGGCAGCAG
ManII(pPICz)_XbaI_Reverse	Primer to introduce a XbaI site for cloning of ManII gene into pPICz plasmid	TGAGCATCTAGATTAACGCAGTTGAATACGGAAGGTAG
ManII(colony PCR)_Forward	Primer to amplify a region within the ManII gene	AGTGCTGAGCGGTGATTTCT
ManII(colony PCR)_Reverse	Primer to amplify a region within the ManII gene	CGTCATGGTGTTGGAACAAG
C.elegans_ManII(colony PCR)_Forward	Primer to amplify a region within the ManII gene	ACAAGTTTCGCCGGTTATTG
C.elegans_ManII(colony PCR)_Reverse	Primer to amplify a region within the ManII gene	TGCCGCATTATTGGTGATTA

Tef_rtPCR_F	Primer to amplify a region within the Tef gene	CGGTGACAACATGATTG AGC
Tef_rtPCR_r	Primer to amplify a region within the Tef gene	ACTGGCACAGTTCCGAT ACC
GFP_BamHI_Forward	Primer to introduce SacI to the GFP for cloning into pet29a	TTCGAGCTCATGGTTTC CAAGGGTGAAGAG
GFP_XhoI_Reverse	Primer to introduce XhoI to the GFP for cloning into pet29a	GTGAGCTCGCTCGAGTT AGGATCCCTTATAAAGCT CGTCCATTCCG
GPR_BamHI_Forward	Primer to introduce XhoI to the GFP for cloning into pet29a	TATGGATCCGGTACCGC CCTCTTGTT
GPR_GPF_Reverse	Primer to introduce XhoI to the GFP for cloning into pet29a	CCTTGCTCACCATTTTGA GTTGGAG
GFP_GPR_Forward	Primer to introduce XhoI to the GFP for cloning into pet29a	CTCCAACTCAAAATGGT GAGCAAGG
CrPV_BamHI_Forward	Primer to introduce XhoI to the GFP for cloning into pet29a	TGAGGATCAAAGCAAAA ATGTGATCTTGC
CrPV_GPF_Reverse	Primer to introduce XhoI to the GFP for cloning into pet29a	CCTTGCTCACCATAGGT AAATTTC
GFP_CrPV_Forward	Primer to introduce XhoI to the GFP for cloning into pet29a	GAAATTTACCTATGGTGAG CAAGG
α-factor(pet28aGFP)_ Forward	Primer to introduce XhoI to the GFP for cloning into pet29a	GGACAGCAAATGGGTCG CGAGCTCATGGTTTCCA AGGGTGAAGAG
α-factor(pet28aGFP)_ Reverse	Primer to introduce XhoI to the GFP for cloning into pet29a	TTAGCAGCCGGATCTCA GTTACTTGTACAGCTCGT C
poly-A (EcoRI)_Forward	Primer for introduction of a Poly-A tail	AATTCAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAA AAAAAAAAAAAAAAAAAA ACG
poly-A (NotI)_Reverse	Primer for introduction of a Poly-A tail	GTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTTCGCCGG

Section 2 Plasmid sequencing information

Sequencing was performed by external company either Eurofins Genomics (Germany) or Source Biosciences (UK). When appropriate generic sequencing primers (e.g. T7 promoter primer) provided by external sequencing companies were utilised, otherwise primers were provided. The list of sequencing primers which were provided to external companies are shown in Section 1 of the appendix.

Sequence of the pGEM-OCH1;Man1a;G418, including sequence alignment score
5'OCH1 CAAAGATTGTAGGAGCGAACCCACGATAAAAGTTTCCCAACCTTGCCAAAGTGTCTA ATGCTGTGACTTGAAATCTGGGTTCCTCGTTGAAGACCCTGCGTACTATGCCCAAAA ACTTTCCTCCACGAGCCCTATTAACTTCTCTATGAGTTTCAAATGCCAAACGGACACG GATTAGGTCCAATGGGTAAGTGAAAAACAACAGAGCAAACCCCAGCTAATGAGCCGG CCAGTAACCGTCTTGGAGCTGTTTCATAAGAGTCATTAGGGATCAATAACGTTCTAAT CTGTTCATAACATACAAATTTTATGGCTGCATAGGGAAAAATTCTCAACAGGGTAGCC GAATGACCCTGATATAGACCTGCGACACCATCATACCCATAGATCTGCCTGACAGCC TTAAAGAGCCCGCTAAAAGACCCGGAAAACCGAGAGCACTCTGGATTAGCAGTCTGA AAAAGAATCTTCACTCTGTCTAGTGGAGCAATTAATGTCTTAGCGGCACTTCCTGCTA CTCCGCCAGCTACTCCTGAATAGATCACATACTGCAAAGACTGCTTGTCGATGACCTT GGGGTTATTTAGCTTCAAGGGCAATTTTTGGGACATTTTGGACACAGGAGACTCAGA AACAGACACAGAGCGTTCTGAGTCCTGGTGCTCCTGACGTAGGCCTAGAACAGGAAT TATTGGCTTTATTTGTTTGTCCATTTCATAGGCTTGGGGTAATAGATAGATGACAGAG AAATAGAGAAGACCTAATATTTTTTGTTCATGGCAAATCGCGGGTTCGCGGTCGGGT CACACACGGAGAAGTAATGAGAAGAGCTGGTAATCTGGGGTAAAAGGGTTCAAAAGA AGGTCGCCTGGTAGGGATGCAATACAAGGTTGTCTTGGAGTTTACATTGACCAGATG ATTTGGCTTTTTCTCTGTTCAATTCACATTTTTCAGCGAGAATCGGATTGACGGAGAAA TGGCGGGGTGTGGGGTGGATAGATGGCAGAAATGCTCGCAATCACCGCGAAAGAAA GACTTTATGGAATAGAACTACTGGGTGGTGTAAGGATTACATAGCTAGTCCAATGGA GTCCGTTGGAAAGGTAAGAAGAAGCTAAAACCGGCTAAGTAACTAGGGAAGAATGAT CAGACTTTGATTTGATGAGGTCTGAAAATACTCTGCTGCTTTTTCAGTTGCTTTTTCCC TGCAACCTATCATTTTCCTTTTCATAAGCCTGCCTTTTCTGTTTTCACTTATATGAGTTC CGCCGAGACTTCCCCAAATTCTCTCCTGGAACATTCTCTATCGCTCTCCTTC
TEF p. CCAATCAGGTAATAACTGTCGCCTCTTTTATCTGCCGCACTGCATGAGGTGTCCCCTT AGTGGGAAAGAGTACTGAGCCAACCCTGGAGGACAGCAAGGGAAAAATACCTACAA CTTGCTTCATAATGGTCGTAAAAACAATCCTTGTCGGATATAAGTGTTGTAGACTGTC CCTTATCCTCTGCGATGTTCTTCCTCTCAAAGTTTGCGATTTCTCTCTATCAGAATTGC CATCAAGAGACTCAGGACTAATTTCGCAGTCCCACACGCACTCGTACATGATTGGCT GAAATTTCCCTAAAGAATTTCTTTTTCACGAAAATTTTTTTTTACACAAGATTTTCAGCA GATATAAAATGGAGAGCAGGACCTCCGCTGTGACTCTTCTTTTTTTTCTTTTATTCTCA CTACATACATTTTAGTTATTCGCCAAC
BIP ss TGTTTTTCAACAGACTAAGCGCTGGCAAGCTGCTGGTACCACTCTCCGTGGTCCTGT ACGCCCTTTTCGTGGTAATATTACCTTTACAGAATTCTTTCCACTCCTCCAATGTTTTA GTTAGAGGTGCCGATGATGGGGATCCACTAGTAGCAACA
Man1a

GAGCCCGCTGACGCCACCATCCGTGAGAAGAGGGCAAAGATCAAAGAGATGATGAC
CCATGCTTGGAATAATTATAAACGCTATGCGTGGGGCTTGAACGAACTGAAACCTATA
TCAAAAGAAGGCCATTCAAGCAGTTTGTTTGGCAACATCAAAGGAGCTACAATAGTAG
ATGCCCTGGATACCCTTTTCATTATGGGCATGAAGACTGAATTTCAAGAAGCTAAATC
GTGGATTAAAAAATATTTAGATTTTAATGTGAATGCTGAAGTTTCTGTTTTTGAAGTCA
ACATACGCTTCGTCGGTGGACTGCTGTCAGCCTACTATTTGTCCGGAGAGGAGATAT
TTCGAAAGAAAGCAGTGGAACTTGGGGTAAAATTGCTACCTGCATTTCATACTCCCTC
TGGAATACCTTGGGCATTGCTGAATATGAAAAGTGGGATCGGGCGGAACTGGCCCT
GGGCCTCTGGAGGCAGCAGTATCCTGGCCGAATTTGGAACTCTGCATTTAGAGTTTA
TGCACTTGTCCCACTTATCAGGAGACCCAGTCTTTGCCGAAAAGGTTATGAAAATTCG
AACAGTGTTGAACAAACTGGACAAACCAGAAGGCCTTTATCCTAACTATCTGAACCCC
AGTAGTGGACAGTGGGGTCAACATCATGTGTCGGTTGGAGGACTTGGAGACAGCTTT
TATGAATATTTGCTTAAGGCGTGGTTAATGTCTGACAAGACAGATCTCGAAGCCAAGA
AGATGTATTTTGATGCTGTTCAGGCCATCGAGACTCACTTGATCCGCAAGTCAAGTG
GGGGACTAACGTACATCGCAGAGTGGAAGGGGGGCCTCCTGGAACACAAGATGGG
CCACCTGACGTGCTTTGCAGGAGGCATGTTTGCACTTGGGGCAGATGGAGCTCCGG
AAGCCCGGGCCCAACACTACCTTGAACTCGGAGCTGAAATTGCCCGCACTTGTCATG
AATCTTATAATCGTACATATGTGAAGTTGGGACCGGAAGCGTTTCGATTTGATGGCG
GTGTGGAAGCTATTGCCACGAGGCAAAATGAAAAGTATTACATCTTACGGCCCGAGG
TCATCGAGACATACATGTACATGTGGCGACTGACTCACGACCCCAAGTACAGGACCT
GGGCCTGGGAAGCCGTGGAGGCTCTAGAAAGTCACTGCAGAGTGAACGGAGGCTA
CTCAGGCTTACGGGATGTTTACATTGCCCGTGAGAGTTATGACGATGTCCAGCAAAG
TTTCTTCCTGGCAGAGACACTGAAGTATTTGTACTTGATATTTTCCGATGATGACCTTC
TTCCACTAGAACACTGGATCTTCAACACCGAGGCTCATCCTTTCCCTATACTCCGTGA
ACAGAAGAAGGAAATTGATGGCAAAGAGAAACATGATGAATTGTGACCGCGGGAATT
CGATTT

AOX Term.

GTTTTAGCCTTAGACATGACTGTTCCTCAGTTCAAGTTGGGCACTTACGAGAAGACCG
GTCTTGCTAGATTCTAATCAAGAGGATGTCAGAATGCCATTTGCCTGAGAGATGCAG
GCTTCATTTTTGATACTTTTTTATTTGTAACCTATATAGTATAGGATTTTTTTTGTCATTT
TGTTTCTTCTCGTACGAGCTTGCTCCTGATCAGCCTATCTCGCAGCTGATGAATATCT
TGTGGTAGGGGTTTGGGAAAATCATTCGAGTTTGATGTTTTTCTTGGTATTTCCCACT
CCTCTTCAGAGTACAGAAGATTAAGTGAGACCTTCGTTTGTGCGGATCCCCCACACA
CCATAGCTTCAAAATG

G418

TGGTGTCCTCGTCCAATCAGGTAGCCATCTCTGAAATATCTGGCTCCGTTGCAACTC
CGAACGACCTGCTGGCAACGTAAAATTCTCCGGGGTAAAACTTAAATGTGGAGTAAT
GGAACCAGAAACGTCTCTTCCCTTCTCTCTCCTTCCACCGCCCGTTACCGTCCCTAG
GAAATTTTACTCTGCTGGAGAGCTTCTTCTACGGCCCCCTTGCAGCAATGCTCTTCCC
AGCATTACGTTGCGGGTAAAACGGAGGTCGTGTACCCGACCTAGCAGCCCAGGGAT
GGAAAAGTCCCGGCCGTCGCTGGCAATAATAGCGGGCGGACGCATGTCATGAGATT
ATTGGAAACCACCAGAATCGAATATAAAAGGCGAACACCTTTCCCAATTTTGGTTTCT
CCTGACCCAAAGACTTTAAATTTAATTTATTTGTCCCTATTTCAATCAATTGAACAACTA
TTTCGAAACGAGGAATTCTTATGAGCCATATTCAACGGGAAACGTCTTGCTCGAGGC
CGCGATTAAATTCCAACATGGATGCTGATTTATATGGGTATAAATGGGCTCGCGATAA
TGTCGGGCAATCAGGTGCGACAATCTATCGATTGTATGGGAAGCCCGATGCGCCAG
AGTTGTTTCTGAAACATGGCAAAGGTAGCGTTGCCAATGATGTTACAGATGAGATGG
TCAGACTAAACTGGCTGACGGAATTTATGCCTCTTCCGACCATCAAGCATTTTATCCG
TACTCCTGATGATGCATGGTTACTCACCACTGCGATCCCCGGGAAAACAGCATTCCA
GGTATTAGAAGAATATCCTGATTCAGGTGAAAATATTGTTGATGCGCTGGCAGTGTTC
CTGCGCCGGTTGCATTCGATTCCTGTTTGTAATTGTCCTTTTAACAGCGATCGCGTAT
TTCGTCTCGCTCAGGCGCAATCACGAATGAATAACGGTTTGGTTGATGCGAGTGATT

TTGATGACGAGCGTAATGGCTGGCCTGTTGAACAAGTCTGGAAAGAAATGCATAAGC
TTTTGCCATTCTCACCGGATTCAGTCGTCACTCATGGTGATTTCTCACTTGATAACCTT
ATTTTTGACGAGGGGAAATTAATAGGTTGTATTGATGTTGGACGAGTCGGAATCGCA
GACCGATACCAGGATCTTGCCATCCTATGGAACTGCCTCGGTGAGTTTTCTCCTTCAT
TACAGAAACGGCTTTTTCAAAAATATGGTATTGATAATCCTGATATGAATAAATTGCAG
TTTCATTTGATGCTCGATGAGTTTTTCTAA

3'OCH1

AGGCATACATTTGAAGGAAGTTGGAAGGACTAAAGAAAGCTAGAGTAAAATAGATATA
GCGAGATTAGAGAATGAATACCTTCTTCTAAGCGATCGTCCGTCATCATAGAATATCA
TGGACTGTATAGTTTTTTTTTTGTACATATAATGATTAAACGGTCATCCAACATCTCGT
TGACAGATCTCTCAGTACGCGAAATCCCTGACTATCAAAGCAAGAACCGATGAAGAA
AAAAACAACAGTAACCCAAACACCACAACAAACACTTTATCTTCTCCCCCCCAACACC
AATCATCAAAGAGATGTCGGAACACAAACACCAAGAAGCAAAAACTAACCCCATATAA
AAACATCCTGGTAGATAATGCTGGTAACCCGCTCTCCTTCCATATTCTGGGCTACTTC
ACGAAGTCTGACCGGTCTCAGTTGATCAACATGATCCTCGAAATGGGTGGCAAGCAT
CGTTCCAGACCTGCCTCCTCTGGTAGATGGAGTGTTGTTTTTGACAGGGGATTACAA
GTCTATTGATGAAGATACCCTAAAGCAACTGGGGGACGTTCCAATATACAGAGACTC
CTTCATCTACCAGTGTTTTGTGCACAAGACATCTCTTCCCATTGACACTTTCCGAATTG
ACAAGAACGTCGACTTGGCTCAAGATTTGATCAATAGGGCCCTTCAAGAGTCTGTGG
ATCATGTCACTTCTGCCAGCACAGCTGCAGCTGCTGCTGTTGTTGTCGCTACCAACG
GCCTGTCTTCTAAACCAGACGCTCGTACTAGCAAAATACAGTTCACTCCCGAAGAAG
ATCGTTTTATTCTTGACTTTGTTAGGAGAAATCCTAAACGAAGAAACACACATCAACTG
TACACTGAGCTCGCTCAGCACATGAAAAACCATACGAATCATTCTATCCGCCACAGAT
TTCGTCGTAATCTTTCCGCTCAACTTGATTGGGTTTATGATATCGATCCATTGACCAAC
CAACCTCGAAAAGATGAAAACGGGAACTACATCAAGGTACAAGATCTTCCACAAGGA
ATTCGTGGTCATTATTCTGCCCAAGATGATTACAATTTGTGTTTATCGGTTCAACCTTT
CATTGAATCTGTAGATGAGACAACAGGCCAAGAATTTTTCAAACCTCTGAAAGGTGTA
TTTGATGACTTGGAATCTCGCTTTCCTCACCATACAAAGACTTCCTGGAGAGACAGAT
TCAGAAAGTTTGCCTCTAAATACGGTGTTCGTCAGTACATCGC

Sequence Alignment; 6261/6261(100%)

Sequence of the pJAN-OCH1;nat1, including sequence alignment score

3'OCH1

CGAAAAGTGCCACCTGGGCCTTATCGTTGGAAGAACGTGATGCCCTCTCTGCCCGC
CCGAACGGAGAGATGTTAAAAGAGAAAGTCCCAGGTAAATATGCACAGCAGGTTGTT
TTATGGGAACAATTTACCAACCTGCGCTCCCCCAAATTAATCGACGATATTCTTATTCT
TCCGATCACCAGCTTCAGTCCAGGGATTGGCCACAGTGGAGCTGGAGATTTGAACCA
TCACCTTGCATATATTAGGCATACATTTGAAGGAAGTTGGAAGGACTAAAGAAAGCTA
GAGTAAAATAGATATAGCGAGATTAGAGAATGAATACCTTCTTCTAAGCGATCGTCCG
TCATCATAGAATATCATGGACTGTATAGTTTTTTTTTTGTACATATAATGATTAAACGGT
CATCCAACATCTCGTTGACAGATCTCTCAGTACGCGAAATCCCTGACTATCAAAGCAA
GAACCGATGAAGAAAAAAACAACAGTAACCCAAACACCACAACAAACACTTTATCTTC
TCCCCCCCAACACCAATCATCAAAGAGATGTCGGAACCAAACACCAAGAAGCAAAAA
CTAACCCCATATAAAAACATCCTGGTAGATAATGCTGGTAACC

5'OCH1

GGATCCTAGCTAGTCCAATGGAGTCCGTTGGAAAGGTAAGAAGAAGCTAAAACCGGC
TAAGTAACTAGGGAAGAATGATCAGACTTTGATTTGATGAGGTCTGAAAATACTCTGC
TGCTTTTTCAGTTGCTTTTTCCCTGCAACCTATCATTTTCCTTTTCATAAGCCTGCCTTT

TCTGTTTTCACTTATATGAGTTCCGCCGAGACTTCCCCAAATTCTCTCCTGGAACATT
CTCTATCGCTCTCCTTCCAAGTTGCGCCCCCTGGCACTGCCTAGTAATATTACCACG
CGACTTATATTCAGTTCCACAATTTCCAGTGTTCGTAGCAAATATCATCAGCCATGGC
GAAGGCAGATGGCAGTTTGCTCTACTATAATCCTCACAATCCACCCAGAAGGTATTAC
TTCTACATGGCTATATTCGCCGTTTCTGTCATTTGCGTTTTGTACGGACCCTCACAAC
AATTATCATCTCCAAAAATAGACTATGATCCATTGACGCTCCGATCACTTGATTTGAAG
ACTTTGGAAGCTCCTTCACAGTTGAGTCCAGGCACCGTAGAAGATAATCTTCGAAGA
CAATTGGAGTTTCATTGATCTAACATCCAAAGACGAAA

Sequence Alignment; 1239/1239(100%)

Sequence of the pKANB TefP;Man1a, including sequence alignment score

TEF p.
CCAATCAGGTAATAACTGTCGCCTCTTTTATCTGCCGCACTGCATGAGGTGTCCCCTT
AGTGGGAAAGAGTACTGAGCCAACCCTGGAGGACAGCAAGGGAAAAATACCTACAA
CTTGCTTCATAATGGTCGTAAAAACAATCCTTGTCGGATATAAGTGTTGTAGACTGTC
CCTTATCCTCTGCGATGTTCTTCCTCTCAAAGTTTGCGATTTCTCTCTATCAGAATTGC
CATCAAGAGACTCAGGACTAATTTCGCAGTCCCACACGCACTCGTACATGATTGGCT
GAAATTTCCCTAAAGAATTTCTTTTTCACGAAAATTTTTTTTTACACAAGATTTTCAGCA
GATATAAAATGGAGAGCAGGACCTCCGCTGTGACTCTTCTTTTTTTTCTTTTATTCTCA
CTACATACATTTTAGTTATTCGCCAAC

BIP ss
TGTTTTTCAACAGACTAAGCGCTGGCAAGCTGCTGGTACCACTCTCCGTGGTCCTGT
ACGCCCTTTTCGTGGTAATATTACCTTTACAGAATTCTTTCCACTCCTCCAATGTTTTA
GTTAGAGGTGCCGATGATGGGGATCCACTAGTAGCAACA

Man1a
GAGCCCGCTGACGCCACCATCCGTGAGAAGAGGGCAAAGATCAAAGAGATGATGAC
CCATGCTTGGAATAATTATAAACGCTATGCGTGGGGCTTGAACGAACTGAAACCTATA
TCAAAAGAAGGCCATTCAAGCAGTTTGTTTGGCAACATCAAAGGAGCTACAATAGTAG
ATGCCCTGGATACCCTTTTCATTATGGGCATGAAGACTGAATTTCAAGAAGCTAAATC
GTGGATTAAAAAATATTTAGATTTTAATGTGAATGCTGAAGTTTCTGTTTTTGAAGTCA
ACATACGCTTCGTCGGTGGACTGCTGTCAGCCTACTATTTGTCCGGAGAGGAGATAT
TTCGAAAGAAAGCAGTGGAACTTGGGGTAAAATTGCTACCTGCATTTCATACTCCCTC
TGGAATACCTTGGGCATTGCTGAATATGAAAAGTGGGATCGGGCGGAACTGGCCCT
GGGCCTCTGGAGGCAGCAGTATCCTGGCCGAATTTGGAACTCTGCATTTAGAGTTTA
TGCACTTGTCCCACTTATCAGGAGACCCAGTCTTTGCCGAAAAGGTTATGAAAATTCG
AACAGTGTTGAACAAACTGGACAAACCAGAAGGCCTTTATCCTAACTATCTGAACCCC
AGTAGTGGACAGTGGGGTCAACATCATGTGTCGGTTGGAGGACTTGGAGACAGCTTT
TATGAATATTTGCTTAAGGCGTGGTTAATGTCTGACAAGACAGATCTCGAAGCCAAGA
AGATGTATTTTGATGCTGTTCAGGCCATCGAGACTCACTTGATCCGCAAGTCAAGTG
GGGGACTAACGTACATCGCAGAGTGGAAGGGGGGCCTCCTGGAACACAAGATGGG
CCACCTGACGTGCTTTGCAGGAGGCATGTTTGCACTTGGGGCAGATGGAGCTCCGG
AAGCCCGGGCCCAACACTACCTTGAACTCGGAGCTGAAATTGCCCGCACTTGTCATG
AATCTTATAATCGTACATATGTGAAGTTGGGACCGGAAGCGTTTCGATTTGATGGCG
GTGTGGAAGCTATTGCCACGAGGCAAAATGAAAAGTATTACATCTTACGGCCCGAGG
TCATCGAGACATACATGTACATGTGGCGACTGACTCACGACCCCAAGTACAGGACCT
GGGCCTGGGAAGCCGTGGAGGCTCTAGAAAGTCACTGCAGAGTGAACGGAGGCTA
CTCAGGCTTACGGGATGTTTACATTGCCCGTGAGAGTTATGACGATGTCCAGCAAAG

TTTCTTCCTGGCAGAGACACTGAAGTATTTGTACTTGATATTTTCCGATGATGACCTTC
TTCCACTAGAACACTGGATCTTCAACACCGAGGCTCATCCTTTCCCTATACTCCGTGA
ACAGAAGAAGGAAATTGATGGCAAAGAGAAACATGATGAATTGTGACCGCGGGAATT
CGATTT

Sequence Alignment; 2025/2025(100%)

Sequence information of DNA elements used in Chapter 5 the development of a Cell-Free Protein Synthesis system based on *Pichia pastoris* cell lysate

GFP (720bp)
ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGC
TGGACGGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGA
TGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCG
TGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC
TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTAC
GTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACT
TCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACACCAGCCAC
AACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATC
CGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACAC
CCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCACTACCTGAGCACCCAGT
CCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC
GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA

IRES(CrPV IGR IRES) (195bp)
AAAGCAAAAATGTGATCTTGCTTGTAAATACAATTTTGAGAGGTTAATAAATTACAAGT
AGTGCTATTTTTGTATTTAGGTTAGCTATTTAGCTTTACGTTCCAGGATGCCTAGTGGC
AGCCCCACAATATCCAGGAAGCCCTCTCTGCGGTTTTTCAGATTAGGTAGTCGAAAA
ACCTAAGAAATTTACCTATG

IRES (GRP1) (403bp)
GCCCTCTTGTTTTTCTTTTTCATTTTTTTAAATCACACTTTTTTTTCTTTTCAAGATTTAT
AAATCTTCCTTCCAGGTTTGAGCCGGACCCATTGTTTAATTAAAGCATAGAAGGCAAC
GAAGGAGTGGGTTATAGAATTAGCATTGCGGCAAAAACGTCATATTTATTAATATTTAT
TTGCGAAGGTCATATTCAACGCAATCGCACGTCCTCTTCCGTCTTTTTTCTAGCTTTT
GTCCTTTTTTATTTGTCTACATCCCTTTCTCTACGGTATCCAGAAGAAGGTGTGTGTGT
GTGTGTGTGTGTGTGTCTATAAAAAGCAGTAAGAGTCACCAAAAAAAAAAAACGACAA
ACAAGTGATCCGAAGTGTGACGAATAAAGCAAACTCTCCAACTCAAAATG

ER signal sequence (alpha factor) (264 bp)
AGATTTCCTTCAATTTTTACTGCTGTTTTATTCGCAGCATCCTCCGCATTAGCTGCTCC
AGTCAACACTACAACAGAAGATGAAACGGCACAAATTCCGGCTGAAGCTGTCATCGG
TTACTCAGATTTAGAAGGGGATTTCGATGTTGCTGTTTTGCCATTTTCCAACAGCACA
AATAACGGGTTATTGTTTATAAATACTACTATTGCCAGCATTGCTGCTAAAGAAGAAG
GGGTATCTCTCGAGAAAAGAGAGGCTGAAGCT

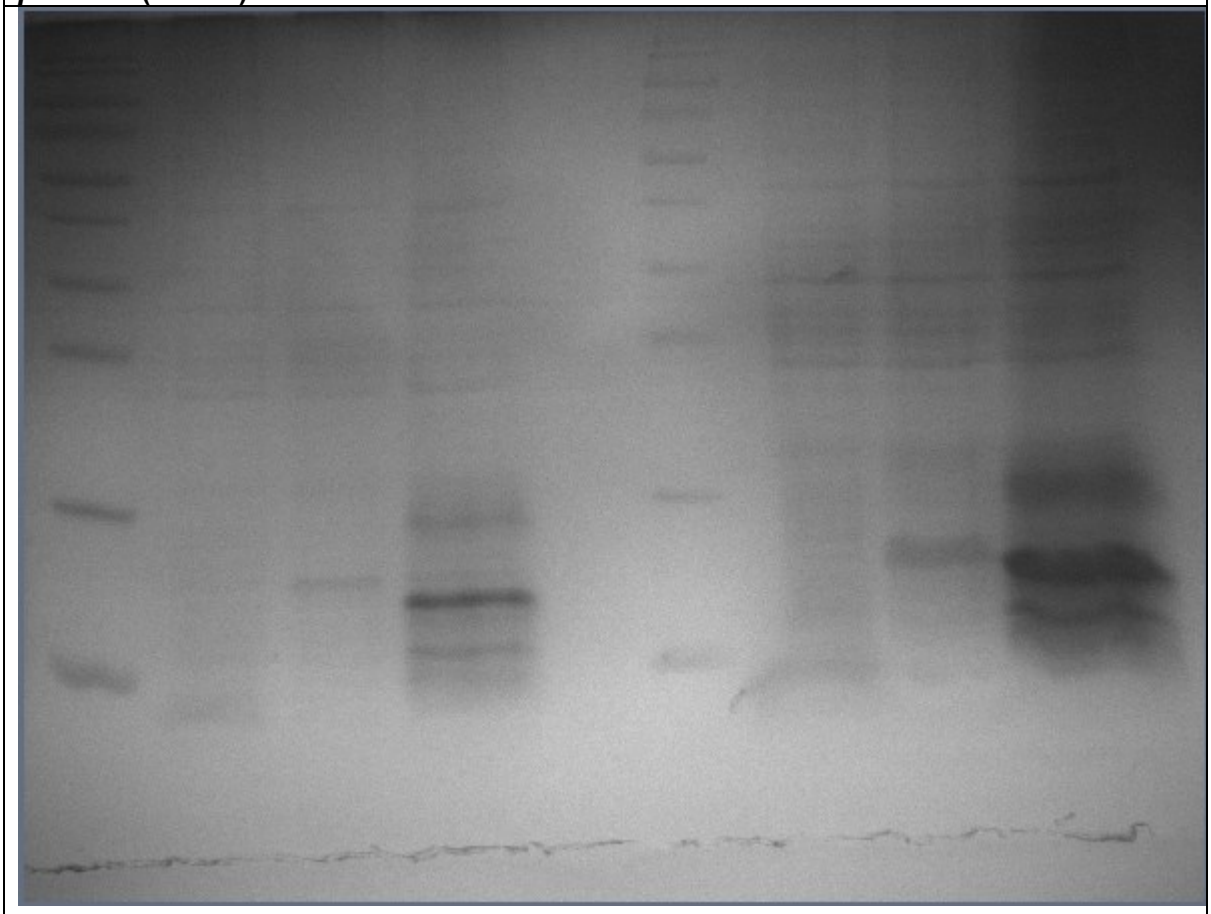
Sequence of the pPICz-Man-II, including sequence alignment score

ATGATCGACCACCTGGAGCGCTTGCTGGCAGAGAATAACGAGATCATCAGCAACATC
CGCGATAGCGTCATCAACCTGTCTGAAAGCGTGGAGGACGGTCCGAAGAGCAGCCA
AAGCAACTTTAGCCAGGGTGCGGGTAGCCACTTGCTGCCGAGCCAGCTGTCTCTGA
GCGTCGATACCGCCGATTGTTTGTTTGCGAGCCAATCGGGCTCCCACAATAGCGACG
TTCAGATGCTGGACGTTTACAGCTTGATCAGCTTTGATAACCCGGATGGTGGTGTTTG
GAAACAGGGCTTCGATATTACCTACGAAAGCAATGAGTGGGACACCGAGCCGCTGC
AAGTTTTTGTGGTCCCGCACAGCCACAATGATCCGGGTTGGCTGAAAACGTTTAACG
ACTACTTTCGCGATAAAACCCAGTATATCTTTAACAACATGGTTCTGAAGCTGAAAGA
GGATAGCCGTCGTAAATTCATTTGGAGCGAGATCAGCTACCTGTCGAAGTGGTGGGA
CATCATTGATATTCAAAAGAAAGACGCAGTTAAAAGCCTGATTGAAAATGGTCAGCTG
GAGATCGTCACGGGTGGCTGGGTGATGCCGGATGAGGCGACCCCGCACTACTTTGC
GCTGATTGATCAGTTGATCGAGGGCCACCAGTGGCTGGAAAACAACATTGGTGTGAA
ACCGCGTAGCGGCTGGGCCATTGACCCGTTTGGTCATAGCCCGACGATGGCCTACC
TGCTGAATCGTGCTGGTCTGTCTCACATGCTGATTCAGCGTGTTCATTACGCAGTGAA
GAAGCATTTCGCACTGCATAAAACCCTGGAGTTCTTTTGGCGCCAAAACTGGGATCT
GGGTTCTGTTACCGACATTTTGTGCCACATGATGCCATTCTACTCCTACGACATCCCG
CACACGTGCGGTCCTGATCCGAAAATCTGTTGCCAATTCGACTTCAAACGCCTGCCG
GGTGGTCGTTTCGGCTGTCCTTGGGGCGTTCCGCCAGAAACTATCCATCCTGGCAAC
GTTCAGAGCCGCGCTCGTATGCTGCTGGATCAGTACCGCAAAAAGAGCAAGCTGTTT
CGCACCAAGGTGCTGCTGGCTCCGTTGGGCGACGACTTCCGCTATTGCGAGTATAC
GGAATGGGATCTGCAATTCAAGAATTACCAGCAGCTGTTTGACTATATGAACAGCCA
GTCTAAATTCAAAGTCAAGATTCAGTTTGGTACGCTGAGCGACTTCTTTGACGCGCTG
GACAAGGCGGACGAAACCCAGCGCGATAAGGGCCAAAGCATGTTTCCAGTGCTGAG
CGGTGATTTCTTTACGTATGCTGATCGTGACGACCACTACTGGAGCGGTTACTTTACC
TCCCGTCCGTTTTACAAACGCATGGACCGTATTATGGAGAGCCACCTGCGTGCAGCG
GAGATTTTGTATTACTTTGCCCTGCGCCAAGCACACAAGTATAAGATCAATAAGTTCC
TGAGCAGCTCCCTGTACACCGCGTTGACCGAGGCGCGTCGTAACCTGGGCTTGTTC
CAACACCATGACGCGATCACCGGCACCGCCAAGGACTGGGTTGTGGTTGATTATGG
TACTCGTCTGTTCCATAGCCTGATGGTTCTGGAGAAGATTATCGGTAATTCTGCTTTT
CTGCTGATCGGTAAGGATAAACTGACCTACGACTCGTACAGCCCGGATACCTTTCTG
GAGATGGATCTGAAGCAGAAATCGCAGGATAGCCTGCCGCAGAAAAACATTATTCGT
TTGAGCGCAGAACCGCGCTACCTGGTCGTGTACAACCCGCTGGAACAAGATCGCAT
CAGCTTGGTGTCCGTGTACGTTTCCAGCCCGACCGTTCAAGTTTTCAGCGCGAGCGG
CAAGCCAGTTGAAGTCCAAGTGAGCGCGGTGTGGGACACGGCCAATACCATCAGCG
AAACCGCCTACGAAATCAGCTTCCGTGCCCATATTCCTCCGCTGGGCCTGAAAGTGT
ACAAGATTCTGGAGAGCGCAAGCTCCAATAGCCATCTGGCAGACTATGTCCTGTACA
AAAACAAAGTCGAAGATAGCGGTATTTTCACGATCAAAAACATGATTAACACCGAAGA
GGGTATTACGCTGGAAAATAGCTTTGTGCTGCTGCGCTTCGATCAGACCGGCCTGAT
GAAACAAATGATGACCAAAGAGGACGGCAAGCATCACGAGGTTAATGTCCAGTTCAG
CTGGTATGGTACGACGATCAAGCGTGACAAGAGCGGTGCGTACTTGTTTCTGCCGG
ATGGTAATGCGAAGCCGTATGTGTATACCACGCCGCCATTCGTCCGTGTCACGCATG
GCCGTATTTACAGCGAGGTGACGTGTTTCTTCGATCATGTGACGCATCGTGTTCGTC
TGTACCACATCCAGGGCATTGAGGGTCAGAGCGTCGAGGTGTCGAATATTGTTGACA
TTCGTAAAGTCTATAATCGTGAGATCGCTATGAAGATCAGCAGCGACATCAAGTCCCA
GAACCGTTTCTATACCGATCTGAACGGTTATCAGATCCAACCTCGTATGACCCTGAG
CAAACTGCCGTTGCAGGCGAATGTTTACCCGATGACCACTATGGCCTATATTCAAGA
TGCCAAGCATCGCCTGACCCTGCTGAGCGCGCAGAGCCTGGGCGTCAGCAGCCTG
AATAGCGGTCAAATTGAAGTCATTATGGACCGTCGTCTGATGCAGGACGACAATCGT
GGTCTGGAGCAAGGTATTCAGGACAACAAGATTACGGCCAATCTGTTCCGCATTCTG
CTGGAGAAACGTTCCGCAGTGAACACGGAAGAAGAGAAGAAAAGCGTTAGCTACCC
GTCTTTGCTGAGCCACATTACTTCCAGCCTGATGAATCACCCGGTGATTCCGATGGC
CAACAAGTTCAGCAGCCCGACGCTGGAACTGCAAGGTGAATTTAGCCCGTTGCAGTC
CAGCCTGCCGTGCGATATTCACCTGGTTAACTTGCGCACCATTCAAAGCAAGGTCGG
TAACGGCCATAGCAACGAGGCGGCATTGATTTTGCACCGTAAAGGTTTCGACTGCCG
TTTCAGCTCCAAGGGCACGGGCCTGTTTTGTTCTACGACCCAAGGTAAGATCCTGGT

GCAGAAACTGCTGAACAAGTTTATTGTGGAGAGCTTGACGCCGTCTAGCCTGTCCTT GATGCACAGCCCTCCGGGTACGCAAAACATTAGCGAAATCAACCTGAGCCCGATGG AGATTTCTACCTTCCGTATTCAACTGCGTTAACTCGAGTAA
Sequence Alignment; 3,285/3,285 (100%)

Section 3 SDS-PAGE assessment of Kringle-3

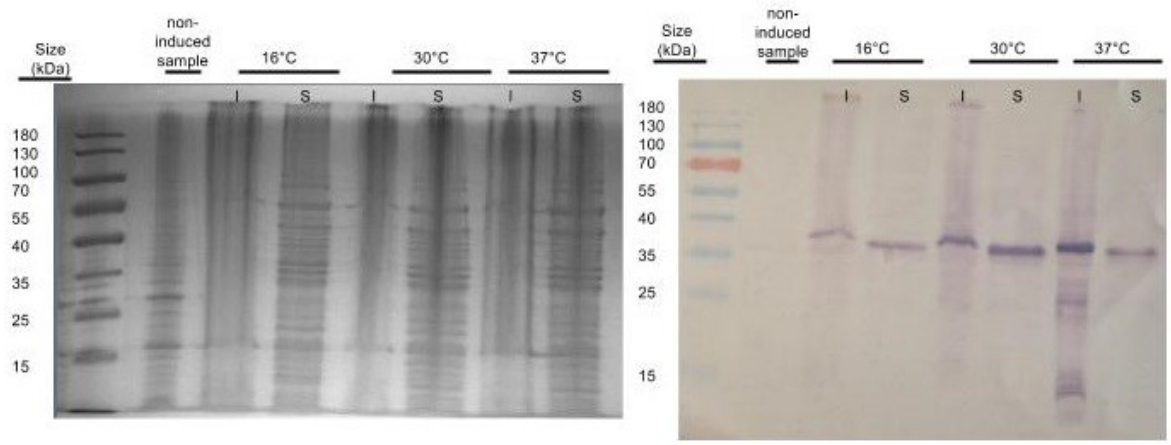
Assessing the protein expression of K3 in wild-type *P. pastoris* (GS115) and *P. pastoris* (*Δoch1*).



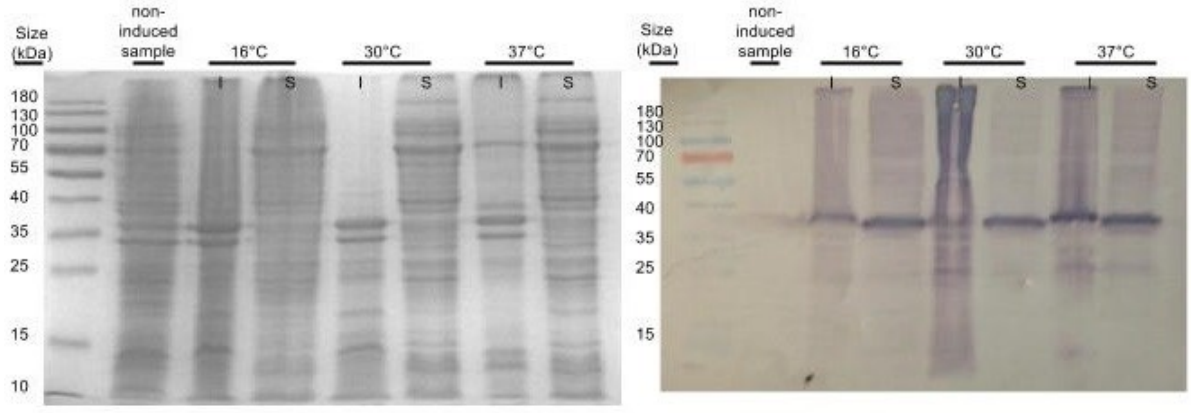
The original unaltered SDS-PAGE of Figure 3.11. Details of the gel can be found within the figure legend of Figure 3.11.

Section 4 SDS-PAGE and Western blot analysis of Golgi Enzymes

Expression of GlcNAcT-I and Man-II in *E. coli* BL21 strain

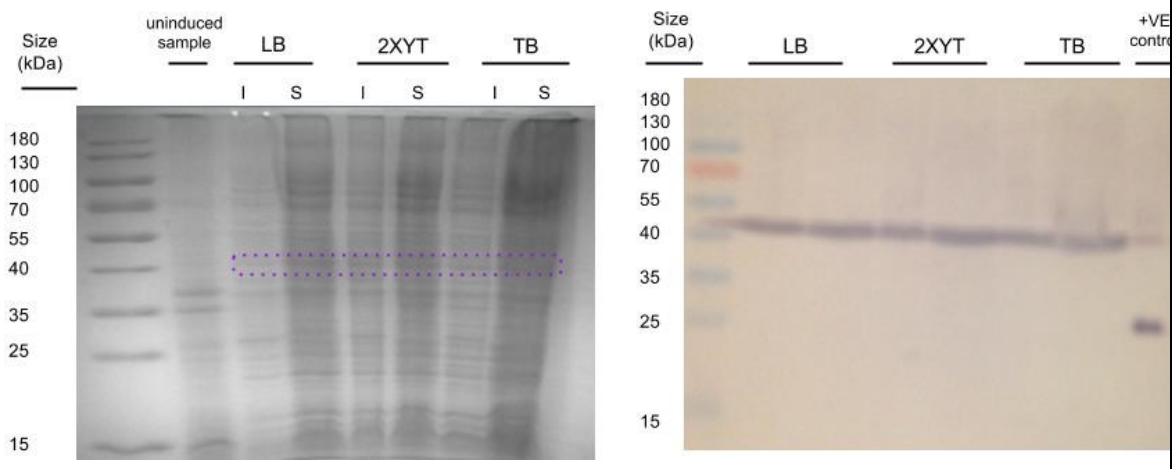


A Coomassie stained SDS-PAGE and Western Blot showing [500μM] IPTG induced expression from *E. coli* strain BL21 containing human GlcNAcT-I recombinant gene in LB media at three temperatures (16°C, 30°C and 37°C). Both the Cell fraction (soluble/insoluble) denoted with the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. No band of expected size 46kDa was present.

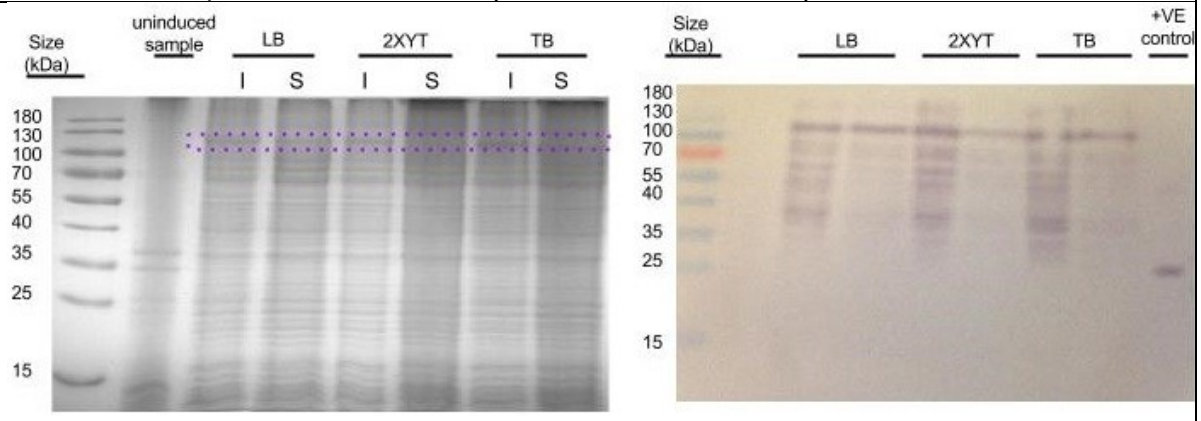


A Coomassie stained SDS-PAGE and Western Blot showing [500μM] IPTG induced expression from *E. coli* strain BL21 containing human Man-II recombinant gene in LB media at three temperatures (16°C, 30°C and 37°C). Both the Cell fraction (soluble/insoluble) denoted with the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. No band of expected size 122kDa was present.

Expression of GlcNAcT-I and Man-II in *E. coli* SHuffle® strain

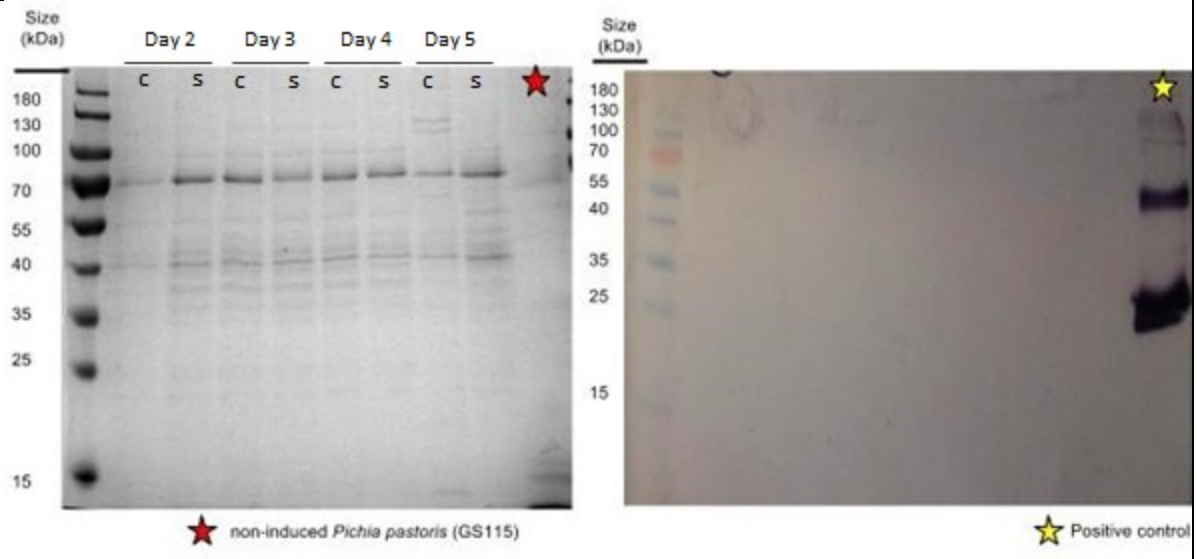


A Coomassie stained SDS-PAGE and Western Blot showing [500µM] IPTG induced expression of Man-II from *E. coli* strain Shuffle®. The expression temperature was 16°C and three culture media were utilised (LB, 2XYT and TB). Both the Cell fraction (soluble/insoluble) denoted with the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. A band of expected size 46kDa was present, in all conditions.



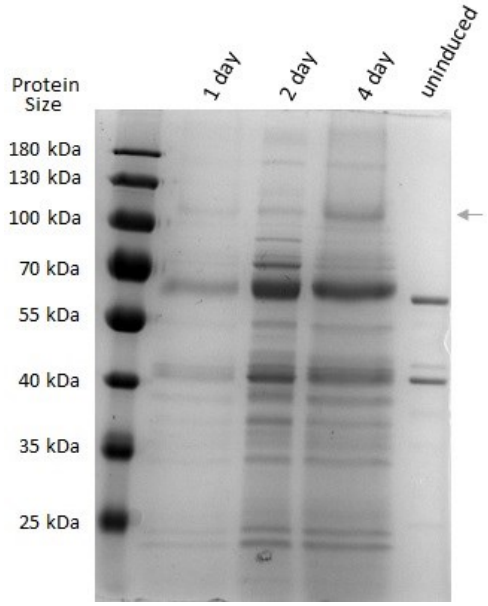
A Coomassie stained SDS-PAGE and Western Blot showing [500µM] IPTG induced expression of Man-II from *E. coli* strain Shuffle®. The expression temperature was 30°C and three culture media were utilised (LB, 2XYT and TB). Both the Cell fraction (soluble/insoluble) denoted with the letter **I**, and the soluble fraction denoted by the letter **S**, were examined for expression. A band of expected size 122kDa was present, in all conditions.

Expression of human Man-II in *P. pastoris GS115* in baffled flasks



A Coomassie stained SDS-PAGE and Western Blot showing expression of Man-II from *P. pastoris (GS115)*. Both the Cell fraction denoted with the letter **C**, and the supernatant denoted by the letter **S**, were examined for expression. No band of expected size 122kDa was present.

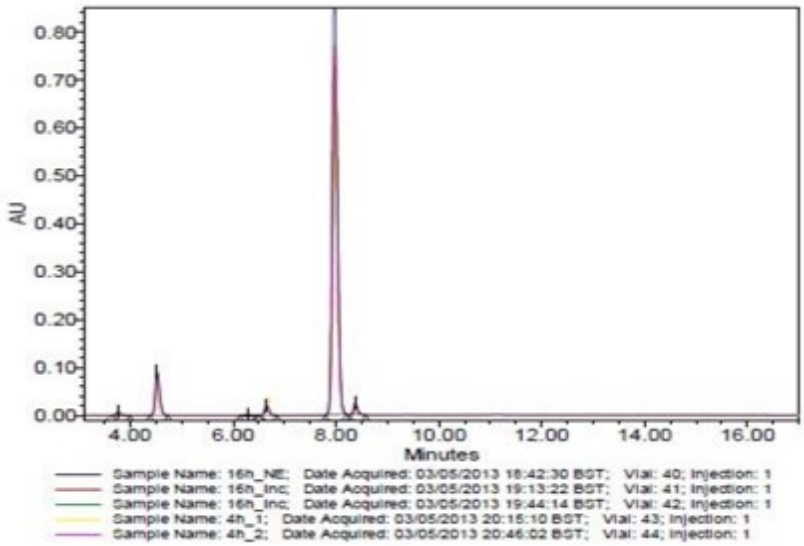
Expression of C. elegans Man-II in *P. pastoris X33*



The *P. pastoris* X33 strain containing the recombinant *C. elegans* Man-II, was induced under optimal growing conditions and in the presence of 0.5% methanol. After 24 hours, 2 day and 4 days, a 50 µL aliquot of the supernatant was removed and protein content was assessed using SDS-PAGE. The supernatant of *P. pastoris* displayed a protein band of 112 kDa, the expected size of ceMan-II. This band was not present within uninduced P. pastoris. Indicating the successful expression of ceMan-II. The SDS-PAGE is courtesy of Dr Kate Royle

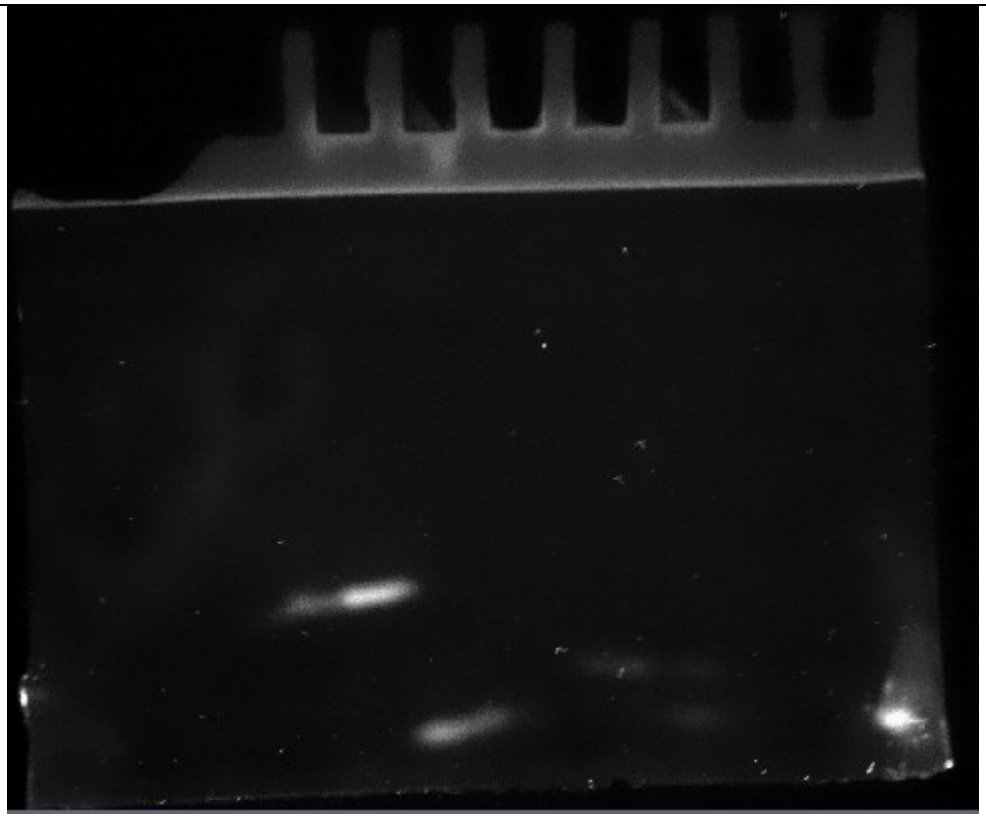
Section 5 GlcNAcT-I activity analysis

The RP-HPLC analysis of UDP and UDP-GlcNAc concentration, from an enzyme reaction containing GlcNAcT-I and a reaction with no enzyme



The RP-HPLC trace generated by Empower 2 software (Waters, UK), utilised to provided data for 'Section 4.6 Assessing the activity of N-acetylglucosaminyltransferase-I'

FACE analysis of $Man_5GlcNAc_2$ with enzyme GlcNAcT-I and donor sugar UDP-GlcNAc



The original FACE image from Figure 4.9

Section 6 Man-II activity analysis

Enzyme activity of purified human Man-II from Section 4.7 against p-nitrophenyl-α-mannopyranoside following the method describe in Section 2.10.3, using differing pH conditions						
	pH condition of reaction					
Substrate Concentration [mM]	pH6.0		pH7.0		pH8.0	
2.5	0.204	0.177	0.172	0.161	0.17	0.314
	0.197	0.161	0.169	0.183	0.15	0.183
5	0.176	0.115	0.155	0.157	0.152	0.152
	0.187	0.186	0.164	0.164	0.161	0.163
10	0.232	0.222	0.178	0.181	0.174	0.172
	0.245	0.238	0.178	0.185	0.192	0.191
0 (Negative control)	0.165	0.165	0.162	0.155	0.171	0.163
Positive control	3.5	3.5	3.5	3.5	2.089	2.106

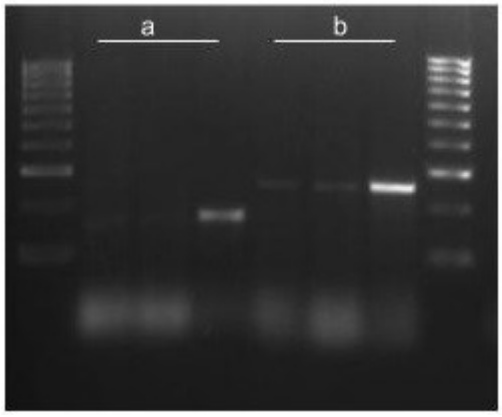
Enzyme activity of purified human Man-II from Section 4.11 against 5 mM of p-nitro-phenyl-α-mannopyranoside following the method described in Section 2.10.3, using differing pH conditions, and divalent ions								
	pH 6.0			pH 7.0			Positive control	Negative control
No divalent ion	0.073	0.076	0.079	0.082	0.07	0.078	2.846	0.053
Cobalt [0.5mM]	0.075	0.074	0.065	0.058	0.057	0.062	1.518	0.051
Zinc [0.5mM]	0.086	0.092	0.079	0.067	0.066	0.075	3.5	0.05

Enzyme activity of purified *C. elegans* Man-II from Section 4.7 against 5 mM of p-nitro-phenyl-α-mannopyranoside following the method described in Section 2.10.3, using differing pH conditions, and divalent ions

	pH 6.0			pH 7.0			Positive control	Negative control
No divalent ion	0.71	0.45	0.36	0.47	0.73	0.44	3.5	0.11
Zinc [0.5mM]	0.49	0.25	0.55	0.58	0.59	0.85	3.5	0.20

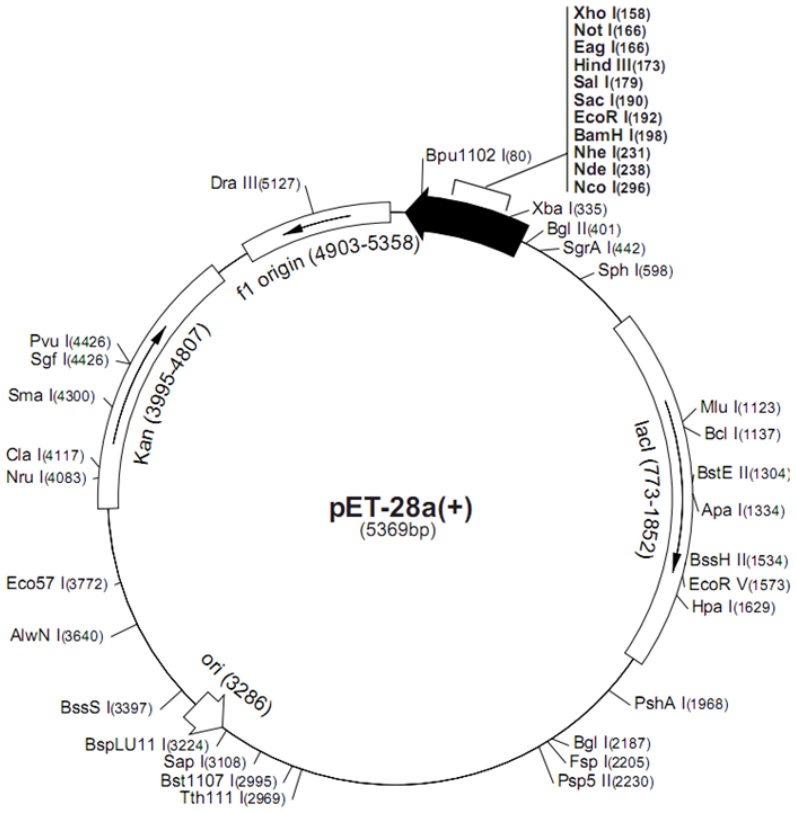
Section 7 RT-PCR of Man-II

Transcript analysis of *Homo sapiens* Man-II (A) and C. elegans Man-II (B) synthesised in *P. pastoris*.



The original image of **Figure 4.12**

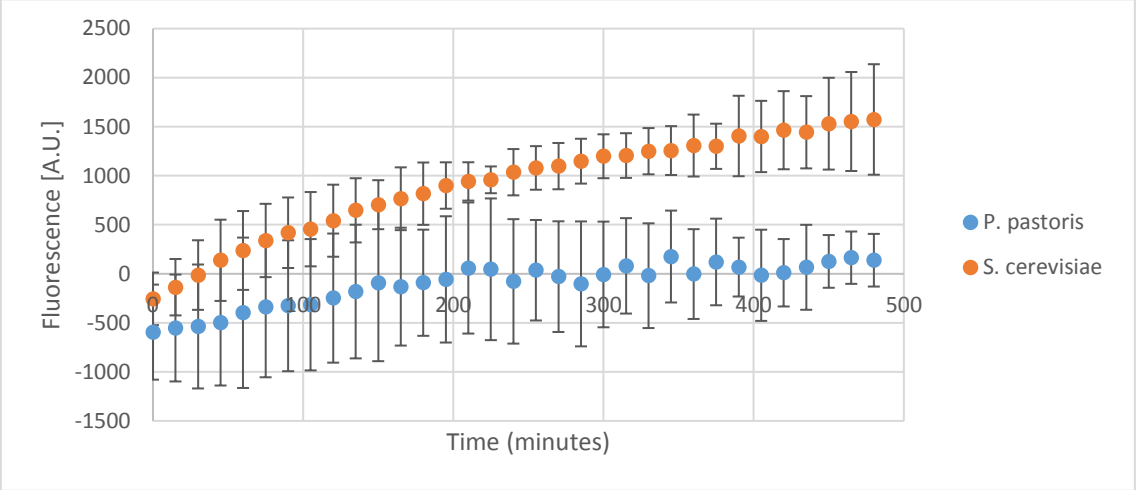
Section 8 pET-28a Vector



The pET-28a (Novagen-Merck, Germany) was utilised throughout plasmid construction in Section 5.2

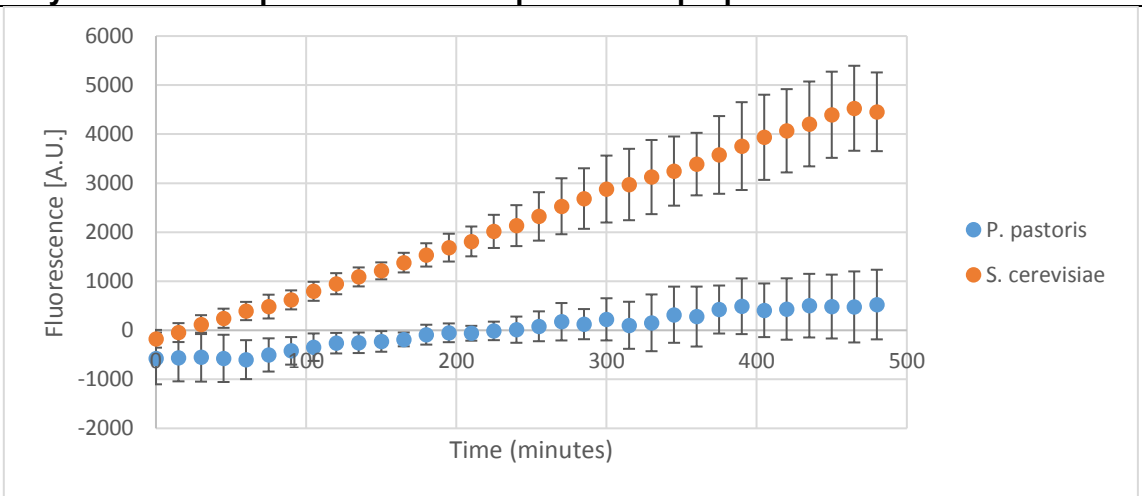
Section 9 Characterisation of GFP expression from *P. pastoris* CFPS and S. cerevisiae CFPS

Analyses of GFP expression from the pCrPV-GFP-pA plasmid within a CFPS



The cell lysate was prepared following the procedure set out in Section 2.9, 25 µL of cell lysate was added to 25 µL of the energy mix from Table 5.3 with 250 ng of the pCrPV-GFP-pA plasmids. A negative control was set-up where no plasmid DNA was added. This was utilised in the analysis to remove the impact of background fluorescence. Experiments were conducted in biological quadruplicate; error bars represent the standard error from the average. The GFP expression was monitored over 8 hours, with fluorescence readings (excitation 485nm / emission 520nm) taken every 15 minutes using a Biotek Synergy HT plate reader, with an incubation temperature of 30°C.

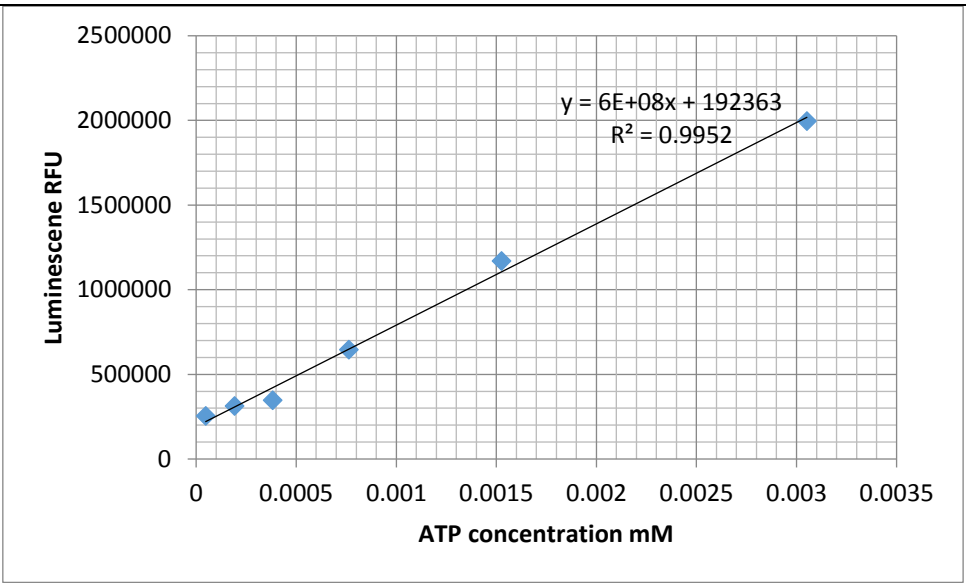
Analyses of GFP expression from the pGPR-GFP-pA plasmid within a CFPS



The cell lysate was prepared following the procedure set out in Section 2.9, 25 µL of cell lysate was added to 25 µL of the energy mix from Table 5.3 with 250 ng of the pCrPV-GFP-pA plasmids. A negative control was set-up where no plasmid DNA was added. This was utilised in the analysis to remove the impact of background fluorescence. Experiments were conducted in biological quadruplicate; error bars represent the standard error from the average. The GFP expression was monitored over 8 hours, with fluorescence readings (excitation 485nm / emission 520nm) taken every 15 minutes using a Biotek Synergy HT plate reader, with an incubation temperature of 30°C.

Section 10 ATP Standard curve produced CellTiter-Glo®

ATP Standard Curve



An ATP standard curve was prepared as according to the of CellTiter-Glo® Manufacturer (Promega, Germany). 100 µL volume of each ATP standard was added to 100 µL CellTiter-Glo® in a microplate, contents were mixed for 2 minutes on an orbital shaker. The microplate was incubated at room temperature for 10 minutes to stabilize the luminescent signal. The luminescence was recorded using Omega plate reader (BMG Labtech, Germany). The relationship between ATP concentration and Luminescence was calculated using Microsoft Excel and data is interrupted in the above line graph. The standard curve was utilised to calculate actual ATP concentrations in Section 5.10

Section 11 Copyright licenses

Copyright license for Figure 1.1

ELSEVIER LICENSE
TERMS AND CONDITIONS

Aug 10, 2017

This Agreement between Mr. Kealan Exley ("You") and Elsevier ("Elsevier") consists of your license details and the terms and conditions provided by Elsevier and Copyright Clearance Center.

License Number	4165290846412
License date	Aug 10, 2017
Licensed Content Publisher	Elsevier
Licensed Content Publication	Seminars in Cell & Developmental Biology
Licensed Content Title	Functional analysis of endoplasmic reticulum glucosyltransferase (UGGT): Synthetic chemistry's initiative in glycobiology
Licensed Content Author	Yukishige Ito,Yoichi Takeda,Akira Seko,Masayuki Izumi,Yasuhiro Kajihara
Licensed Content Date	May 1, 2015
Licensed Content Volume	41
Licensed Content Issue	n/a
Licensed Content Pages	9
Start Page	90
End Page	98
Type of Use	reuse in a thesis/dissertation
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Format	both print and electronic
Are you the author of this Elsevier article?	No
Will you be translating?	No
Original figure numbers	Fig. 1. (A) Overview of glycoprotein folding and quality control process in the ER.
Title of your thesis/dissertation	Towards the cell-free expression of glycoproteins
Expected completion date	Aug 2017
Estimated size (number of pages)	200
Requestor Location	Mr. Kealan Exley 57 cedar road

	Brentwood, Essex cm13 1nd United Kingdom Attn: Mr. Kealan Exley
Publisher Tax ID	GB 494 6272 12
Total	0.00 GBP

Copyright license for Figure 1.2

Copyright © 2015 Vasconcelos-dos-Santos, Oliveira, Lucene, Mantuano, Whelan, Dias and Todeschini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms

Copyright license for Figure 1.5

JOHN WILEY AND SONS LICENSE TERMS AND CONDITIONS

Aug 10, 2017

This Agreement between Mr. Kealan Exley ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4165401380864
License date	Aug 10, 2017
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	FEBS Letters
Licensed Content Title	Engine out of the chassis: Cell-free protein synthesis and its uses
Licensed Content Author	Gabriel Rosenblum, Barry S. Cooperman
Licensed Content Date	Oct 22, 2013
Licensed Content Pages	8
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic

Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 1 Cell-free protein synthesis and its functionalities.
Will you be translating?	No
Title of your thesis / dissertation	Towards the cell-free expression of glycoproteins
Expected completion date	Aug 2017
Expected size (number of pages)	200
Requestor Location	Mr. Kealan Exley 57 cedar road Brentwood, Essex cm13 1nd United Kingdom Attn: Mr. Kealan Exley
Publisher Tax ID	EU826007151
Billing Type	Invoice
Billing Address	Mr. Kealan Exley 57 cedar road Brentwood, United Kingdom cm13 1nd Attn: Mr. Kealan Exley
Total	0.00 USD