The cloning, expression and characterisation of bacterial chitin-binding proteins from *Pseudomonas aeruginosa, Serratia marcescens, Photorhabdus luminescens and Photorhabdus asymbiotica*

Thesis

submitted for the degree of

Doctor of Philosophy

by

Ruth Larragy, B.Sc.

Supervised by

Brendan O'Connor, B.Sc., Ph.D.

and

Michael O'Connell, B.Sc., Ph.D.

School of Biotechnology

Dublin City University

Ireland

April 2011

Declaration

I herby certify that this material, which I now submit for assessment on the programme of study leading to the award of Degree of Doctor of Philosophy, is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No: _____

Date: _____

Acknowledgements

I would like to take this opportunity to thank....

My mother and father for all of their support throughout this process. It would not have been possible without you.

My brother for his words of encouragement from afar - “I know nothing about science so anything you say impresses me”

Paul and Roisin for all of your wisdom and guidance, for giving me a taste of your passion for science and for introducing me to scuba diving so I could forget all of my troubles at the bottom of the sea.

Brendan and Mick for the opportunity to pursue a PhD and for your enthusiasm, for making me feel like everything would work out in the end.

To everyone in the lab past and present, it’s been fun but im glad it’s over. Special thanks to Mary for being there to listen and offer words of encouragement when it was just the two of us.

To Emmet and Sarah, for being there, for understanding, I couldn’t have done it without ye!

Abbreviations

2D	Two dimensional
aa	Amino acids
Amp^r	Ampicillin resistant
APS	Ammonium persulfate
Asn	Asparagine
ATCC	American type culture collection
bp	Base pairs
BSA	Bovine serum albumin
CBM	Carbohydrate-binding module
CBP	Carbohydrate-binding protein
cDNA	Copy DNA
CHO	Chinese hamster ovary
dH_2O	Distilled water
DMSO	Dimethyl sulfoxide
DNS	Dinitrosalicylic acid
DSL	*Datura stramonium* Lectin
ECL	*Erythrina cristagalli* Lectin
EDTA	ethylene diamine tetra acetic acid
ELLA	Enzyme linked lectin assay
EPO	Erythropoietin
ERAD	Endpoplasmic reticulum-associated protein degradation pathway
FPLC	Fast protein liquid chromatography
GFP	Green fluorescent protein
GlcNAc	N-acetyl-D-glucosamine
GSL II	*Griffonia simplicifolia* Lectin II
HPLC	High performance liquid chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
K_d	Dissociation constant
LC	Liquid chromatography
MBP	Maltose-binding protein
MW	Molecular weight

MS	Mass spectrometry
MSC	Multiple cloning site
NCBI	National Centre for Biotechnology Information
NCIMB	National Collection of Industrial, Marine and Food Bacteria
NMR	Nuclear magnetic resonance
NP-HPLC	Normal Phase HPLC
NPL	*Narcissus pseudonarcissus* Lectin
OD	Optical Density
PAGE	Poly-acrylamide gel electrophoresis
PBS	Phosphate buffered saline
Rif^r	Rifampican resistant
RP-HPLC	Reverse Phase HPLC
Ser	Serine
SDS	Sodium dodecyl sulphate
SNA	*Sambucus nigra* Lectin
ss	Signal sequence
SSC	Saline Sodium Citrate
TEMED	Tetramethylethylenediamine
TMB	3,3,5,5-Tetramethylbenzidine
Thr	Threonine
TOF	Time-of-flight
Tris	tris(hydroxymethyl)aminomethane
WGA	Wheat germ agglutinin
WT	Wild type

Table of Contents

1.0	**Introduction**	**1**
1.1	Glycobiology	2
1.2	Glycosylation	2
1.2.1	Eukaryotic N-linked glycosylation	3
1.2.2	N-linked glycosylation quality control	8
1.2.3	Eukaryotic O-linked glycosylation	8
1.2.4	Bacterial glycosylation	11
1.3	Lectins	12
1.4	Significance of glycoforms in eukaryotic systems	14
1.4.1	Significance of terminal residues	14
1.4.2	Carbohydrate-specific receptors	15
1.4.3	Abberant glycosylation	17
1.4.4	Abberant glycosylation in cancer	18
1.5	Glycoprotein analysis and separation	19
1.5.1	Two dimensional (2D) gel electrophoresis	21
1.5.2	Mass Spectrometry (MS)	21
1.5.3	Nuclear Magnetic Resonance (NMR)	22
1.5.4	High Pressure Liquid Chromatography (HPLC)	22
1.5.5	Lectin analysis	23
1.5.6	Carbohydrate arrays	25
1.5.6.1	Problems associated with carbohydrate arrays	26
1.6	Glycosylation – Implications for the biotechnology industry	28
1.7	Novel carbohydrate binding molecules	31
1.8	Chitin	32
1.9	Chitin active/binding enzymes	34
1.10	Chitin Synthases	35
1.11	Glycoside Hydrolases	36
1.11.1	Glycosyl Hydrolases – family 18	36
1.11.2	Glycosyl Hydrolases – family 19	38
1.11.3	Carbohydrate-binding modules	40
1.12	Fungal chitin active enzymes	41
1.13	Plant chitin active enzymes	42

1.14	Invertebrate chitin active enzymes	43
1.15	Mammalian chitin active enzymes	44
1.16	Bacterial chitin active enzymes	47
1.16.1	Chitin-binding proteins utilized in this project	49
1.16.2	The chitinolytic system of *Serratia marcescens*	49
1.16.3	CbpD from *Pseudomonas aeruginosa*	54
1.16.4	Chitin-binding proteins of *Photorhabdus*	56
1.17	Uses of chitin active proteins	57
1.18	Project aims and objectives	59
2.0	**Materials and Methods**	**60**
2.1	Bacterial strains, primers and plasmids	61
2.2	Microbiological Media	67
2.3	Solutions and Buffers	67
2.4	Antibiotics	74
2.5	Storing and culturing of bacteria	75
2.6	Isolation and purification of DNA	75
2.6.1	Isolation of Genomic DNA	75
2.6.2	Isolation of Plasmid DNA	76
2.6.2.1	1-2-3 Method	76
2.6.2.2	Sigma GenElute Plasmid Mini Prep Kit	77
2.7	Agarose gel electrophoresis	77
2.8	Isolation of DNA from agarose gels	78
2.9	Preparation of high efficiency competent cells	79
2.9.1	The RF method	79
2.9.2	The TB method	79
2.9.3	Transformation of competent cells	80
2.9.4	Determination of cell efficiency	80
2.10	Enzymatic reactions	80
2.10.1	Polymerase chain reaction	81
2.10.1.1	Standard PCR reaction mixture	81
2.10.1.2	Standard PCR programme cycle for Phusion *Taq* polymerase reactions	81

2.10.2	Ligation reaction	82
2.10.3	Glycosyl hydrolase reaction	82
2.11	Site specific mutagenesis	82
2.12	DNA sequencing	83
2.12	Insilico analysis of DNA sequences	83
2.14	Standard expression culture	83
2.15	Preparation of cleared lysate for protein purification	84
2.15.1	Cell lysis by sonication	84
2.15.2	Cell lysis using lysozyme	84
2.15.3	Solubilisation of insoluble inclusion bodies	85
2.16	Membrane isolation by water lysis method	85
2.17	Colony blot procedure	86
2.18	Protein purification	87
2.17.1	Standard IMAC procedure	87
2.18.2	IMAC using FPLC and GE nickel resin	88
2.18.3	Purification and on-column re-folding of insoluble $(His)_6$ protein	88
2.18.4	Desalting of purified protein using HiPrep 26/10 desalting column	89
2.19	Recharging of IMAC resin	89
2.20	Protein quantification	90
2.20.1	Quantitative determination of protein by Bradford assay	90
2.20.1	Quantitative determination of protein by 280 nm readings	90
2.21	Quantitative determination of reducing sugars by DNS assay	91
2.22	Lyophilisation	91
2.23	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis	92
2.23.1	Preparation of SDS gels	92
2.23.2	Sample Preparation	92
2.23.3	Sample application	93
2.23.4	Gel analysis	94
2.24	Western blot	95
2.25	Size exclusion chromatpgraphy	96
2.25.1	Superdex 75	96
2.25.2	Toyopearly HW-55S	97

2.26	Insoluble substrate assays	97
2.27	Enzyme linked lectin assay	98
2.28	ElectroSpray ionization mass spectrometry	98
2.29	Immobilisation of protein onto cyanogen bromide activated sepharose	99
3.0	**Cloning, Expression and Characterisation of CBP21**	**101**
3.1	Overview	102
3.2	Cloning and small scale expression of *cbp21* from *Serratia marcescens*	102
3.2.1	Initial cloning of the *S. marcescens* gene *cbp21*	103
3.2.2	Small scale expression of CBP21	107
3.2.3	Sub-cloning and small scale expression of CBP21	109
3.2.4	Cellular compartment analysis of pCBP21_60 lysate	112
3.3	Expression of recombinant CBP21	113
3.3.1	Selection of an *E. coli* expression strain from recombinant CBP21 expression	113
3.3.2	Optimisation of expression conditions for recombinant CBP21	116
3.4	Purification of recombinant CBP21	121
3.4.1	Purification of recombinant CBP21 using IMAC	121
3.4.2	Protein purification using fast protein liquid chromatography	123
3.4.3	Buffer exchange of CBP21	126
3.5	Total yield of recombinant CBP21	128
3.6	Determination of protein size using size exclusion chromatography	128
3.6.1	Validation of CBP21 conformation using size exclusion chromatography	129
3.6.2	Size exclusion chromatography using Superdex 75 column	129
3.6.3	Investigation of protein compatibility with Superdex 75 SEC column	133
3.6.4	Investigation of protein compatibility with Toyopearl HW-55S	136
3.6.5	Investigation of CBP21 compatibility with Toyopearl HW-55S	141
3.7	Mass Spectrometry	142

3.8	Insoluble substrate assays	144
3.9	Discussion	147
4.0	**CBP21 Characterisation**	**153**
4.1	Investigation of recombinant CBP21 affinity for glycoproteins	154
4.1.1	ELLA analysis of CBP21	154
4.1.2	Glycoprotein treatments	156
4.1.3	Glycoprotein affinity chromatography	161
4.1.4	Invertase affinity chromatography	162
4.2	Establishment of a CBP21 activity assay	164
4.2.1	Determining the limit of detection for assay substrates	165
4.2.2	Assay format development	169
4.2.3	Blocking agent optimisation	172
4.2.4	CBP21 characterisation using the $(GlcNAc)_N$-binding assay	177
4.2.5	Validation of the chitin activity assay using wheat germ agglutinin	179
4.3	CBP21 affinity chromatography	181
4.4	Sugar Inhibition studies	187
4.5	CBP21 mutagenesis	191
4.5.1	Site directed mutagenesis of CBP21	193
4.5.2	Expression and purification of CBP21 mutants	200
4.5.3	CBP21 mutant characterisation	204
4.5.4	Random mutagenesis of CBP21	213
4.6	Discussion	217
5.0	**Cloning, Expression, Purification and characterisation of CbpD, CbpA and CbpL**	**225**
5.1	Overview	226
5.2	Initial cloning and small scale characterisation of the CBP21 chitin-binding protein homologues CbpD, CbpA and CbpL	226
5.2.1	Initial cloning of CbpD, CbpA and CbpL	228
5.2.2	Small scale expression of CbpD, CbpA and CbpL	232

5.3	Sub-cloning of CbpD	236
5.3.1	Cloning of untagged CbpD	236
5.3.2	Cloning of CbpD with alternative N-terminal signal sequences	238
5.3.3	Cloning of CbpD sub-domains	244
5.4	Purification of CBP21 homologues	248
5.4.1	Solubilisation, refolding and purification of CbpD, CbpA and CbpL	248
5.4.2	IMAC purification of CbpDb	254
5.5	Protein characterisation	256
5.5.1	Insoluble substrate assays	256
5.5.2	ELLA analysis	261
5.6 `	Discussion	262
6.0	**Final Discussion and Conclusions**	**266**

List of Figures

Chapter 1

Fig 1.1	Structure of the lipid dolichol molecule	4
Fig 1.2	Pre cursor oligosaccharide for N-linked glycosylation	4
Fig 1.3	Biosynthesis of the precursor oligosaccharide molecule for N-linked glycan synthesis	6
Fig 1.4	Processing of the precursor oligosaccharide for N-linked glycosylation to a complex glycan	7
Fig 1.5	H, A and B antigens that form the O, A and B blood group determinants	9
Fig 1.6	Mucin synthesis	10
Fig 1.7	Structure of the mannose receptor	16
Fig 1.8	Structure of the asialoglycoprotein recptor	17
Fig 1.9	Cracking the glycocode: emerging technologies underpinning glycomics	20
Fig 1.10	Enzyme linked lectin assay (ELLA)	24
Fig 1.11	Conduct of microarray experiments exemplified for protein binding	26
Fig 1.12	Chemical structure of chitin	32
Fig 1.13	Structure of α-chitin	33
Fig 1.14	Structure of β-chitin	33
Fig 1.15	Chemical structure of chitosan	34
Fig 1.16	The crystal structure Chtinase A (ChiA) from *Serratia marcescens* in complex with tetra-*N*-acetylchitotriose	37
Fig 1.17	Crystal structure of wheat germ agglutinin (WGA) in complex with GlcNAc	38
Fig 1.18	Proposed hydrolysis mechanism for family 18 and 19 chitinases from the glycoside hydrolase family	39
Fig 1.19	The steady-state model of apical wall growth	42
Fig 1.20	Chitin recycling system employed by microbial organisms	47
Fig 1.21	Crystal structure of CBP21	50

Fig 1.22	Location of surface patch residues of CBP21 thought to be involved in chitin binding	52
Fig 1.23	Multiple sequence alignment of the N-terminal region of family 33 bacterial CBPs	53
Fig 1.24	Multiple sequence alignment of CbpD homologous proteins	55

Chapter 2

Fig 2.1	The pQE30 vector (Qiagen)	65
Fig 2.2	The pQE60 vector	65
Fig 2.3	The pQE_PelB vector	66
Fig 2.4	The pQE_DsbA vector	66
Fig 2.5	1kb Molecular Marker (Invitrogen)	78
Fig 2.6	A; NEB broad range protein marker, B; Pre-stained protein marker and C; Color plus pre-stained protein marker.	94
Fig 2.7	Schematic of Western blot	96

Chapter 3

Fig 3.1	*S. marcescens* BJL200 *cbp21* coding sequence	104
Fig 3.2	Schematic of the cloning of *cbp*21 into commercial pQE expression vectors pQE30 and pQE60	105
Fig 3.3	DNA sequence alignment of the *cbp21* gene from *S.marcescens* BJL200 and *S. marcescens* koln	106
Fig 3.4	Amino acid sequence alignment of the *cbp21* gene from *S. marcescens* BJL200 and *S. marcescens* koln	106
Fig 3.5	Expression analysis of N-terminally $(His)_6$ tagged CBP21 in *E. coli* XL10-Gold	107
Fig 3.6	Expression analysis of C-terminally $(His)_6$ tagged CBP21 in *E. coli* XL10-Gold	108
Fig 3.7	Schematic of Phusion^{TM} whole vector amplification strategy	109
Fig 3.8	Whole vector amplification strategy for the generation of pCBP2130_N	110
Fig 3.9	Whole vector amplification strategy for the generation of pCBP2160_N 109	110

Fig 3.10	Expression analysis of C-terminally $(His)_6$ tagged CBP21 expressed without its native leader sequence in *E. coli* XL10-Gold	111
Fig 3.11	Soluble cell fraction analysis of C-terminally $(His)_6$ tagged CBP21 expressed from pCBP21_60	112
Fig 3.12	Effect of strain selection on CBP21 expression in *E. coli*	114
Fig 3.13	SDS-PAGE analysis of CBP21 expression in *E. coli* strains BL21, JM109, KRX and XL-10 Gold	115
Fig 3.14	Effect of varying IPTG concentration on expression of CBP21 in KRX	117
Fig 3.15	Effect of varying IPTG concentration on the growth rate of *E. coli* KRX	118
Fig 3.16	Effect of variation of growth temperatures and harvest time on CBP21 expression from KRX. A: 37°C incubation, B: 37°C incubation prior to induction, 30°C incubation following induction	119
Fig 3.17	Effect of temperature variation on CBP21 expression in *E. coli* KRX	120
Fig 3.18	Purification of C-terminally tagged CBP21 soluble cell lysate using Amersham Ni-NTA resin	122
Fig 3.19	Purification of CBP21 over HisTrap Crude Column by FPLC	124
Fig 3.20	SDS-PAGE analysis of elution fractions resulting from the purification of CBP21 by FPLC-IMAC	125
Fig 3.21	FPLC trace of CBP21 desalting and buffer exchange using the HiPrep 26/10 desalting column (GE Healthcare). A; CBP21 desalting into dH_2O, B; CBP21 – desalting and buffer exchange into PBS (Section 2.3)	127
Fig 3.22	Schematic view through a section of a bead of superdex	130
Fig 3.23	Determination of Superdex 75 total column and void volumes	130
Fig 3.24	Elution of protein standards from the Superdex 75 size exclusion chromatography column	131
Fig 3.25	Development of size exclusion chromatography standard curve using the Superdex 75 SEC column	132

Fig 3.26	The adsorption A_{280nm} of CBP21 through the Superdex 75 column in the presence of PBS and PBS with 0.5 M NaCl	134
Fig 3.27	The adsorption A_{280nm} of CBP21 through the Superdex 75 column in the presence of galactose, glucose and GlcNAc	135
Fig 3.28	Polymeric structure of methyl methracylate	137
Fig 3.29	Determination of Toyopearl HW-55S column efficiency	137
Fig 3.30	Elution of protein standards from the Tyopearl HW-55S size exclusion chromatography column	139
Fig 3.31	Development of size exclusion chromatography standard curve using the Toyopearl HW-55S SEC column	140
Fig 3.32	Elution profile of CBP21 in PBS pH 7.2 from the Toyopearl -55S SEC matrix	141
Fig 3.33	Manually de-convoluted mass spectrum of CBP21 in 50% ACN, 50% H_2O and 0.2% formic acid, using the nano flow Z-spray source	143
Fig 3.34	Efficiency of binding of CBP21 to various insoluble substrates A; Crustacean shell chitin, B; Avicel, C; Chitosan from crustacean shells	145
Fig 3.35	Efficiency of binding of CBP21 to various insoluble substrates A: squid pen chitin, B; chitosan from carapacea skin	146
Fig 3.36	A 3-D structure of the theorized active site of CBP21	149

Chapter 4

Fig 4.1	Comparison of CBP21 and GSLII ELLA binding profiles	155
Fig 4.2	Structures of the two most abundant bi-antennary *N*-linked glycans present in human transferrin	157
Fig 4.3	Structure of unit-B type glycan structure present on porcine thyroglobulin	157
Fig 4.4	Glycoprotein terminal structures following glycosidase treatments.	158
Fig 4.5	Quantitative detection of carbohydrate-binding protein interaction with transferrin and thyroglobulin	159

Fig 4.6	β-N-acetylglucosaminidase and mannosidase treatment of invertase	160
Fig 4.7	Reaction activation scheme of sepharose by cyanogen bromide and protein coupling through primary amines	161
Fig 4.8	Application of carbohydrate-binding proteins to sepharose bound invertase	163
Fig 4.9	The general structure of PAA-linked multivalent biotinylated polymers	165
Fig 4.10	Determination of the limit of detection for CBP21	166
Fig 4.11	Calculating the limit of detection for chitobiose	167
Fig 4.12	Calculating the limit of detection for chitotriose	168
Fig 4.13	Potential assay formats for the detection of CBP21 interaction with PAA-linked $(GlcNAc)_N$ oligos	170
Fig 4.14	Development of CBP21 – $(GlcNAc)_N$ Binding assay	171
Fig 4.15	Comparison of blocking reagents	174
Fig 4.16	Comparison of blocking reagents	175
Fig 4.17	Comparison of blocking reagents using an immobilised biotin labelled lectin	176
Fig 4.18	Quantitative detection of biotinylalted PAA-linked $(GlcNAc)_N$ oligos bound to immobilised CBP21, using assay format A	177
Fig 4.19	Detection of binding of $(His)_6$ tagged CBP21 to immobilised PAA-linked $(GlcNAc)_N$ oligos, using assay format B	178
Fig 4.20	Binding of immobilised unlabelled wheat germ agglutinin to immobilised PAA-linked un-conjugated chitobiose	180
Fig 4.21	Binding of biotinylated wheat germ agglutinin to immobilised PAA-linked un-conjugated chitobiose	180
Fig 4.22	Release and isolation of PNGase F liberated glycans using a C-20 column	182
Fig 4.23	Location of surface exposed lysine and asparagine residues on the CBP21 molecule	184
Fig 4.24	Determining the limit of detection of un-conjugated chitobiose using the DNS assay for reducing sugars	185
Fig 4.25	Analysis of chitobiose bound to sepharose immobilised CBP21	186

Fig 4.26	Analysis of CBP21 affinity column bound chitobiose by boiling and DNS assay	186
Fig 4.27	Analysis of PNGase F released glycans by modified ELLA	187
Fig 4.28	Sugar inhibition study of CBP21 using biotin labelled PAA-linked chitobiose	189
Fig 4.29	Sugar inhibition study of CBP21	190
Fig 4.30	Application of CBP21 to mannan affinity agarose	191
Fig 4.31	Structure of CBP21 with residues thought to be involved in chitin binding	192
Fig 4.32	Schematic of Phusion^{TM} site directed mutagenesis	193
Fig 4.33	Outline of construction of $CBP21_{Y54A}$	194
Fig 4.34	Outline of construction of $CBP21_{E55A}$	194
Fig 4.35	Outline of construction of $CBP21_{P56A}$	195
Fig 4.36	Outline of construction of $CBP21_{Q57A}$	195
Fig 4.37	Outline of construction of $CBP21_{S58A}$	196
Fig 4.38	Outline of construction of $CBP21_{E60A}$	196
Fig 4.39	Outline of construction of $CBP21_{T111A}$	197
Fig 4.40	Outline of construction of $CBP21_{H114A}$	197
Fig 4.41	Outline of construction of $CBP21_{D182A}$	198
Fig 4.42	Structure of CBP21	199
Fig 4.43	Purification of C-terminally tagged CBP21 mutants by IMAC	201
Fig 4.44	Purification of C-terminally tagged CBP21 mutants by IMAC	202
Fig 4.45	Purification of C-terminally tagged CBP21 mutants by IMAC	203
Fig 4.46	CBP21 solid substrate assays using β-chitin from squid pen	206
Fig 4.47	CBP21 solid substrate assays using α-chitin from crustacean shell	207
Fig 4.48	CBP21 solid substrate assays using β-chitosan from carapacea skin	208
Fig 4.49	CBP21 solid substrate assays using α-chitosan from crustacean shell	209
Fig 5.50	CBP21 solid substrate assays using crystalline cellulose	210
Fig 4.51	ELLA analysis of CBP21 mutants using glycosidase treated thyroglobulin as a glycoprotein source	212

Fig 4.52	Outline of the construction of $CBP21_{E55A}$ random mutants	214
Fig 4.53	Outline of the construction of $CBP21_{H114A}$ random mutants	215
Fig 4.54	ELLA analysis of $CBP21_{E55A}$ random mutants bound to asialo agalacto thyroglobulin	215
Fig 4.55	ELLA analysis of $CBP21_{E55A}$ random mutants G3, G10, B1 and B10	216
Fig 4.56	ELLA analysis of $CBP21_{H114A}$ random mutagenesis using asialo agalacto thyroglobulin	216
Fig 4.57	Reduced detection of CBP21 by anti-His antibody caused by polymer based blocking agents PVA and PVP	219
Fig 4.58	Immobilisation of a $(His)_6$ tagged protein onto a NUNC Immobilizer™ Ni-chelate plate	221
Fig 4.59	Comparison of Cyanogen bromide and covalent coupling of protein molecules to an agarose support column	223

Chapter 5

Fig 5.1	Sequence alignment of CBP21 with homologues CbpD, CbpA and CbpL	227
Fig 5.2	*P. aeruginosa cbpd* coding sequence	229
Fig 5.3	*P. asymbiotica cbpA* coding sequence	230
Fig 5.4	*P. luminescens cbpL* coding sequence	230
Fig 5.5	Outline of the cloning strategy for the genes encoding CbpD, CbpA and CbpL	231
Fig 5.6	Expression of C-terminally $(His)_6$ tagged CbpD in *E. coli* XL-10 Gold	233
Fig 5.7	Time course expression analysis of C-terminally $(His)_6$ tagged CbpA in *E. coli* XL-10 Gold	233
Fig 5.8	Time course expression analysis of C-terminally (His)6 tagged CbpL in *E. coli* XL-10 Gold	234
Fig 5.9	Western blot analysis of C-terminally $(His)_6$ tagged CbpA and CbpL	234
Fig 5.10	Analysis of the expression of CbpD under varying temperature and induction conditions	235

Fig 5.11	Cloning of untagged CbpD	237
Fig 5.12	Analysis of the expression of untagged CbpD	238
Fig 5.13	Schematic representation of single peptide interactions of the Sec system	240
Fig 5.14	Schematic of the cloning of CbpD into pPelB and pDsbA	241
Fig 5.15	Amino acid alignment of the N-terminal domains of CbpD with WT, PelB and DsbA signal sequences	242
Fig 5.16	Time course expression analysis of CbpD expressed with PelBss	242
Fig 5.17	Time course expression analysis of CbpD expressed with DsbAss	243
Fig 5.18	The cloning of CbpD sequences CbpDa, CbpDb and CbpDc	245
Fig 5.19	Amino acid alignment of CbpD and sub-clones CbpDa, CbpDb, And CbpDc	246
Fig 5.20	Western blot analysis of CbpDa, CbpDb and CbpDc expressed in *E. coli* XL-10 Gold	247
Fig 5.21	Solubilisation of inclusion bodies from a pellet of *E. coli* cells expressing CbpD	250
Fig 5.22	On-column refolding and purification of CbpD	250
Fig 5.23	On-column refolding and purification of CbpA and CbpL	251
Fig 5.24	Analysis of fractions from the insoluble purification of CbpD, CbpA and CbpL	252
Fig 5.25	Analysis of the elution fractions from purification of CbpD from insoluble cell pellet	253
Fig 5.26	On-column refolding and purification of CbpD	253
Fig 5.27	IMAC purification optimisation of CbpDb from soluble *E. coli* cell lysate	255
Fig 5.28	Analysis of CbpD bound to α-chitin, α-chitosan and crystalline cellulose	257
Fig 5.29	Analysis of CbpDb bound to α-chitin, α-chitosan and Cellulose	258
Fig 5.30	Analysis of CbpA bound to β-chitin, α-chitin, Chitosan from carpacea skin, α-chitosan and Cellulose	259

Fig 5.31	Analysis of CbpL bound to β-chitin, α-chitin, Chitosan from carpacea skin, α-chitosan and Cellulose	260
Fig 5.32	Quantitative detection of CbpD, CbpA and CbpL interaction with thyroglobulin	261

List of Tables

Chapter 1

Table 1.1	Lectin-carbohydrate complexes	13
Table 1.2	Chitin active proteins and associated human diseases	46
Table 1.3	Binding preferences of family 33 non-hydrolytic bacterial chitin binding proteins	48
Table 1.4	Binding of CBP21 mutants to β-chitin	51

Chapter 2

Table 2.1	Bacterial strains	61
Table 2.2	Plasmids	62
Table 2.3	Primer sequences	63
Table 2.4	Incubation of nitrocellulose membrane for in situ lysis	87
Table 2.5	Preparation of SDS gels	92
Table 2.6	Silver staining of SDS-PAGE Gels	95

Chapter 3

Table 3.1	Construction of a protein molecular weight standard curve for the Superdex 75 SEC column	132
Table 3.2	Molecular mass calculation using the Superdex 75 SEC column	134
Table 3.3	CBP21 molecular mass calculations under sugar replete conditions, using the Superdex 75 SEC column	135
Table 3.4	Construction of a protein molecular weight standard curve for the Tyopearl HW-55S SEC column	140

Chapter 4

Table 4.1	Immobilisation of invertase to cyanogen bromide activated sepharose 4B	163
Table 4.2	Final procedure for PAA-linkede $(GlcNAc)_N$ assay	181
Table 4.3	Immobilisation of CBP21 onto cyanogen bromide activated sepharose 4B	184

Table 4.4 Average yield of recombinant CBP21 mutant varients from a 500 mL expression in LB medium 200

Table 4.5 Overview of the marked changes observed between the CBP21 wild type protein and its mutants 211

Chapter 5.0

Table 5.1 Relative molecular masses of chitin-binding proteins CbpD, CbpA and CbpL 232

Table 5.2 Properties of CbpD sub-clones CbpDa, CbpDb and CbpDc 246

Table 5.3 Average yield of recombinant proteins CbpD, CbpA and CbpL following expression from a 400 mL *E. coli* culture and on-column refolding 254

Abstract

It is well recognised that most proteins are subject to post translational modifications and that these modifications can have specific effects on the biological properties and functions of these proteins. The majority of proteins secreted by cells are modified by the attachment of oligosaccharide chains. This glycosylation event has been shown to impact correct protein folding, protein stability, solubility, to aid in cell recognition and to help regulate cell processes. In order to gain a deeper understanding into the impact of altered glycosylation patterns on cellular processes and cell recognition it is necessary to develop new technologies to profile the glycan species displayed on the surface of protein molecules.

The present study was dedicated to the development of prokaryotic chitin-binding proteins as novel carbohydrate-binding molecules. Prokaryotic chitin-binding proteins from *Serratia marcescens, Pseudomonas aeruginosa, Photorhabdus asymbiotica* and *Photorhabdus luminescens* were cloned, over-expressed in *E. coli* and purified to homogeneity via $(His)_6$ affinity tags. The activity and specificity of these proteins was tested using a number of insoluble substrates; chitin, chitosan and crystalline cellulose. The ability of these proteins to bind to protein linked glycans was tested using Enzyme linked lectin assays (ELLAs). None of the proteins exhibited any ability to bind glycoproteins in this assay format. A novel N-acetylglucosamine binding assay was developed using CBP21 and the ability to immobilise active CBP21 on a sepharose surface was also demonstrated. Sugar inhibition studies indicated that CBP21 may be capable of binding to mannan and galactan polymers. A site-directed mutagenesis of CBP21 was carried out on the putative binding domain residues to alter the affinity of CBP21. Residues Y54, E55, P56, Q57, S58, E60, T111, H114 and D182 were mutated to alanine, expressed, purified and characterised. The mutation H114A was shown to negatively impact on β-chitin affinity, the Q57A mutant had an increased affinity for chitosan with the proteins Y54A, T111A and D182A displaying an increased affinity for Cellulose. Furthermore it was shown that the putative C-terminal binding domain of CbpD is a chitin-binding domain and that the putative chitin-binding proteins CbpA and CbpL are capable of binding to both α- and β-chitin.

1.0 Introduction

1.1 Glycobiology

The term glycobiology was coined in 1988 to recognize the coming together of the traditional disciplines of carbohydrate chemistry and biochemistry (Rademacher *et al.,* 1988). Glycobiology is defined in the broadest sense as the study of the structure, biosynthesis, and biology of saccharides (sugar chains or glycans) that are widely distributed in nature. Glycobiology has come to the forefront of the community in the past number of years with the discovery that many proteins, especially those destined for secretion or insertion into cell membranes are modified by the attachment of carbohydrates in a process known as glycosylation.

1.2 Glycosylation

Glycosylation is most often classified as a form of post translational modification, where proteins are adapted by addition of carbohydrate chains. Among the different types of covalent modifications that proteins can undergo, none are as diverse as glycosylation. Unlike proteins glycans are not encoded on a DNA or RNA-like sequence and monosaccharides have the capacity to combine with each other in a variety of ways that differ not only in the sequence and chain length but also in anomery, the position of linkages and branching points (i.e. not template driven). Further structural diversity can arise from the attachment of sulphate, phosphate, acetyl or methyl groups to the sugars. In the words of Winterburn and Phelps (1972) 'carbohydrates are ideal for generating compact units with explicit informational properties, since the permutations on linkages are larger than can be achieved by amino acids, and, uniquely in biological polymers, branching is possible'. A staggering possible number of isomer permutations for a hexamer with an alphabet of just 20 letters have been calculated as 1.44×10^{15} (Laine, 1997).

The glycosylation process takes place in the endoplasmic reticulum of eukaryotic cells, giving rise to protein-carbohydrate conjugates, or glycoproteins. The most common forms of glycosylation take place through N- or O-linkages on the polypeptide chain (see section 1.2.1 and 1.2.3). From analysis of the SWISS-PROT database Apweiler *et al.,* (1999) have predicted that more than half of all eukaryotic proteins are glycosylated, up to 90% being N-linked with the remainder decorated

with O-linkages. The nature of the glycan attached to an individual glycoprotein is dependent on the protein and the cell in which it is expressed. The structure of the glycoprotein can also influence the glycosylation pattern, the position of certain glycosylation sites within some proteins favouring highly branched structures while other sites remain decorated with low complexity structures. An important challenge in the field of glycobiology is to understand why eukaryotes are endowed with such extensive glycosylation mechanisms and the effects these glycosylation patterns exert.

1.2.1 Eukaryotic N-linked glycosylation

The biosynthesis of N-linked glycans in eukaryotes is a task shared by the endoplasmic reticulum (ER) and the Golgi apparatus. N-linked Glycans are formed by a series of complex pathways, beginning with the en bloc attachment of a pre-synthesised core oligomer (see Fig 1.2) to the nitrogen of an asparagine residue within a tripeptide consensus sequence, Asn-X-Ser/Thr, where X can be any amino acid except proline. The 14-unit core glycan structure is first assembled on a lipid molecule, dolichol phosphate (see Fig 1.1), at the membrane of the ER (see Fig 1.3), for review see Burda and Aebi (1999**)**. An enzyme known as oligosaccharyl transferase (OT) transfers the lipid linked oligosaccharide to the amide nitrogen of an Asn within the consensus sequence of a growing peptide chain. This co-translational event takes place within the lumen of the ER. The attached oligosaccharide subsequently undergoes a number of enzymatic modifications as the protein moves through the ER. Beginning with the removal of the 3 glucose residues, by glycosidases, the mannose structure is trimmed with the removal of some or all of the four mannose residues in α1-2 linkages until the core GlcNAc unit with attached mannose residues remain (see Fig 1.4).

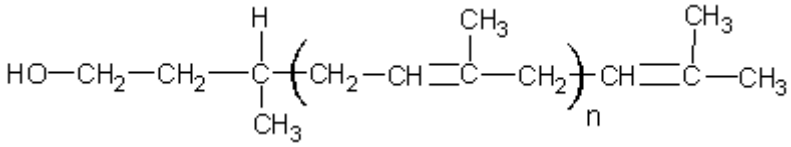


Fig 1.1: Structure of the lipid dolichol molecule. The precursor oligosaccharide is linked by a phosphoryl group to dolichol. It is a highly hydrophic molecule, long enough (75-95 carbon atoms) to span the ER membrane 3-4 times. Image produced using Chemsketch.

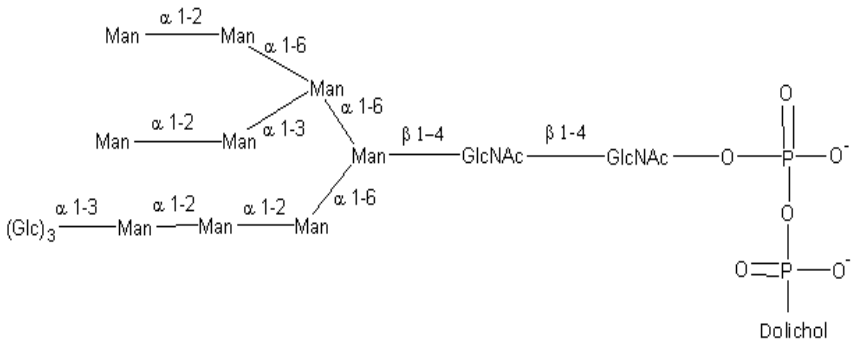


Fig 1.2: Pre cursor oligosaccharide for N-linked glycosylation. Schematic of the dolichol lipid linked oligosaccharide common to all N-linked glycans. Image produced using Chemsketch.

This assembly of the oligosaccharide precursor and the initial processing of the oligosaccharide in the ER are similar in almost all eukaryotes but the later stages of glycosylation may differ. Once the glycoproteins have folded and oligomerized properly they move into the golgi apparatus, where some glycans remain in the mannosidase treated state while others are processed into more complex forms (see Fig 1.4). In mammals the branched structures may be re-elongated with the attachment of a GlcNAc residue to the 1-3 arm of the core, with further extensions including the addition of galactose, either as a single residue at the end of a branch or as a longer polylactosamine chain. A variety of terminal elaborations can then be added to this chain, which acts as a scaffold. These elaborations can include the simple addition of sialic acid through a number of linkages or additionally the

attachment of sulphate groups. Bi-antennary glycans are most common but rarer glycans containing up to 5 or more have been documented. Complex-type plant N-linked glycans can be characterised by the absence of sialic acid and the presence of fucose and/or xylose residues linked respectively to the proximal *N*-acetylglucosamine and to the β-mannose residues of the core (Lerouge, *et al.*, 1998). Yeast N-glycan maturation invariably involves the addition of further mannose residues directly to the core, termed core maturation, or in some instances outer chains of variable sizes may also be added. The majority of these glycoproteins in yeasts are incorporated into the rigid cell wall structure as mannans (Herscovics and Orlean, 1993).

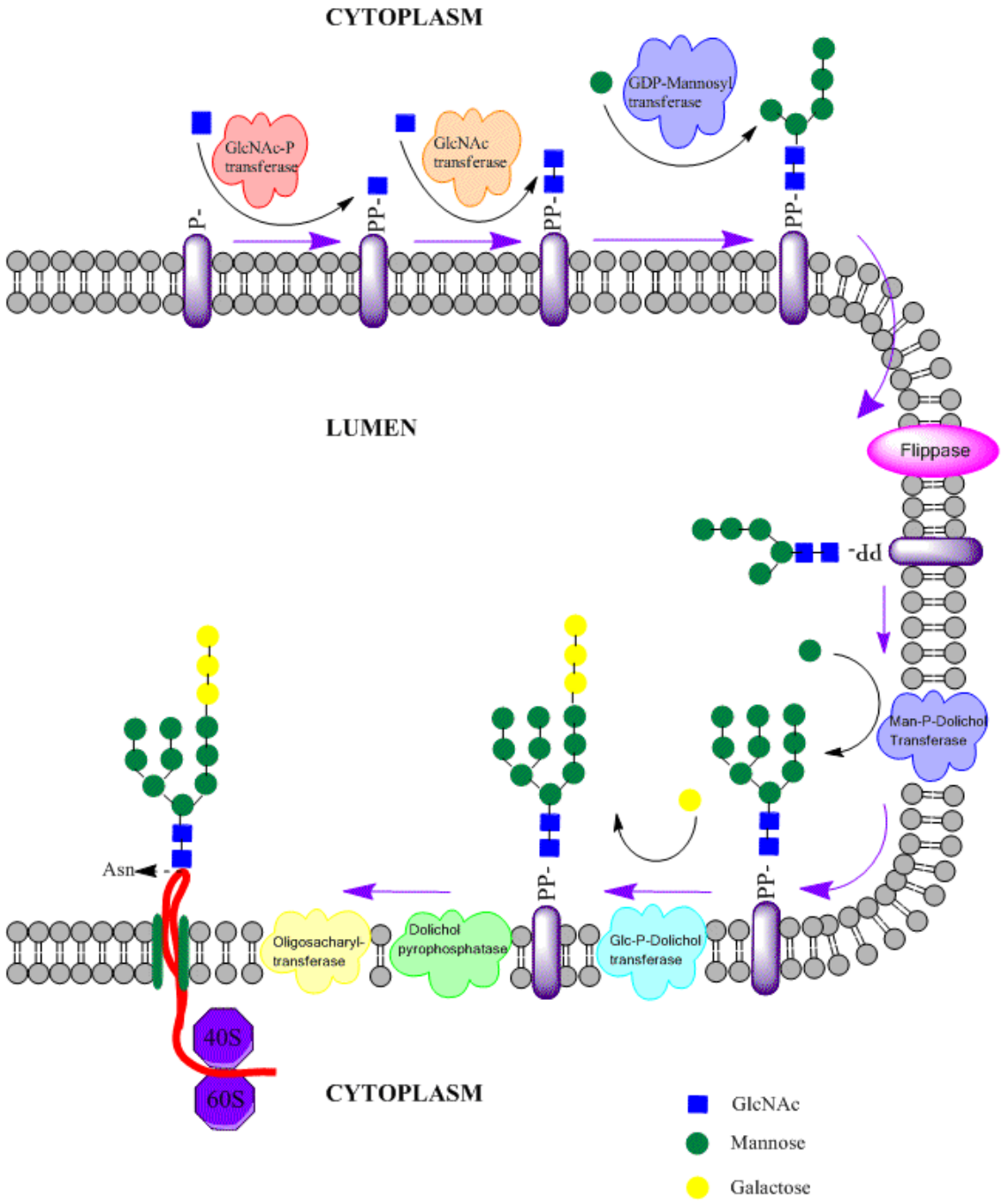


Fig 1.3: Biosynthesis of the precursor oligosaccharide molecule for N-linked glycan synthesis. N-glycan synthesis is initiated at the cystolic face of the ER with the sequential addition of GlcNAc and mannose residues to a dolichol phosphate molecule. Membrane flippase proteins facilitate the transfer of the oligosaccharide as well as the subsequently required free mannose and glucose residues across the ER membrane. The oligosaccharide is finally transferred co-translationally to an Asn residue on the polypeptide by the action of an oligosaccharide transferase. Image produced using ChemBioDraw 12.0**.**

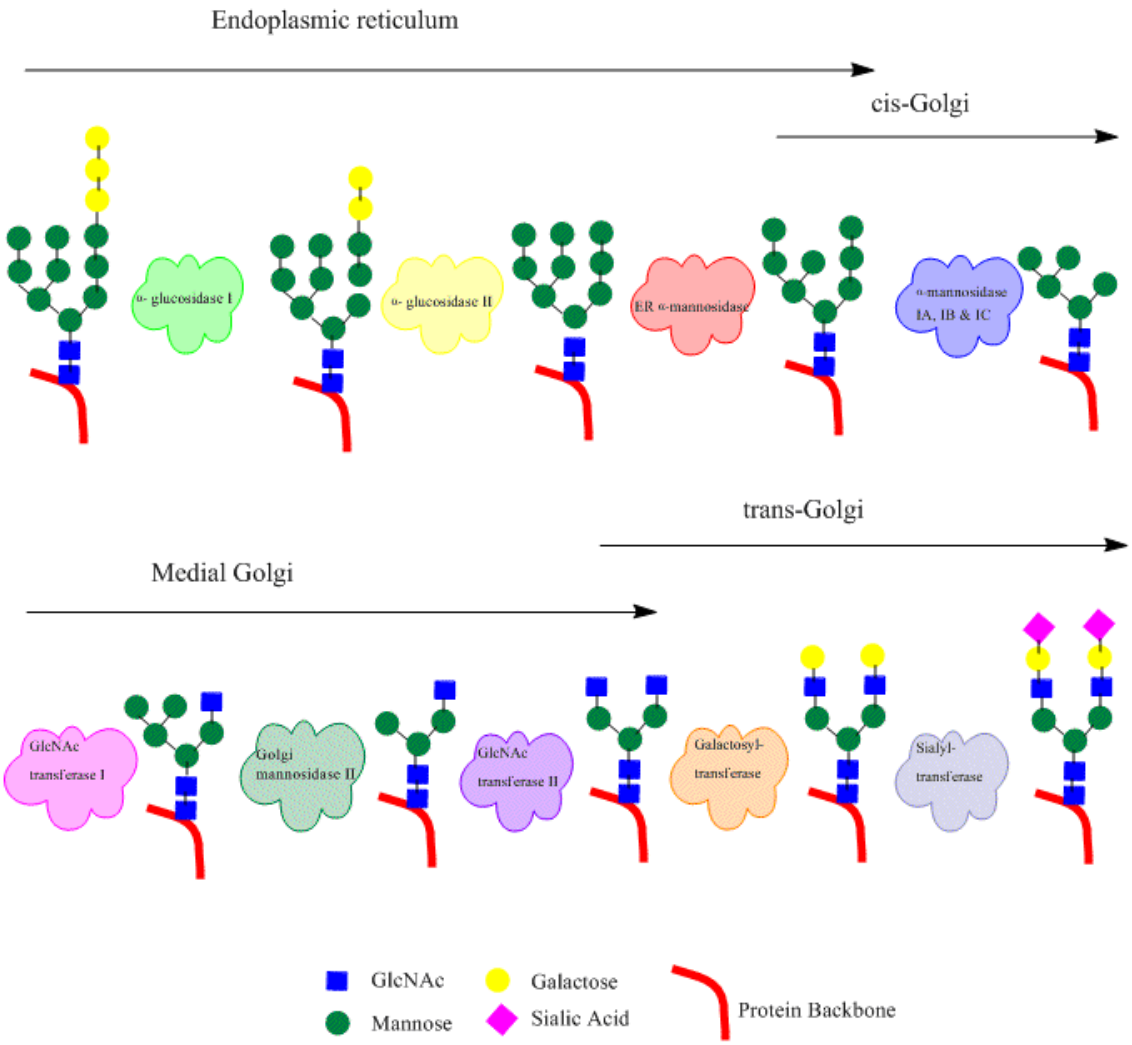


Fig 1.4: Processing of the precursor oligosaccharide for N-linked glycosylation to a complex glycan. Following transfer to the polypeptide, the N-linked glycan is processed, first by removal of the three glucose residues by glucosidases 1 and 2 in the ER. Some or all of the four α1-2 linked mannose residues are then trimmed by the action of mannosidases, as the peptide moves into the cis portion of the Golgi apparatus. Re-elongation is initiated by the attachment of GlcNAc with galactosyltransferase and sialytransferase responsible for attachment of galactose and sialic acid residues in the trans portion of the golgi apparatus. Image produced using ChemBioDraw 12.0.

1.2.2 N-linked glycosylation quality control

Akin to the proof reading abilities of DNA polymerases, mammalian glycosylation systems also have a stringent regulatory pathway or 'quality control' system to ensure that transport is limited to properly folded and assembled proteins. Beginning with the en bloc attachment of the core oligosaccharide to the growing peptide chain, the first and second glucose units are removed by glucosidases I and II respectively (see Fig 1.4). The remaining monoglucosylated ligand binds the glycoprotein transiently to resident ER proteins calnexin or calreticulin (Hammond *et al.*, 1994). Calnexin (transmembrane protein) and calreticulin (soluble luminal protein) are molecular chaperones responsible for preventing aggregation or premature departure of immature oligomers until correct folding has taken place. Both also promote correct disulphide bond formation through interaction with ERp57. To release the bound chains from calnexin and calreticulin glucosidase II removes the remaining glucose residue. Correctly folded glycoproteins can now move from the ER to the golgi apparatus for further processing, whereas incompletely folded glycoproteins are bound by a folding sensor, a soluble enzyme known as UDP-glucose glycoprotein glycosyltransferase (GT). GT reglucosylates incompletely folded glycoproteins with the transfer of glucose residues from UDP-glucose to protein-bound high mannose glycans. Once complete, the proteins are re-bound by calnexin or calreticulin and stay in this cycle until they are either properly folded and oligomerized or degraded. Degradation takes place through the endpoplasmic reticulum-associated protein degradation pathway or ERAD, which involves the translocation of the misfolded peptides to the cytoplasm where, digestion by proteasomes occurs (Helenius and Aebi, 2004).

1.2.3 Eukaryotic O-linked glycosylation

O-linked glycans represent the second most common form of glycosylation in eukaryotes. O-glycans are extremely diverse in both structure and function, with the full extent of this diversity not yet well established. While N-linked glycans share a common core structure, O-linked glycans are built on a different set of protein-glycan linkages. Attachment of O-linked glycans to the peptide backbone can take place via the hydroxyl group of serine, threonine, tyrosine, hydroxyproline,

hydroxylysine or other hydroxylated amino acids in a stepwise series of reactions. Unlike N-linked glycosylation, no consensus sequence has been identified for O-glycosylation, although proline residues are often close to O-glycosylated Ser/Thr residues (Christlet and Veluraja, 2001). Sugar transfer to the protein occurs post-translationally in the Golgi apparatus, with the most common types of O-linked glycans containing an initial GalNAc residue. Other initial sugars include galactose, mannose, fucose, GlcNAc, xylose or glucosamine. Due to the heterogeneity of O-linked glycans they are generally classified by their core structure.

The structure of O-glycans is most often reflected by the levels of enzymes in the cell. For example, the differences between A, B and O blood groups is brought about by the variation in glycan motifs of the A, B and H antigens expressed on the surface of erythrocytes. These antigens differ only in the sugar residues at the non-reducing ends of the glycans (see Fig 1.5). The H antigen occurring in type O individuals is a precursor oligosaccharide of the A and B antigens. Type A individuals express a transferase that specifically adds an *N*-acetylgalactosamine (GalNAc) residue to the terminal position of the H antigen, whereas type B individuals add a galactose to the terminal position.

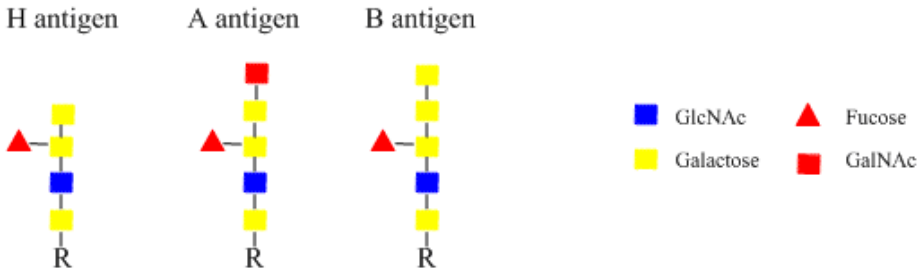


Fig 1.5: H, A and B antigens that form the O, A and B blood group determinants. The H antigen forms the O blood group, the A antigen the A blood group and the B antigen the B blood group. Image created using ChemBioDraw 12.0

The most abundant and widely characterised type of O-glycans in eukaryotes are the mucins (Gum, 1992). The primary purpose of many mucins is to retain water at surfaces that are exposed to the environment but are not sealed by moisture permeable layers, such as the surfaces of the digestive and respiratory tracts (Perez-

Vilar and Hill, 1999). Mucins contain extended stretches of peptide sequence with repeating serine, threonine and proline residues, O-glycan decoration of these proteins can account for up to 80% of the overall mass. Mucin glycosylation, containing initial GalNAc residues, is based on a relatively simple set of core structures with step-by-step addition of further glycans (see Fig 1.6). The overall pathway of mucin glycosylation is regulated by the branchpoint enzyme T-synthase (Ju *et al.*, 2008). T-synthase is a resident golgi protein, a β3-galactosyltransferase that adds galactose to the Tn antigen, generating mucin core structure 1. T-synthase requires a unique, client specific-chaperone, core 1 β3-Gal-T specific molecular chaperone or Cosmc, which is required for the formation of active enzyme When Cosmc, a resident ER protein, is absent, T-synthase is retained in the lumen of the ER as an enzymatically inactive complex. Thus Cosmc serves a unique function in the ER as the key post-translational regulator for expression of T-synthase and consequently regulates mucin O-glycosylation. Cosmc orthologues have been identified in all vertebrates and are highly conserved across species. T-synthase function has been found to be developmentally important in mice (Wu *et al.*, 2004).

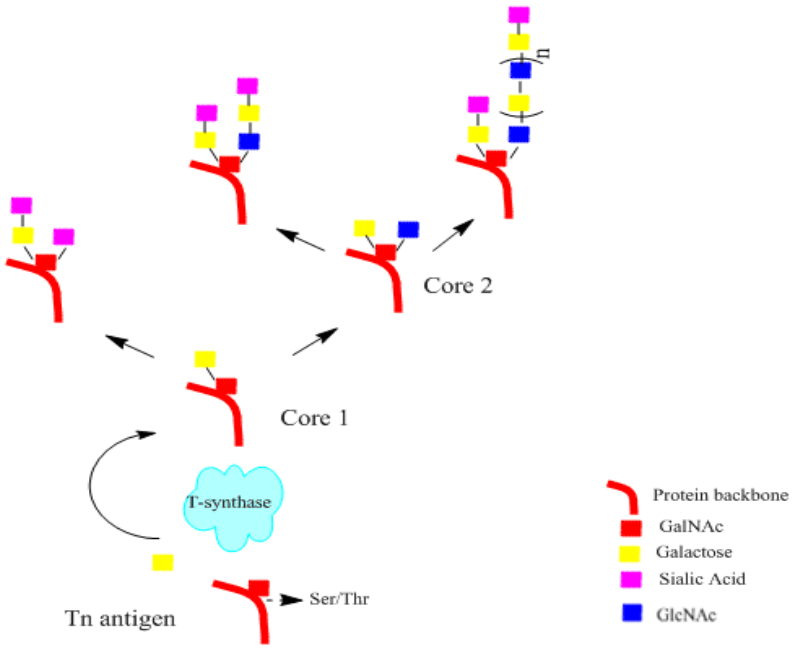


Fig 1.6: Mucin synthesis. *N*-acetylgalactosamine is linked to the side chain of serine or threonine residue. A single β1-3 linked galactose is attached forming Core structure 1. This core structure can be disialyated or further processed by elongation with the addition of *N*-acetyllactosamine and sialic acid. Image produced using ChemBioDraw 12.0.

Although the method of formation differs, terminal structures on some O-linked glycans have been found to be identical to structures on N-linked glycans, suggesting that there may be a functional overlap between the two types of glycosylation. It is also known that glycoproteins can be glycosylated with both N-linked and O-linked glycan structures.

1.2.4 Bacterial glycosylation

Until the 1970s, protein glycosylation was thought to be a task limited to eukaryotes. With evidence to the contrary growing every year it is now a widely accepted that glycoproteins are a common feature in all domains of life. The first examples of non-mammalian glycosylation came with the discovery of sugar structures on surface layer (S-layer) proteins belonging to archaeal and bacterial species, the halophile *Halobacterium halobium* (*salinarum)* and the thermophilic *Clostridium thermosaccharolyticum* (Mescher and Strominger, 1976, Sleytr and Thorne, 1976). It is now know that S-layer proteins can undergo glycosylation to an overall degree of up to 15% (Schaffer and Messner 2001). In the case of *Halobacterium halobium* S-layer N-and O-linked glycoproteins account for up to 50% of the total cell envelope protein (Mescher *et al.,* 1974). From the information available on the possible structures of the prokaryotic glycoproteins, it is obvious that while they share some characteristics, they are far more diverse than the structures found in eukaryotes (Moens and Vanderleyden, 1997). This diversity constitutes differences in consensus sequences, sugar constituents and linkage units, as reviewed Schaffer *et al.* (2001).

Since the initial discovery of S-layer protein glycosylation, it has become more apparent that bacterial glycosylation is not limited in this regard. Non S-layer glycoproteins have been found increasingly in insect and mammalian bacterial pathogens. The flagellin of a number of bacterial species are known to be extensively glycosylated, including *Halobacter* subspecies (Serganova *et al.,* 1995), *Campylobacter jejuni* (Doig *et al.,* 1996), *Clostridium* subsp. (Arnold *et al.,* 1998, Lyristis *et al.,* 2000), *Helicobacter pylori* (Schirm *et al.,* 2003) and *Listeria monocytogenes* (Schirm *et al.,* 2004). Pili are also known to be decorated with sugar moieties that constitute major virulence factors such as those of *Pseudomonas*

aeruginosa (Castric *et al.*, 2001) and *Neisseria meningitides* (Virji *et al.*, 1993). *P. aeruginosa* pili are known to be composed of pilin subunits that are glycosylated with a sugar similar in structure to sialic acid, pseudaminic acid, while both N- and O-linked glycan structures have been identified on *N. meningitides* pili that incorporate galactose residues (Virji, 1997).

Speculations as to the possible role of protein glycosylation in prokaryotic systems vary widely and are reviewed by Upreti *et al.*, (2003). Their prominent location on the cell surface or surface associated organelles suggests a role in host interaction. Carbohydrate modification has also been shown to affect protein stability and function, including secretion and subunit interactions and assembly (Schirm *et al.*, 2003, Guerry 2007). Decoration with typical mammalian glycans and mammalian-like sugar units points towards a role in the evasion of the immune response, while expression of sugar moieties is also known to be a requirement for effective host colonisation (Lyristis *et al.*, 2000, Duck *et al.*, 2007).

1.3 Lectins

In vivo recognition of glycan structures occurs via specific molecules termed 'lectins' (Boyd and Shapleigh, 1954). The word lectin stems from the latin word *legere* meaning to select or to choose and has been defined as a carbohydrate binding protein other than an enzyme or antibody. Knowledge of the presence of lectins in nature began in 1888 with Peter Hermann Stillmark's discovery that certain plant proteins had the ability to agglutinate erythrocytes. It wasn't until 1936, when Concanavalin A (Con A), the first pure lectin was isolated from the Jack Bean (*Canavalia ensiformis)*, that the sugar specificity was demonstrated (James and Stacey, 1936). With much foresight they suggested that the hemaglutination induced by Con A might be a consequence of a reaction from the plant protein with carbohydrates on the surface of the red cells. In the subsequent decades lectins were isolated not only from plants but from mammals, viruses and bacteria, encompassing proteins from a wide range of protein families holding diverse functionalities. There are numerous comprehensive reviews delineating lectin structures, functions and incidences in nature (Van Damme *et al.*, 1998, Kasai and Hirabayashi, 1996,

Rapoport *et al.*, 2008, Sharon and Lis, 2004, Leffler *et al.*, 2002, Loris, 2002, Singh *et al.*, Imberty *et al.*, 2004).

The most striking feature of lectin-sugar interactions is that the binding affinities for monosaccharides are relatively quite low. Many lectins are made up of multiple subunits that increase the capability of binding to disaccharides and more complex oligosaccharide structures with significantly higher affinities and exquisite specificity. This ability is a prerequisite for their ability to function as recognition molecules in biological processes (Ambrosi *et al*., 2005). In some cases the oligosaccharide-binding specificity is dependent not only on the composition of an oligosaccharide chain or on the presence or absence of specific terminal residues but on the overall structure of the glycan. This may include the positions and anomeric (α or β) configurations of the linkages between the constituent monosaccharide subunits (Weis and Drickamer, 1996, Ambrosi *et al.*, 2005, Geyer and Geyer, 2006). Table 1.1 summarises a number of lectin-carbohydrate complexes.

Table 1.1: Lectin-carbohydrate complexes

Lectin Family	Lectin	Abbreviation	Sugar specificity
Plant Lectins			
Legume lectins	Concanavalin A	ConA	Man/α1-Me
	Griffonia simplicifolia lectin 1	GSLI	α-gal
	Griffonia simplicifolia lectin II	GSLII	β1-4 GlcNAc
Cereal lectins	Wheat germ agglutinin	WGA	NeuNAc, GlcNAc
Plant toxins	Ricin	RCA	β-gal/ lactosamine
Bulb lectins	Snowdrop lectin	GNL	Man
Animal lectins			
Galectins	Galectin 1	Gal-1	Lactose
	Galectin 2	Gal-2	Fucose
	Galectin 3	Gal-3	Fucose
P-type lectin	Mannose-6 phosphate receptor	M6P	Man-6 phosphate
Bacterial lectins			
	P. aeruginosa Lectin 1	PA-IL	β-gal
	P. aeruginosa Lectin 1	PA-IIL	Fucose/Man
Viral lectins			
Influenza	Hemagglutinin	HA	NeuNAc

1.4 Significance of glycoforms in eukaryotic systems

Glycoproteins are known to be of fundamental importance to many biological processes as reviewed by Paulson (1989) and Varki (1993). Glycan roles include immune defence, cell growth, cell-cell adhesion, inflammation and fertilisation. Attachment of glycoforms to proteins also functions to stabilise proteins against denaturation and proteolysis, confer structural rigidity, enhance solubility, increase blood retention time and reduce immunogenicity. Unfortunately, our greatest understanding of the significance of glycans to biological processes arises from the study and genetic engineering of abnormal glycosylation and disease states.

1.4.1 Significance of terminal residues

The multistep process of glycan formation, whether N- or O-linked, involves several enzymes that are competing with one another for a single substrate and/or acceptor molecule. As a result glycoproteins contain a micro heterogeneous array of oligosaccharides (Kobata, 1992). Hence glycoproteins are often truncated containing terminal sialic acids (Sias), galactose and GlcNAc residues. Mammalian glycoproteins generally terminate in galactose or sialic acid. Sialic acid, the more abundant of the two, is found on vertebrate cell surface glycoproteins and glycolipids. Mannose and GlcNAc are not commonly found as terminal residues on mammalian glycoproteins, though they are frequently found on the glycoproteins that decorate the surface of yeasts and many microorganisms.

Most mammalian cell surfaces display two major sialic acids (Sias), N-acetylneuramic acid (Neu5Ac) and N-glycolylneuraminic acid (Neu5Gc). Humans however lack Neu5Gc due to a hydroxylase mutation that occurred after the divergence from great apes (Muchmore *et al.*, 1998). The presence of sialic acid as a terminal moiety is known to greatly increase the blood retention time of many glyoproteins and glycosylated therapeutics (Sorensen *et al.*, 2009, Darling*et al.*, 2002). Bisecting fucose is present in a wide variety of organisms. In mammals fucose containing glycans play many important roles, most notably in blood transfusion reactions, host-microbe interactions (Cambi *et al.*, 2005), cell signalling and development, for review see Becker and Lowe (2003).

Although the terminal residues expressed on glycoprotein surfaces are deemed optimal for each biological process, they can also facilitate host pathogen interactions, e.g. mammalian-expressed Sias are known to facilitate cell recognition by organism-extrinsic receptors, such as the viral haemagglutinin influenza via specific viral lectins (Janakiraman *et al.,* 1994). Similarly fucosylation of blood group antigens facilitates *Helicobacter pylori* cell-lectin interaction, leading to *H.pylori* mediated peptic ulcer disease (Mahdavi *et al.,* 2002).

Human and animal studies of glycoprotein clearance have shown that alterations in the glycan structure can significantly affect its pharmacokinetic properties (Jones *et al.,* 2007). This metabolic clearance of glycoproteins is caused by specific receptors which recognise certain structural features of the sugar moiety.

1.4.2 Carbohydrate-specific receptors

The revelation that some de-sialylated serum glycoproteins had drastically reduced survival times in circulation compared to native forms of the same proteins (terminal sialic acid) led to the discovery of the asialoglycoprotein receptor. A specific receptor on hepatocytes, this receptor was found to mediate their clearance by recognition of terminal galactose residues (Morell *et al.,* 1971, Dini *et al.,* 1992). Similarly, the mannose receptor was identified in the late 1970s as a receptor involved in the clearance of certain internal glycoproteins (Stahl P., 1979). It was observed that the exposure of terminal GlcNAc on glycans was accompanied by rapid clearance from circulation. Subsequently it was discovered that this receptor recognised both high mannose N-linked glycans, terminal GlcNAc and fucose residues (Taylor *et al.*, 2005, Jones *et al.*, 2007).

The most extensively studied glycoprotein receptors are the asialoglycoprotein and mannose receptors and are reviewed by Ashwell (1982). Other carbohydrate specific receptors have also been described, including those specific for fucose and 6-phosphomannose (Shepherd *et al.*, 1982, Dahms *et al.*, 2008). The carbohydrate-specific receptors vary in structure and cellular location but all contain multi-subunit lectin-like carbohydrate recognition domains, which can in themselves be

glycosylated. These lectin-like receptors bind preferentially to different branches of target glycans (see Fig 1.7 and 1.8).

Since their initial discovery the carbohydrate-specific receptors have furthermore been implicated not only in the clearance of intracellular glycoproteins but in the recognition of exposed glycan residues on the surface of selected pathogens and the internalisation of glycosylated antigens (Taylor *et al.*, 2005, Schwartz, 1984). Although built from the same building blocks, glycans on bacteria, yeast and fungi can differ widely but many share a few key characteristics. As the glycans usually serve structural roles they tend to have repetitive structures and these structures often present terminal glucose, GlcNAc and mannose residues. Recognition of these features allows the carbohydrate receptors to distinguish between self and non-self glycoproteins.

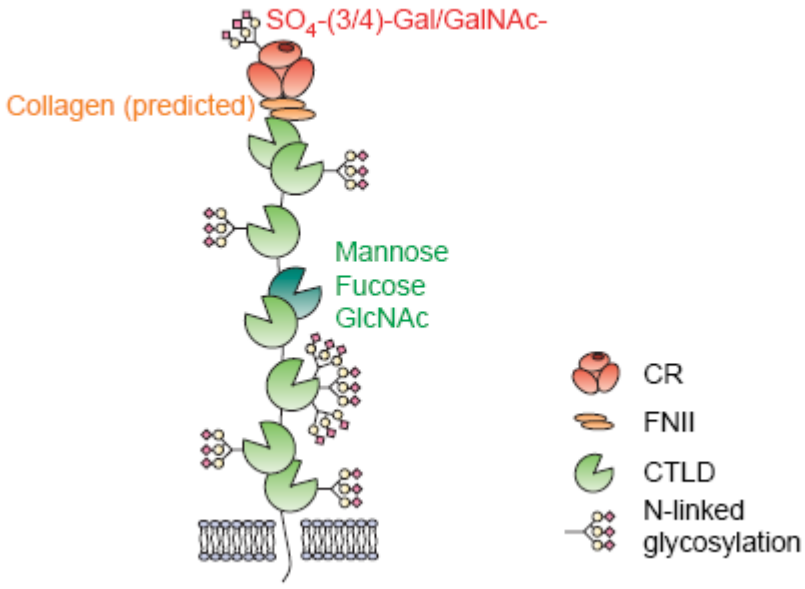


Fig 1.7: Structure of the mannose receptor. Domain structure of the mannose receptor showing the proposed extended conformation and predicted N-linked glycosylation. The cysteine domain (CR) is shown in red, the fibronectin type II domain (FNII) in orange, and the lectin like carbohydrate recognition domains (CTLD) in green. (Taylor *et al.*, 2005)

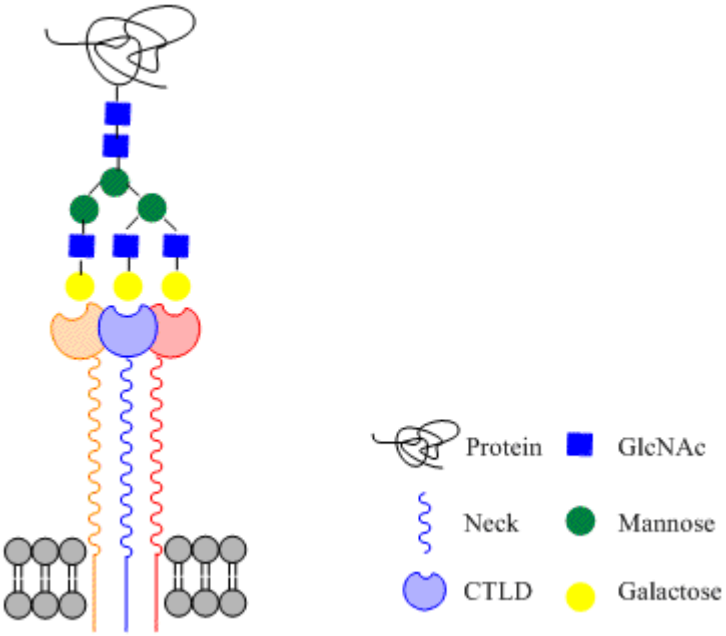


Fig 1.8: Structure of the asialoglycoprotein recptor. The interaction of the lectin-like carbohydrate recognition domains (CTLDs) of the asialoglycoprotein receptor with the terminal galactose residues (yellow) of an asialo N-linked glycan. Image produced using ChemBioDraw 12.0.

1.4.3 Aberrant glycoglycosylation

Firstly, it is important to note that eukaryotic cells require N- and O-linked carbohydrates for survival. At least thirteen congenital disorders of glycosylation have been identified to date and they almost invariably affect N-glycan assembly or processing. These disorders are characterized most prominently by neurological and developmental deficiencies, (Freeze, 1998, Jaeken and Matthijs, 2001, Spiro 2002). Mice lacking *N*-acetylglucosaminyltransferase 1 activity die at mid-gestation, showing a pronounced lack of development in neural tissues (Ioffe and Stanley, 1994). Transgenic mice, born devoid of calreticulin or calnexin, responsible for N-glycan quality control, have embryonic lethal phenotypes, or are born with strong debilitation phenotypes including myelinopathy (Kraus *et al.*, 2010).

Post-translational modifications not only play a crucial role in normal nervous system development but also in brain injury and neurodegeneration, specifically Alzheimers disease. More than 30% of all detectable proteins in the human frontal cortex appear to be glycosylated, with decreased O-GlcNAc addition of tau proteins

mediating the progression of the neurodegeneration that leads to Alzheimers (Kanninen *et al.*, 2004). Muscular dystrophies, including muscle-eye brain disease, Duchenn, Becker and Fukuyama congenital muscular dystrophy have also been shown to be facilitated by an altered glycosylation pattern on α-dystroglycan (Muntoni *et al.*, 2002).

More recently it has been shown that an altered glycosylation phenotype is present on mucins of the cystic fibrosis airway epithelial cells. This altered phenotype, decreased sialylation and/or increasd fucosylation, facilitates the colonisation of the lung by bacteria such as *Haemophilus influenza* and *Pseudomonas aeruginosa* by means of asialo-substrate and fucose-binding lectins, (Scanlin and Glick, 2000, Rhim *et al.*, 2001, Stoykova and Scanlin, 2008). There are multiple other instances in which altered glycoforms have been implicated, such as cardiac conditions, diabetes, stress, nephropathy (Smith *et al.*, 2006), some auto-immune diseases (Hirschberg, 2001), and arthritis (Rudd *et al.*, 2001). Of the diseases caused or signified by a difference in glycosylation patterns, however, none are more studied than cancer.

1.4.4 Aberrant glycosylation in cancer

Aberrant glycosylation occurs in essentially all types of experimental and human cancers, with many glycosyl epitopes constituting tumor-associated antigens. What is not presently understood is whether aberrant glycosylation is a result or cause of initial oncogenic transformation (Hakomori, 2002). Aberrant glycosylation has been implicated as an essential mechanism in defining stage, direction and fate of tumor progression. Studies have shown that there are clear correlations between altered glycan patterns, status of primary tumors, invasive/metastatic potentials of the cancer and ultimately, patient survival rates (Ludwig and Weinstein, 2005, Moiseeva *et al.*, 2005). Altered carbohydrate structures in tumor cells can influence cell recognition, signalling and adhesion allowing tumor cell growth in addition to altering tumor cell motility, promoting or inhibiting invasion and metastasis (Hakomori, 1996).

Changes in glycosylation and cancer include both the under and over expression of naturally occurring glycans, as well as neoexpression of glycans normally restricted to embroyonic tissues (Hakomori, S., 1985) . These structures most often arise from changes in the expression levels of glycosyltransferases in the golgi compartment of cancerous cells. One of the most common changes is an increase in size and branching of N-linked glycans, the increased branching facilitates additional sites for terminal sugar residues, that in conjunction with an up-regulation of sialyltransferase leads to an increase in overall sialyation (Brockhausen *et al.,* 1995). In addition altered terminal structures can also be associated with malignancy, again most notably with the addition of terminal sialic acid structures e.g. Neu5Gc and Lewis structures (sialylated fucosylated structures). These are reviewed by Kim and Varki (1997) and Dube and Bertozzi (2005).

A deeper understanding of the glycosylation processes and disease states has allowed for the development of a number of glycosylated tumor marker candidates (Reis *et al.,* 2010). Miyoshi *et al.* (2006) have identified an increase in α1-3, α1-4, and α1-6 fucosylation in the haptoglobin purified from the sera of patients with pancreatic cancer. Subsequent assay development led to the development of a novel method for the prediction of pancreatic cancer, using the increase in fucosylation as a tumor marker (Matsumoto *et al.,* 2010). Similarly Fukushimo *et al.,* (2010) have revealed that there is elevated expression of α1-2 fucosylated and β-*N*-acetylgalactosaminylated prostate-specific antigen (PSA) in cases of prostate cancer. They have proposed that this increase in glycosylation may be exploited as a potential marker of prostate cancer in the future.

1.5 Glycoprotein analysis and separation

The total glycome of a cell is characteristic of a particular cell type at a specific state of development. Glycosylation patterns are known to vary between individuals, with some variation associated with disease conditions as previously discussed. Thus, analysis of glycomes provides a basis for understanding the functions of glycans in cell differentiation, disease and as targets for receptor binding. Although it has been estimated that only 36 sugar building blocks are required to construct 75% of the 3299 mammalian oligosaccharides identified to date (Werz, *et al.*, 2007), the

characterisation of glycans attached to proteins still presents a more daunting task than the sequencing of proteins and nucleic acids. Difficulties lie not only in the branched nature of the structures but also in extensive heterogeneity of the glycans (Taylor and Drickamer, 2006). Projects to catalogue the structures of all glycans associated with particular cells are underway with databases containing published glycan sequences, providing a clue as to the scale of glycomes. Allowing for structures missing from these databases and for others that have not yet been examined, it seems reasonable to estimate that there may be around 500 endogenous mammalian glycan structures in glycoproteins and glycolipids (Drickamer and Taylor, 2002). In order to crack the glycocode, a diverse range of technologies are coming into play (see Fig 1.9). Primary current analysis methods include the use of 2D gel electrophoresis, Mass spec, NMR, HPLC, lectin chromatography and oligosaccharide arrays with an extensive review of all strategies available by Geyer and Geyer (2006).

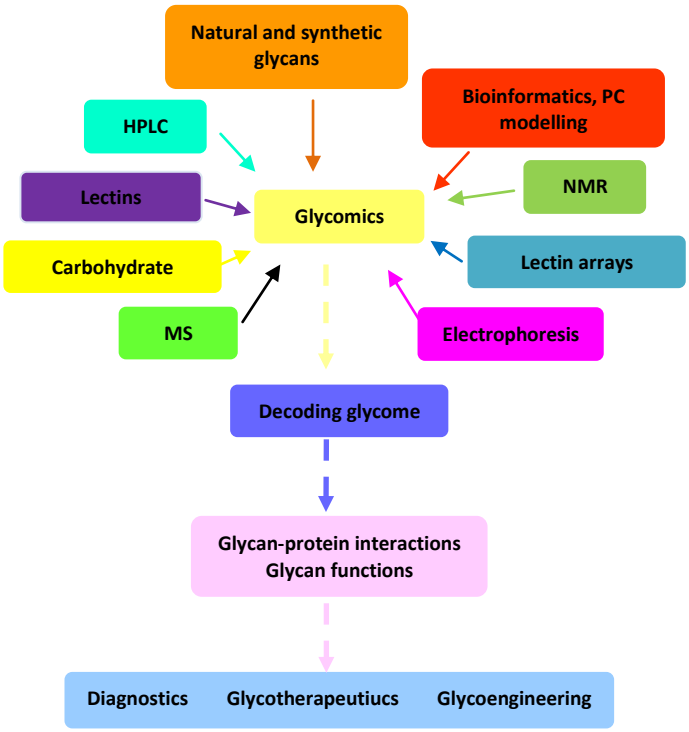


Fig 1.9: Cracking the glycocode: emerging technologies underpinning glycomics. A range of technologies are now being exploited to undertake large-scale analyses of the structure–function relationships of the glycome. These approaches hold the potential to decode the glycome, leading to new insights and biomedical applications, adapted from Mislovicova *et al*, (2009)

1.5.1 Two dimensional (2D) gel electrophoresis

As one of the most efficient protein separation/analytical techniques, gel electrophoresis is often the first step employed for glycan analysis. 2D gel electrophoresis can be used to separate proteins, reflecting both size and/or isoelectric points. Once separated the glycosylation of electroblotted proteins can be characterised by probing with carbohydrate-specific lectins or antibodies. A significant drawback to this technique is the frequent under-representation of membrane proteins, due to the high degree of insolubility in the solubilisation detergents (Geyer and Geyer, 2006). Another is that this technique merely gives an indication of whether glycans are present or not, it does not allow the elucidation of definitive carbohydrate structures or bonds beyond the terminal elaborations.

1.5.2 Mass Spectrometry (MS)

Mass spectrometry is an analytical technique that detects samples that can be successfully converted to gas-phase ions. The resulting ions are accelerated out of the ionisation source into a mass analyser, where they are separated according to their mass to charge ratio and detected to produce a mass spectrum (Hitchen and Dell, 2006). Mass spectrometric identification of glycopeptides has proven to be an important tool. Due to its high degree of sensitivity, it can be used to resolve the components of a glycan mixture and derive information about the structures of individual glycans. There are several ways in which mass spectrometry can be used to analyse individual structures; accurate mass determinations of intact glycans can be used to derive their compositions, as only a limited number of combinations will be consistent with observed masses. Alternatively, glycans can be fragmented, with the differences between the masses of the glycans providing information about the composition (Taylor and Drickamer, 2006).

For mass determination, glycans must first be separated from the protein backbone. Separation is typically carried out using enzymatic or chemical techniques. Enzymes such as trypsin and PNGaseF, a glycosidase that cleaves between the innermost GlcNAc and asparagine residues of N-linked glycans, are most often used to separate glycans prior to analysis. Drawbacks exist with both release techniques in

that the enzymatic release of glycans is dependent on the protein-glycan linkage and that not all glycans may be released, while chemical release may lead to the destruction of the non-carbohydrate constituents of the protein backbone. Where there are multiple glycosylation sites in a single protein the information as to the specific site and the exact protein-glycan linkage will be lost. For the analysis of a heterogenous glycan mixture, an initial separation step is necessary. This can be achieved by using MS/MS or sample enrichment using lectin-specific matrices (Dalpathado and Desaire, 2008, Budnik *et al.*, 2006).

1.5.3 Nuclear Magnetic Resonance (NMR)

NMR analysis of glycoproteins is based on the extent that a glycan distorts in a magnetic field. It is a non-destructive technique and can provide full structural information and so serves as the ultimate source of definitive information on glycan structures. There are some major limitations associated with this technique, including not only the high cost of equipment and trained personnel but also the amount of glycan required, concentrations of 3-4 orders of magnitude (10-100 ng) of the purified glycan (Budnik *et al.*, 2006). NMR is also less well adapted to work on mixtures. Although NMR can provide information that cannot be obtained from other approaches and is well suited to the analysis of small sugars, interpretation of the spectra can be made difficult by the similar environments of many of the protons (Taylor and Drickamer, 2006).

1.5.4 High Pressure Liquid Chromatography (HPLC)

Due to their structural heterogeneity, glycoprotein glycans represent often complex mixtures of closely related isomeric compounds. HPLC has proven to be of great value in the separation and profiling of glycans due to the wide range of adsorbents and solvent systems available, in addition to the speed and reproducibility of separation. Anion exchange chromatography can be used to separate glycans predominantly on the basis of the number of charged groups present. Similar systems can be used to separate glycans on the basis of hydrophilicity (NP-HPLC) and hydrophobicity (RP-HPLC). One of the major advantages of this technique is the ability to predict possible glycan structures through the comparison of elution

times with standard glucose oligomers using only pico moles of sample. A HPLC database for sugar maps has been constructed based on a large number of well-defined standard oligosaccharides http://www.gak.co.jp/ECD/Hpg_eng/hpg_eng.htm (Geyer and Geyer, 2006).

1.5.5 Lectin analysis

Lectins are emerging at the forefront of new methods in glycoprotein analysis. As discussed in section 1.3 lectins have been identified and characterised from a broad range of sources and have been shown to have wide ranging sugar binding specificities. Numerous techniques that employ glycan specific lectins have been developed over a number of years (Mislovicova and Danica, 2009) such as the ability of lectins to distinguish between subtle variations of oligosaccharide structure, making them perfectly suitable as decoders for carbohydrate information.

Lectin affinity techniques so far seem to potentially offer the most comprehensive approach to glycan analysis. Lectin affinity chromatography is especially useful where a heterogeneous mixture of oligosaccharides is passed over a lectin column. Based on the relative affinity of the lectin for sugars, the glycans may be tightly bound to the column, requiring haptenic sugars for elution, thus separating them from the unbound material (Geyer and Geyer, 2006). This technique can be expanded indefinitely using multiple columns with several specificities represented, and also in conjunction with other analysis methods such as MS and NMR, where separation or sample enrichment/preparation is usually necessary to achieve the required result.

The enzyme-linked lectin assay (ELLA) is a potentially useful technique for the generation of glycan profiles. The ELLA, analogous to an ELISA in the immunology field, has been reported numerous times over the past 3 decades as a useful tool for both glycan and lectin profiling (Matsumoto *et al.*, 2010, McCoy *et al.*, 1983, Andrade *et al.*, 1988, Lambre *et al.*, 1991). The ELLA involves the immobilisation of a protein-glycan conjugate on a plate surface. Subsequent to blocking, the carbohydrate can be probed by specific labelled lectins; binding is then detected using colourmetric techniques (see Fig 1.10). This method may also be

used in conjunction with the traditional ELISA. Used in combination an immobilised antibody can be used to capture an antigen before the carbohydrate structure is probed with labelled lectins (Matsumoto *et al.*, 2010).

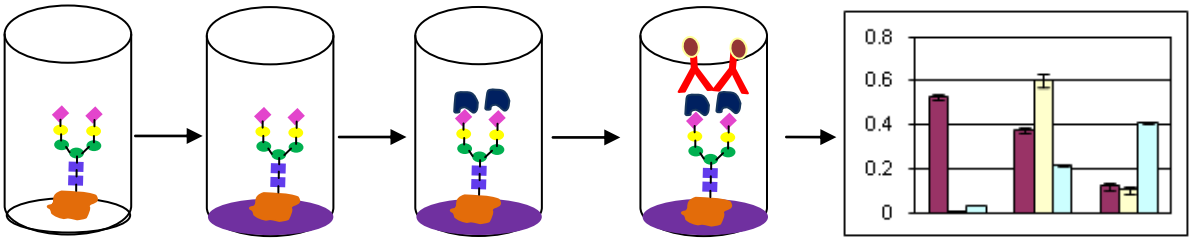


Fig 1.10: Enzyme linked lectin assay (ELLA). A protein carbohydrate conjugate is immobilised on the bottom of a 96 well plate. The plate is blocked with appropriate blocking solution. The glycan complex is probed with labelled lectin(s). The lectins are detected using specific, peroxidase labelled antibodies. The antibody-peroxidase conjugate can be detected using colormetric techniques.

The main advantage of ELLA analysis of glycoconjugates compared with some of the aforementioned techniques, is the lack of a need for protein denaturation or peptide cleavage. Exposure of underlying residues and specific linkages can be achieved by the glycosidase-mediated trimming of terminal glycans and the use of sugar and linkage specific lectins. Due to the highly specific nature of lectin-glycan interactions, high concentrations of glycan are not a prerequisite. In addition to the uses already mentioned lectins can also be applied to in vitro and in vivo studies. With the greater understanding of lectin glycan interactions commercial lectins can be used to localise specific glycans in cells by conjugating them to probes such as fluorescent tags e.g. GFP (Fitches *et al.*, 2001, Krist *et al.*, 2004, Wanchoo *et al.*, 2009), and radiolabelling (Smart *et al.*, 2002). The biggest drawback to the use of lectins is the cost and limited availability of fully characterised lectins.

1.5.6 Carbohydrate arrays

The biggest drawback to the aforementioned analysis techniques is not just the cost and high degree of personnel expertise required but the lack of large-scale applicability and automation. The call for increased time and cost-efficient methods has been answered with the development of oligosaccharide arrays. The first high throughput sugar array assemblies were described in 2002 (Bryan *et al.*, 2002, Fukui *et al.*, 2002, Houseman and Mrksich, 2002), with much effort invested in the development of these glycan arrays in the latter half of the last decade. A paradigm of DNA and protein arrays carbohydrate microarrays primarily consist of sugars that are attached to a surface in a spatially defined and miniaturised fashion. A spacer between the sugar and the surface ensures that the binding partner, usually a lectin, can gain access to the immobilised carbohydrate (see Fig1.11). The microarray format minimises the amount of carbohydrate needed for each binding experiment and makes the most of the expensive materials. The dense presentation of the sugars on the surface mimics the situation encountered on cell surfaces that allows for multivalent interactions of relatively weak binding sugars (Horlacher and Seeberger, 2008).

Researchers have developed a number of carbohydrate array formats to accommodate multipurpose applications in carbohydrate research. Oligosaccharide array technologies can be classified into monosaccharide chips, oligosaccharide chips and microarrays of carbohydrate-containing macromolecules, including polysaccharides and various glycoconjugates. The mono- and disaccharide arrays are suitable for screening and characterising novel carbohydrate-binding proteins (CBPs) or carbohydrate catalysing/binding enzymes and for the identification of novel inhibitors of carbohydrate protein interactions. Arrays containing more complex carbohydrate ligands are suited for the screening of lectins with intricate binding specificities and antibodies with anticarbohydrate binding (Wang, 2003). Glycan arrays have been exploited not just for the characterisation of novel carbohydrate structures or carbohydrate binders but for investigating biological systems. A microarray system was used to analyse the glycan dependent interactions of two HIV-1 envelope proteins which are responsible for initiating the uptake of the virus into the host cell. Blocking these proteins allows for the prevention of virus

internalisation and replication (Adams *et al.*, 2004). Glycan array technologies also have applications in sport; in 2010, Hardy *et al.*, developed a glycoprotein microarray assay to detect antibodies produced in response to recombinant human EPO (rHuEPO) use/abuse in animal based sports. These anti-rHuEPO antibodies remained in circulation for considerably longer than EPO itself, thus prolonging the window in which dopers can be caught.

Carbohydrate or oligosaccharide standards for array immobilisation are either isolated from natural sources or chemically synthesised, with access to these structures the major bottle-neck for the production of carbohydrate arrays (Horlacher and Seeberger, 2008). For glycan array fabrication a number of different immobilisation technologies have been developed, which can be divided into three classes: on-chip synthesis, covalent conjugation and physical adsorption (Lepenies and Seeberger, 2010, Wang, 2003). Recent advances in glycan array fabrication and glycoprofiling have most recently been reviewed by Liang and Wu (2009).

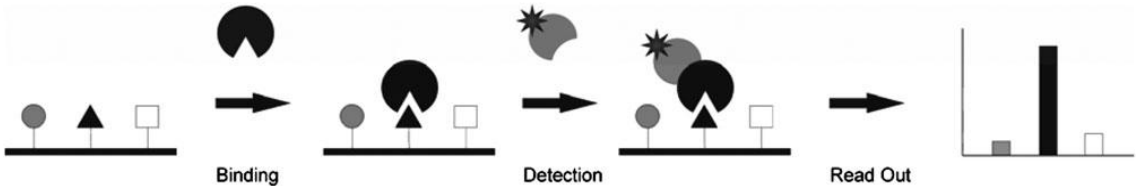


Fig 1.11: Conduct of microarray experiments exemplified for protein binding. Binding of a protein to arrayed sugars, binding of the fluorescently labelled detection protein, read out by a fluorescence scanner and analysis (Horlacher and Seeberger, 2008)

1.5.6.1 Problems associated with carbohydrate arrays

A key outstanding issue in the glycan-screening field is the importance of the context in which glycans are normally seen by lectins. In contrast to cDNA arrays, where on-chip denaturation is a prerequisite, carbohydrate arrays require preservation of the molecular structures, particularly the 3D conformations. Oligosaccharides are usually present on the surfaces of glycoproteins and lipid bilayers rather than free in solution or on synthetic chemical surfaces.

Immobilisation of glycans to array surfaces can have at least three different effects on their interactions with lectins. Firstly, lectins are usually oligomeric and bind with highest avidity when making multiple interactions with appropriately spaced oligosaccharides. The tightest binding may be achieved when multiple glycans are present on the surface of a specific glycoprotein at appropriate spacings. Secondly, some lectins actually bind to protein-carbohydrate co-determinants rather than to glycans alone. Finally, the proximity to protein or polysaccharide surfaces can increase the affinity of lectin- binding to glycoproteins or to terminal structures on large oligosaccharides, without the need for extended, highly specific binding sites (Drickamer and Taylor, 2002). A number of additional technical difficulties will also have to be overcome in order to establish this carbohydrate microarray technology. Difficulties include considering whether carbohydrate macromolecules of hydrophilic character can be immobilised on a chip surface by methods that are suitable for high-throughput production of microarrays. Whether immobilised carbohydrate-containing macromolecules preserve their immunological properties, (such as expression of carbohydrate-epitopes or antigenic determinants and their solvent accessibility) or whether the carbohydrate microarray system can reach the sensitivity and capacity to detect a broad range of antibody specificities in clinical specimens. Finally whether this technology can be applied to investigate the carbohydrate-mediated molecular recognition on a scale that was previously impossible is another consideration (Wang, 2003).

To circumvent the problems associated with the synthesis of carbohydrate polymers and the immobilisation of oligosaccharides in appropriate orientations, methods based on the reverse methodology, lectin arrays, have recently been developed. Of those described none are more successful than GlycoScope^{TM} and Qproteome^{TM}Glycoarray kits marketed by the registered company Procognia. These rapid kit-based assays consist of an array of multiple immobilised plant lectins with overlapping specificities that can be used to profile intact glycoproteins. The binding of a glycoprotein to the array results in a characteristic fingerprint that is highly sensitive to changes in the proteins glycan composition (Rosenfeld *et al.*, 2007).

The biggest advantage of lectin arrays is that no pre-treatment of the samples are necessary, contrary to what is the case for oligosaccharide arrays and the other profiling techniques discussed. However, it is likely that not all glycan substructures are readily accessible for lectin recognition without appropriate treatment of the underlying glycoproteins. Sometimes glycans may be shielded by other glycan structures but, by contrast, it can be advantageous for some research questions to only screen the accessible glycans on the surface. Certain other issues arise when working with lectin microarrays. For example, washing steps are necessary to remove, e.g., unbound sample, which can be detrimental for the relatively low affinity binding of lectins with their ligand. Also, absolute amount of glycans cannot be directly derived from the obtained signals and relative quantification can sometimes be complex since a fraction of the applied lectins have overlapping substrate specificities thus making it difficult to normalise. The presence of interfering free sugars could also leade to false positive results. Finally, some lectins are glycosylated and could potentially be bound by proteins present in the sample (Vanderschaeghe *et al.*, 2010).

1.6 Glycosylation – Implications for the biopharmaceutical industry

Commercial interest in new high-throughput methods for automated, quantitative analysis of glycosylation patterns are indicative of a growing biotech industry wide focus on glycan importance and analysis (Sheridan, 2007). The focus of this industry has dramatically shifted in the past number of years with the majority of new drugs entering the market today recombinant glycoprotein molecules. As of 2006 70% of the therapeutic proteins approved by the European and US regulatory authorities were glycosylated with some 500 candidates at the clinical development stage (Sethuraman and Stadheim, 2006). Due to the nature of glycosylation, the specific glycan moieties on the surface of these therapeutic molecules can vary dramatically. Thus variation can have a large impact on the efficacy of the drug by affecting plasma half life, and can affect tissue targeting and impact a drug's biological activity. In order to meet the demands of the regulatory authorities, biopharmaceutical companies require rapid throughput techniques to enable detailed glycoprotein analysis, not only for initial product characterisation and process

validation but for continuous process monitoring, to ensure product consistency from batch-to-batch (Paul Clarke *et al.,* 2011 manuscript in preparation).

With the emergence of genetic engineering techniques and a proficiency in recombinant protein production, therapeutic product production from mammalian cell lines, yeast, bacteria and insect cells has become the norm. These expression systems vary widely with regard to the ability and proficiency of the cells to incorporate glycans. The choice of particular cell expression systems and the influence of the manufacturing process are important considerations when producing any recombinant products (Jenkins *et al.*, 1996). Adequate control of the microheterogeneity of glycoprotein pharmaceuticals is critical for the production of batches possessing reproducible pharmacokinetic properties. Studies analysing glycoproteins from batch samples demonstrate that the culture environment can influence both the macroheterogeneity and microheterogeneity of oligosaccharides in recombinant glycoproteins (Jenkins *et al.,* 1996). From the known properties of glycoprotein clearance, the main focus of batch assessment is often the sialic acid content. While many non-mammalian cell based expression systems offer clear advantages over mammalian cell lines for recombinant protein expression, cell line choices are limited by the glycosylation machinery. For example, yeast-based expression systems carry significant advantages over mammalian cell lines, such as higher protein titers, shorter fermentation times and the ability to grow in chemically-defined media. Yeast has been used for the production of several approved therapeutics including recombinant insulin, the ability of yeast to carry out N- glycosylation offering a major advantage. The presence of yeast-specific high mannose sugars has, however, limited the use of recombinant proteins produced in yeast to those that do not require N-glycosylation for therapeutic efficacy e.g. insulin (Sethuraman and Stadheim, 2006).

Advances in attempts to humanise cell lines are providing major advantages to this industry, with the possibility that the availability of such cell lines may eliminate the need for mammalian cell culture in the future. The biggest success story in this regard is the humanisation of yeast by GlycoFi. By altering the N-linked glycosylation pathway, Gerngross and colleagues at Darthmoth and GlycoFi

succeeded in generating a yeast cell line capable of secreting terminally-sialylated, complex, bi-antenary glycoproteins (Hamilton *et al.*, 2006, Beck *et al.*, 2010). Merck parted with $400 million in 2006 to acquire GloyoFi, giving an indication of the importance of such technologies to this industry (Sheridan, 2007).

Although great strides have been made in altering cellular machineries for increased sialylation for some glycoprotein therapeutics, it is not always sialic acid that is the key player for the improvement of drug efficiencies. As lectins of the immune system play an essential role in the recognition of glycan structure on the surface of pathogens and autoantigens, they represent valuable targets for cell-specific delivery of immunomodulatory drugs (Lepenies and Seeberger, 2010). For example, glucocerebrosidase (GBA) is a therapeutic molecule approved for the treatment of Gaucher's disease. GBA requires $GlcNAc_2Man_3$ (paucimannose) for targeting through the mannose-binding receptor to macrophages in the liver, where the recombinant enzyme metabolises accumulated glucocerebroside (Friedman *et al.*, 1999). GBA is currently produced in CHO cells, resulting in terminal sialylation; thus, it requires a series of enzymatic digests to produce paucimannose through the sequential removal of sialic acid, galactose and GlcNAc (Sethuraman and Stadheim, 2006).

An increased awareness of the importance of product glycosylation in the biotechnology industry has not only led to more detailed carbohydrate analysis at earlier stages of product development and the glycoengineering of cell expression systems but has also led to an increase in the requirement to provide evidence of glycan elaborations to both regulatory agencies and legal departments. These bodies will use the resulting information to assess product safety and in patent applications to substantiate claims and to distinguish their organisations material from that of their competitors (Jenkins *et al.*, 1996).

1.7 Novel carbohydrate binding molecules

It is now more than evident that progress in the understanding of cell glycosylation machinery, cell-glycan interactions, host-microbe interactions and, aberrant glycosylation etc., is limited primarily by the analytical techniques accessible for glycan profiling. While major strides have been made over the past 40 years it appears the biggest limiting factor in years to come may be the availability of glycan binding molecules. Lectins that bind glycoforms are employed in almost all analysis techniques whether for glycan enrichment, separation or detection. The cost and availability of these molecules are the major limiting factors for use and development. Traditionally, lectins have been isolated whole from plant seeds, fruit, stems and bark. Lectins, from these sources, are typically multi-subunit proteins in which the subunits themselves may vary. This makes it difficult to produce large quantities of plant lectins recombinantly and also gives rise to a wide degree of variation depending on the method of isolation. An explosion in advances of genome sequencing and bioinformatics has meant that prokaryotes have emerged over the past number of years as an almost limitless source of novel carbohydrate-binding molecules (Imberty *et al.*, 2004, Loris *et al.*, 2003). Unlike plant lectins, by their very nature, bacterial lectins are more likely to be amenable to overproduction in simple bacterial expression systems. DNA methods can be used to further enhance carbohydrate-binding molecules and facilitate their implementation in glycoanalytical platforms. A number of protein sequence affinity tags, such as His-, Strep- and poly-lysine could be engineered to enhance not only downstream purification of recombinant binding proteins but to simplify their incorporation onto chromatography or array based methodologies. These tags could also facilitate a more efficient orientation specific immobilisation. Recombinant carbohydrate-binding molecules could also be subjected to mutagenesis in order to enhance or change their binding specificities. In addition to the aforementioned benefits, the ability to scale production of these molecules and simplified downstream purification methodologies could enable the use of these lectins for wide reaching applications at low costs. Potential useful directions for the development and expansion of novel carbohydrate-binding molecules include those which target terminal elaborations that when present on circulatory molecules, are cleared by the mannose receptor - terminal GlcNAc and terminal mannose.

New carbohyrate-binding molecules with novel sugar-binding properties would have a number of applications in all industries, whether for glycan array methodologies or for use in therapeutic molecule enrichment or removal to enhance product efficiencies.

1.8 Chitin

Chitin is a naturally occurring glucose derivative, a homopolymer consisting of variable length linear chains of β1-4 linked GlcNAc (see Fig 1.12). It is structurally identical to cellulose apart from the hydroxyl (-OH) group in cellulose at C_2 which is replaced by an acetamide group ($NH.CO.CH_3$). Chitin is a common constituent of arthropod exoskeletons, shells of crustaceans and the cell walls of fungi and yeast. The derivative is one of the most abundant naturally occurring materials in nature, second to only cellulose (Gooday 1990). In the aquatic atmosphere alone $>10^{11}$ metric tons of chitin are produced annually (Meibom *et al.,* 2004). It is a possibility that proteins that are capable of interacting with chitin may be capable of binding to the GlcNAc elaborations present on glycan structures.

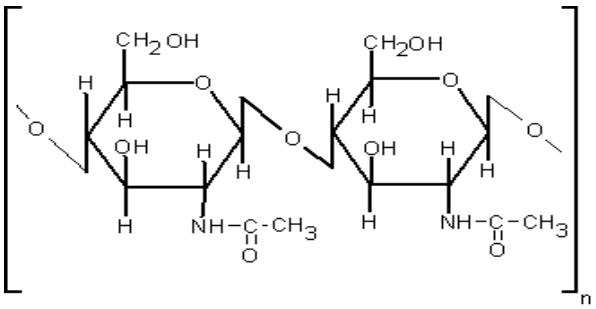


Fig 1.12: Chemical structure of chitin. Image generated using ChemDraw

Chitin occurs in nature as ordered insoluble crystalline microfibrils, occurring in two polymorphs, α- and β-chitin. In both structures the chitin chains are organised in sheets where they are held by a number of intra-sheet hydrogen bonds, but differ in the packing and polarities of adjacent polymer chains (Sikorski *et al.*, 2009). In α-chitin the piles of chains are arranged alternately antiparallel (see Fig 1.13), whereas they are all parallel in β-chitin (see Fig 1.14). A second member of the α family, γ-

chitin, has also been identified, when two chains run in one direction and another chain in the opposite direction. α-chitin is the most abundant of the chitin polymoprhs and is found in fungal and yeast cell walls, lobster and crab tendons, and also shrimp shells. It is also the most thermodynamically stable, with strong inter-sheet and intra-sheet hydrogen bonding. In contrast the rarer β-chitin, derived from squid pen, exhibits no inter-sheet hydrogen bonding with weak hydrogen bonding by intra-sheets, (Jang *et al.*, 2004). Conversion from the β-form to the α-form is possible, but not in the reverse (Khoushab and Yamabhai, 2010).

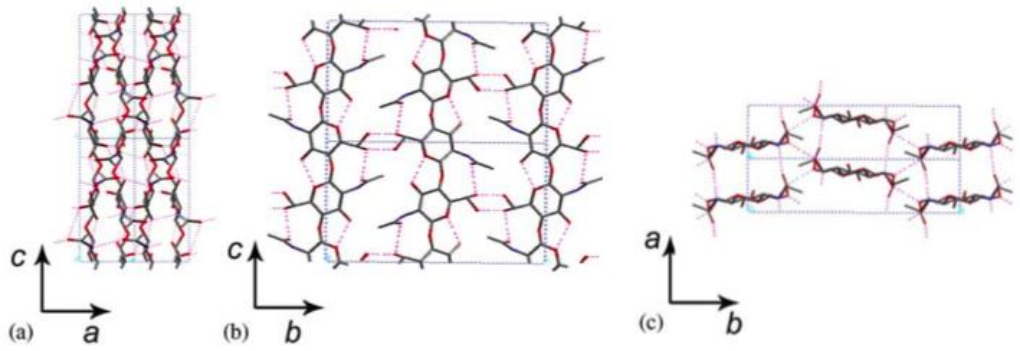


Fig 1.13: Structure of α-chitin. The structure of α-chitin in the (a) *ac* projection; (b) *bc* projection; (c) *ab* projection (Rinaudo, M., 2006).

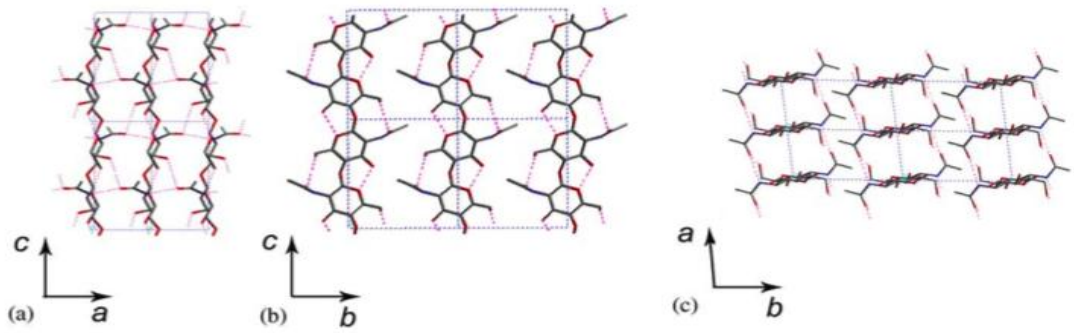


Fig 1.14: Structure of β-Chitin. The structure of β-chitin in the (a) *ac* projection; (b) *bc* projection; (c) *ab* projection (Rinaudo, M. 2006).

Chitin naturally occurs partially deacetylated, depending on the source. When the degree of deacetylation reaches about 50% it becomes soluble in aqueous acidic media and is called chitosan (see Fig 1.15). Commercially chitosan is obtained by deacetylation of chitin in its solid state under alkaline conditions (concentrated NaOH) or by enzymatic hydrolysis in the presence of a chitin hydrolase (Rinaudo 2006). Despite the abundant production and inherent insolubility, chitin does not accumulate in most ecosystems due to the presence of chitinolytic organisms.

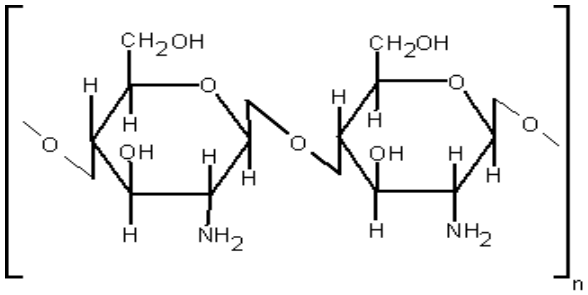


Fig 1.15: Chemical structure of chitosan. Image produced using ChemDraw

1.9 Chitin active/binding enzymes

Chitin active enzymes and nonhydrolytic proteins serve a wide variety of purposes, depending on the host organism. In bacteria they play a role in nutrition, infection and parasitism (Keyhani and Roseman, 1999, Oh *et al.*, 2007, Yu *et al.*, 1991, Joshi *et al.*, 2008, Manos *et al.*, 2009, Meibom *et al.*, 2004, Shen and Jacobs-Lorena, 1998). In fungi, protozoa and invertebrates they are also involved in morphogenesis (Merzendorfer and Zimoch, 2003, Wessels, 1993, Sahai and Manocha, 1993) , and are used as defence mechanisms in plants, invertebrates and mammals (Hejgaard *et al.*, 1992, Millar*et al.*, 1992, Kawabata *et al.*, 1996, Brameld and Goddard, 1998, Okawa *et al.*, 2003, Sanders *et al.*, 2007).

Carbohydrate-active enzymes can exhibit modular structures, when a module is defined as a functional and structural unit. In the classification of Carbohydrate-Active enZYmes (CAZy) (http://www.cazy.org/), each family in the CAZy database is classified depending on a common segment within the gene that ultimately contains the catalytic or binding module (Cantarel *et al.*, 2009). At present CAZy provides details of four classes of enzyme activities, Glycoside Hydrolases,

Glycosyltransferases, Polysaccharide lyases and Carbohydrate esterases. CAzy also provides family classification of carbohydrate-binding modules, classified as a contiguous amino acid sequence within a carbohydrate-active enzyme with a discrete fold having carbohydrate-binding activity. Chitin synthases are classified as belonging to family 2 of the glycosyltransferases. Proteins and protein domains with chitin-binding properties are classified by CAZy into several sub-families e.g. families 18, 19 and 48 of the glycoside hydrolases and families 1, 2, 3, 5, 12, 14, 18, 19, 33, and 37 of the carbohydrate-binding modules.

1.10 Chitin Synthases

Chitin synthases (E.C 2.4.1.16) catalyse the formation of chitin through the transfer of *N*-acetyl-D-glucosamine residues to the chitin polymer, using uridine diphosphate *N*-acetyl-D-glucosamine (UDPGlcNAc) and chitin as substrates. This produces a longer chitin chain and UDP as the reaction products (reviewed by Cohen, E. 2001). Chitin synthesis is a necessary event in the life cycle of many organisms. As previously mentioned, chitin serves as the major cell wall/exoskeleton scaffolding in many species of crustaceans, arthropods, yeasts and fungi (Mellado *et al.,* 1996). Chitin synthases have also been described in nematodes (Veronico *et al.,* 2001) and amoebae (Campos-Gongora *et al.,* 2004), where chitin synthesis is necessary for cyst development. Most organisms contain two or more chitin synthase genes which fall into one of two categories; the first is a zymogen that is activated in vitro by trypsin treatment, while the second type does not require proteolytic activation. The zymogen form can be further subdivided into five categories based on sequence similarity (Bowen *et al.,* 1992, Ruiz-Herrera *et al.*, 2002, Ruiz-Herrera and Ortiz-Castellanos, 2010). Although chitin is not present in vertebrates, chitin metabolism by glycosyl hydrolases is important as a control strategy against chitinous organisms.

1.11 Glycoside Hydrolases

Chitinases (E.C. 3.2.1.14) that catalyse the hydrolysis of chitin may be divided into three categories, classified as endo-, exochitinases or β-*N*-acetylglucosaminidase. Endochitinases act to catalyse the hydrolysis of chitin at random, internal sites along the polymer, where as exochitinases remove single *N*-acetyl-D-glucosamine residues from the non-reducing ends of chitin chains. β-*N*-acetylglucosaminidases cleave GlcNAc units sequentially from the non reducing end of the substrate. Chitinases are classified as belonging to families 18, 19 and 48 of the glycoside hydrolases. They do not share amino acid similarity between families and have completely different 3D structures. Chitinases characteristically contain both chitin binding and catalytic domains, where the chitin binding domain helps anchor the catalytic unit near the substrate, thus enhancing its activity (Ubhayasekera *et al.*, 2007). The presence of separate carbohydrate binding and catalytic domains is a generally observed feature in enzymes that hydrolyse solid carbohydrate substrates and is also observed in cellulases and amylases (Beintema, 1994).

1.11.1 Glycosyl Hydrolases – family 18

Family 18 glycosyl hydrolases encompass chitinases from most living organisms and according to a recent study by Karlsson and Magnus (2009) may be further characterised into three separate clusters, namely, A, B and C. Clusters A and B both contain representatives from bacteria, fungi and plants. Cluster C consists of hydrolytic domains with bacterial origin, with the exception of the carboxy-terminal domain of an archean ChiA protein. No obvious domain patterns separate the three main clusters but subgroups corresponding to taxonomic groupings within the clusters display similar length and analogous domain patterns in addition to the chitinolytic domain (Karlsson and Stenlid, 2009). This suggests functional differentiation of family 18 glycosyl hydrolases and that proteins with similar domain structure have adapted to a specific substrate and possibly to a specific biological function. Typical family 18 glycosyl hydrolases can comprise up to five distinct domains: an N-terminal signal peptide region, a catalytic domain, a serine/threonine-rich region, a chitin binding domain and a C-terminal extension region (Perrakis, *et al.*, 1994). The signal peptide preceding the N-terminal domain

of the mature protein presumably mediates secretion of the enzyme and is cleaved off by signal peptidases after the protein has been transported across the membrane. The catalytic domain of family 18 chitinases consists of an $(\alpha/\beta)_8$-barrel with a deep substrate-binding cleft formed by loops positioned between the carboxy-terminal end of β-strands and the amino terminal end of the helices (Davies and Henrissat, 1995, Henrissat *et al.*, 1995, Perrakis *et al.*, 1994). The active site of GH18s contain a conserved consensus of amino acids residues, (LIVMFY)-(DN)-G-(LIVMF)-(DN)-(LIVMF)-(DN)-X-E, which are essential for catalytic activity (Watanabe, *et al.*, 1993). The chitin-binding domain is not always present and is not necessary for chitinase activity but it is thought to enhance the efficiency of the enzymes by acting as an anchor thus keeping the chitinase in contact with the substrate (Suzuki *et al.*, 1999, Kuranda and Robbins, 1991).

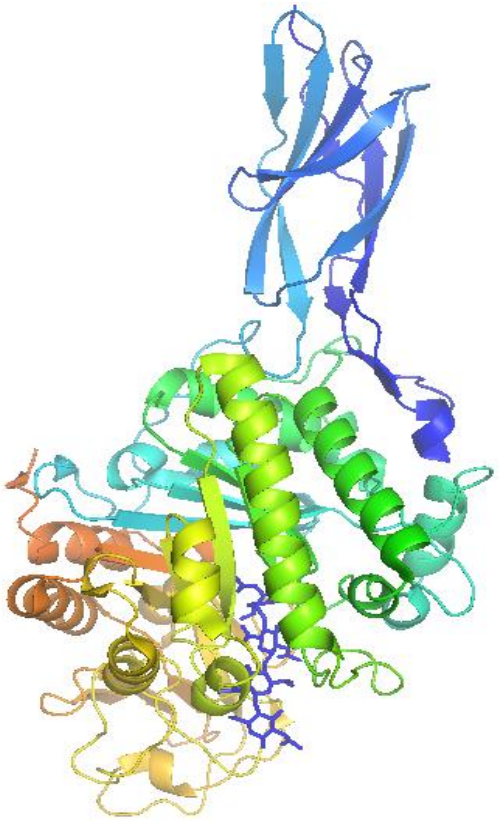


Fig 1.16: The crystal structure Chtinase A (ChiA) from *Serratia marcescens* in complex with tetra-*N*-acetylchitotriose. An image of the crystal structure of ChiA from *S. marcescens*, a member family 18 of the glycosyl hydrolases, in complex with tetra-*N*-acetylchitotriose (PDB Code: 1K9T). ChiA comprises three domains: an all β-strand amino-terminal domain, a catalytic α/β-barrel domain, and a small α- and β-fold domain. Image generated using Pymol.

1.11.2 Glycosyl Hydrolases – family 19

Family 19 chitinases had until recently only been isolated from higher plants but have since been reported in a number of bacterial species, (Brameld and Goddard, 1998, Watanabe *et al.*, 1999, Sasaki *et al.*, 2006). A variety of putative family 19 chitinases from eukaryotes, bacteria and viruses are also reported in the CAzY database. With few exceptions family 19 chitinases are synthesised with an N-terminal signal sequence. After removal of the signal the mature protein consists of two domains, a chitin-binding domain and the catalytic module. In contrast to the family 18 chitinases, the catalytic domain of family 19 chitinases comprise of two lobes, each of which is rich in α-helical structure. A conserved motif to characterise the substrate binding site of family 19 chitinases has been proposed by Huet *et al.*, (2008), defined between Tyr111 and Tyr125 as ^{111}Y-(FHY)-G-R-G-(AP)-x-Q-(IL)-(ST)-(FHYW)-(HN)-(FY)-N-Y^{125}. Only one member of family 48 has been described as having chitinase activity.

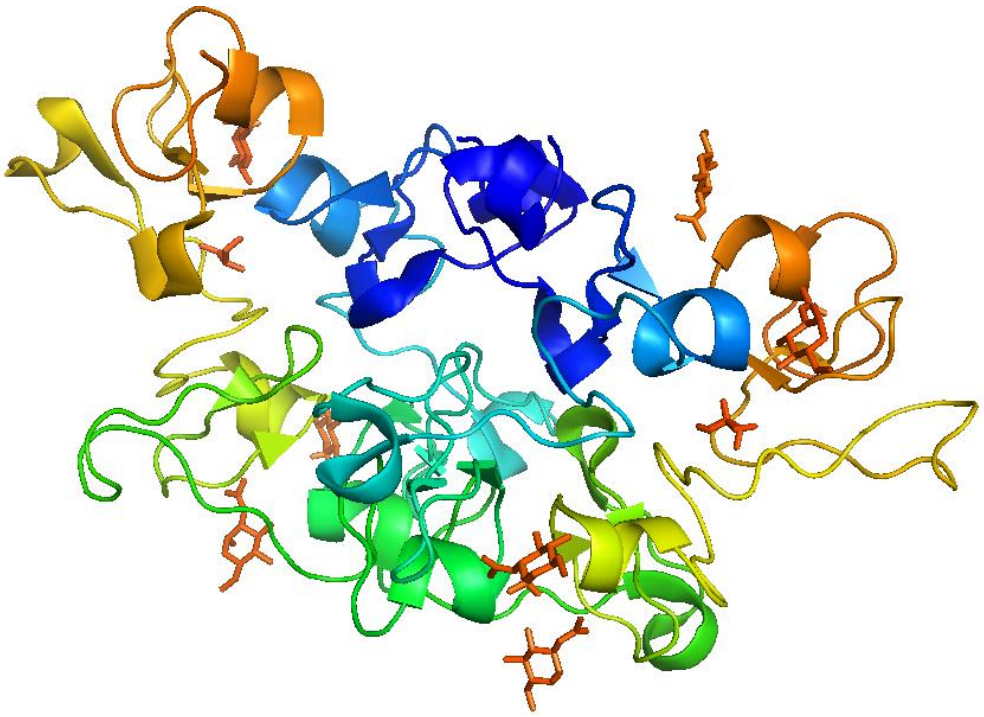


Fig 1.17: Crystal structure of wheat germ agglutinin (WGA) in complex with GlcNAc. An image of the crystal structure of WGA, a member of family 19 of the glycosyl hydrolases, in complex with GlcNAc. GlcNAc residues are highlighted in orange (PDB Code: 2UVO). Image generated using Pymol.

In addition to a lack of homology, family 18 and 19 endo- and exo- acting chitinases also differ in their mechanisms of action. The family 18 chitinases reported to date yield a β-anomer hydrolysis product, whereas family 19 chitinases give rise to α-anomer configuration. This dissimilarity is explained by a proposed difference in the mechanism of hydrolysis employed by the two distinct families. Family 18 chitinases are thought to utilise a double displacement mechanism, where protonation of a GlcNAc residue in a boat configuration leads to an oxazoline intermediate, which may be hydrolised to form a product with retention of the anomeric configuration (see Fig 1.18a). A single displacement mechanism has been proposed for family 19 chitinases where two acidic residues are required in the active site and the hydrolysis product has inversion of the anomeric configuration (see Fig 1.18b) (Brameld and Goddard, 1998).

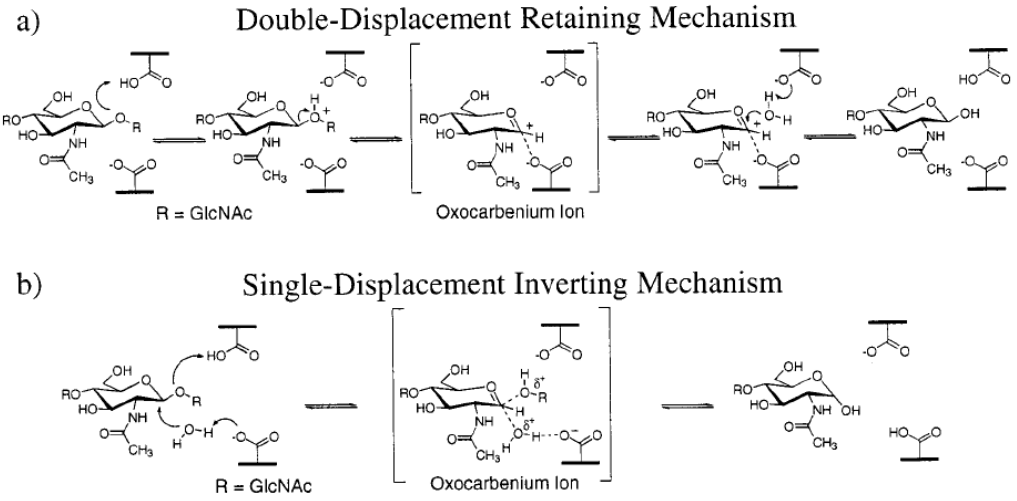


Fig 1.18: Proposed hydrolysis mechanism for family 18 and 19 chitinases from the glycoside hydrolase family. (*a*) The double-displacement hydrolysis mechanism proposed for family 18 chitinases. Protonation of a GlcNAc residue in a boat conformation leads to an oxazoline intermediate, which may be hydrolyzed to form a product with retention of the anomeric configuration. (*b*) The single displacement hydrolysismechanismproposed for family 19 chitinases. Two acidic residues are required in the active site and the hydrolysis product has inversion of the anomeric configuration (Brameld and Goddard, 1998).

1.11.3 Carbohydrate-binding modules

The families of carbohydrate binding modules that display chitin binding activity (families 1, 2, 3, 5, 12, 14, 18, 19, 33 and 37) as described by CAZy, are simple units, approx 40-150 residues long, that appear as discrete domains within larger carbohydrate active proteins. The one exception to this is the family 33 CBMs. Some members of this family are distinct i.e. they also exist as independent, non-catalytic chitin-binding proteins, mainly found in bacteria and viruses. Genome analysis indicates that secreted family 33 CBPs are produced by most chitin-degrading organisms (Vaaje-Kolstad *et al.,n*2005b), though only a few have been characterised to date. Independent family 33 CBMs have been reported and isolated from *Streptomycetes* (Saito *et al.,* 2001, Schnellmann *et al.,* 1994, Kolbe *et al.,* 1998, Svergun *et al.,* 2000, Schrempf.H, 2001), *Pseudoalteromonas sp.* (Techkarnjanaruk and Goodman, 1999), *Pseudomonas aeruginosa* (Folders *et al.,* 2000), *Bacillus amyloliquefaciens* (Chu *et al.,* 2001), *Serratia* sp. (Fuchs *et al.,* 1986, Watanabe *et al.,* 1997), *Thermobifida fusco* (Moser *et al.,* 2008) and *Lactococcus lactis* spp (Vaaje-Kolstad *et al.,* 2009). These nonhydrolytic chitin-binding proteins (CBPs) are thought to act synergistically with chitinases by increasing substrate accessibility thus increasing chitinase activity (Moser *et al.,* 2008, Vaaje-Kolstad *et al.,* 2009, Vaaje-Kolstad *et al.,* 2005a). As is a common feature of the chitinases discussed previously, the family 33 CBPs have N-terminal signal sequences, which are cleaved from the protein during secretion of the proteins to the extracellular medium, suggesting a requirement for export and localisation outside the cell. The combination of sequence and structural information from site-directed mutagenesis studies have shown that the surface of family 33 CBPs contain a patch of highly conserved mostly polar residues which are important for binding to chitin and for the positive effect on chitinase efficiency (Vaaje-Kolstad *et al.,* 2005b, Vaaje-Kolstad *et al.,* 2005a). The exact mechanism of interaction, however, remains unknown.

1.12 Fungal chitin active enzymes

Fungal chitin synthases and chitinases are involved in fungal growth and development, spore formation, pathogenicity and parasitism. The rigid chitinous cell wall of fungi provides both protective and aggressive functions. In this way, it acts as an initial barrier to the outside environment while harbouring many hydrolytic and toxic molecules. With the exception of unicellular yeasts, fungi typically grow by means of hyphae that extend only at their apices and ramify into a mycelium. Analysis of the chemistry of the fungal wall and its biosynthesis by Joseph Wessels (1993) has disclosed a simple mechanism explaining the transition from a newly formed expandable wall at the apex to a more rigid wall at the base of the hyphal extension zone. Two individual wall polymers, chitin and β-glucan, are extruded at the apex and modified within the domain of the wall. Among the modifications observed are the formation of covalent crosslinks between these two polymers and hydrogen bonds between the homologous polymer chains, leading to the formation of chitin microfibrils crosslinked to a glucan matrix (see Fig 1.19). By extending at their hyphae in this way, fungi can penetrate solid substrata such as wood, allowing for the excretion of lytic enzymes, including chitinases to digest the substrate polymers and convert them to nutrients.

In yeast cells, chitin produced by synthase genes accounts for 1% of the cell wall (Orlean, 1987). Located in the septum chitin is important for the mechanical stability of the temporary junction between mother and daughter cells during budding. Separation of the progeny is dependent on expression of a chitinase to break down the chitin in the primary septum (Kuranda and Robbins, 1991). Most fungal chitinases belong to family 18 of the glycosyl hydrolases (reviewed by Duo-Chuan, 2006) and therefore are typically composed of up to five domains, an N-terminal signal peptide region, a catalytic domain, a serine/threonine-rich region, a chitin binding domain and a C-terminal extension region. Generally fungal chitinases lack the last three domains but are still catalytically active. Fungal chitinases lacking an N-terminal signal peptide have been shown to be located intracellularly and may function during morphogenesis (Takaya *et al.,* 1998, Seidl *et al.,* 2005). The serine/threonine-rich region of fungal chitinases is usually glycosylated with sugar chains, thought to be necessary for secretion and stability of the protein.

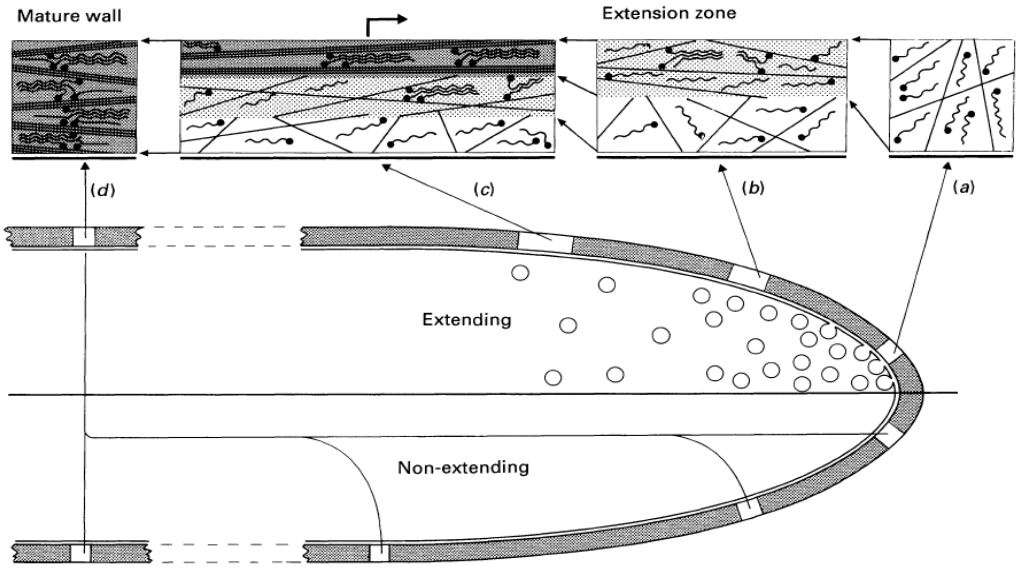


Fig 1.19: The steady-state model of apical wall growth. For the extending apex the schematic drawing shows the stretching of the wall and the addition of new wall material from the cytoplasmic side, maximally at the extreme tip. Newly added wall consists of chitin chains (straight lines) and (1→3)-β-glucan chains (wavy lines). With time, these two polymers interact to form chitin microfibrils and β-glucan triple helices while covalent linkages occur between chitin and β-glucan chains, which also become branched. At the very tip (a) the wall is minimally crosslinked and supposed to be most plastic. Subapically (b) a wall volume added at the very tip is stretched and partially crosslinked (increased shading) while new wall material is being added from the inside to maintain wall thickness. At (c) the average crosslinking has proceeded to the point that the wall hardly yields to turgor pressure but it has not yet maximal strength. Crosslinking continues in the tubular part of the wall till completion far behind the apex (d). (Wessels, 1993).

1.13 Plant chitin active enzymes

The β1-4 GlcNAc linkages of chitin do not occur in plants, whereas multi-domain enzymes consisting of several chitinase and chitin-binding domains occur frequently. The chitinous components of the cell wall of fungal pathogens have been regarded as the target of plant chitinases for defence purposes (Werner *et al.,* 2002). Present knowledge also suggests that they are differentially expressed in response to different forms of environmental stress such as mechanical wounding, drought, salinity and frost (Hejgaard *et al.,* 1992, Millar *et al.,* 1992, Sasaki *et al.,* 2006).

Plant chitinases are synthesised at ribosomes attached to the ER, before entering the lumen of the ER where they are transported further through the golgi apparatus to a vacuole. Some plant chitinases are modified during the synthesis pathway, undergoing cleavage of their N-terminal signal sequences and the addition of N-linked glycan chains. Others may undergo rearrangements, deletions and reglucosylation to ensure proper folding. Prior to the development of the CAzY database, plant chitinases had been categorized as either b-type or h-type with further subdivision into five classes based on the amino acid sequence and domain features. According to the classification of glycoside hydrolases by Henrissat and Davies (1995), b-type enzymes of classes I, II, and IV are included in family 19, whereas h-type classes III , and V are included in family 18. Class I chitinases have an N-terminal cysteine-rich chitin binding domain and a C-terminal catalytic domain which are connected by a short linker peptide of about 10–20 amino acid residues. On the other hand, class II chitinases consist only of a catalytic domain homologous to that of class I chitinases. Class IV chitinases share homology with class I chitinases. Class III and V chitinases show no homology with classes I, II, and IV but have distinct sequence similarity to bacterial and fungal chitinases (Beintema, 1994, Graham and Sticklen, 1994).

1.14 Invertebrate chitin active enzymes

Arthropod growth and development is strictly dependent on the capability to remodel chitinous structures. For example, insect cuticles form an exoskeleton that is limited in its capacity to keep pace with body growth due to the rigidity of the structure caused by the presence of chitin and other sclerotised proteins. In order for insects to develop, they are periodically forced to replace their old cuticle with a new, looser one during ecdysis (molting) (Merzendorfer and Zimoch, 2003). Ecdysis is initiated by apolysis, where the epidermal cells are separated from the old cuticle by molting fluid secretion and ecdysial membrane formation. The molting fluid contains multiple chitinases to aid in the digestion of the old endocuticle. Once degraded, the fluid is then re-absorbed, recycling the old cuticle components. Secretion of cuticle proteins and chitin fibers from epidermal cells initiates formation of the new cuticle. Chitin synthesis is also necessary for formation of the

Peritrophic Matrix (PM). The PM is an extracellular chitin-containing envelope that completely surrounds the food bolus in the guts of most arthropods. It is thought to be crucial for protection of the digestive cells from bacterial damage and parasitic invasion, in addition to facilitating the digestive processes. All PMs are composed of chitin, proteins and proteoglycans (Elvin *et al.,* 1996). Arthropod chitinases belong to family 18 of the glycosyl hydrolases and contain the typical modular stucture, an N-terminal signal peptide region, the mature catalytic amino-terminal domain, a serine/threonine-rich domain and a chitin binding domain. Like some fungal chitinases insect chitinases can also be extensively glycosylated, with a size range of 40-85 kDa.

While chitin-containing organisms use chitin synthases and chitinases to remodel their own structures, the same classes of enzymes are exploited for colonisation by viruses and parasitic organisms. For instance insect viruses produce chitinases to disintegrate the PM of their target organism, thereby increasing viral infectiousness. The malaria parasite *Plasmodium falciparum* however, uses chitinases to cross the chitinous PM of the mosquito host (Shen and Jacobs-Lorena, 1998, Mitsuhashi *et al.,* 2007, Vaaje-Kolstad Horn *et al.,* 2005). Anopheline mosquitoes are the sole vectors for human malaria, causing about 2 million deaths annually. Upon ingestion of an infected blood meal, the mosquito forms a PM around it. Inside the PM *Plasmodium* gametes mate, develop into ookinetes, and about one day later, the ookinetes penetrate the PM and the gut epithelial cells through the secretion of chitinases. After traversing the gut epithelium and reaching the gut basal lamina, the ookinetes develop into oocysts. Each oocyst eventually releases thousands of sporozoites that invade the salivary glands and infect the next person when the mosquito bites again.

1.15 Mammalian chitin active enzymes

To date, chitin has not been found in mammals. However several proteins showing homology to chitinases (all members of family 18 of the glycosyl hydrolases) have been reported in a number of species. The first mammalian chitinase to demonstrate chitinolytic activity was a tetra-*N*-acetylchitotetraoside isolated from man (Boot *et*

al., 1995). Chitotriosidase was found to be synthesised by activated macrophages and its activity was found to be markedly increased in plasma and tissue of patients with Gauchers disease (Hollak *et al.,* 1994). This marked increase was due to the massive production and secretion by the lipid laden macrophages which accumulate as a consequence of this disease. The presence of a chitotriosidase gene in humans suggests an anti-microbial function, however people showing a mutation in the gene encoding chitotriosidase produce no functional protein and yet have no greater susceptibility to infection (Boot *et al.*, 1998).

A second catalytically active human chitinase, namely, acidic mammalian chitinase (AMCase), with a homologue in cynomolgus monkeys, was also recently identified (Boot *et al.,* 2001, Krykbaev *et al.,* 2010). Like chitotriosidase, AMCase contains an N-terminal catalytic core domain and C-terminal chitin binding domain, separated by a hinge region. Unlike chitotriosidase, AMCase is expressed in the stomach and lung tissue, where it is thought that it may play a dual role in food assimilation and defence respectively (Boot, 2001).

A third GlcNac active enzyme, Di-*N*-Acetyl-Chitobiase has also been identified in humans (Fisher and Aronson, 1992). It is a reducing end glycosidase that functions along with a number of other acid hydrolases to split the GlcNAcβ-D-(1-4)GlcNAc chitobiose core of asparagine-linked glycoproteins, completely degrading these glycoproteins and related glycoconjugates.

A number of chitinase-like proteins, termed chi-lectins, Oviductin (Buhi 2002), human cartilage protein gp-39 (YKL40) (Hakala *et al.*, 1993), human condrocyte protein 39 (YKL39) (Hu *et al.*, 1996) , and stablin-1 interacting chitinase-like protein (SI-CLP) (Kzhyshkowska *et al.,* 2006) , have also been identified as capable of binding chitin. These chi-lectins lack enzymatic activity due to substitution of active-site catalytic residues (Bussink *et al.,* 2007). While the precise functions of these chi-lectins and mammalian chitinases are not yet precisely understood, most of the human family 18 chitinases have been linked to a variety of diseases and could potentially be used as diagnostic and prognostic markers for those diseases (see Table 1.2) (Guan *et al.,* 2009).

Table 1.2: Chitin active proteins and associated human diseases. (Guan *et al.*, 2009)

Chitinase	Character of Disease	Disease
AMCase	Th2-mediated inflammation	Asthma Rhinosinusitis with nasal polyps Vernal keratoconjunctivitis and season allergic conjunctivitis
Chitotriosidase	Inherited lysosomal storage disorders/Sphingolipids	Gaucher's disease or β-glucocerebrosidase deficiency Niemann Pick GMI-gangliosidosis Krabbe's disease
YLK-40	Infection	Human endotaxemia *Streptococcus pneumoniase* pneumonia *S.pneumonia* bacteraemia Purulent meningitis and encephalitis
	Joint Disease	Rheumatoid arthritis Osteoarthritis Ankylosing spondylitis
	Chronic inflammation	Giant cell arthritis Inflammatory bowel disease Sarcoidosis Systemic sclerosis
	Liver fibrosis	Alcoholic liver disease Chronic hepatitis C virus infection
	Advanced cancer	Adencarcinoma Small cell lung cancer Glioblastoma Myxoid chondrosarcoma Papillary thyroid carcinoma Melanoma
		Arthersclerosis
YLK-39	Joint disease	Rheumatoid arthritis Osteoarthritis

1.16 Bacterial chitin active enzymes

Chitin turnover is an essential process for recycling carbon and nitrogen, with bacteria the primary degraders of chitin in the biosphere. Bacteria capable of degrading chitin usually produce an array of chitinases; *Serratia marcescens* produces three chitinases, a hexosaminidase and a non-catalytic chitin-binding protein, when grown on chitin (Suzuki *et al.*, 2002, Fuchs *et al.*, 1986). Similarily, *Streptomyces olivaceoviridis* produces five chitin active proteins (Romaguera *et al.*, 1992). This multi-protein approach allows microorganisms to launch a synergistic attack on insoluble chitin matrices, with endochitinases attacking the chitin chain at random intervals, exochitinases attacking at either the reducing or non-reducing ends and the hexosaminidases converting the chitobiose products to GlcNAc. These compounds may then enter the periplasm where chitodextrinases (EC 3.2.1.14) and N-acetylglucosaminidases (E.C. 3.2.1.52) act to form a pool of GlcNAc and chitobiose (Howard *et al.*, 2003). Following transport to the cytoplasm, the GlcNAc and chitobiose are metabolised or modified for use in cell wall biogenesis (see Fig 1.20).

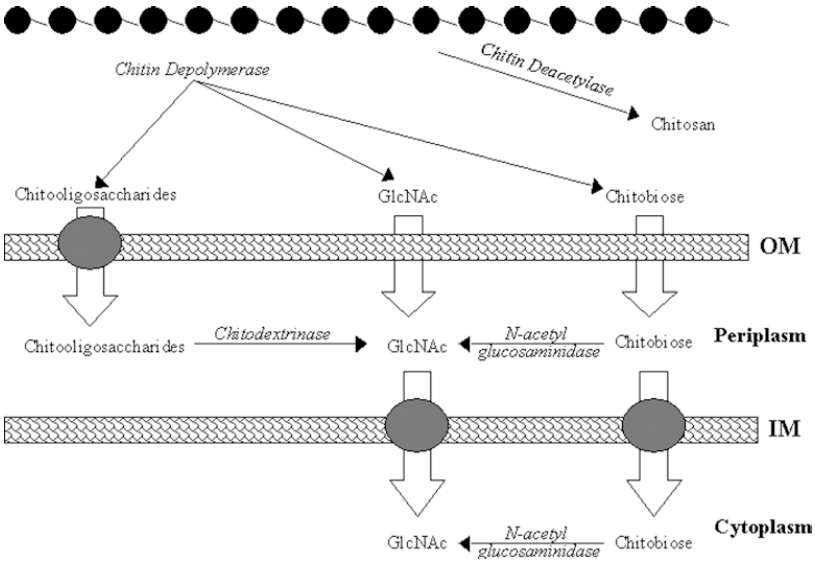


Fig 1.20: Chitin recycling system employed by microbial organisms. Multiple enzymes are required for the degradation of chitin. *Shaded ovals* Transporters, *OM* outer membrane, *IM* inner membrane, (Howard *et al.*, 2003).

In bacteria an alternative strategy for chitin degradation is also sometimes employed, involving nonhydrolitic chitin binding proteins (family 33 of the carbohydrate binding modules) that act synergistically with chitinases, as mentioned in section 1.11.3. Genome analyses indicate that these proteins are secreted by most chitin-degrading microorganisms but few have been characterised biochemically. Biochemical studies have revealed a large diversity in binding preferences (see table 1.3).

Table 1.3: Binding preferences of family 33 non-hydrolytic bacterial chitin binding proteins

Organism	Name	Binding preference	Reference
Serratia marcescens	CBP21	β-chitin cellulose	(Watanabe *et al.*, 1997)
Streptomyces coelicolor	Chb3	α-chitin, β-chitin, chitosan	(Saito *et al.* 2001)
Streptomyces olivaceoviridis	CBH1	α-chitin	(Schnellmann *et al.* 1994)
Sterptomyces reticuli	CHB2	α-chitin	(Kolbe *et al.*, 1998)
Streptomyces tendae	AF1	α-chitin, chitosan, fungal cell walls	(Bormann *et al.* 1999)
Pseudomonas aeruginosa	CbpD	α-chitin	(Folders, *et al.*, 2000)
Pseudoalteromonas piscicida	Cbp1	α-chitin, cellulose, β-chitin	(Tsujibo, *et al.*, 2002)
Bacillus amyloliquefaciens	ChbB	β-chitin, cellulose, α-chitin, β-glucans	(Chu *et al.*, 2001)
Thermobifidia fusca	E7	β-chitin, α-chitin, cellulose	(Moser *et al.*, 2008)
	E8	β-chitin, α-chitin, cellulose	
Lactocossus lactis	*LI*CBP33A	β-chitin, α-chitin, cellulose	(Vaaje-Kolstad *et al.*, 2009)

1.16.1 Chitin-binding proteins utilized in this project

In the search for novel carbohydrate-binding proteins, particularly those with preference for GlcNAc, four family 33 chitin-binding proteins were chosen; CBP21 from *S. marcescens*, CbpD from *P. aeruginosa* and two previously uncharacterised proteins that show homology to CBP21. They have been named CbpA and CbpL from *Photorhabdus asymbiotica* and *Photorhabdus luminescens* respectively, for the purposes of this project. The chitinolytic system of *S. marcescens* is one of the most extensively studied. The previously characterised CBP21 is a good candidate on which to model expression and binding studies of the three other molecules.

1.16.2 The chitinolytic system of *Serratia marcescens*

Serratia marcescens is a species of gram-negative bacteria that belongs to the family Enterobacteriacae. A human pathogen, it is involved in nosocomial infections, particularly urinary tract infections and wound infections, and as discussed previously, is one of the most efficient and well studied chitinolytic organisms. *S. marcescens* is capable of producing 3 chitinases, ChiA, ChiB and ChiC, a β-*N*-acetylglucosaminidase, and a non catalytic chitin binding protein CBP21 (belonging to family 33 of the carbohydrate binding modules) when grown in the presence of chitin.

CBP21 was first described by Fuchs, R.L *et al.*, in 1986 as a 21 kDa catalytically active chitinase. It was later shown by Watanabe *et al.*, (1997) that this 21kDa protein, found in the culture supernatant of *Serratia marcescens* 2170, cultivated in the presence of 0.5% colloidal chitin, did not have any chitinase activity but did adsorb chitin. Chitin-binding activity and the fact that only 5 proteins are produced by this strain in medium containing colloidal chitin suggested that CBP21 participated in the chitin degradation system of this bacterium. The nucleotide sequence of CBP21 was later elucidated by Suzuki *et al.*, (1998), who also demonstrated that the production of chitinases and CBP21 from *S. marcscens* 2170 are regulated in parallel, but only in the sense that the substrates that induce production of chitinases also induce production of CBP21. In addition, they

examined the binding of CBP21 to various insoluble substrates. It was demonstrated that the highest binding activity of CBP21 was observed when β-chitin was used as a substrate, followed by regenerated chitin and colloidal chitin. Binding of CBP21 to cellulose was less significant and binding to chitosan was almost negligible (Suzuki *et al.*, 1998). Although the function of chitin-binding proteins like CBP21 are relatively not well understood, Vajee-Kolstad *et al.*, (2005a) have shown that the binding of CBP21 to chitin leads to the disruption of the crystalline substrate structure and to a dramatic increase in chitinase efficiency. This is thought to be achieved through a number of specific polar interactions which alter the substrate structure.

The crystal structure of CBP21 was solved using rhenium single-wavelength anomalous dispersion in 2005. The structure consists of a three-stranded and a four-stranded β-sheet that form a β-sandwich. There is a bud like 65-residue pseudo domain between β-strands one and two which consist of loops, one short α-helix, and two short 3_{10} helices. The four cysteine residues appear as two disulfide bridges, one in the loop/helical region and one joining β-strands 4 and 5 (see Fig 1.21) (Vaaje-Kolstad *et al.*, 2005b).

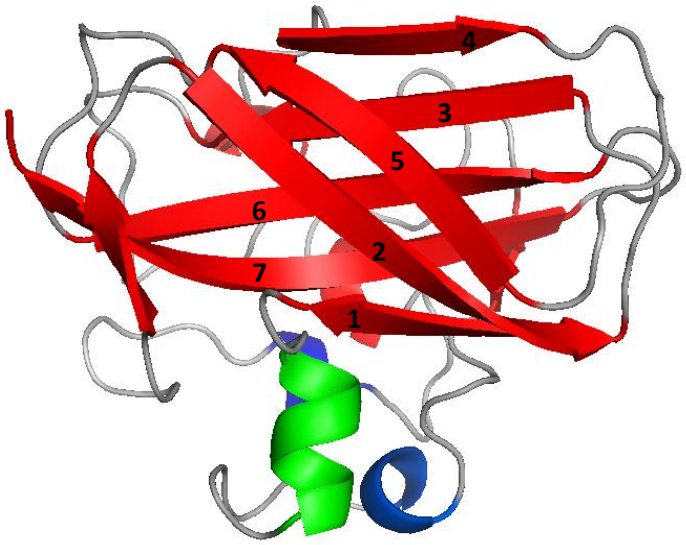


Fig 1.21: Crystal structure of CBP21. The structure of CBP21 (PDB code: 2EBM). The FnIII-like fold of CBP21 consists of two β-sheets, the first β-strands 1, 2, and 5, and the second β-strands 3, 4, 6, and 7. The loop/helical domain consists of one α-helix and two 3_{10} helices. Image generated using PyMol.

Attempts to solve the crystal structure of CBP21 in complex with GlcNAc by Vaaje-Kolstad *et al.,* (2005b) have so far proved fruitless, in part perhaps because CBP21 is not capable of binding to a single GlcNAc unit. A number of studies have demonstrated the importance of particular residues for chitin binding, although, none have definitively characterised the residues involved in chitin interaction. Zeltins and Schrempf generated a number of mutant chb1 genes and concluded that a Trp residue in position 57 was essential for binding of CHB1 to α-chitin, (Zeltins and Schrempf, 1997). Tryptophan, although present in a number of species in this position, is not ubiquitously conserved among all chitin binding proteins thus far characterised (see Fig 1.23). Vaaje-Kolstad *et al.,* (2005b) have shown that a conserved aromatic residue in this position (Tyr-54) does not contribute to the chitin affinity of CBP21 but they have identified a number of other potentially important residues through site directed mutagenesis studies (see Table 1.4).

Table 1.4: Binding of CBP21 mutants to β-chitin. (Vaaje-Kolstad *et al.,* 2005b)

CBP21	Kd (μM)	βmax (μmol of CBP21/g)
Wild type	1.4 ± 0.4	5.9 ± 0.4
Y54A	11.0 ± 3.1	4.0 ± 0.6
E55A	5.1 ± 1.4	2.1 ± 0.2
E60A	8.2 ± 2.6	7.9 ± 1.4
H114A	6.9 ± 1.7	3.0 ± 0.4
D182A	7.4 ± 1.8	5.3 ± 0.6
N185A	6.6 ± 2.0	7.3 ± 1.1

In nature chitin is often found as a composite where layers/sheets of chitin are interwoven with proteins and/or minerals in a recalcitrant heteropolymer (Vaaje-Kolstad *et al.,* 2009). The crystalline chitin used in most experiments has been treated by strong acids and bases in order to remove the protein and/or mineral fraction. It is conceivable that the real natural substrates of the CBP proteins differ from chitin/cellulose and chitosan. There may exist composite natural chitinous substrates that are more susceptible to the action of CBPs. Through the mutagenesis of highly conserved residues it is possible that those necessary for binding can

finally be elucidated, and give direction to possible substitutions that may alter the proteins binding preference (see Fig 1.23).

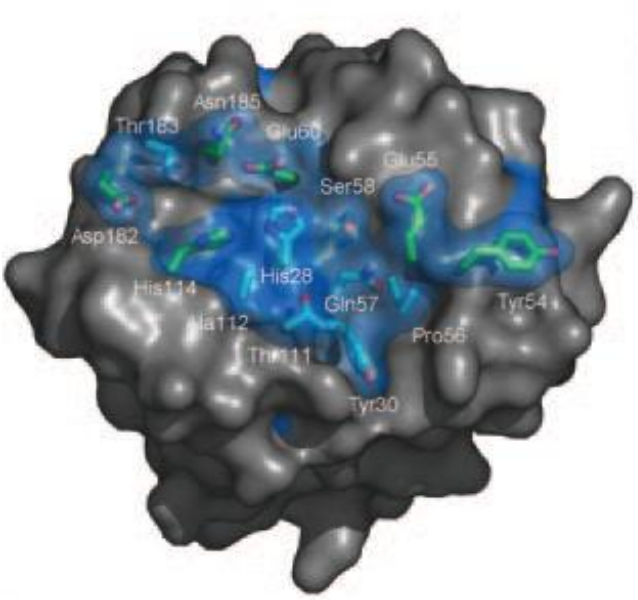


Fig 1.22: Location of surface patch residues of CBP21 thought to be involved in chitin binding. A surface representation of CBP21 illustrating the patch of conserved residues. (Vaaje-Kolstad *et al.*, 2005b)

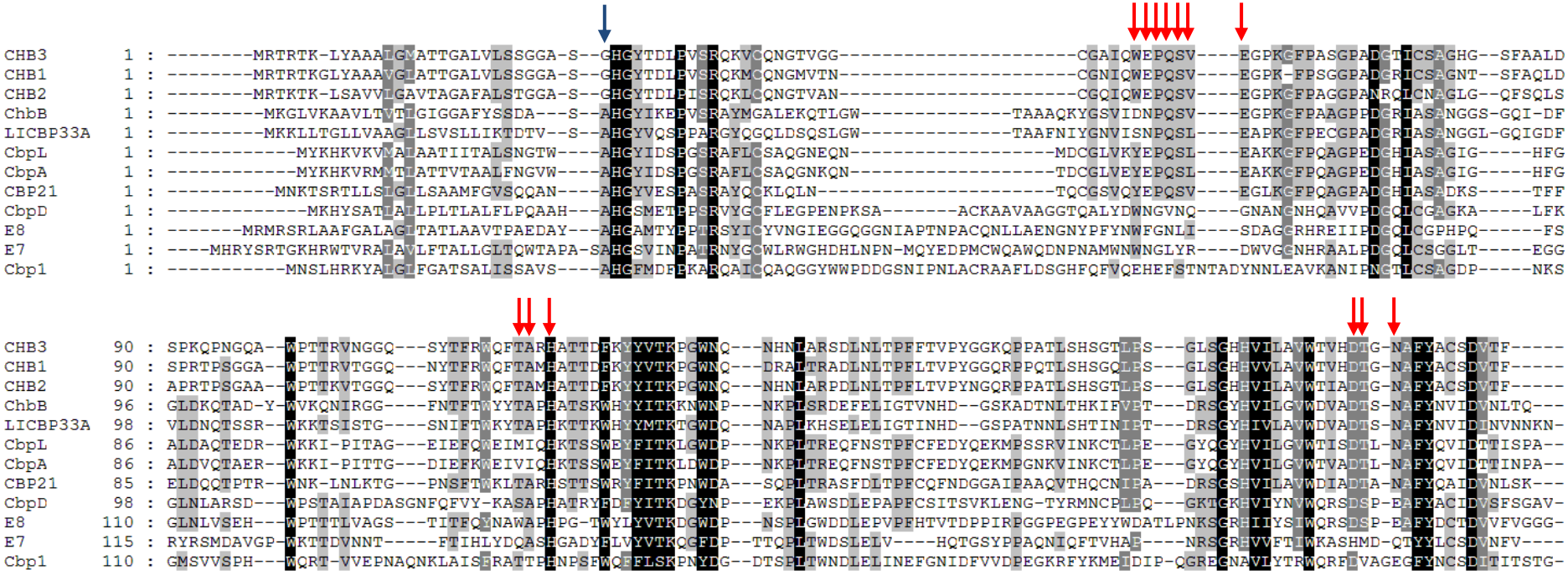


Figure 1.23: Multiple sequence alignment of the N-terminal region of family 33 bacterial CBPs. The blue arrow indicates the first residue after the signal sequence, the red arrows indicate conserved surface patch residues thought to be involved in chitin binding through hydrophobic interactions. CHB3; *Streptomyces coelicolor*, CHB1; *Streptomyces olivaceoviridis*, CHB2; *Streptomyces reticuli*, ChbB; *Bacillus amyloquefaciens*, LICBP33A; *Lactococcus lactis*, CBP21; *Serratia marcescens*, CbpL; *Photorhabdus luminescens*, CbpA; *Photorhabdus asymbiotica*, CbpD; *Pseudomonas aeruginosa*, E7; *Thermobifidia fusca*, E8; *Thermobifidia fusca* and Cbp1; *Pseudoalteromonas piscicida*. Alignment generated using ClustalW and Gendoc 2.6 (section 2.13).

1.16.3 CbpD from *Pseudomonas aeruginosa*

Pseudomonas aeruginosa is a pathogenic gram-negative bacterium. It is a ubiquitous and opportunistic organism capable of inhabiting a large range of environments. It poses significant problems as a human pathogen, particularly in hospitals, where it accounts for more than 10% of hospital-acquired infections (Thompson *et al.*, 2001). *P. aeruginosa* is able to secrete many proteins, including exoenzymes, S and T, exotoxin A, proteases and elastases into the extracellular medium. Most of these proteins contribute to the virulence of the bacteria, as they are associated with epithelial cell and tissue damage or dys-functioning of infected host cells (Folders *et al.*, 2000).

In 2000 Folders *et al.,* identified a chitin-binding protein, with homology to Chb1 and CBP21 of *S. olivacerviridis* and *S. marcescens* respectively, in the culture supernatant of *P. aeruginosa*. It was shown that like Chb1 the 43 kDa protein CbpD consists of two domains and that the CbpD is a chitin-binding protein (Folders *et al.*, 2000). As with all family 33 chitin-binding proteins the physiological function of CbpD remains unclear. It is known that CbpD is produced by a type II secretion system that is under control of the quorum-sensing system (Folders, *et al.*, 2000), indicating that CbpD is only needed under high-density conditions, such as in biofilms and during pathogenesis. In this study Folders and co-workers carried out a search for homologues of CbpD in different *Pseudomonas* spp. , and discovered that while present in many clinical isolates it was not in the soil pseudomonads tested. They surmised that this suggested a role for CbpD in pathogenicity, speculatively that a glycosylated epithelial cell surface molecule might exist as a receptor for CbpD. While both domains of CbpD cluster with proven and putative *N*-acetylglucosamine binding proteins (see Fig 1.24) it is still not known if both are capable of binding to chitin or whether both have diverse binding preferences.

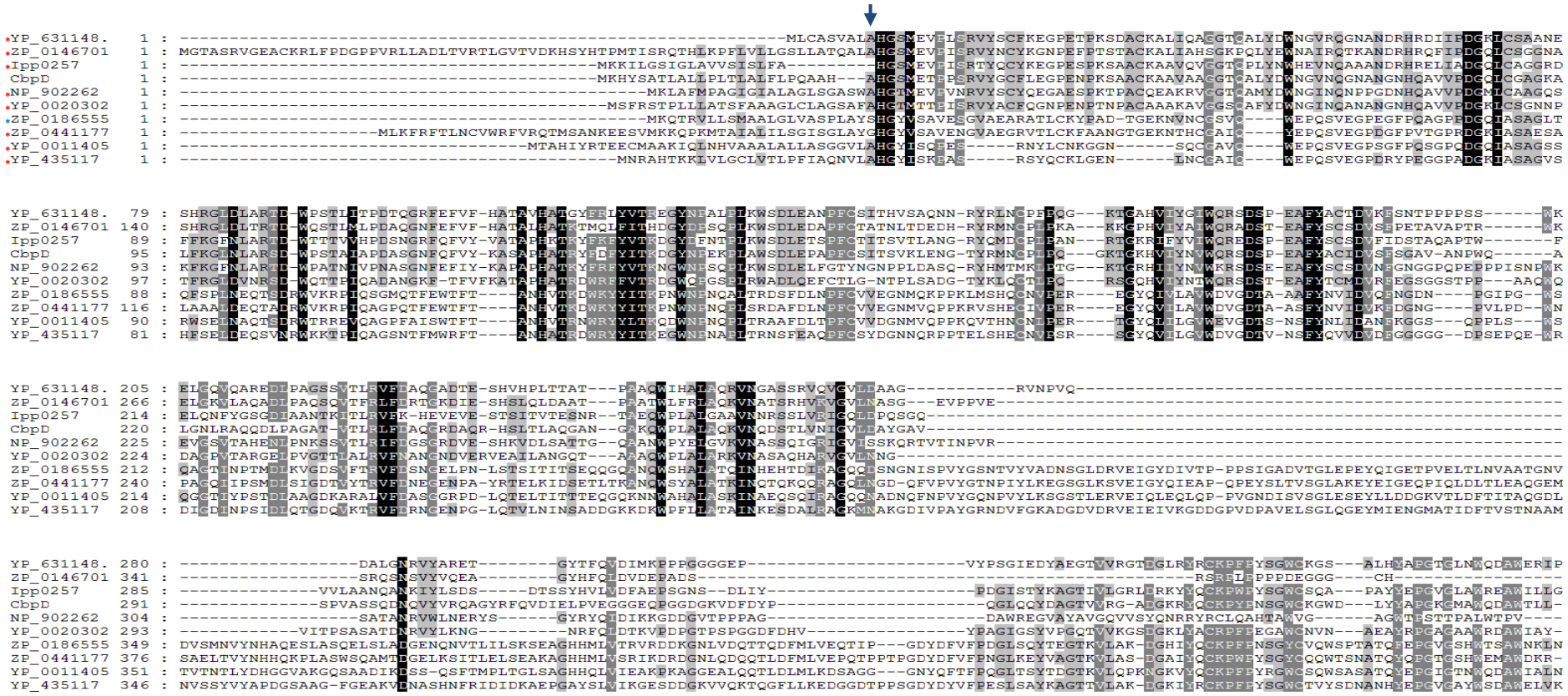


Fig 1.24: Multiple sequence alignment of CbpD homologous proteins. Multiple sequence alignment of proteins that show homology with CbpD when residues 211-389 are submitted to a BLAST search. The blue arrow indicates the first residue after the signal sequence. YP_631148.1; *Myxococcus Xanthus*, ZP_01467014; *Stigmatella aurantiaca*, Ipp0257; *Legionella pneumophila*, CbpD from *P. aeruginosa*, NP_902262; *Chromobacterium violaceum*, YP_002030259; *Stenotrophomonas maltophilia*, ZP_01865556; *Vibrio shilonii*, ZP_0441177; *Vibrio cholerae*, YP_001140518; *Aeromonassalmonicida* YP_43511776; *Vibrio cholera.* *Putative GlcNAc binding proteins, *Putative chitinase. Alignment generated using ClustalW and Gendoc 2.6 (section 2.13).

1.16.4 Chitin-binding proteins of *Photorhabdus*

Photorhabdus organisms are γ-proteobacteria that display the curious property of bioluminescence. Three species are currently recognized: *P. asymbiotica*, *P. luminescens*, and *P. temperata* (Fischer-Le Saux *et al.,* 1999). The latter 2 species have been intensively studied by entomologists because they are virulent insect pathogens. They form a symbiotic relationship with nematodes (*Heterorhabditis* sp.) that invade the larvae of insects. The nematodes regurgitate the bacteria, which kill the insects and provide a food source for the nematodes. Insect-pathogenic nematodes are thought to be harmless to vertebrates and are used in horticulture for biologic control of insects (Duchaud *et al.,* 2003, Gerrard *et al.,* 2006). Gerrard *et al.,* (2006) have shown that *P. asymbiotica* is also found in a symbiotic association with an insect-pathogenic soil nematode of the genus *Heterorhabditis*.

Metabolites of the species *Photorhabdus* are known to lyse fungal mycelia that contain chitin in the cell wall, implying the existence of chitinases (Chen *et al.,* 1994). This observation led to an investigation of exoenzymatic activity of strains of *Photorhabdus* species by Chen *et al.,* 1996. They demonstrated the chitinase activity of various *Photorhabdus* strains by the fact that the proteinaceous preparations of the culture broth hydrolyzed glycol chitin after gel electrophoresis (Chen *et al.,* 1996). In 2003 Duchand *et al.,* published the genome sequence of *Photorhabdus luminescens* and predicted a chitin-binding-like protein Plu2352, (CbpL) and chitinase-like proteins Plu2235, Plu2458 and plu2461. As chitin and its oligomers appear to be predominant surface molecules in invertebrates they may be specifically recognized by microbial symbionts, much as hosts recognize microorganisms through patterned surface molecules such as the peptidoglycans (Chaston and Goodrich-Blair, 2010). While the genome of *P. asymbiotica* has been sequenced, no chitin-binding molecules have thus far been identified.

1.17 Uses of chitin active proteins

As reviewed by Rinuado (2006) and Khoushab and Montarop (2010), chitin and its derivatives have a number of high value added applications in industrial and applied fields due in part to the biodegradability and biocompatibility of these compounds.

The biggest emerging industry use for chitin-based products is in the medical field, where uses for chitin-derived products include drug and vaccine delivery systems, wound and burn treatments (Dai *et al.*, 2009), blood cholesterol lowering agents (Maezaki *et al.,* 1993), anti-clotting agents, and enhanced dissolution of certain drugs such as ibuprofen (Okawa *et al.,* 2003). The key property of chitin-derived products for application in biomedical fields is the immuno-modulating effect. Chitin and its polymers have the ability to enhance macrophage production, activate complement, and have been shown to be effective agents for hemostasis of blood coagulation, thus promoting wound healing (Muzzarelli, 2009, Muzzarelli, 2010, Muzzarelli, 1997, Khoushab and Yamabhai, 2010). Chitin derivates furthermore have antibacterial activity, thought due in part to the positive charge of C2 amino group in glucosamine, a monomer of chitosan. This positive charge interacts with negatively charged microbial cell membrane and lead to leakage of the intracellular constituents of the microorganisms (Dutta *et al.,* 2009). This property has promoted the use of chitosan based films in the medical field, as wound dressings (Dai *et al.*, 2009) and in the food preservation industry where it can delay the growth of microorganisms (Dutta *et al.*, 2009).

Due to their rigid nature chitin and chitosan are widely used to immobilise enzymes and whole cells, allowing for improvements in heterogeneity of the immobilised enzymes. This facilitates easier recovery of both enzyme and product, in addition to increasing enzyme robustness and reusability of supports. This technology has applications in the food industry, such as organic contaminant removal from wastewaters to sophisticated biosensor design (for review see Krajewska, 2004). Rigid chitinous structures in addition, have applications as scaffolds for bone and other natural tissue regeneration, as well as

structures by which three-dimensional formation of tissues are supported (Tsioptsias *et al.,* 2009).

Chitin and chitosan are often used in applications such as bioremediation of heavy metal contamination due to their ability to form specific complexes with a number of ions or dyes and with small organic molecules (Mazeau *et al.*, 1994). Both have also been exploited in the field of paper and textile production, as additives in animal feed and as components of cosmetics. Chitin and chitosan derivatives have wide applications in the field of cosmetics due to their high water-holding capacity and positive anti-microbial characteristics (Jayakumar *et al.,* 2010). The patents for the use of chitin and chitosan derivatives in skin care cosmetics cover strong moisturizing activity, photo-protection from UV and sunlight, anti-aging, skin elasticity-increasing activity, antibacterial and anti-inflammatory effects.

Since chitin metabolism is crucial for fungal and arthropod development, inhibition or de-regulation of the key enzymes are important objectives for the development of fungicides and insecticides, including anti-malaria agents, not least because chitin polymers are absent in vertebrates. Chitinases have been shown to act as biocontrol and anti-fungal agents in addition to having anti-biofouling applications (Bormann *et al.*, 1999, Chang *et al.*, 2010, Dehestani *et al.,* 2010). Recombinant bacterial chitinases can also be used to produce transgenic plants with enhanced resistance against pathogenic fungi (Dehestani *et al.,* 2010, Chang *et al.*, 2010 and Andersen *et al.*, 2005). Chitin-binding modules have furthermore been incorporated into protein expression vectors to facilitate the purification of recombinant proteins (Chong *et al.,* 1997). Chitin has been used to prepare affinity chromatography columns to isolate lectins with GlcNAc affinities and to determine their structure (Vaaje-Kolstad *et al.,* 2005b, Vaaje-Kolstad *et al.*, 2005a) .

Extraction of chitin from crustacean shells is routinely carried out using a chemical demineralisation and deproteinisation process (Goycoolea *et al.,* 2000). The build up of waste chemicals from this process creates an expensive disposal

problem, not least caused by enforced environmental controls and disposal measures. An alternative and inexpensive, environmentally-friendly process for the demineralisation of crab shell has been proposed using the chitin degrading organisms *Pseudomonas aeruginosa* F722 by Oh *et al.* (2007). It has also been proposed that the enzymatic conversion of structural polysaccharides, namely cellulose to biofuel could be improved by exploiting chitin-binding-like proteins to improve substrate accessibility, thus reducing the need for processive enzymes (Eijsink, *et al.*, 2008) .

1.18 Project aims and objectives

The aim of this project was to identify, clone, express, purify and characterise prokaryotic carbohydrate-binding proteins for potential use in GlcNAc detection in important therapeutic glycoproteins. Chitin-binding proteins from four different organisms; *S. marcescens*, *P. aeruginosa*, *P. luminescens* and *P. asymbiotica* were selected for this study. Recombinant proteins would be produced in *Escherichia coli* based expression systems, with particular focus on the selection of an appropriate vector system for each, capable of producing soluble, active protein. Following successful expression of these molecules, protocols would be developed to yield relatively large quantities of highly purified protein, for sugar/oligosaccharide binding characterisation studies. CBP21 would be mutated within the proposed sugar-binding site to confirm if the sugar specificity of these proteins could be manipulated and ultimately altered.

2.0 Materials and Methods

2.1 Bacterial strains, primers and plasmids

Table 2.1: Bacterial Strains

Strain	Genotype	Features/Uses	Source
Escherichia Coli			
XL10 Gold	Δ(*mcr*A)183 Δ(*mcr*CB-*hsd*SMR-mrr)173 *end*A1 *rec*A1 *rel*A1 *gyr*A96 *sup*E44 *thi*-1 *lac* [FΔ *pro*AB *lac*IqZΔM15 ::Tn10(tetR) Amy (CmR)]	All purpose cloning strain. Produces stable plasmid DNA.	Stratagene
JM109	FΔtraD36, *pro*AB+ *lac*Iq, Δ*lac*Z M15 *end*A1 *rec*A1 *hsd*R17(rk-, mk+) *mcr*A *sup*E44 - *gyr*A96 *rel*A1 Δ *(lacproAB)* *thi-1*	All purpose cloning strain. Produces stable plasmid DNA.	Sigma
KRX	[F´, *tra*D36, Δ*omp*P, *pro*A+B+, *lac*Iq, Δ(*lac*Z)M15] Δ*omp*T, *end*A1, *rec*A1, *gyr*A96 (Nalr), *thi*-1, *hsd*R17 (rk–, mk+), *rel*A1, *sup*E44, Δ(*lac*-*pro*AB), Δ(*rha*BAD)::T7 RNA polymerase	Expression strain	Promega
BL21 (DE3)	FΔ *dcm* *omp*T *hsd*SB(rB-,mB-) *gal*	Expression strain	Novagen
Serratia marcescens			
Klon	Wild Type		NCIMB no 2302
Pseudomonas aeruginosa			
PAO1	Wild type		Dr Keith Poole
Photorhabdus luminescens			
TT01	Wild type	Rif^R	Dr David Clarke
Photorhabdus asymbiotica			
ATCC43949	Wild type		Dr David Clarke

Table 2.2: Plasmids

Plasmid	Description	Source
Vectors		
pQE30	Amp^R Expression vector for N-terminally-tagged $(His)_6$ proteins, T5 promoter/lac operon	Qiagen
pQE60	Amp^R Expression vector for C-terminally tagged $(His)_6$ proteins, T5 promoter/lac operon	Qiagen
pQE_DsbA	pQE60 (Qiagen) derived plasmid with DsbA signal sequence for periplasmic expression of C-terminally $(His)_6$ tagged proteins. Amp^R, T5 promoter/lac operon	Ryan, 2006 (Ph.D Thesis)
pQE_PelB	pQE60 (Qiagen) derived plasmid with pelB signal sequence for periplasmic expression of C-terminally $(His)_6$ tagged proteins. Amp^R, T5 promoter/lac operon	Ryan, 2006 (Ph.D Thesis)
Constructs		
pCBP21_30	pQE30 containing cbp21 ORF	This Work
pCBP21_60	pQE60 containing cbp21 ORF	This Work
pCBP2130_N	pQE30 containing cbp21 ORF without N-terminal leader sequence	This Work
pCBP2160_N	pQE60 containing cbp21 ORF without N-terminal leader sequence	This Work
pCBP21wt	pQE60 containing cbp21 with a stop codon encoded before the $(His)_6$ tag	This Work
pCBP21Y54A	pQE60 containing cbp21 with Y54A mutation	This Work
pCBP21E55A	pQE60 containing cbp21 with E55A mutation	This Work
pCBP21P56A	pQE60 containing cbp21 with P56A mutation	This Work
pCBP21Q57A	pQE60 containing cbp21 with Q57A mutation	This Work
pCBP21S58A	pQE60 containing cbp21 with S58A mutation	This Work
pCBP21E60A	pQE60 containing cbp21 with E60 mutation	This Work
pCBP21T111A	pQE60 containing cbp21 with T111A mutation	This Work
pCBP21H114A	pQE60 containing cbp21 with H114A mutation	This Work
pCBP21D182A	pQE60 containing cbp21 with D182A mutation	This Work
pCbpD_60	pQE60 vector containing cbpD ORF	This Work
pCbpD_UN	pQE60 containing cbpD with a stop codon encoded before the $(His)_6$ tag	This Work
pPelBCbpD	pQE60 containing cbpD ORF with PelB periplasmic leader sequence	This Work
pDsbACbpD	pQE60 containing cbpD ORF with DsbA periplasmic leader sequence	This Work

Table 2.2 continued……..

Plasmid	Description	Source
pCbpda	pQE60 containing cbpD ORF residues 1-212	This Work
pCbpDc	pQE60 containing cbpD ORF residues 117-389	This Work
pCbpDb	pQE60 containing cbpD ORF residues 212-389	This Work
pCbpL_60	pQE60 vector containing ORF *P.luminescens*	This Work
pCbpA_60	pQE60 vector containing ORF *P.asymbiotica*	This Work

Table 2.3: Primer sequences (Synthesised by Sigma-Aldrich, U.K.)

Name	Primer Sequence (5'-3')	T_M (°C)
Pseudo-R2	GTCAGTAGATCTCAGCAGGGTCCAGGCGTCC	72.6
Pseudo-R4	P-TTACAGCAGGGTCCAGGCGTCCTG	69.8
Pseudo-R5	GTCAGTAGATCTGGCGCCGCTGAAGCTCACGTC	72.1
Pseudo-R6	P-CAGCAGGGTCCAGGCGTCCTGCCA	75.1
pQE60_F1	P-TAAAGATCTCATCACCATCACCATCACTAAGCTTAATTA	61.5
Pseudo-F3	GTCAGTCCATGGGAAAACACTACTCAGCCACCC	78.7
Pseudo-F4	P-CACGGCTCGATGGAAACGCC	65.5
Pseudo-F7	GTCAGTGCGGCCGCCCACGGCTCGATG	76.8
Pseudo-F8	CAGTCCATGGGAGCCGTCGCCAACCCCTGG	78.9
Pseudo-F9	CAGTCCATGGGAAGCGGCAACTTCCAGTTCGTCTAC	72.2
Lumin-F3	GTCAGTCCATGGGATATAAACATAAAGTGAAAGTG	75.5
Lumin-F4	GTCAGTCCATGGGACATGGTTATATTGATAGCCCA	61.2
Lumin-R2	GTCAGTAGATCTAGCAGGGCTAATTGTTGTGTC	63.8
Asymb-F3	GTCAGTCCATGGGATACAAACATAAAGTGAGA	62.6
Asymb-F4	P-CATGGTTATATTGATAGTCCAGGAAG	74.3
Asymb-R2	GTCAGTAGATCTGGCTGGGTTAATGGTTGT	64.8
Marces-F1	GTCAGTGGATCCATGAACAAAACTTCCCGTACC	67.5
Marces-R1	GTCAGTAAGCTTTTATTTGCTCAGGTTGACGTCGATCGCCTG	69.1
Marces-F3	GTCAGTCCATGGGAAACAAAACTTCCCGTACC	76.8
Marces-R2	GTCAGTAGATCTTTTGCTCAGGTTGACGTC	81.7
Marces-F4	P-CACGGTTATGTCGAATCGCCGG	76.8

Table 2.3 contd....

Marces-R3	P-TCCCATGGTTAATTTCTCCTCT	70.9
30-R1	P-CATGGATCCGTGATGGTGATGGTG	64.5
Marces-F5	GTCAGTGCGGCCGCCCACGGTTATGTCGAAT	69.7
Marces-R4	P-TTTGCTCAGGTTGACGTCGATCGCCTG	67.9
55-60A_R	P-GCTGCCGCACTGCGTGTTGAGTTGCAG	70.2
Y54A_F	P-GTGCAGGCGGGACCGCAGAGCGTCGAA	75.5
E55A_F	P-GTGCAGTACGCACCGCAGAGCGTCGAA	70.4
P56A_F	P-GTGCAGTACGAAGCGCAGAGCGTCGAA	68.3
Q57A_F	P-GTGCAGTACGAACCGGCGAGCGTCGAA	70.7
S58A_F	P-GTGCAGTACGAACCGCAGGCGGTCGAA	71.2
E60A_F	P-GTGCAGTACGAACCGCAGAGCGTCGCAGGCCTG	75.3
E55_R_F	P-GTGCAGTACNNNCCGCAGAGCGTCGAA	66.3-73.8
T111A_F	P-TTTACCTGGAAGCTGGCCGCCCGTCAC	72.6
T111A_R	P-GGAGTTCGGGCCGGTTTTCAGGTTGAGCTT	71.6
H114A F	P-TTTACCTGGAAGCTGACCGCCCGTGCCAGCACAACCAGCT GGCGC	80.5
H114_R_F	P-TTTACCTGGAAGCTGACCGCCCGTNNNAGCACAACC	71.1-76.7
T111A_R	P-GGAGTTCGGGCCGGTTTTCAGGTTGAGCTT	71.6
D182A_F	P-GTGTGGGACATAGCCGCCACCGCCAACGCCTTCTAT	75.6
D182A_R	P-GGCAAGGATCACGTGCGAACCGCTGCGATC	72.6
Sequencing Primers		
QE-F	CCCGAAAAGTGCCACCTG	61.6
QE-R	GTTCTGAGGTCATTACTGG	53.5

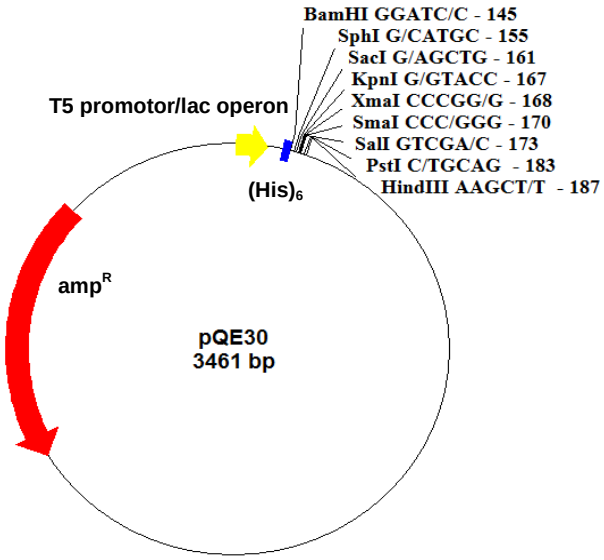


Figure 2.1: The pQE30 vector (Qiagen). The 3461bp pQE30 expression vector contains the following features; the MSC is located before the $(His)_6$ amino acid sequence (blue) which allows for the expression of C-terminally $(His)_6$ tagged proteins. This is under control of the T5 promotor/lac operon (yellow). The *bla* gene encodes beta-lactamase which confers ampicillin resistance to the bacteria (red). This image was generated using pDRAW32.

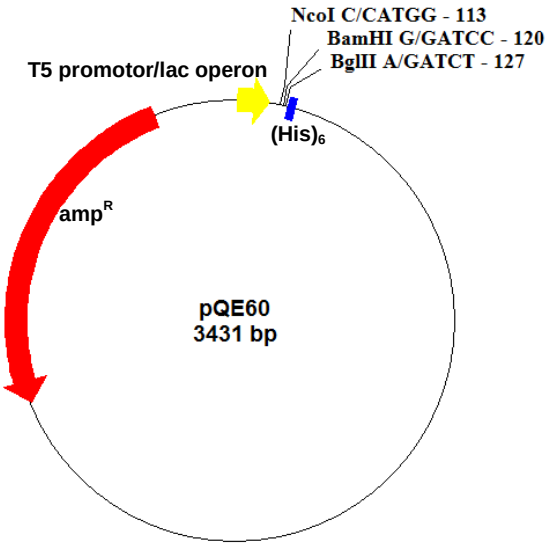


Figure 2.2: The pQE60 vector. The 3431bp pQE60 vector from Qiagen contains the following features. The MSC is located before the $(His)_6$ amino acid sequence (blue) which allows for the expression of C-terminally $(His)_6$ tagged proteins. This is under control of the T5 promotor/lac operon (yellow). The *bla* gene encodes beta-lactamase which confers ampicillin resistance to the bacteria (red). This image was generated using pDRAW32.

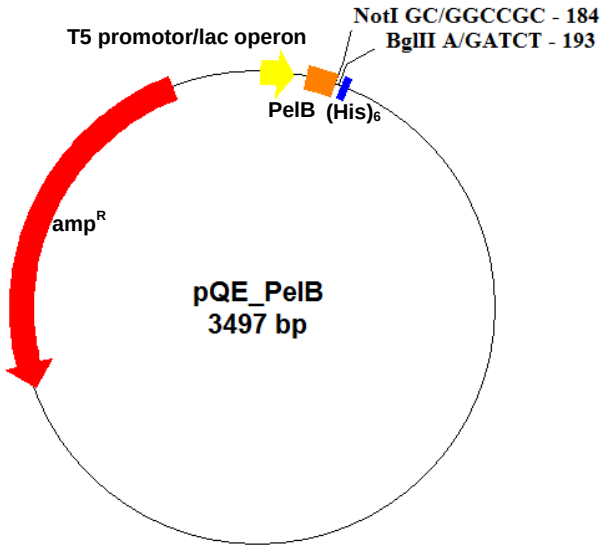


Figure 2.3: pQE_PelB vector. pQE60-PelB derived vector contains the following features. The MSC is located downstream from the PelB leader sequence (orange). The $(His)_6$ amino acid sequence (blue) allows for the expression of C-terminally $(His)_6$ tagged proteins. This is under control of the T5 promotor/lac operon (yellow). The *bla* gene encodes beta-lactamase which confers ampicillin resistance to the bacteria (red). This image was produced using pDRAW32.

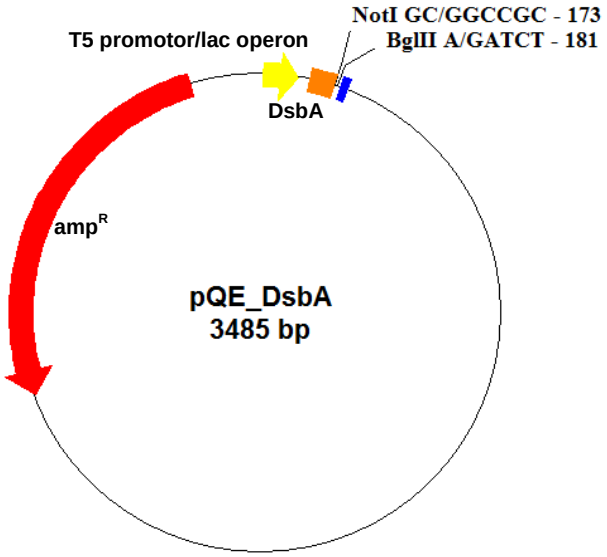


Figure 2.4: The pQE_DsbA vector. pQE60-DsbA derived vector contains the following features. The MSC is located downstream from the DsbA leader sequence (orange). The $(His)_6$ amino acid sequence (blue) allows for the expression of C-terminally $(His)_6$ tagged proteins. This is under control of the T5 promotor/lac operon (yellow). The *bla* gene encodes beta-lactamase which confers ampicillin resistance to the bacteria (red). This image was generated using pDRAW32.

2.2 Microbiological Media

Microbiological media was obtained from LAB M. Sterilisation was achieved by autoclaving at 121°C for 20min, unless otherwise stated.

Luria Bertani Broth (LB)

Tryptone	10 g/L
NaCl	10 g/L
Yeast Extract	10 g/L
pH	7.0

NaOH was used to adjust to pH 7.0 and the volume brought to 1 L using dH_2O. The solution was then sterilized by autoclaving. For the production of solid agar plates agar (15 g/L) was added prior to sterilization.

Super Optimal Broth (SOB)

Tryptone	20 g/L
NaCl	500 mg/L
Yeast Extract	5 g/L
KCl	2.5 mM
pH	7.0

The solution was autoclaved and allowed to cool to 55ºC before $MgCl_2$ and $MgSO_4$ were added to a final concentration of 10 mM from sterile 1 M stock solutions.

2.3 Solutions and Buffers

All chemicals and solutions were obtained from Sigma Aldrich unless otherwise stated. All chemicals were ACS grade.

TB Buffer

PIPES	10 mM
$CaCl_2$	15 mM
KCl	250 mM
pH	6.7

Following pH adjusted with KOH $MnCl_2$ was added to a final concentration of 50 mM. The solution was filter sterilised through a 0.22µm filter membrane and stored at 4ºC.

RF1

RbCl	100 mM
$CaCl_2$	10 mM
Potassium Acetate	30 mM
Glycerol	15% (v/v)
pH	5.8

When the pH had been adjusted with HCl $MnCl_2$ was added to a final concentration of 50 mM. The solution was filter sterilised through a 0.22 µm filter membrane and stored at 4ºC.

RF2

RbCl	10 mM
$CaCl_2$	75 mM
MOPS	10 mM
Glycerol	15% (v/v)
pH	6.8

The solution was filter sterilised through a 0.22µm filter membrane and stored at 4ºC.

Solution 1 of the 1-2-3 method

Glucose	0.5 M
Na_2-EDTA	0.1 M
Tris-HCl	1 M

Solution 2 of the 1-2-3 method

NaOH	1 M
SDS	1%

Solution 3 of the 1-2-3 method

KOAc	3 M
pH	4.8

To 60mL of 5M potassium acetate, 11.5mL of glacial acetic acid and 28.5mL of dH_2O was added. The resulting solution was 3M with respect to potassium and 5M with respect to acetate.

TE Buffer

Tris	10 mM
Na_2-EDTA	1 mM
pH	8.0

TAE Buffer (50X)

Tris	242 g/L
Glacial Acetic Acid	57.1 mL/L
EDTA	100 mL/L (of 0.5M stock)
pH	8.0

Western Blot Transfer Buffer

Tris	25 mM
Glycine	150 mM
Methanol	10% (v/v)

TBS Buffer

Tris	20 mM
NaCl	150 mM
$CaCl_2$	1 mM
$MgCl_2$	1 mM
pH	7.6

When the pH had been adjusted with HCl $MnCl_2$ was added to a final concentration of 1mM. For TBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v)

PBS (10x)

Na_2HPO_4	10.9 g/L
NaH_2PO_4	3.2 g/L
NaCl	90 g/L

A 1X has a pH of 7.2. For PBST, the detergent Triton X-100 was added to a final concentration of 0.1% (v/v)

Agarose Gel loading Dye

Bromophenol Blue	0.25% (w/v)
Xylene Cyanol	0.25% (w/v)
Ficoll (Type 400)	15% (w/v)

Bromophenol Blue and/or Xylene Cyanol were used as appropriate. On a 1% agarose gel bromophenol blue and Xylene Cyanol migrate approximately with the 300bp and 4,000bp fragments respectively.

Ethidium Bromide Stain

A 10mg/mL stock solution in dH_2O was stored at 4ºC in the dark. For the staining of agarose gels, 100 µL of stock solution was mixed into 1 L of dH_2O. The staining solution was kept in a plastic tray and covered to protect against light. Used ethidium bromide was collected and the ethidium bromide was extracted by mixing with de-staining bag (GeneChoice) overnight. The clear liquid was disposed of routinely, and the ethidium was incinerated.

SDS-PAGE Buffer (5X)

Tris	15 g/L
Glycine	72 g/L
SDS	5 g/L
pH	8.3

SDS-PAGE Sample Buffer (10X) 8mL

SDS	3.2 mL 10% Stock
2-Mercaptoethanol	0.8 mL
Tris	2.0 mL mM Stock, pH 6.8
Glycerol	1.6 mL
Bromophenol Blue	0.5% (w/v)
H_2O	0.4 mL

Coomassie blue stain solution

dH_2O	50% (v/v)
Methanol	40% (v/v)
Acetic Acid	10% (v/v)
Coomassie Blue	0.25% (w/v)

Coomassie blue de-stain solution

dH_2O	50% (v/v)
Methanol	40% (v/v)
Acetic Acid	10% (v/v)

SDS solution

SDS	10% (w/v)

Denaturing Solution

NaOH	0.5 M
NaCl	1.5 M

Neutralisation solution

NaCl	1.5 M
Tris	0.5 M
pH	7.4

20x Saline Sodium Citrate (SSC)

NaCl	3 M
Sodium Citrate	300 mM

Lysis Buffer

NaH_2PO_4	20 mM
NaCl	0.5 M
Imidazole	20 mM
pH	7.4

Resuspension Buffer

Tris	20 mM
pH	8.0

Isolation Buffer

Urea	2 mM
Tris	20 mM
NaCl	0.5 M
TritonX-100	2% (v/v)
pH	8.0

Binding Buffer

Guanidine Hydrochloride	6 M
Tris	20 mM
NaCl	0.5 M
Imidazole	10 mM
2-mercaptoethanol	1 mM
pH	8.0

Wash Buffer

Urea	6 M
Tris	20 mM
NaCl	0.5 M
Imidazole	20 mM
2-mercaptoethanol	1 mM

Refolding Buffer

Tris	20 mM
NaCl	0.5 M
Imidazole	20 mM
2-mercaptoethanol	1 mM

Elution Buffer

Tris	20 mM
NaCl	0.5 M
Imidazole	0.5 M
2-mercaptoethanol	1 mM

Column Stripping Buffer

NaH_2PO_4	20 mM
NaCl	0.5 M
EDTA	50 mM
pH	7.4

TMB Solution (10mL)

Citric Acid	1.37 mL 0.1M Stock
Sodium Citrate	3.63 mL 0.1M Stock
dH_2O	5 mL

2 mg of TBM powder was dissolved in 200 µL DMSO and added to 9.8 mL TMB solution. 2 µL of H_2O_2 was added directly before use.

DNS reagent

Solution A

Sodium potassium tartrate	300 g

The sodium potassium tartrate was dissolve in 500 mL dH_2O and heated to 40ºC.

Solution B

NaOH	16 g
DNS	10 g

The NaOH was dissolved in 200 mL dH_2O and heated to 90ºC before adding 10 g DNS.

Solution A and B were mixed together and allowed to cool. Phenol was added to a final concentration of 0.2% The solution was made up to 1 L with dH_2O. The solution was stored in the dark at room temperature.

2.4 Antibiotics

Ampicillin was prepared in dH_2O at a concentration of 100 mg/mL and stored at -20ºC. The working concentration in *E.coli* cultures was 100 µg/mL.

2.5 Storing and culturing of bacteria

Bacterial strains were stored as 33% glycerol stocks. 1000 µL of an overnight culture was mixed with 500 µL sterile 80% glycerol in a microfuge tube. If the bacterial strains contained plasmids, the selective antibiotic was included in the culture. Duplicate stocks were stored at –20°C and –80°C. Working stocks streaked on LB agar plates, containing antibiotics where appropriate, were stored at 4ºC.

2.6 Isolation and purification of DNA

2.6.1 Isolation of Genomic DNA

Genomic DNA was isolated using the Promega Wizard genomic DNA purification kit. This kit was used according to the manufacturer's instructions: 1mL of a bacterial culture in a microfuge tube was centrifuged at 13,000 rpm for 2 min (using a LabNet Spectrafuge, fixed rotar). The supernatant was discarded and the cells were re-suspended in 600 µL of nuclei lysis solution. Samples were incubated at 80°C for 5 min to lyse the cells. Cell lysate was cooled to room temperature. Next, 3 µL of RNase solution was added to the cell lysate and mixed by inversion. Samples were incubated at 37°C for 15-60 min and then cooled to room temperature. Subsequently 200 µL of protein precipitation solution was added to the RNase treated cell lysate. Samples were vortexed vigorously for 20 s to mix the protein precipitation solution with the cell lysate. Samples were incubated on ice for 5 min, followed by centrifugation at 13,000 rpm for 3 min. The supernatant was removed and transferred to a fresh tube. 500 µL of phenol chloroform isoamylalcohol (25:24:1) was added and the sample was centrifuged at 13,000 rpm for 5 min. Supernatant was removed a second time and again transferred to a fresh microfuge tube containing 500 µL of phenol chloroform isoamylalcohol (25:24:1) and centrifuged at 13,000 rpm for 5 min. The supernatant was transferred to a fresh microfuge tube containing 600 µL of room temperature isopropanol, and mixed by inversion. The sample was centrifuged at 13,000 rpm for 2 min to pellet the genomic DNA. The supernatant

was removed and the pellet was washed by the addition of 600 µL of 70% ethanol, and spun at 13,000 rpm for 2 min. The ethanol was removed carefully so as not to disturb the pellet and the pellet was allowed to air dry for 10-15 min. Finally, 100 µL of DNA rehydration solution was added to the tube. Samples were left to rehydrate overnight at 4°C. Genomic DNA was stored at -20°C.

2.6.2 Isolation of Plasmid DNA

Two procedures for the isolation of plasmid DNA were variably employed. The 1-2-3 method was used for convenient plasmid isolation from large numbers of samples, mostly for the purposes of screening. The Sigma Plasmid Mini Prep Kit was used to prepare consistently pure and supercoiled DNA mostly for the purpose of DNA sequencing.

2.6.2.1 1-2-3 Method

Bacterial growth was taken off an agar culture plate with a sterile loop and re-suspended in 200 µL of Solution 1. Alternatively, 1.5 mL of a bacterial culture in a microfuge tube was centrifuged at 13,000 rpm for 2 min to collect the cells. The supernatant was discarded and the cell pellet re-suspended in 200 µL of Solution 1 (Section 2.3). The re-suspension was left for 5 min at room temperature. Subsequently 200 µL of Solution 2 (Section 2.3) was added and the tube was mixed by inversion and left for 5 min at room temperature. Finally 200 µL of Solution 3 (Section 2.3) was added, the tube was mixed by inversion and left at room temperature for 10 min. Cell debris was collected by centrifugation at 13,000 rpm for 10 min. The supernatant was transferred to a fresh microfuge tube with 500 µL of phenol chloroform isoamylalcohol (25:24:1) and mixed by brief vortexing. Upon centrifugation at 13,000 rpm for 5 min the mixture is divided into an upper aqueous and lower organic layer. The aqueous (top) layer was removed to a new microfuge tube with an equal volume of isopropanol and mixed by inversion. The tube was then centrifuged at 13,000 rpm for 5 min to pellet the plasmid DNA. The pellet was washed with 70% ethanol and then dried briefly in a SpeedVac (Savant) vacuum centrifuge. The plasmid DNA was

resuspended in 30 µL of TE buffer (Section 2.3) and 1 µL of Ribonuclease A was added to digest co-purified RNA. Plasmid DNA was stored at –20ºC.

2.6.2.2 Sigma GenElute Plasmid Mini Prep Kit

This kit was used according to the manufacturer's instructions: 1.5 mL of bacterial culture in a microfuge tube was centrifuged at 13,000 rpm for 2 min to pellet the bacterial cells. The supernatant was discarded and the pellet was re-suspended in 200 µL of re-suspension solution. To lyse the cells 200 µL of lysis solution was added and mixed by inversion, the mixture was incubated for 5min at room temperature. Following lysis 350 µL of neutralisation solution was added and mixed by inversion to precipitate the cell debris, lipids, proteins and chromosomal DNA. The mixture was incubated for 10 min at room temperature. The precipitate was collected by centrifugation at 13,000 rpm for 10 min. The supernatant was transferred to spin column in a microfuge tube and centrifuged for 30 s to bind the plasmid DNA. The flowthrough was discarded and 750 µL of wash solution was added and centrifuged at 13,000 rpm for 30 s, this step was repeated. The supernatant was discarded and the column was spun at 13,000 rpm for 2 min to dry the matrix. The spin column was transferred to a fresh microfuge tube, up to 100 µL of DNA elution/TE buffer was added. The DNA was eluted by centrifugation at 13,000 rpm for 30 s.

2.7 Agarose gel electrophoresis

DNA was analysed by electrophoresing through agarose gels in a BioRad horizontal gel apparatus. Agarose was prepared by adding the required amount (concentration typically 0.7-2%) to TAE buffer and was dissolved by boiling. Agrose was stored at 60°C. The agarose solution was poured into plastic trays and allowed to set with a plastic comb fitted to create sample wells. TAE buffer was used as the running buffer. Agarose gel loading dye (Section 2.3) was mixed with the DNA samples to facilitate loading and to give indication of migration distance during electrophoresis. Gels were run at 140 volts for 20-40 min depending on size of gel. Gels were stained for 20 min by immersion in an

ethidium bromide staining solution. Gels were visualised using a UV transilluminator coupled with an image analyser to capture the image to a PC. On every gel 0.5 µg of 1 Kb Plus DNA Ladder (Invitrogen) was run as a molecular size marker.

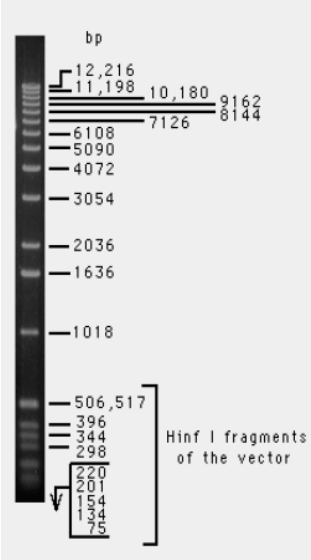


Fig 2.5: 1kb Molecular Marker (Invitrogen). A representative image of the 1 kb DNA molecular marker used in this study.

2.8 Isolation of DNA from agrose gels

The Hi Yield^{TM} Gel/PCR DNA extraction kit from RBC Bioscience was used to isolate DNA from agrose gels. The kit was used according to the manufacturer's instructions: The DNA band to be isolated from the agrose gel was excised using a scalpel. Up to 300 mg of gel slice was transferred to a microfuge tube. 500 µL of Detergent free (DF) buffer was added to the sample and mixed by vortexing. The sample was incubated at 55°C for 10-15 min until the gel slice was completely dissolved. During incubation the tube was inverted every 2-3 min. A DF column was placed in a collection tube. Up to 800 µL of the dissolved gel sample was transferred to the DF column. The DF column was centrifuged at 13,000 rpm for 30 s. The flowthrough was discarded and 750 µL of wash solution was added to the DF column which was then centrifuged at 13,000 rpm for 30 s. The flowthrough was again discarded and the column was dried by centrifugation at 13,000 rpm for 2 min. 30 µL of DNA elution buffer was added

to the centre of DF column matrix. The column was left to stand for 2 min at room temperature and subsequently centrifuged at 13,000 rpm for 30 s to elute the isolated DNA. Isolated DNA was stored at -20°C.

2.9 Preparation of high efficiency competent cells

Two methods were used to prepare competent cells, the TB method and the RF method. The RF method was preferred as it is the easier of the two methods and was found to give the best efficiencies.

2.9.1 The RF method

5 mL of LB broth containing the relevant antibiotics was inoculated with a single colony of the desired bacterial strain from a plate stock and cultured overnight at 37°C. A 1 L flask with 200 mL of SOB broth was inoculated with 2 mL of the overnight culture and incubated at 37ºC shaking at 180-220 rpm. When the culture had reached an OD_{600} of ~0.5 (early-mid exponential phase) the flask was cooled in ice water. All subsequent transactions took place at 4ºC. The culture was transferred to a sterile centrifuge bottle. The cells were collected by centrifugation at 3,000 rpm for 5 min (using a Beckman JA-14 rotor). The supernatant was decanted and the cells gently re-suspended in 60 mL of chilled RF1 buffer (Section 2.3). The suspension was left on ice for 90 min. The cells were again collected by centrifugation at 3,000 rpm for 5 min. The supernatant was decanted and the cells gently re-suspended in 8 mL of chilled RF2 buffer (Section 2.3). Aliquots of 400 µL were prepared in sterile 1.5 mL microfuge tubes and flash frozen using –80°C ethanol. The competent cells were stored at –80°C.

2.9.2 The TB Method

5 mL of LB broth containing the relevant antibiotics was inoculated with a single colony of the desired bacterial strain from a plate stock and cultured overnight at

37ºC. A 1 L flask with 200 mL of SOB broth was inoculated with 2 mL of the overnight culture and incubated at 37ºC shaking at 180-220 rpm. When the culture had reached an OD_{600} of ~0.4 (early-mid exponential phase) the flask was cooled in ice water. All subsequent transactions took place at 4°C. The culture was transferred to a sterile centrifuge bottle. The cells were collected by centrifugation at 4,500 rpm for 5 min.The supernatant was decanted and the cells gently re-suspended in 80 mL of chilled TB buffer (Section 2.3). The suspension was left on ice for 10 min. The cells were again collected by centrifugation at 5,500 rpm for 5 min. The supernatant was decanted and the cells gently re-suspended in 15 mL of chilled TB buffer. 7% (v/v) DMSO was slowly added to the suspension. The suspension was incubated on ice for 10 min. Aliquots of 800 µL were prepared in sterile 1.5 mL microfuge tubes and flash frozen using –80°C ethanol. The competent cells were stored at –80°C.

2.9.3 Transformation of competent cells

An aliquot of competent cells was thawed on ice, with 200 µL of the cell suspension gently mixed with 2 µL of plasmid DNA in a sterile 1.5 mL microfuge tube. The mixture was incubated on ice for 30 min to allow the DNA to bind to the cells. The cells were heat-shocked at 42°C for 45 s and placed back on ice for 2 min. Subsequently 800 µL of LB broth was added to the cells and they were incubated for 1 hour at 37°C. 200µL of the suspension was plated on LB agar containing appropriate antibiotics and incubated at 37°C overnight.

2.9.4 Determination of cell efficiency

Competent cell efficiency is defined in terms of the number of colony forming units obtained per µg of transformed plasmid DNA. A 10 ng/µL stock of pUC18 plasmid DNA was diluted to 1 ng/µL, 100 pg/µL and 10 pg/µL. 1 µL of each dilution was transformed as described above. The cell efficiency was calculated from the number of colonies obtained, taking into account the dilution factor and the fraction of culture transferred to the spread plate.

2.10 Enzymatic reactions

All enzymes and relevant buffers were obtained from Invitrogen Life Technologies®, New England Biolabs® or Sigma Corporation and were used according to the manufacturer's instructions.

2.10.1 Polymerase chain reaction

PCR reactions (Mullis and Faloona 1987) were carried out using a Veriti Thermocycler (Applied Biosystems).

2.10.1.1 Standard PCR reaction mixture

Template DNA	1 µL
dNTP's	1 µL
Primers (10 µM)	1 µL of each
Buffer (10X)	5 µL
dH_2O	40 µL
Taq polymerase	1 µL

2.10.1.2 Standard PCR programme cycle for Phusion *Taq* polymerase reactions

Stage 1:	Step 1: 95ºC for 10 min
Stage 2:	Step 1: 95ºC for 30 s
	Step 2: Annealing temperature for 30 s
	Step 3: 72ºC for 30 s per kb to be synthesised
Stage 3:	Step 1: 72ºC for 10 min

Stage 2 was typically carried out for 30 cycles as per manufacturer's guidelines. A 1:3 dilution of Phusion *Taq* polymerase was carried out before use in PCR.

2.10.2 Ligation reaction

PCR product	6 µL
Vector	10 µL
T4 DNA Ligase	1 µL
Ligase Buffer (10X)	2 µL
dH_2O	1 µL

Reactions were incubated for 3 hrs at room temperature or for 16 hrs at 4ºC.

2.10.3 Glycosyl hydrolase reaction

Glycoprotein	10 µL (100 µg)
Buffer	2 µL
Enzyme	2 µL
dH_2O	6 µL

Reactions were incubated overnight at 37ºC.

2.11 Site specific mutagenesis

Mutations were introduced into open reading frames on plasmid constructs by PCR amplification using complementary phosphorylated primers carrying the desired mutation. A standard PCR reaction mixture as show in Section 2.10.1.1 was set up. The PCR was carried out using the high fidelity Phusion *Taq* polymerase. The standard program cycle for Phusion *Taq* polymerase (Section 2.10.1.2) was used. Typically an extension time of 2 min was used to amplify plasmids of 3.5-4 kb. The template DNA was eliminated by digestion with *Dpn*I restriction endonuclease. *DpN*I is biased towards a methylated recognition sequence. It selectively digests the template DNA from dam^+ *E.coli* strains over the newly synthesized DNA. The amplified plasmid, with the incorporated mutation, is circularized by ligation which was facilitated by the use of phosphorylated primers.

2.12 DNA Sequencing

Recombinant clones and potential mutants were verified by DNA sequencing. Commercial sequencing services were provided by MWG Biotech/Eurofins and Source BioScience gene service. Suitable sequencing primers for standard vectors were sent with samples to be analysed. Samples were sent as plasmid preparations (see section 2.6.2.2).

2.13 *In silico* analysis of DNA sequences

To identify homologous protein and DNA sequences deposited in GenBank, the BLAST programme (Altschul, *et al.* 1998) available at NCBI (www.ncbi.nlm.nih.gov) was used. DNA sequences for the strains used in this study were obtained from the following sequencing studies; *Pseudomonas aeruginosa* PAO1 (NC_002516), *Photorhabdus luminescens* (NC_005126), *Photorhabdus asymbiotica* (NC_012962) and *Serratia marcescens* BJL200 (AY665558). DNA and protein sequences were aligned using the ClustalW (Thompson, *et al.*, 1994) available at http://www.ebi.ac.uk/Tools/clustalw/ and Gendoc programme, available to download at http://www.nrbsc.org/. DNA sequences were analysed for restriction sites using Webcutter 2.0 programme (http://rna.lundberg.gu.se/cutter2/). Protein imaging software used was Deepview Swiss PdbViewer from http://www.expasy.org/spdbv/ and PyMol from Delano Scientific available from http://pymol.sourceforge.net/.

2.14 Standard expression culture

An LB plate with the appropriate antibiotic was streaked from a glycerol stock of the strain containing the expression plasmid. One colony was selected and used to inoculate 5 mL of LB broth containing the appropriate antibiotic, and grown over night at 37°C while shaking. A 1 L baffled conical flask containing 500 mL of LB broth was inoculated with 5 mL of the overnight culture and the appropriate antibiotic added. The culture was incubated at 37°C, shaking at 180-220 rpm, until an optical absorbance (A_{600}) of 0.4-0.6 was reached. IPTG was

added to a final concentration of 50 µM to induce expression. The culture was allowed to incubate at 30°C for a defined period. The culture was then centrifuged at 4,000 rpm for 10 min (using a Beckman JA-14 rotor) to pellet the cells. The supernatant was autoclaved and discarded, and the pellets were stored at -20°C.

2.15 Preparation of cleared lysate for protein purification

Cell lysis was carried out using sonication and/or lysozyme.

2.15.1 Cell lysis by sonication

A cell pellet from a 500 mL expression culture was re-suspended in 50 mL lysis buffer containing 20 mM Imidazole, pH 7.6. The cells were disrupted with a 3 mm micro-tip sonicator (Sonics & Materials Inc.) using 2.5 s, 40 kHz pulses for 60 s. The lysate was centrifuged at 4,000 rpm for 20 min at 4ºC (using a Rotanta 460 R rotar) to pellet any insoluble material. The cleared lysate was transferred to a fresh universal container and passed through a 0.2 µM filter to remove any remaining cell debris and stored at 4°C.

2.15.2 Cell lysis using lysozyme

A cell pellet from a 500 mL expression culture was re-suspended in 50 mL lysis buffer containing 20 mM Imidazole, 10 mg/mL lyzozyme and 10 µg/mL DNase, pH 7.6. The cells were incubated at 37°C for 30 min. The lysate was centrifuged at 4,000 rpm for 20 min at 4°C to pellet any insoluble material. The cleared lysate was transferred to a fresh universal container and passed through a 0.2 µM filter to remove remaining cell debris and stored at 4°C.

2.15.3 Solubilisation of insoluble inclusion bodies

A cell pellet from 100 mL expression culture was re-suspended in 4 mL resuspension buffer. The cells were disrupted with sonication on ice with a 3 mm micro-tip sonicator using 2.5 s, 40 kHz pulses for 60 s. The lysate was centrifuged at 8,000 rpm for 10 min at 4ºC. The supernatant was discarded and the cell pellet was re-suspended in 3 mL cold isolation buffer. The cells were again disrupted with sonication on ice as above and centrifuged at 8,000 rpm for 10 min at 4ºC. The re-suspension, sonication and spin steps were repeated, as above. At this stage the pellet material was either re-suspended in 5 mL binding buffer or washed once in a buffer lacking urea and stored at -20ºC for later processing.

2.16 Membrane isolation by water lysis method

This method was performed as described by (Ward, *et al.* 2000) and the principle of the method involved lysis of the cells by osmotic pressure. A 50 mL expression culture was centrifuged at 4,000 rpm for 10 min at 10^{o}C. The supernatant was discarded and the pellet re-suspended in 10 mL 0.2 M Tris pH8.0 and left stirring for 20 min at room temperature. At this point 9.6 mL of deionised H_2O should be aliquoted in anticipation of the water lysis step. At time zero, 4.85 mL of 1 M sucrose/ 0.2 M Tris pH8.0/ 1 mM EDTA was added. At time 1.5 min, 65 µL of 10 mg/mL lysozyme solution prepared in 1 M sucrose/ 0.2 M Tris pH8.0/ 1 mM EDTA buffer was added. At time 2 min, 9.6 mL of deionised H_2O, which had been prepared earlier was added, and left to stir for 20 min. The preparation was examined under oil immersion for spheroplasts. Spheroplasts were sedimented at 18,000 rpm for 20 min at 4^{o}C. The supernatant was retained and stored at -20^{o}C (periplasmic fraction) and the spheroplasts were re-suspended in 15 mL deionised water using a 15 mL homogeniser. The suspension was fully homogenised and was left stirring for 30 min at room temperature. The suspension was sedimented at 18,000 rpm for 20 min at 4^{o}C. The supernatant was retained (cytoplasmic fraction) and stored at -20^{o}C. The preparation was then washed twice with 30 mL 0.1 M NaPhosphate buffer/ 1 mM

Mercaptoethanol pH7.2. The homogeniser was used to re-suspend the pellet each time and sedimented at 18,000 rpm for 20 min at 4^{o}C. The final pellet (mixed membrane fraction) was re-suspended in ice cold re-suspension buffer 0.1 M NaPhosphate/ 1 mM Mercaptoethanol pH7.2 using a 1 mL homogeniser.

2.17 Colony blot procedure

Using a multichannel pipette, 300 µL of LB containing the required antibiotic was aseptically aliquoted into an autoclaved deep well 96-well plate. Each well was inoculated with a single colony obtained from a transformation plate. The 96 well plate was sealed with a Breathe-easy Sealing Membreane and incubated at 37°C, shaking at 800-1000 rpm overnight. Using a multichannel pipette 10 µL of growth from each well was aseptically transferred to a fresh autoclaved deep well 96-well plate. The second 96-well plate was sealed and grown up as described above to ensure that there was sufficient growth to proceed. An LB-amp agar plate containing 50 µg/mL IPTG was overlaid with a nitrocellulose membrane and an equal volume of each overnight culture was spotted from the 96-well plate on to the membrane. The plate was incubated at 37°C for eight hours. A set of dishes for colony lysis was prepared. Each dish contained a sheet of 3 mM paper soaked in one of the following;

- 10% SDS solution (w/v)
- Neutralisation solution
- Denaturing solution
- 2x SSC

Excess fluid was discarded from the dish so that the filter was moist but not wet. The nitrocellulose membrane was removed from the LB-amp plate carefully and placed colony side up, on top of each filter, taking care to exclude air bubbles. The filters were incubated at room temp, as outlined in Table 2.4.

Table 2.4: Incubation of nitrocellulose membrane for in situ lysis. Steps carried out on the nitrocellulose membrane of a 96-well colony blot. For denaturing, neutralization and 2x SSC buffer preparation, see Section 2.3.

Solution	Time
SDS	10 min
Denaturing	5 min
Neutralisation	5 min
Neutralisation	5 min
2x SSC	15 min

2.18 Protein purification

Immobilised Metal Affinity Chromatography (IMAC) was used to purify recombinant proteins with terminal $(His)_6$ tag.

2.18.1 Standard IMAC procedure

A 1 mL Ni-Agarose (Qiagen) column was washed with 5-10 column volumes of dH_2O. The column was equilibrated with 5-10 column volumes of lysis buffer containing 20 mM Imadizole. The cleared lysate was gently mixed with nickel-nitrilotriacetic acid resin (Ni-NTA, Qiagen) in a column. The column was washed with 10 column volumes of lysis buffer containing 20 mM Imadizole. A gradient of imidazole in lysis buffer was used to elute contaminating proteins from the column. Target protein was eluted from the column using a high concentration of imidazole in lysis buffer (depending on protein – 120-500 mM Imidazole). The column was washed with 5-10 column volumes of water followed by 3-5 column volumes of 20% ethanol. Samples were taken at each step and analysed by SDS-PAGE electrophoresis.

2.18.2 IMAC using FPLC and GE nickel resin

A HisTrap FF 1mL column was washed with 5 column volumes of dH_2O, and equilibrated with 5 column volumes of lysis buffer. Prior to the attachment of the column to the AKTA purifier 100 the maximum back pressure was set to 0.5 MPa and the flow set to 1 mL/min. The top of the affinity column was attached to the pump outlet no 1 and the bottom of the column attached directly to the FPLC unit. The sample injection loop was filled with cleared lysate (Section 2.15). A reservoir of lysis buffer containing 20 mM imidazole was attached to pump A, with lysis buffer containing 200 mM imidazole attached to pump B. The purification strategy involved

- Column equilibration with 5 mL lysis buffer
- Sample application from the sample injection loop
- A wash of 10 mL lysis buffer
- Application of an imidazole gradient to remove non-specifically bound proteins, by applying 10 mL of a 70 mM imidazole wash
- Purified protein was then eluted by applying 20 mL of a 200 mM imidazole wash

Samples were collected in 1mL fractions in 96-well plates. The FPLC software was used to monitor column back pressure, OD_{280}, OD_{540}, conductivity, flow rate and fraction collection through-out the run.

2.18.3 Purification and on-column refolding of insoluble $(His)_6$ protein

Resolubilised cell lysate (Section 2.15.3) was passed through a 0.45 µm filter and loaded into the sample injection loop. The purification strategy involved;

- Column equilibration with 5 mL binding buffer
- Sample application from the sample injection loop
- A wash of 10 mL binding buffer
- A wash of 10 mL wash buffer

- Refolding of the bound protein through the application of a linear 30 mL 6 – 0 M urea gradient, starting with wash buffer and finishing with refolding buffer.
- A wash of 10 mL refolding buffer
- Elution through the application elution buffer

Samples were buffer exchanged into PBS using Vivaspin 500 (mass cut off at 10,000 Da) micro-concentrators (Sartorius) and analysed using SDS-PAGE (Section 2.23).

2.18.4 Desalting of purified protein using HiPrep 26/10 desalting column

The FPLC back pressure limit was set to 0.35 MPa and the flow rate to 1 mL/min before attachment of the HiPrep 26/10 desalting column. The top of the column was attached to pump outlet no 1 and the bottom of the column attached to the FPLC using green peek tubing (0.75 mm I.D.). Before the first sample application the column storage buffer (20%) ethanol was removed, and the column equilibrated with sample buffer. This was done by washing with 2 column volumes of water (106 mL) followed by 5 column volumes (265 mL) of PBS. Equilibration was not necessary between runs using the same buffer. Washing was carried out at the flow rate intended for chromatography (5 mL/min). Samples were collected in 0.5 mL fractions in 96-well plates.

2.19 Recharging of IMAC resin

This column was routinely recharged prior to re-use of the Ni-NTA resin. The used resin was poured into a column and washed with 5 column volumes of dH_2O. The resin was then stripped by washing with 5 column volumes of stripping buffer (Section 2.3). Ionically bound proteins were removed by washing with 5 column volumes of 1.5M NaCl. Precipitated proteins, hydrophobically bound proteins, lipoproteins and lipids were removed by washing with 5 column volumes of 1M NaOH, contact time 1-2 hours, followed by 10 column volumes of dH_2O. Remaining impurities were removed by washing with 10 column

volumes of 30% isoproponal for 30 min, followed by 10 column volumes of dH_2O. The resin was charged by adding 2 column volumes of 100 mM $NiSO_4$. The resin was washed again with 2 column volumes of dH_2O before storage in 20% ethanol.

2.20 Protein quantification

Two procedures for the quantification of protein were variably employed. The Bradford protein as described by Bradford M.M. ((Bradford 1976)) assay is a colorimetric method, when the coomasie dye binds protein in an acidic medium an immediate shift in absorption maximum occurs from 465 nm to 595 nm with a concomitant color change from brown to blue. Quantification of the protein concentration by 280 nm reading is based on the proteins ability to absorb ultraviolet light in solution with the absorbance maxima at 280 nm.

2.20.1 Quantitative determination of protein by Bradford assay

The Bradford assay using BradfordUltra™ reagent (Expedeon) was used to quantify total protein in the range of 1 µg/mL – 1.5 mg/mL. For low range protein samples (1-25 µg/mL) 150 µL of sample was mixed with 150 µL Bradford Ultra, for high protein range samples (0.1 -1.5 mg/mL) 20 µL of sample was mixed with 300 µL of BradfordUltra™ reagent. All samples, standards and blanks were prepared in triplicate. The absorbance was read at 600 nm. Bovine serum albumin (BSA) was used as the reference protein. BSA standards were prepared in PBS and assayed in triplicate to yield a standard curve. Protein concentration of samples was determined from this standard curve.

2.20.2 Quantitative determination of protein by 280 nm readings

Prior to reading all samples were spun down at 13,000 rpm for 10 min to minimize any debris present in solution that may interfere in absorbance readings. The UV lamp was set to 280 nm and allowed to warm up for 15 min. The zero absorbance was calculated with buffer solution. The absorbance of the

protein solution was read in triplicate. The concentration of the protein was then calculated using the Beer-Lambert law, Equation 2.1;

$$A = \varepsilon cl$$

Equation 2.1: Beer-Lambert equation for the quantitative determination of protein concentration from absorbance readings at 280 nm. A = absorbance at 280nm, ε = molar extinction coefficient, c = concentration (in the units corresponding to ε) and l = path length.

2.21 Quantitative determination of reducing sugars by DNS assay

Reducing sugars, under alkaline conditions reduce 3.5-dinitrosalicylic acid to 3-amino-5-nitrosalicylic acid. A red colour is produced on boiling, which absorbs maximally at 540-550 nm (Miller 1959). The DNS assay was used to quantify reducing sugars in the range 0.2-1 mg/mL. The carbohydrate solution was mixed with DNS reagent in the ratio of 1:2 and the samples were boiled for 10 min. Samples were allowed to cool for 5 min before aliquoting 250 µL of each solution into a microtitre plate. All samples, standards and blanks were prepared in triplicate. The absorbance was read at 570 nm.

2.22 Lyophilisation

Samples were aliquoted (0.1-0.5 mL) in a 96-well plate and sealed with adhesive film, before being snap frozen in -80ºC ethanol. A pin hole was made in the film adhesive over the top of each well prior to lyophilisation. Samples were lyophilized at -45ºC and at a vacuum of 25 bar. Once lyophilized, samples were stored at -20ºC.

2.23 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.23.1 Preparation of SDS gels

Protein samples were analysed by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), based on the method outlined by LaemmLi (Laemmli 1970).

10%, 12.5% and 15% resolving gels were prepared as outlined in table 2.5

Table 2.5: Preparation of SDS gels

Resolving Gel	**7.5%**	**10%**	**12.5%**	**15%**	**4%**
30% Acrylamide (mL)	1.875	2.501	3.123	3.750	0.325
H_2O (mL)	3.633	3.007	2.385	1.758	1.540
1.5M Tris-HCl pH8.8 (mL)	1.875	1.875	1.875	1.875	-
0.5M Tris-HCl pH 8.8 (µL)	-	-	-	-	625
10% APS (µL)	37.5	37.5	37.5	37.5	12.5
10% SDS (µL)	75	75	75	75	25
Temed (µL)	3.75	3.75	3.75	3.75	2.5

2.23.2 Sample Preparation

To a microfuge tube 18 µL of sample was added to 2 µL of 10X SDS-PAGE sample buffer (Section 2.3). Samples were boiled for 5 min and applied to wells that were flushed of un-polymerized acrylamide. To the insoluble pellet obtained from a 1 mL sample, 20 µL of 10X gel loading dye was added, and boiled for 5 min. The sample was then diluted with water up to 100 µL.

2.23.3 Sample application

A total of 15 µL of the prepared sample was applied to each SDS-PAGE well. One lane on each gel was kept for the relative molecular weight protein marker (Broad range protein marker, NEB, Fig 2.5). A 20 µL aliquot of the marker was applied for coomassie staining (Section 2.13.3), a 10 µL aliquot for silver staining (Section 2.12.3). When western blot analysis was required, the NEB prestained and color plus prestained protein markers were used *(see Fig* 2.6)

The Broad range protein marker consisted of Rabbit Muscle Myosin (212 kDa), MBP-β-galactosidase from *E.coli* (158 kDa), β-galactosidase from *E.coli* (116 kDa), Rabbit muscle Phosphorylase B (97 kDa), Bovine Serum Albumin (66 kDa), Bovine liver Glutamic dehydrogenase (55 kDa), MBP2 from *E.*coli (42 kDa), Thioredoxin reductase from *E.coli* (34 kDa), Soybean Trypsin inhibitor (20 kDa), Chicken egg white lysozyme (14 kDa), Bovine lung Aproptinin (6.5 kDa), Bovine pancreas insulin (3.4 kDa) and Bovine pancreas B chain (2.34 kDa). The prestained protein marker consisted of MBP-β-galactosidase from *E.coli* (175 kDa), MBP-paramyosin from *E.coli* (80kDa), MBP-CBD from *E.coli* (58 kDa), CBD-Mxe Intein-2CBD from *E.coli* (46 kDa), CBDBmFKBP13 from *E.coli* (25kDa), Chicken egg white lysozyme (17 kDa), Bovine lung aprotinin (7 kDa).

The color plus protein marker consisted of MBP-β-galactosidase from *E.coli* (175 kDa), MBP-truncated-β-galactosidase *E.coli* (80 kDa), MBP-CBDE *E.coli* (58 kDa), CBD-*Mxe* Intein-2CBDE *E.coli* (46 kDa), CBD-*Mxe* Intein *E.coli* (30 kDa), CBD *E.coli* ParE (23 kDa), Chicken egg white lysozyme (17kDa), Bovine lung aprotinin (7 kDa).

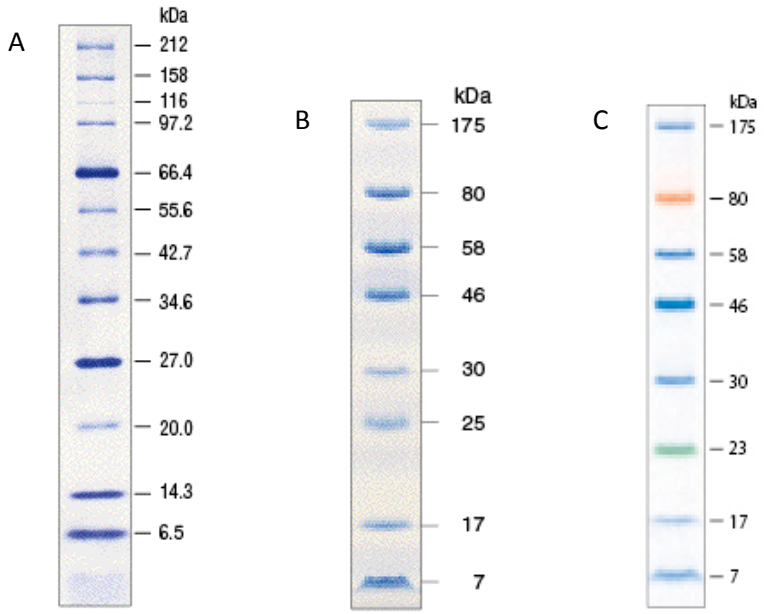


Figure 2.6: A; NEB broad range protein marker, B; Pre-stained protein marker and C; Color plus pre-stained protein marker. Representative images of the NEB broad range protein marker, pre-stained protein marker and color plus pre-stained protein marker used in this study. Images obtained from www.NEB.com.

2.23.4 Gel analysis

Polyacrylamide gels were removed from the electrophoresis chamber and washed with dH_2O. Gels were routinely stained for one hour with Coomassie blue stain solution (Section 2.3). An overnight de-stain followed using Coomassie blue de-stain solution (Section 2.3). Subsequent soaking in dH_2O enhanced the protein bands further. For gels which required a greater degree of sensitivity the silver staining method (Blum, Beier and Gross 1987) was used as outlined in Table 2.6.

Table 2.6: Silver staining of SDS-PAGE Gels

Step	Reagent	Time	Per Gel (50 mL)
Fix	30% Ethanol 10% Acetic acid	>1 hour	15 mL 5 mL
Wash 1	20% Ethanol	15 min	10 mL
Wash2	Water	15 min	
Sensitisation	0.1% Sodium Thiosulphate	1 min	111 µL 4.5% Stock
Rinse	Water	2 x 20 s	
Silver Nitrate	0.1% $AgNO_3$ 0.026% Formaldehyde	30 min	0.05 g 35 µL 37% Stock
Rinse	Water	2 x 20 s	
Development	3% Sodium Carbonate 0.019% Formaldehyde Sodium Thiosulphate	Until bands appear	1.5 g 25 µL 37% Stock 2.2 µL 4.5% Stock
Stop	5% Tris	1 min	2.5 g 1.25 mL

2.24 Western blot

An SDS-PAGE gel was run (Section 2.23.2) using NEB pre-stained molecular weight markers. Four pieces of blotting paper and a piece of nitrocellulose membrane were cut to the same dimensions as the SDS-PAGE gel. The blotting paper, nitrocellulose and gel were then soaked in transfer buffer (Section 2.3), and arranged on the semi-dry electroblotter as per Fig 2.7. Transfer then occurred at a constant 20 V for 15 min.

To detect recombinant $(His)_6$ tagged proteins, the membrane was blocked with 3% BSA-TBST (Section 2.3) for one hour. The membrane was washed four times for 5 min with TBST, and then incubated with 1:10,000 murine anti-His antibody in 3% BSA-TBST. The membrane was washed four times for 5 mins with TBST and incubated in 15 mL dH_2O containing SIGMAFAST^{TM} 3,3-Diaminobenzidine tablets. Development was stopped by washing with dH_2O.

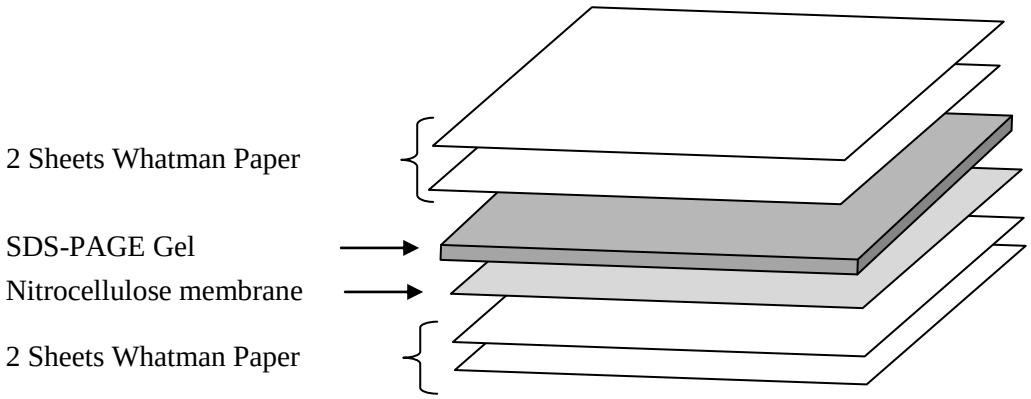


Fig 2.7: Schematic of Western blot. Sheets of Whatman paper and nitrocellulose are cut to the same dimensions as the SDS-PAGE gel, as any overlap over the gel surface can lead to inefficient protein transfer. The transfer stack is laid down from bottom to top as follows, two sheets blotting paper, nitrocellulose membrane, SDS-PAGE gel, two sheets blotting paper.

2.25 Size exclusion Chromatography

The native molecular mass of recombinant proteins under native conditions were determined by size exclusion chromatography (Whitaker 1963 and Andrews 1964). Size exclusion chromatography was carried out using FPLC.

Before attachment of the gel filtration column to the AKTA purifier the maximum back pressure was set according to the manufacturers instructions. The column storage valve was disconnected from the gel filtration column, and the top was attached to pump outlet no 1. The bottom of the column was attached directly to the FPLC. To remove the column storage buffer (20% ethanol) two column volumes of water were pumped through the column at a flowrate of 1

mL/min (unless otherwise stated). The column was then equilibrated with 5 column volumes of sample buffer, at the same flowrate, before sample application.

2.25.1 Superdex 75

With pump A attached to a reservoir of degassed sample buffer 100 µL of a 1 mg/mL protein sample was applied through the sample injection port. The run commenced, with a flow rate of 0.5 mL/min of buffer A. Following sample injection 2.5 column volumes of buffer was passed through the column. The proteins retention time was measure through eluent absorbance at 280 nm.

2.25.2 Toyopearl HW-55S

With pump A attached to a reservoir of degassed sample buffer 500 µL of a 5 mg/mL protein sample was applied through the sample injection port. The run commenced, with a flow rate of 1 mL/min of buffer A and a maximum back pressure of 0.5 MPa. Following sample injection 2.5 column volumes of buffer was passed through the column. The proteins retention time was measured using the online Monitor U-900 (GE Healthcare), which read the eluent absorbance at 280 nm.

2.26 Insoluble substrate activity assay

A 1 mL volume (30 µg/mL) of chitin binding protein in PBS was mixed with 0.05g of insoluble substrate (crustacean shell chitin, avicel (crystalline cellulose), crustacean shell chitosan (>85% deacetylation), squid pen chitin (BioLog), chitosan powder from carapacea skin (88-96% deacetylation - BioLog) in a column. The mixture was incubated with end-over stirring for a defined period of time. Unbound protein was collected in the eluent and subsequent washes with 0.9% NaCl (2 x 1 mL). Bound protein was eluted from the insoluble substrates by boiling for 10 min in 50 µL SDS-PAGE sample buffer. Bound and unbound protein samples were visualized by 12% SDS-PAGE analysis and silver staining.

2.27 Enzyme linked lectin assay

A 50 µL volume of glycoprotein was immobilized in each well of a NUNC MaxiSorp ELISA plate at 4ºC overnight. Each sample was assayed in triplicate, at a concentration of 10 µg/mL. The unbound glycoprotein was removed by inverting the plate and the wells were blocked with 150 µL 2.5% BSA in TBS for one hour at 25°C. The solution was then removed by inverting the plate and washing with TBS supplemented with 0.1% Triton X-100 four times. A 50 µL aliquot of lectin in TBS supplemented with 10 mM $CaCl_2$ was then added at a concentration of 5 µg/mL and let incubate at 25ºC for one hour. This was removed by inversion and washed with TBST as before. This was followed with 50 µL of 1:10,000 murine anti-histidine or anti-biotin antibody, as appropriate. Antibody was created fresh, diluted in TBST, and was incubated for one hour at 25ºC. Unbound antibody was removed by inversion and washed four times with TBST, before the addition of 100 µL TMB substrate (Section 2.3). The reaction was stopped after a specified time by the addition of 50 µL 10% H_2SO_4. The absorbance was read at 450 nm.

2.28 ElectroSpray ionization mass spectrometry

Protein samples were desalted using Vivaspin 500 (mass cut off at 10,000 Da) micro-concentrators (Sartorius) and diluted 1 mg/mL in mass spectrometry grade solvents (50% H_2O, 50% Methanol, 0.2% Formic acid).

Electrospray ionization mass spectrometry (Q-TOF Ultime Global™, Micromass, Manchester, UK) coupled with a nanoflow Z-spray source. The nanoflow source interface was operated at room temperature and was equipped with PicoTip™ emitter with distal coating (20 µm i.d with 10 µm Silica tip™, New Objective, Woburn, MA, USA). Ions were focused by a radio frequency (RF) lens before transmission to the quadropole. The ions were then transmitted through the hexapole collision cell and were pulsed into the TOF analyzer. The TOF analyzer was set in the positive ion mode at 9.10 kV when the reflectron was used in V configuration at 35.6 V and pusher was in auto mode. Acquisition

was achieved by using a time-to-digital convertor operating at 4 GHz. Mass spectra were recorded using a micro channel plate (MCP) detector at the exit of the TOF analyzer, after calibration with NaI solution (2 mg/mL) in isopropanol/water (50/50). A mass range (*m/z)* of 500 – 5000 Th was scanned over 2.4 s with an inter scan delay time of 0.1 s. Samples were analysed in "denaturing" conditions. Samples were introduced into the source by direct infusion using a syringe pump (Cole-Parmer, Vernon Hills, Illinois, USA) at a rate of 200 nl/min for the nanoflow source.

Calculations and data analyses were performed using MassLynx software version 4.0 (Waters, Manchester, UK). Deconvolution of spectra was accomplished either manually or by using a transform algorithm. Mass spectra were background subtracted, smoothed using the Savitzky Golay method, and centered before final calculations of the molecular mass. Molecular species were represented by an envelope of a series of peaks corresponding to multiple charged ions that deconvoluted to a mean mass ± standard deviation.

2.29 Immobilisation of protein onto cyanogen bromide activated sepharose

The protein to be immobilized was suspended in 0.1 M $NaHCO_3$ coupling buffer, pH 8.5, which contained 500 mM NaCl. The capacity for the gel was 5-10 mg per mL of gel. The resin was then washed in cold 1 mM HCl for 30 min, at 200 mL/g of gel. The supernatant was removed and the resin was washed with 10 column volumes of dH_2O, followed by 1 column volume of coupling buffer. The protein solution was immediately added to the column, and the flow stopped. The solution was allowed to mix by inversion overnight at 4ºC, after it was washed with coupling buffer to remove unbound protein. The resin was then blocked by the addition of 2 column volumes of 0.2M glycine, pH 8.0, and was allowed to mix by inversion for 2 hours at room temperature. The unbound glycine was removed by washing with 2 column volumes of coupling buffer, followed by washing with 2 column volumes of 0.1 M Sodium acetate, pH 4.0, containing 500 mM NaCl. Alternate washing steps using coupling buffer and acetate buffer

were repeated 5 times. The resin was then used immediately or stored at 4ºC in 1 M NaCl with 0.05% NaN_3.

3.0 Cloning, Expression and Characterisation of CBP21

3.1 Overview

This chapter describes the cloning of the *cbp*21 gene from *Serratia marcescens* that encodes the chitin binding protein CBP21, and the subsequent development of a recombinant expression system, which was utilised to prepare active recombinant protein of a high level of purity. Functional analysis of CBP21 was carried out, which included gel filtration (see Section 3.6), mass spectrometry (see Section 3.7) solid substrate assays (see Section 3.8) and the development of a novel chitin activity assay (Chapter 4). Mutant derivatives of the CBP21 protein were also constructed and analysed (Chapter 4). Additionally alternative sources of chitin binding proteins were investigated (Chapter 5).

3.2 Cloning and small scale expression of *cbp*21 from *Serratia marcescens*

The utilisation of prokaryotic systems for protein expression is generally considered to be the most effective in terms of simplicity and reproducibility. Over expression of specific proteins in the natural prokaryotic host cell however is difficult to achieve. The gram negative bacterium *Escherichia coli* is one of the most commonly used hosts for the expression of recombinant heterologous proteins. This is due in part to its simplicity, safety, rapid high density growth rate on inexpensive substrates and well defined genetics, allowing for large scale protein production.

There are a number of variable *E. coli* host strains available as well as a broad range of compatible cloning and expression vectors. Depending on the system employed expression of recombinant proteins in *E. coli* can facilitate protein labelling, the addition of fusion tags, the incorporation of unnatural amino acids, in addition to the direction of proteins to particular cellular compartments. The biggest disadvantage associated with recombinant protein expression in *E. coli* is its inability to perform many of the post-translational modifications necessary for eukaryotic protein functionality, mainly glycosylation. Carbohydrate binding proteins from prokaryotic sources, such as the chitin binding proteins, do not

require post translational modifications; therefore an *E. coli* based expression system can be exploited

The Qiagen pQE expression vectors, pQE30 (see Fig 2.1) and pQE60 (see Fig 2.2) have been designed to produce high yields of recombinant protein. These vectors also contain various desired elements including an optimised promoter consisting of the phage T5 transcriptional promoter and a *lac* operator sequence, extensive multiple cloning sites, and $(His)_6$ tag sites at either the N- or C-terminus. They also express the β-lactamase gene which confers ampicillin resistance to cells harbouring this plasmid.

To study the effect of the $(His)_6$ position, plasmids that express *N*-terminally and *C*-terminally $(His)_6$ tagged proteins were desired. In order to determine the optimal system for the expression and purification of CBP21 in *E. coli* a number of constructs were examined.

3.2.1 Initial cloning of the *S .marcescens* gene *cbp21*

The *S. marcescens* strain Koln was obtained from NCIMB (Accession no 2302). The sequence for the related organism *Serratia marcescens* BJL200 (see Fig 3.1) was obtained from the NCBI data bank (Accession no AY665558). On the basis of the BJL200 *cbp21* gene sequence primers were designed to amplify the *cbp21* gene for cloning into both the pQE30 and pQE60 vectors. The primers used for the amplification of N-terminally $(His)_6$ tagged *cbp21* were Marces-F1 and Marces-R1, which contained *Bam*HI and *Hind*III restriction sites, and those for the amplification of *C*-terminally $(His)_6$ tagged CBP21 were Marces-F3 and Marces-R2, which contained *Nco*I and *Bgl*II sites.

ATGAACAAAA CTTCCCGTAC CCTGCTCTCT CTGGGCCTGC TGAGCGCGGC CATGTTCGGC
GTTTCGCAAC AGGCGAATGC CCACGGTTAT GTCGAATCGC CGGCCAGCCG CGCCTATCAG
TGCAAACTGC AGCTCAACCG CAGTGCGGCA GCGTGCAGTA CGAACCGCAG AGCGTCGAGG
GCCTGAAAGG CTTCCCGCAG GCCGGCCCGG CTGACGGCCA TATCGCCAGC GCCGACAAGT
CCACCTTCTT CGAACTGGAT CAGCAAACGC CGACGCGCTG GAACAAGCTC AACCTGAAAA
CCGGTCCGAA CTCCTTTACC TGGAAGCTGA CCGCGCGTCA CAGCACCACC AGCTGGCGCT
ATTTCATCAC CAAGCCGAAC TGGGACGCTT CGCAGCCGCT GACCCGCGCT TCCTTTGACC
TGACGCCGTT CTGCCAGTTC AACGACGGCG GCGCCATCCC TGCCGCACAG GTCACCCACC
AGTGCAACAT ACCGGCAGAT CGCAGCGGTT CGCACGTGAT CCTTGCCGTG TGGGACATAG
CCGACACCGC TAACGCCTTC TATCAGGCGA TC**GACGTCAA CCTGAGCAAA** TAA

Fig 3.1: *S. marcescens* BJL200 *cbp21* coding sequence. Nucleotide sequence AY665558. Annealing sites of primers Marces-F1, Marces-F3 and Marces-R1 and Marces-R2 are indicated with red. There are no *Bam*HI, *Hind*III, *NcoI* or *BglII* restriction sites within this gene.

The PCR products were analysed by agarose gel electrophoresis and bands corresponding to the expected size were present. The PCR products were then restricted using either *Bam*H1-*Hind*III or *Nco*I-*Bgl*II, gel extracted and ligated to *Bam*H1-*Hind*III digested pQE30 or *Nco*I-*Bgl*II digested pQE60 vector respectively (see Fig 3.2). Following transformation into *E. coli* XL10-Gold cells (see Section 2.9.3), plasmid DNA was isolated and screened for inserts using gel electrophoresis and restriction analysis. The presence of the correct insert was further confirmed by DNA sequencing (see Section 2.12).

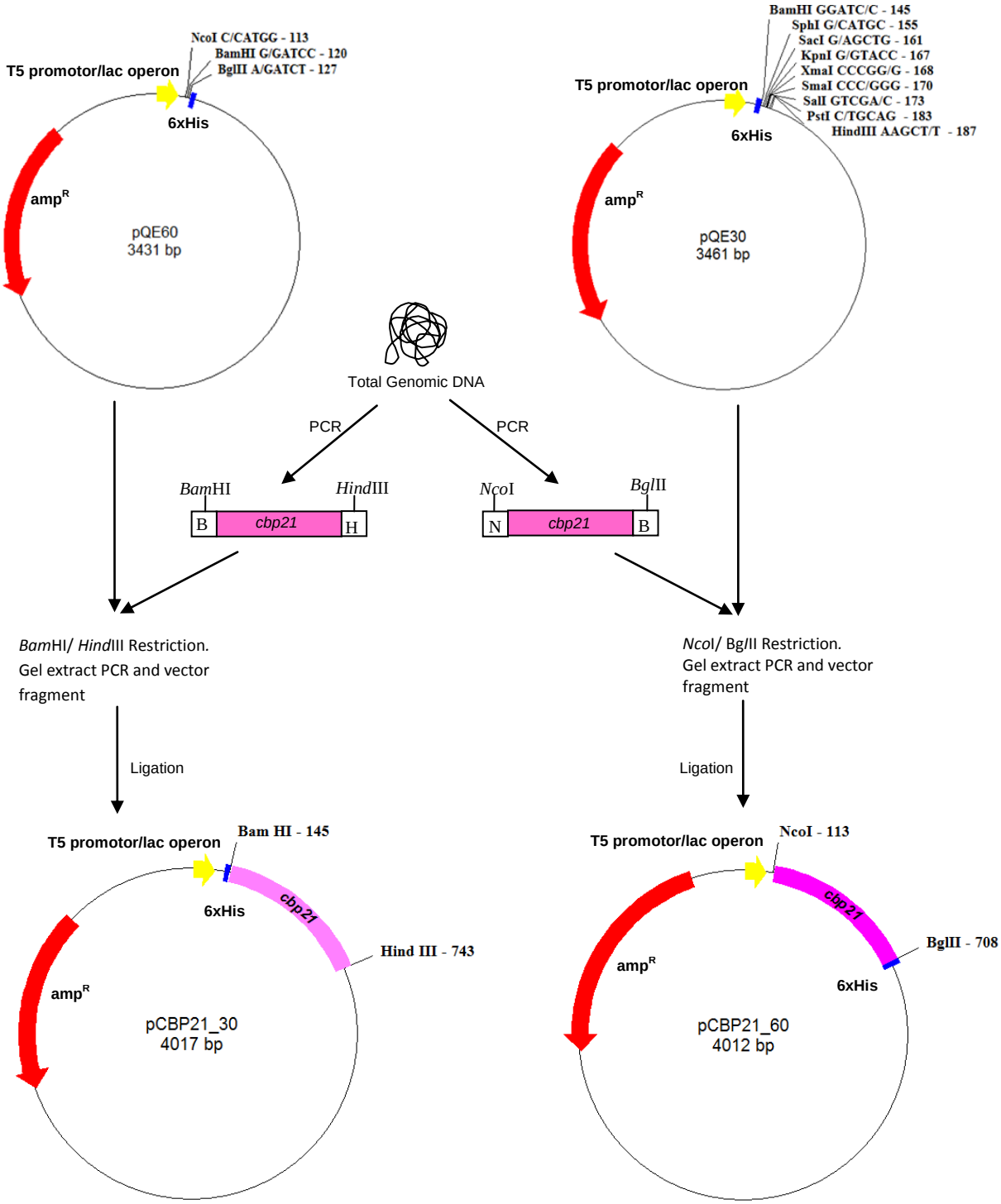


Fig 3.2: Schematic of the cloning of *cbp*21 into commercial pQE expression vectors pQE30 and pQE60. Outline of the positions of the relevant restriction sites and $(His)_6$ tags involved in the generation of p30-CBP21 and p60-CBP21. The primers Marces-F1 and Marces-R1 amplified the *cbp21* gene with the restriction sites *Bam*HI and *Hind*III. Marces-F3 and Marces-R2 amplified the *cbp21* gene, with restriction sites *NcoI* and *BglII* at the 5' and 3' ends respectively.

The cbp21 gene was successfully cloned into the pQE30 and pQE60 vectors, creating pCBP21_30 and pCBP21_60 respectively. Following sequencing it was observed that sequence variations existed between the *cbp21* gene as amplified from *S. marcescens* koln genomic DNA and the previously reported sequence for *cbp21* from *S. marcescens* BJL200 (see Fig 3.3). The changes in the gene sequence did not correspond to any changes in amino acid sequence (see Fig 3.4)

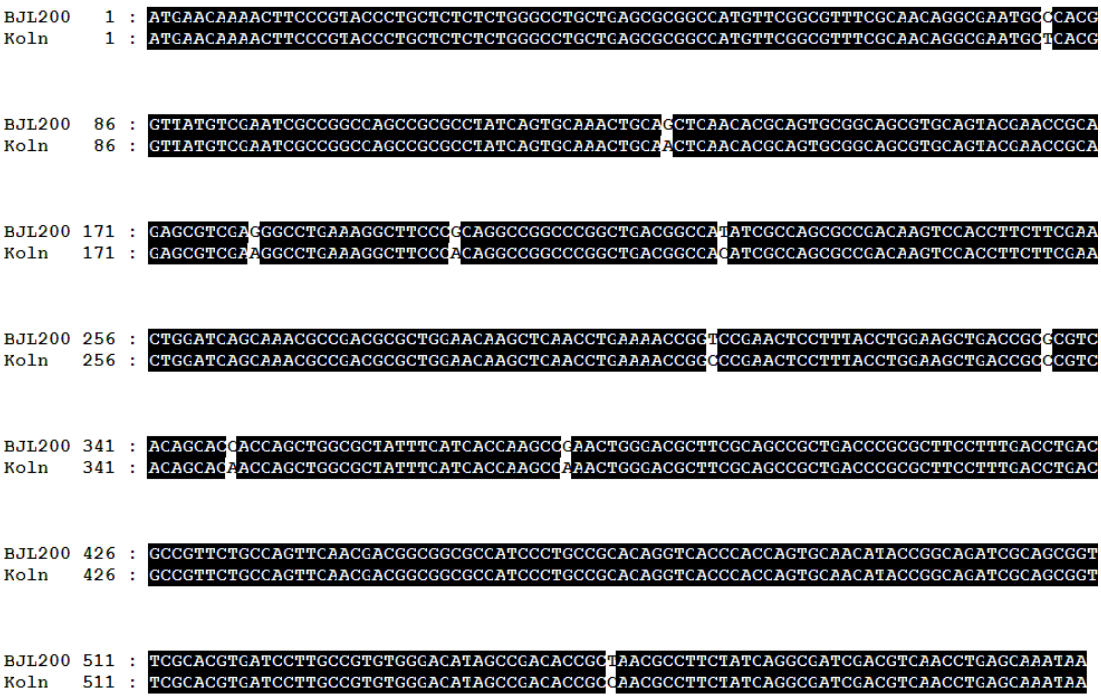


Fig 3.3: DNA sequence alignment of the *cbp21* gene from *S. marcescens* BJL200 and *S. marcescens* koln. Alignment generated using Clustal W and GenDoc 2.6 (see Section 2.13)

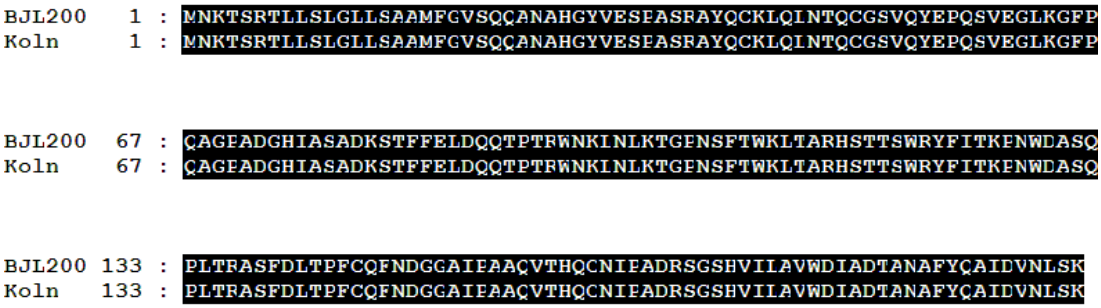


Fig 3.4: Amino acid sequence alignment of the *cbp21* gene from *S. marcescens* BJL200 and *S. marcescens* koln. Alignment generated using ClustalW and Gendoc 2.6 (see Section 2.13).

3.2.2 Small scale expression of CBP21

The *cbp21* gene was subsequently expressed in *E. coli* from pCBP21_30 and pCBP21_60 in 100 mL cultures (see Section 2.14), and the cell lysate examined by SDS-PAGE (see Section 2.23) and western blot (see Section 2.24), to determine if CBP21 is expressed as a soluble protein by *E. coli.* An over-expressed protein of the correct size was visible in the insoluble protein fraction of N-terminal $(His)_6$ tagged CBP21 (see Fig 3.5), while CBP21 cloned with a C-terminal $(His)_6$ tagged expressed to the soluble fraction (see fig 3.6).

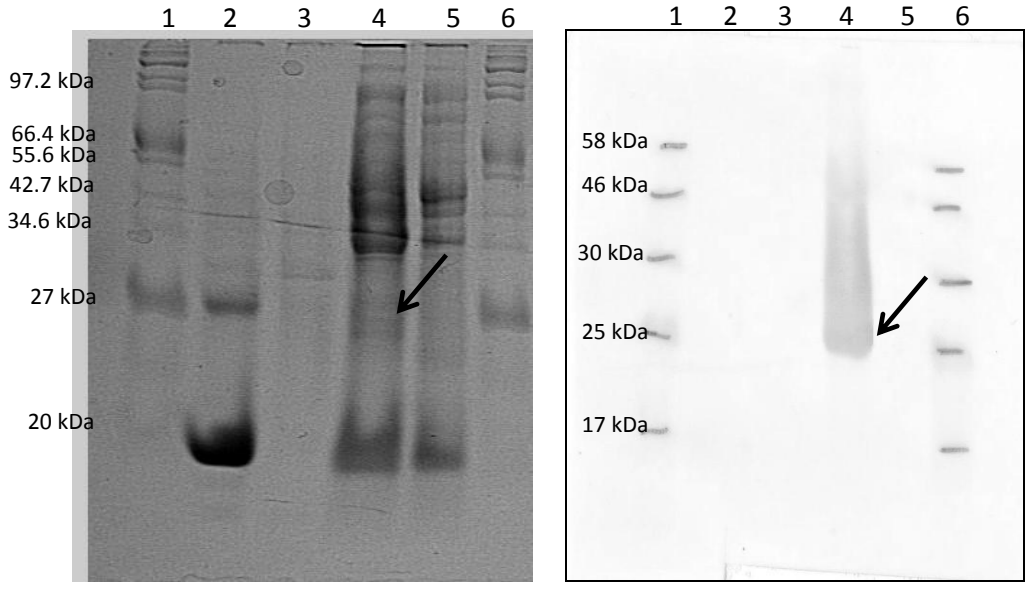


Fig 3.5: Expression analysis of N-terminally $(His)_6$ tagged CBP21 in *E. coli* XL10-Gold. Analysis by 12% SDS-PAGE and western blot of N-terminally $(His)_6$ tagged CBP21, expressed from pCBP21_30 in *E. coli* XL10-Gold. Lane 1; Protein ladder (see Section 2.23.3), Lane 2; pCBP21_30 soluble cell extract, Lane 3; pQE30 empty vector soluble cell extract, Lane 4; pCBP21_30 insoluble cell extract, Lane 5; pQE30 empty vector insoluble cell extract, Lane 6; Protein Ladder (see Section 2.23.3). Over-expressed CBP21 bands are indicated with an arrow.

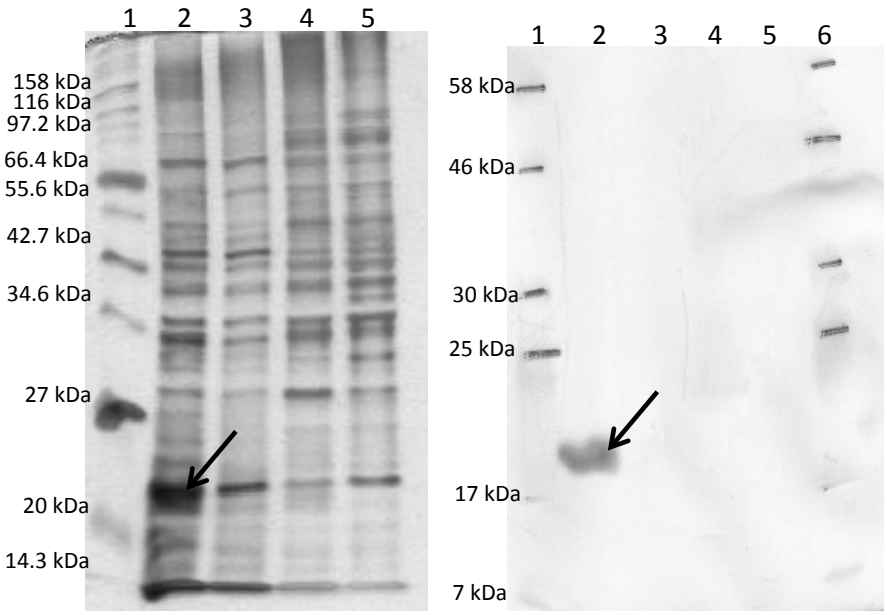


Fig 3.6: Expression analysis of C-terminally $(His)_6$ tagged CBP21 in *E. coli* XL10-Gold. Analysis by 12% SDS-PAGE and western blot of C-terminally $(His)_6$ tagged CBP21 expression from pCBP21_60 in *E. coli* XL10-Gold. Lane 1; Broad range protein ladder (see Section 2.23.3), Lane 2; pCBP21_60 soluble cell extract, Lane 3; pQE60 empty vector soluble cell extract, Lane 4; pCBP21_60 insoluble cell extract, Lane 5; pQE60 empty vector insoluble cell extract. Over-expressed CBP21 bands are indicated with an arrow.

In nature, in order to facilitate the recognition of CBPs within the cell and secretion to the extracellular environment these proteins have signal sequences at the N-terminus of the protein (~25aa) (see Fig 1.23) (Brurberg, *et al.* 2000). CBP21 was cloned into the pQE30 vector (pCBP21_30) with a $(His)_6$ tag at the N-terminus, upstream of the native signal sequence. Expression of the protein from this plasmid in *E. coli* resulted in the localisation of CBP21 to the insoluble protein fraction (see Fig 3.5).

3.2.3 Sub-cloning and small scale expression of CBP21

The addition of a $(His)_6$ tag to the N-terminus of the protein, upstream of the signal sequence, may be interfering with the recognition of the native signal sequence by *E. coli*, thus preventing the protein from being transported to the correct cellular location for accurate protein folding to occur. It is also possible that the $(His)_6$ tag is not interfering with the recognition of the signal sequence and is simply preventing correct protein folding due to is position at the N-terminus. In order to assess these theories *cbp*21 was amplified both with and without its native signal sequence and re-cloned into the pQE30 and pQE60 vectors.

cbp21 was amplified through PCR, without its native N-terminal signal sequence using a Phusion™ directed whole vector amplification strategy (see Fig 3.7). Primers Marces-F4 and 30-R1 were used to amplify N-terminally $(His)_6$ tagged *cbp21* without its native N-terminal signal sequence, with pCBP21_30 used as the template (see Fig 3.8). For the amplification of C-terminally $(His)_6$ tagged *cbp21* without its native N-terminal signal sequence primers Marces-F4 and Marces-R3 were used, the template plasmid was pCBP21_60 (see Fig 3.9).

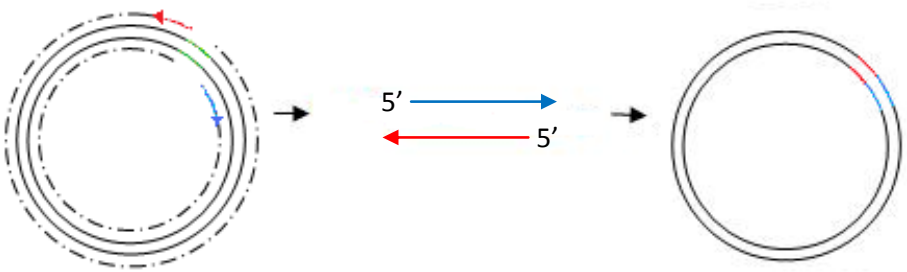


Fig 3.7: Schematic of Phusion™ whole vector amplification strategy. The undesired DNA is shown in green, with forward and reverse primers shown in blue and red respectively. The two phosphorylated primers are designed to amplify the entire expression plasmid, each with an annealing site at either side of the undesired DNA. Following PCR (see Section 2.10.1) a *Dpn*I restriction digest can be carried out to restrict any of the methylated template DNA. The phosphorylated ends are then ligated (see Section 2.10.2). Subsequently the ligation is transformed directly into an *E. coli* cloning strain (see Section 2.9.3).

GTGGTAGT GGTAGTGCCT AGGTAC

ATTAACTATG AGAGGATCGC ATCACCATCA CCATCACGGA TCC**ATGAACA AAACTTCCCG**

TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC GGCGTTTCGC AACAGGCGAA

CACGGT TATGTCGAAT CGCCGG

TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT CAGTGCAAAC TGCAACTCAA

CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC GAAGGCCTGA AAGGCTTCCC

ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC AAGTCCACCT TCTTCGAACT

GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG AAAACCGGCC CGAACTCCTT

TACCTGGAAG CTGACCGCCG TCACAGCACA ACCAGCTGGC GCTATTTCAT CACCAAGCCA

AACTGGGACG CTTCGCAGCC GCTGACCCGC GCTTCCTTTG ACCTGACGCC GTTCTGCCAG

TTCAACGACG GCGGCGCCAT CCCTGCCGCA CAGGTCACCC ACCAGTGCAA CATACCGGCA

GATCGCAGCG GTTCGCACGT GATCCTTGCC GTGTGGGACA TAGCCGACAC CGCCAACGCC

TTCTATCAGG CGATCGACGT CAACCTGAGC AAATAA

Fig 3.8: Whole vector amplification strategy for the generation of pCBP2130_N. *cbp21* was amplified without its native N-terminal signal sequence from pCBP21_30 using the primers Marces-F4 and 30-R1. *The cbp21* coding sequence is highlighted in purple, with the native leader sequence underlined in red. Primer annealing sites are shown in blue.

TCTCCT CTTTAATTGG

GATTCAATTG TGAGCGGATA ACAATTTCAC ACAGAATTCA TTAAAGAGGA GAAATTAA**CC**

TACCCT

ATGGGA**AACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC**

CACGGT TATGTCGAAT CGCCGG

GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT

CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC

GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC

AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG

AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCG TCACAGCACA ACCAGCTGGC

GCTATTTCAT CACCAAGCCA AACTGGGACG CTTCGCAGCC GCTGACCCGC GCTTCCTTTG

ACCTGACGCC GTTCTGCCAG TTCAACGACG GCGGCGCCAT CCCTGCCGCA CAGGTCACCC

ACCAGTGCAA CATACCGGCA GATCGCAGCG GTTCGCACGT GATCCTTGCC GTGTGGGACA

TAGCCGACAC CGCCAACGCC TTCTATCAGG CGATCGACGT CAACCTGAGC AAAAGATCTC

ATCACCATCA CATCACTAA

Fig 3.9: Whole vector amplification strategy for the generation of pCBP2160_N. *cbp21* was amplified without its native N-terminal signal sequence from pCBP21_60 using the primers Marces-F4 and Marces-R3. *The cbp21* coding sequence is highlighted in purple, with the native leader sequence underlined in red. Primer annealing sites are shown in blue.

The resulting PCR products were analysed by agarose gel electrophoresis and bands corresponding to the correct size were present. The bands were gel extracted, subjected to digestion by *Dpn*I, ligated and transformed into *E. coli* XL10-Gold, generating the plasmids pCBP2130_N and pCBP2160_N. These plasmids were subsequently expressed in 100 mL cultures (see Section 2.14) to assess the level of protein solubility in *E. coli*. No over expressed or $(His)_6$ positive bands were visible on SDS-PAGE or western blot analysis for N-terminally $(His)_6$ tagged CBP21, expressed from pCBP2130_N. An over-expressed band of the correct size was visible in the insoluble protein fraction of C-terminally $(His)_6$ tagged CBP21, expressed from pCBP2160_N (see Fig 3.10).

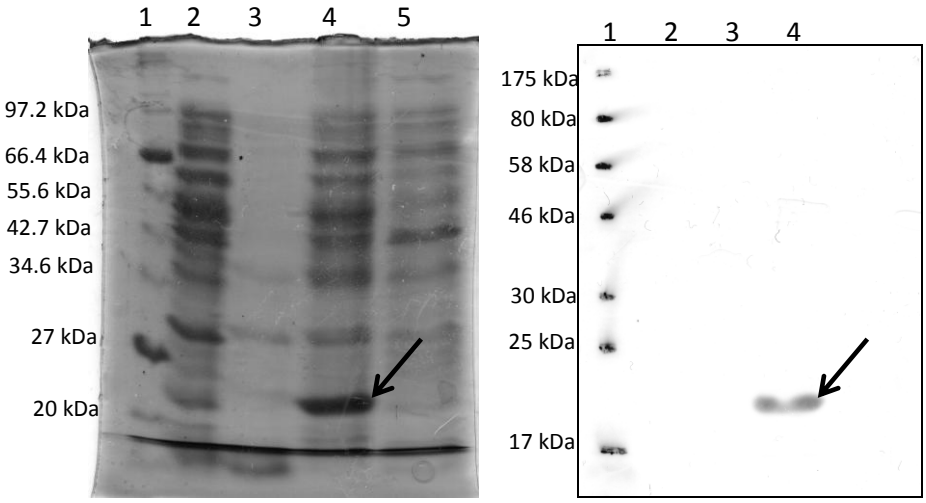


Fig 3.10: Expression analysis of C-terminally $(His)_6$ tagged CBP21 expressed without its native leader sequence in *E. coli* XL10-Gold. 12% SDS-PAGE and the corresponding western blot analysis of C-terminally $(His)_6$ tagged CBP21 over-expression in *E. coli* XL10-Gold. Lane 1; Protein ladder (see Section 2.23.3), Lane 2; pCBP2160_N soluble cell extract, Lane 3; pQE60 empty vector soluble cell extract, Lane 4; pCBP2160_N insoluble cell extract, Lane 5; pQE60 empty vector insoluble cell extract. Over-expressed CBP21 bands are indicated with an arrow.

3.2.4 Cellular compartment analysis of pCBP21_60 lysate

Expression of CBP21 from pCBP2160_N to the insoluble protein fraction lends weight to the theory that the native N-terminal signal sequence is functioning in *E. coli* through recognition and transport of the protein from the reducing environment of the cytoplasm to the periplasm, resulting in a soluble protein form. Although no CBP21 expression from pCBP21_30 was detectable, one could still surmise that it is the position of the N-terminal $(His)_6$ tag that resulted in whole protein expression from pCBP21_30 to the insoluble fraction. To determine if *E. coli* was able to recognise the native signal sequence, and transport the CBP21 protein to the periplasm, cell fractionation was undertaken using the water lysis method, as outlined in section 2.16. Periplasmic and cytoplasmic fractions were isolated and analysed by SDS-PAGE and western blot (see Fig 3.11).

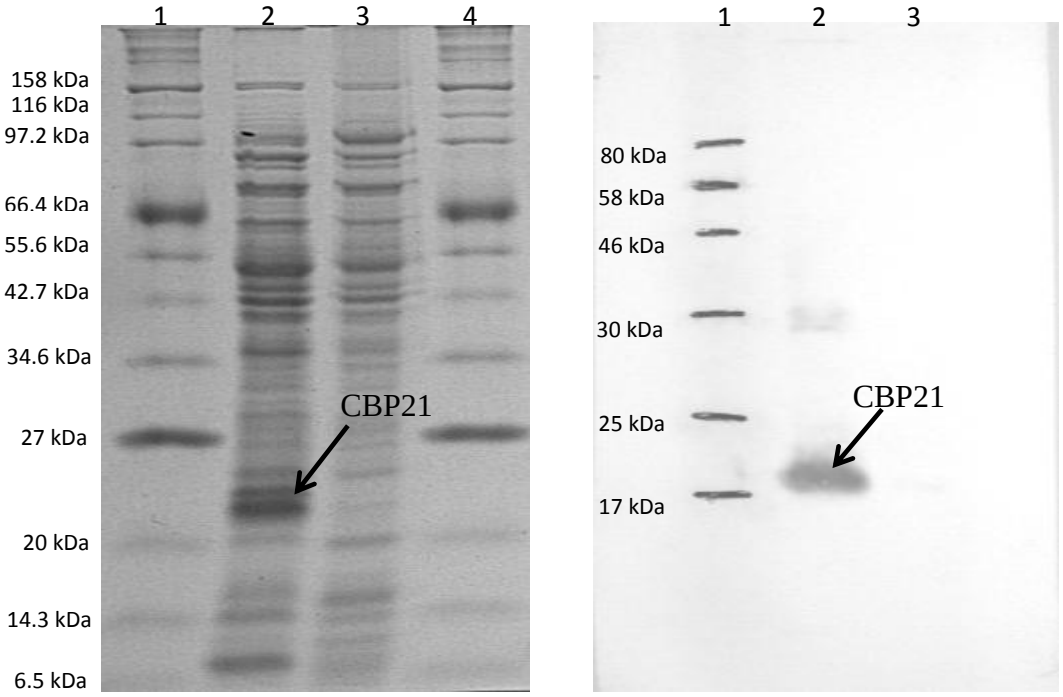


Fig 3.11: Soluble cell fraction analysis of C-terminally $(His)_6$ tagged CBP21 expressed from pCBP21_60. 12% SDS-PAGE and the corresponding western blot analysis of the periplasmic and cytoplasmic fractions from C-terminally $(His)_6$ tagged CBP21 over-expression, from p60_CBP21 in *E. coli*. Lane 1; Protein ladder (see Section 2.23.3), Lane 2; periplasmic fraction, Lane 3; cytoplasmic fraction, Lane 4; Broad Range protein ladder**.**

Analysis of the periplasmic and cytoplasmic fractions from soluble *E. coli* lysate expressed from pCBP21_60 (see Fig 3.11) illustrated that *E. coli* is capable of recognising the native CBP21 N-terminal signal sequence, and transporting the protein from the cytoplasm to the periplasm. Further size estimation techniques will have to be undertaken to determine if the protein is further processed through cleavage of this signal sequence (see Section 3.6 and 3.7).

3.3 Expression of recombinant CBP21.

Due to the high quantity and quality of purified protein required by analytical methods such as size exclusion chromatography and mass spectrometry, optimisation of expression conditions is an essential step for the production of large scale recombinant protein production. The ability to produce considerable amounts of protein is also advantageous when looking to exploit molecules for novel analytical purposes.

3.3.1 Selection of an *E. coli* expression strain for recombinant CBP21 expression.

The optimal *E. coli* strain for the expression of CBP21 was investigated. The plasmid pCBP21_60 (see Section 3.2), which expressed soluble C-terminally $(His)_6$ tagged CBP21 was transformed into four different *E. coli* strains to ascertain the optimal expression strain for recombinant CBP21. The *E. coli* strains used were XL10-Gold, JM109, KRX and BL21 (DE3) (Table 2.1).

XL10-Gold and JM109 cells are endonuclease and recombination deficient which makes them ideal cloning strains. The XL10-Gold and KRX strains also contain the *lac*Iq allele. This genotype results in a high expression level of the LacI repressor protein, which in turn strongly represses the activity of the Ptac promoter in the absence of IPTG, allowing for tight control over protein expression. BL21 and KRX are commonly used expression strains that are OmpT

protease deficient, which should allow for higher recovery of recombinant protein. Culturing conditions for the four strains were outlined in section 2.14. The optimal expression strain was determined using the Colony blot procedure as outlined in section 2.17 and SDS-PAGE analysis (see Fig 3.12 and 3.13).

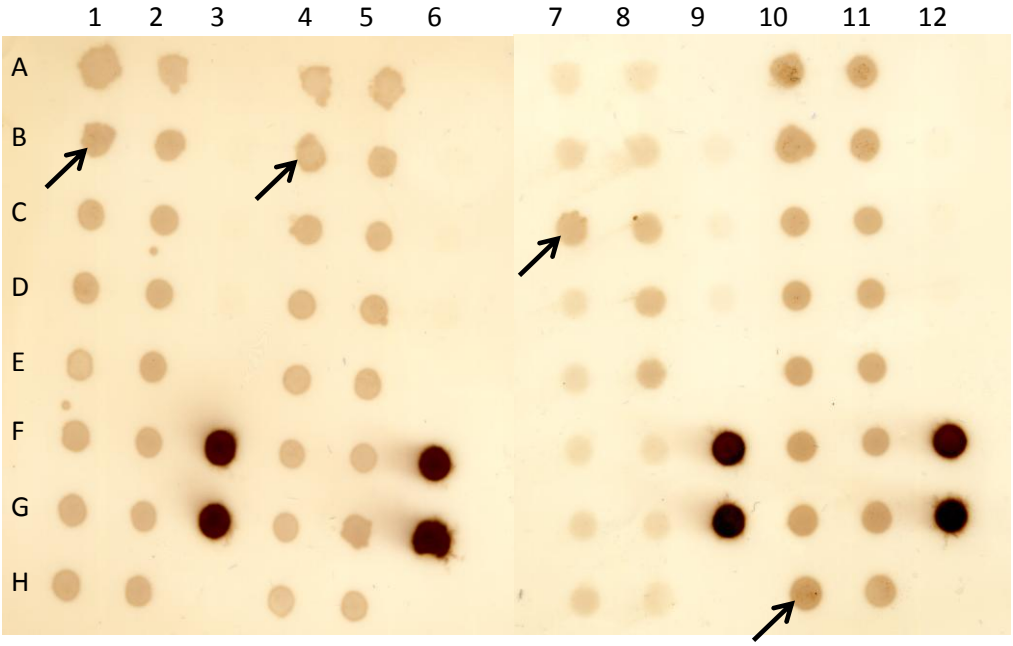


Fig 3.12: Effect of strain selection on CBP21 expression in *E. coli*. Colony blots of CBP21 constructs expressed in *E. coli*. Column 1 and 2 (A-H); BL21, column 3 B-D; BL21 negative controls, F-G; $(His)_6$ positive controls, column 4 and 5 (A-H); XL10-Gold, column 6 B-D; XL10-Gold negative controls, F-G; $(His)_6$ positive controls, column 7 and 8 (A-H); KRX, column 9 B-D; KRX negative controls, F-G; $(His)_6$ positive controls, column 10 and 11 (A-H); JM109, column 12 B-D; JM109 negative controls, F-G; $(His)_6$ positive controls. Clones selectede for further analysis are indicated with an arrow.

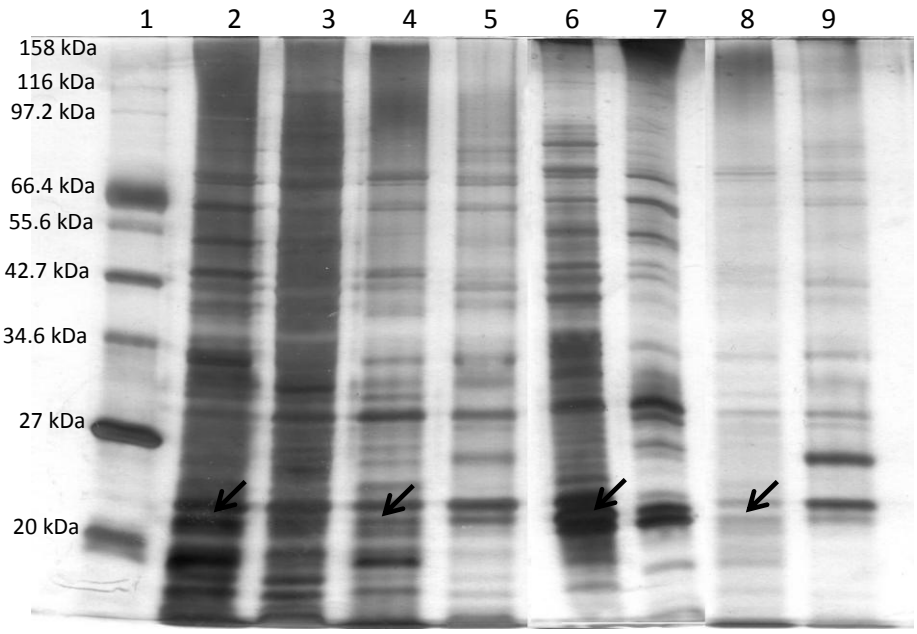


Fig 3.13: SDS-PAGE analysis of CBP21 expression in *E. coli* strains BL21, JM109, KRX and XL-10 Gold. Analysis by 12% SDS-PAGE of soluble and insoluble protein fractions at T = 4 hrs, of CBP21 over expression in four different *E. coli* strains, BL21, JM109, KRX and XL-10 Gold. *Lane* 1; Protein ladder (see Section 2.23.3), Lane 2; BL21 soluble fraction, Lane 3; BL21 insoluble fraction, Lane 4; JM109 soluble fraction, Lane 5; JM109 insoluble fraction, Lane 6; KRX soluble fraction, Lane 7; KRX insoluble fraction, Lane 8; XL-10 Gold soluble fraction, Lane 9; XL-10 Gold insoluble fraction. Overexpresssed CBP21 bands are indicated with an arrow.

In this optimisation study the CBP21 protein was expressed in a number of different *E. coli* strains to determine which strain yielded the highest amount of soluble protein. The most efficient clone from each of four different strains was selected using colony blot analysis (see Fig 3.12). The amount or protein expressed from each clone was then determined using SDS-PAGE analysis (see Fig 3.13). Upon selection of the most affluent KRX clone (see Fig 3.12), a glycerol stock of the clone was created, which could then be referred to when the protein was desired.

3.3.2 Optimisation of expression conditions for recombinant CBP21

Having determined in the previous section that *E. coli* KRX was the optimal strain for CBP21 protein expression, the most favourable conditions for induction and harvesting of the protein were subsequently examined (see Fig 3.14, 3.15, 3.16 and 3.17).

The CBP21 protein is derived from a pQE expression plasmid which is under the control of the P_{TAC} promoter system. The tac promoter/operator (P_{TAC}) is one of the most widely used expression systems, and is a strong hybrid promoter composed of the -35 region of the trp promoter and the -10 region of the lacUV5 promoter/operator. Expression of P_{TAC} is repressed by the LacI protein. The $lacI^q$ allele is a promoter mutation that increases the intracellular accumulation of LacI repressor, resulting in strong repression of P_{TAC}. Addition of the inducer molecule IPTG inactivates the LacI repressor. Thus the amount of expression from P_{TAC} is proportional to the amount of IPTG added: low levels of IPTG result in low expression from P_{TAC} and high concentrations of IPTG result in high expression from P_{TAC}. The IPTG concentration was varied in CBP21 expression and the amount of protein expressed from P_{TAC} in soluble form was examined by densitometry (see Fig 3.14), the effect of IPTG on growth rate was also examined (see Fig 3.15). Following optimisation of IPTG concentration the effect of incubation temperatures and harvest time were also examined (see Fig 3.16 and 3.17).

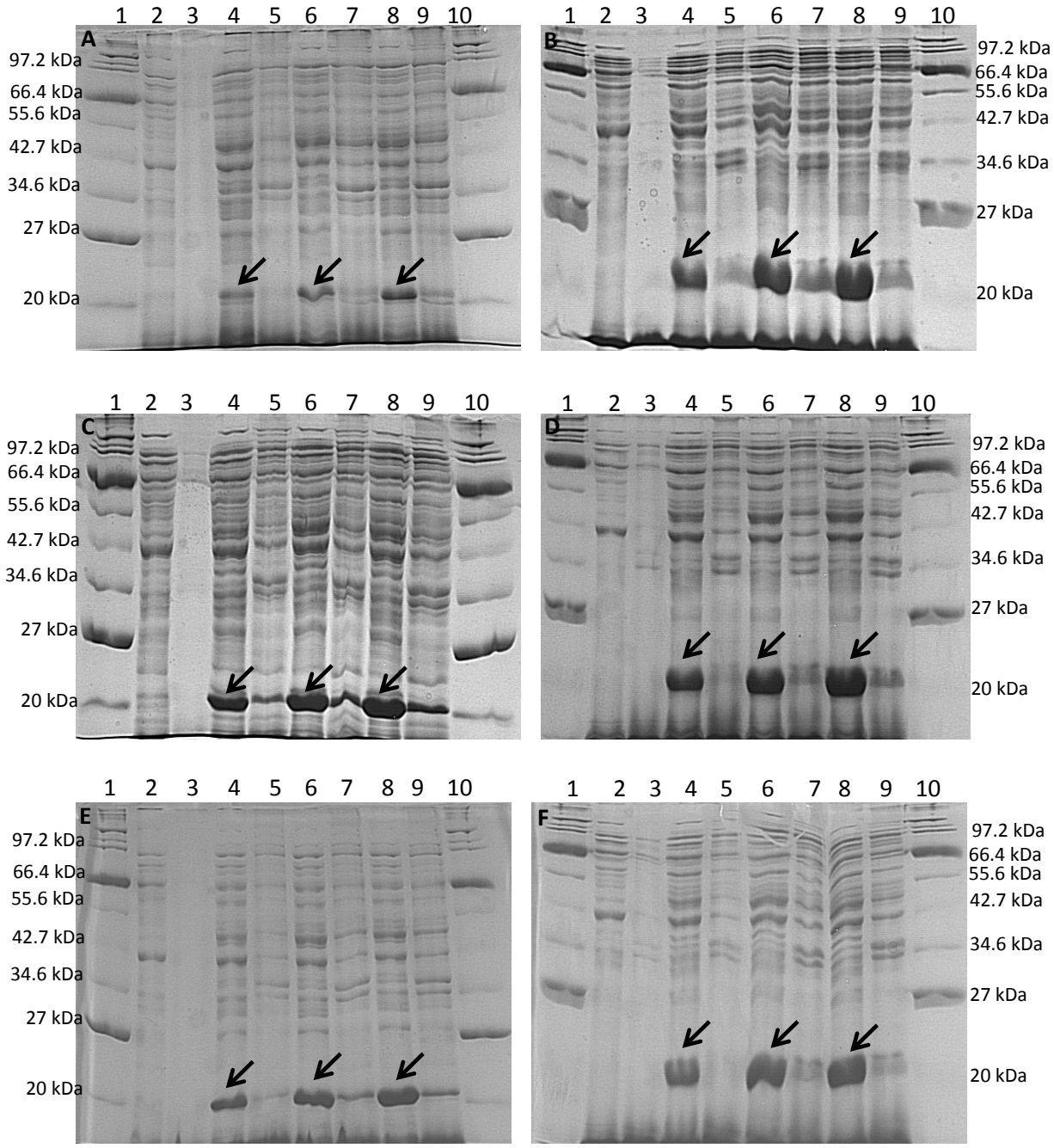


Fig: 3.14: Effect of varying IPTG concentration on expression of CBP21 in KRX. Analysis of increasing IPTG concentration on the expression of CBP21 from pCBP21_60 in *E. coli* KRX by 15% SDS-PAGE. A; 0 µM, B; 5 µM, C: 50 µM, D: 100 µM, E; 500 µM and F; 1000 µM. Lane 1; Broad range protein ladder (see Section 2.23.3), Lane 2; Soluble fraction 3 hours, Lane 3; Insoluble fraction 3 hours, Lane 4; Soluble fraction 5 hours, Lane 5; Insoluble fraction 5 hours, Lane 6; Soluble fraction 7 hours, Lane 7; Insoluble fraction 7 hours, Lane 8; Soluble fraction 9 hours, Lane 9; Insoluble fraction 9 hours, Lane 10; Broad range protein marker. Over-expressed CBP21 bands are indicated with an arrow.

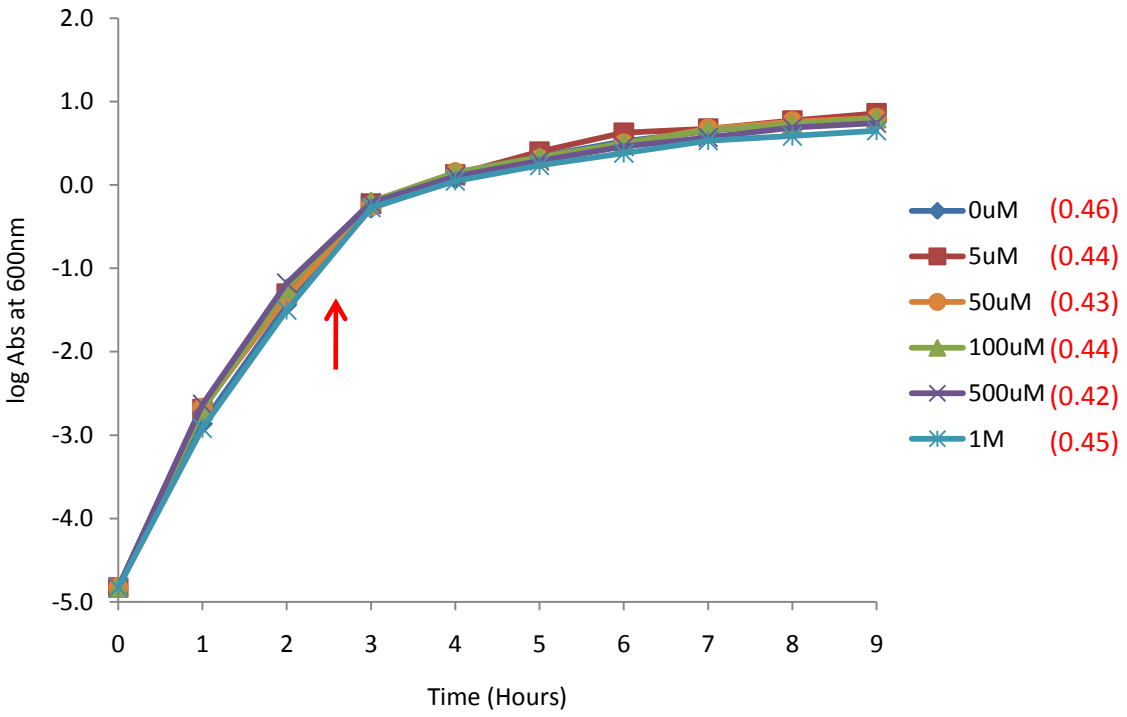


Fig: 3.15: Effect of varying IPTG concentration on the growth rate of *E. coli* KRX. Absorbance readings taken during expression of CBP21 from pCBP21_60 in *E. coli* KRX using varying IPTG concentrations. The point of induction for all cultures is indicated with a red arrow. The growth rate for the cultures (increase in OD_{600} per hour during exponential phase) is indicated in red. Cell growth was not affected by the increasing concentrations of IPTG.

It was deduced from densitometry that there was no significant difference between the amount of soluble protein produced upon induction with 100 µm IPTG (30.89%), 500 µM IPTG (28.55%) and 1 mM IPTG (33.1%). Increasing the concentration of IPTG from 100 µM to 1 mM was not found to significantly reduce the growth rate of the cells (see Fig 3.15). To save on reagents 100 µM IPTG was determined to be the optimal induction concentration. The effect of incubation temperature on CBP21 expression was then examined by measuring the amount of CBP21 in the total cell lysate at set time points, post induction, of active cell growth (see Fig 3.16 and 3.17). The optimal harvest time was also analysed (see Fig 3.16).

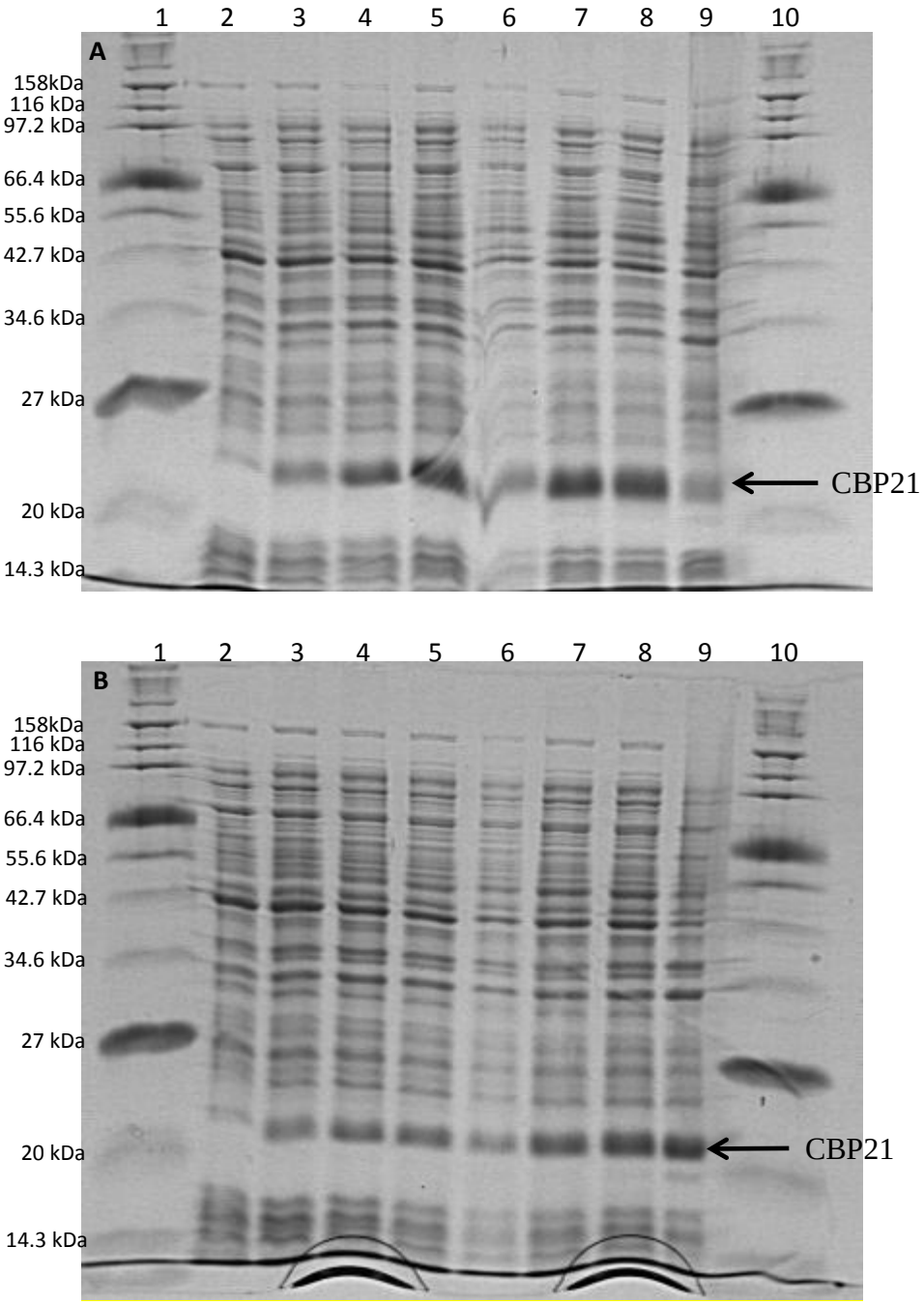


Fig 3.16: Effect of variation of growth temperatures and harvest time on CBP21 expression from KRX. A: 37°C incubation, B: 37°C incubation prior to induction, 30°C incubation following induction. Lane 1; Broad range protein ladder (see Section 2.23.3), Lane 2; soluble fraction T = 0, Lane 3; soluble fraction T = 1, Lane 4; soluble fraction T = 2, Lane 5; soluble fraction T = 3, Lane 6; soluble fraction T = 4, Lane 7; soluble fraction T = 5, Lane 8; soluble fraction T = 6, Lane 9; soluble fraction T = O/N.

Densitometry and SDS-PAGE were used to deduce optimal incubation temperatures and harvest time for the expression of recombinant CBP21 in *E. coli* KRX. From SDS-PAGE analysis it was clear that variations existed in the amount of CBP21 visible in the total cell lysate (soluble and insoluble) under both conditions. At time point T = 6 22.065% for 37°C and 12.481% for 30°C, at time point T = O/N 10.671% for 37°C and 13.324% for 30°C.

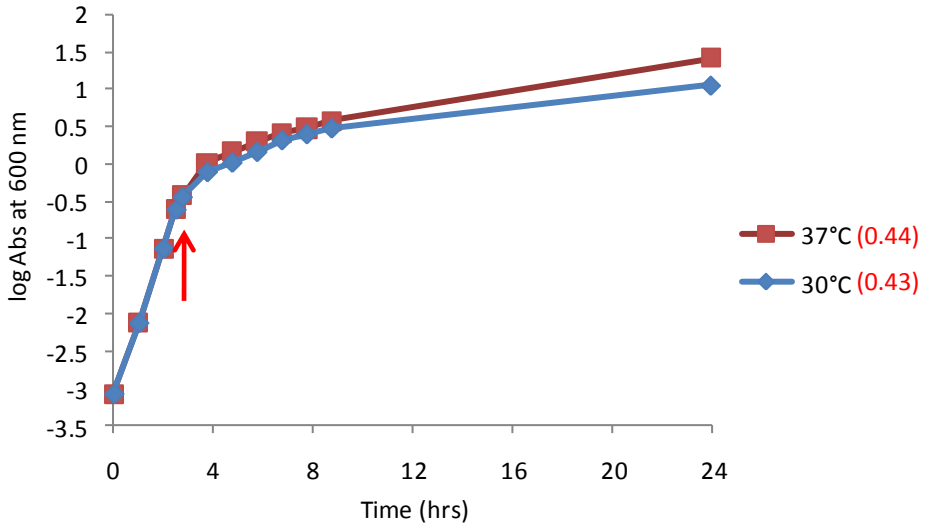


Fig 3.17: Effect of temperature variation on CBP21 expression in *E. coli* KRX. The log of absorbance readings against time recorded during CBP21 expression from pCBP21_60 in *E. coli* KRX, with two different incubation temperatures following induction. The induction time for both cultures is indicated with a red arrow. The growth rate for the cultures (increase in OD_{600} per hour during exponential phase) is indicated in red.

3.4 Purification of recombinant CBP21

3.4.1 Purification of recombinant CBP21 using IMAC

Immobilized metal ion affinity chromatography (IMAC) exploits a molecule's affinity for chelated metal ions. In this case the purification of a histidine tagged protein involves the exploitation of the strong interaction between the six histidine residues in the introduced tag and the two ligand binding sites of the Ni^{2+} ion, which has been immobilised on a solid matrix, in this case nitrilotriacetic acid linked to sepharose (Ni-NTA).

As many cellular proteins will display histidine residues on their surface, several wash steps are employed using imidazole, the active side chain of histidine, which displaces loosely bound proteins, and when used in high concentrations displaces the $(His)_6$ tagged recombinant protein. Depending on the position of the $(His)_6$ affinity tag, purification conditions will vary, as the strength of binding between the tag and the nickel matrix is largely dependent on the availability of the histidine residues. The strength of binding will affect the strength of wash tolerated by the interaction, and the point at which the target protein will be eluted. This section outlines the investigation into whether CBP21 can be purified using IMAC (see Section 2.18.2) and the wash stringencies tolerated by CBP21 (see Fig 3.18).

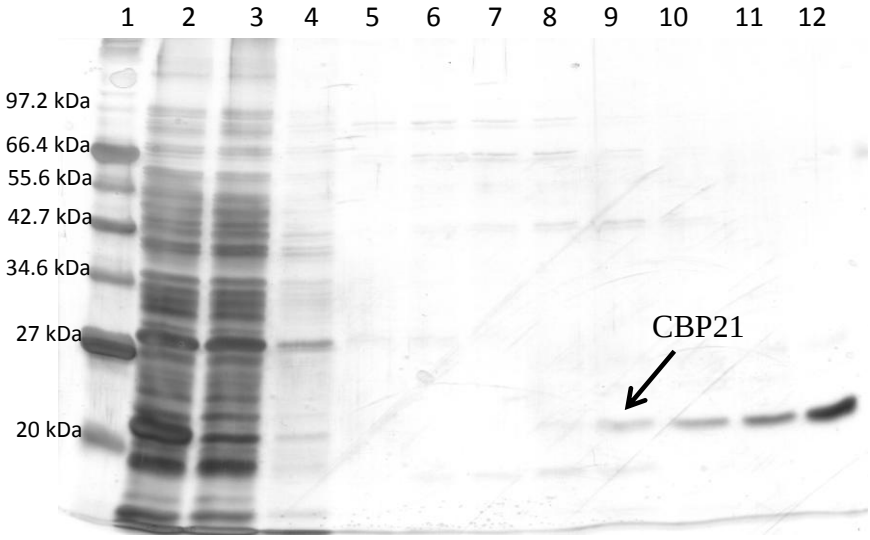


Fig 3.18: Purification of C-terminally tagged CBP21 soluble cell lysate using Amersham Ni-NTA resin. Analysis by 12% SDS-PAGE of fractions collected from the purification of CBP21 containing soluble cell lysate on an Ni-NTA column. Lane 1; Broad range protein ladder (see Section 2.23.3), Lane 2; Crude lysate, Lane 3; Column flowthrough, Lane 4; 20mM Imidazole wash, Lane 5; 30mM Imidazole wash, Lane 6; 40mM Imidazole wash, Lane 7; 50mM Imidazole wash, Lane 8; 60mM Imidazole wash, Lane 9; 70mM Imidazole wash, Lane 10; 80mM Imidazole wash, Lane 11; 90mM Imidazole wash, Lane 12; 100mM Imidazole wash.

It was evident from SDS-PAGE analysis that C-terminally $(His)_6$ tagged CBP21 protein could be purified to a high level of homogeneity using IMAC. It was observed that CBP21 could tolerate imidazole washes of up to 70mM before leaching from the column.

3.4.2: Protein purification using fast protein liquid chromatography

Fast protein liquid chromatography (FPLC) is a form of high-performance chromatography, commonly employed for the separation of proteins from complex mixtures. First developed by Pharmacia in 1982 it features high loading capacity, biocompatible aqueous buffer systems, fast flow rates and a wide availability of stationary phases. The self-contained system comprises a liquid mobile phase, with a solid stationary phase. The buffer mobile phase is controlled by pumps, with the buffers accessed through tubing from an external reservoir. The sample is introduced into the system through an injector and carried onto the column by the mobile buffer phase. Once in the column the sample mixture separates as a result of different components adhering to or diffusing into the stationary phase.

Having determined the optimal IMAC purification strategy for CBP21 using Amersham Ni-NTA resin, large scale purification could be carried out using FPLC-compatible Fast Flow HisTrap columns (GE Healthcare). This automated approach is preferable for large scale purifications as it makes reproducible separation possible by incorporating a high level of automation, including a gradient control programme and peak collection, as well as offering high levels of control over variables such as flow-rate and pressure, while allowing for online monitoring of temperature, conductivity and absorbance.

A sample FPLC purification profile of CBP21 purification is shown in Fig 3.19. The HisTrap column was washed with 10 column volumes of 20mM imidazole, followed by a second 10 column volume wash of 70mM imdazole. The protein was eluted using a 200mM imidazole wash. The elution profile can be visualised at OD_{280nm} as shown in Fig 3.19 (blue line), with the corresponding SDS-PAGE analysis of each fraction shown in Fig 3.20

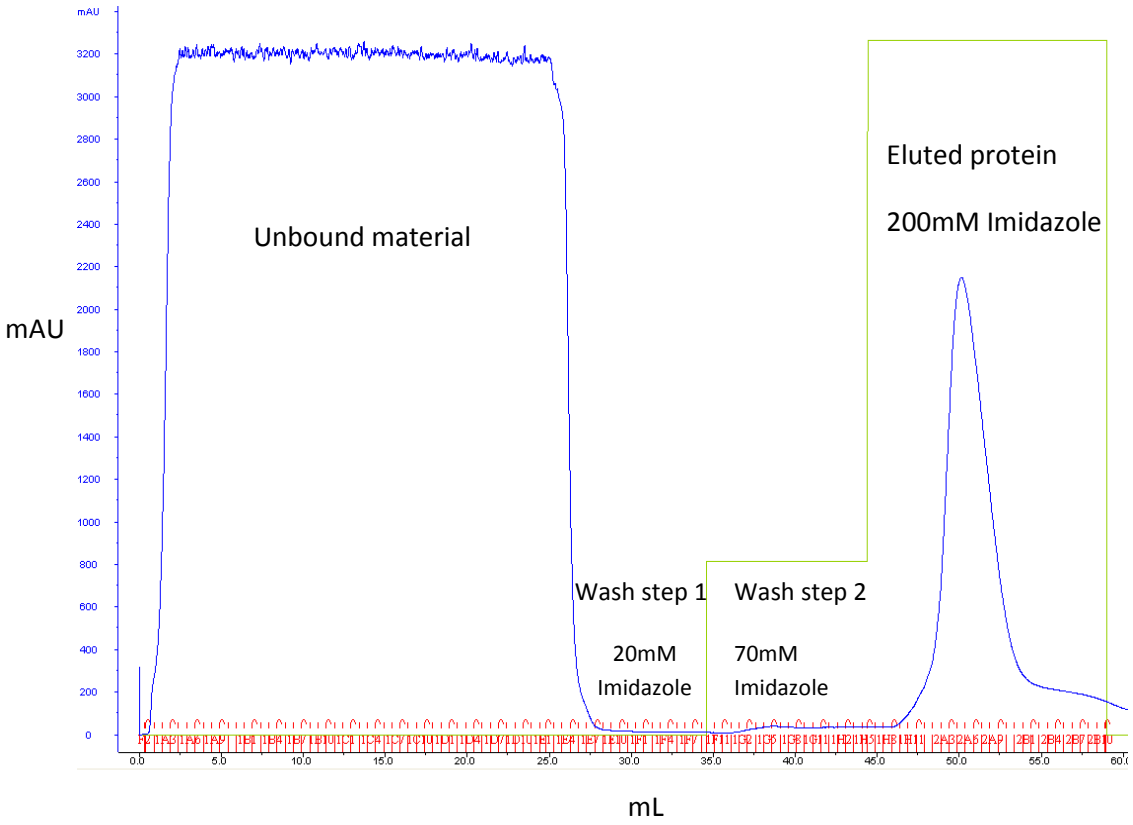


Fig 3.19: Purification of CBP21 over HisTrap Crude Column by FPLC. Purification profile of cleared cell lysate from a 500 mL expression culture for *E. coli* KRX expressing the protein CBP21 from the plasmid pCBP21_60, passed through the Amersham HisTrap FF crude 1 mL column, using FPLC (see Section 2.18.2). The blue line indicates the absorbance at A_{280}, the green line indicates the percentage of imidazole contained in the running buffer, while the red markers signify the fractionation of the eluent.

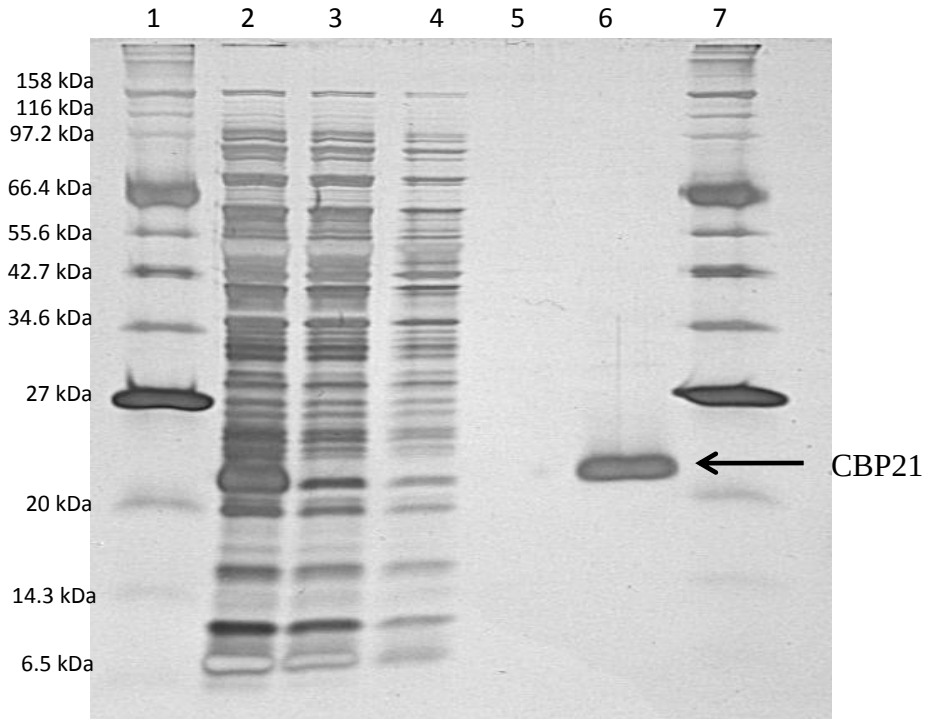


Fig 3.20: SDS-PAGE analysis of elution fractions resulting from the purification of CBP21 by FPLC-IMAC. Analysis by 12.5% SDS-PAGE and coomassie staining of fractions collected during CBP21 purification by IMAC using Ni^{2+} charged sepharose 1 mL FF column. Lane 1; Broad range protein ladder (see Section 2.23.3), Lane 2; Crude lysate, Lane 3; Column flowthrough, Lane 4, 20 mM Imidazole wash, Lane 5; 70 mM Imidazole wash, Lane 6; 200 mM Imidazole elution, Lane 7; Broad range protein ladder.

It was found that CBP21 could be purified successfully on the FPLC using the strategies found optimal in Section 3.4.1. The advantages of protein purification using FPLC include the ability to control parameters such as flow rate, whereas in manual purification one has to rely on gravity flow. It also allows for instant analysis of flow-through fractions using A_{280} readings, and for greater control of buffer concentrations and gradients. In addition FPLC allows for the accurate quantitation of purified protein as well as the calculation of peak heights and peak areas. This saves time by eliminating the requirement for subsequent protein quantification assays.

3.4.3 Buffer exchange of CBP21

Desalting and buffer exchange are necessary steps following IMAC purification. Downstream processes, such as protein quantification assays and subsequent activity assays can be adversely affected by the presence of excess salts, specifically imidazole. There are a number of techniques available to achieve these goals, namely dialysis, ultra filtration membranes and size exclusion chromatography. Both desalting and buffer exchange can be achieved simultaneously using size exclusion chromatography (see Section 2.18.4).

Desalting is accomplished by first equilibrating the chromatography column in dH_2O, whereas buffer exchange is achieved by first equilibrating the buffer exchange column in the replacement buffer. When the protein sample to be desalted/buffer exchanged is loaded onto the column, the protein macromolecules will be too large to enter the pores of the resin and will quickly pass through the column. On the contrary the buffer salts will enter the pores of the resin, thus slowing their rate of migration through the column. As a result the protein sample will be buffer exchanged into the pre-equlibration buffer. Consequently the protein macromolecules can be recovered separately from the smaller salt molecules.

Buffer exchange using a desalting column on an FPLC system saves time and labour when compared to the use of dialysis tubing or ultra filtration membranes. It allows for online analysis or OD_{280} readings and conductivity measurements, therefore giving a more accurate picture of complete buffer exchange as well as a brief insight into protein behaviour in alternative buffers (see Fig 3.21).

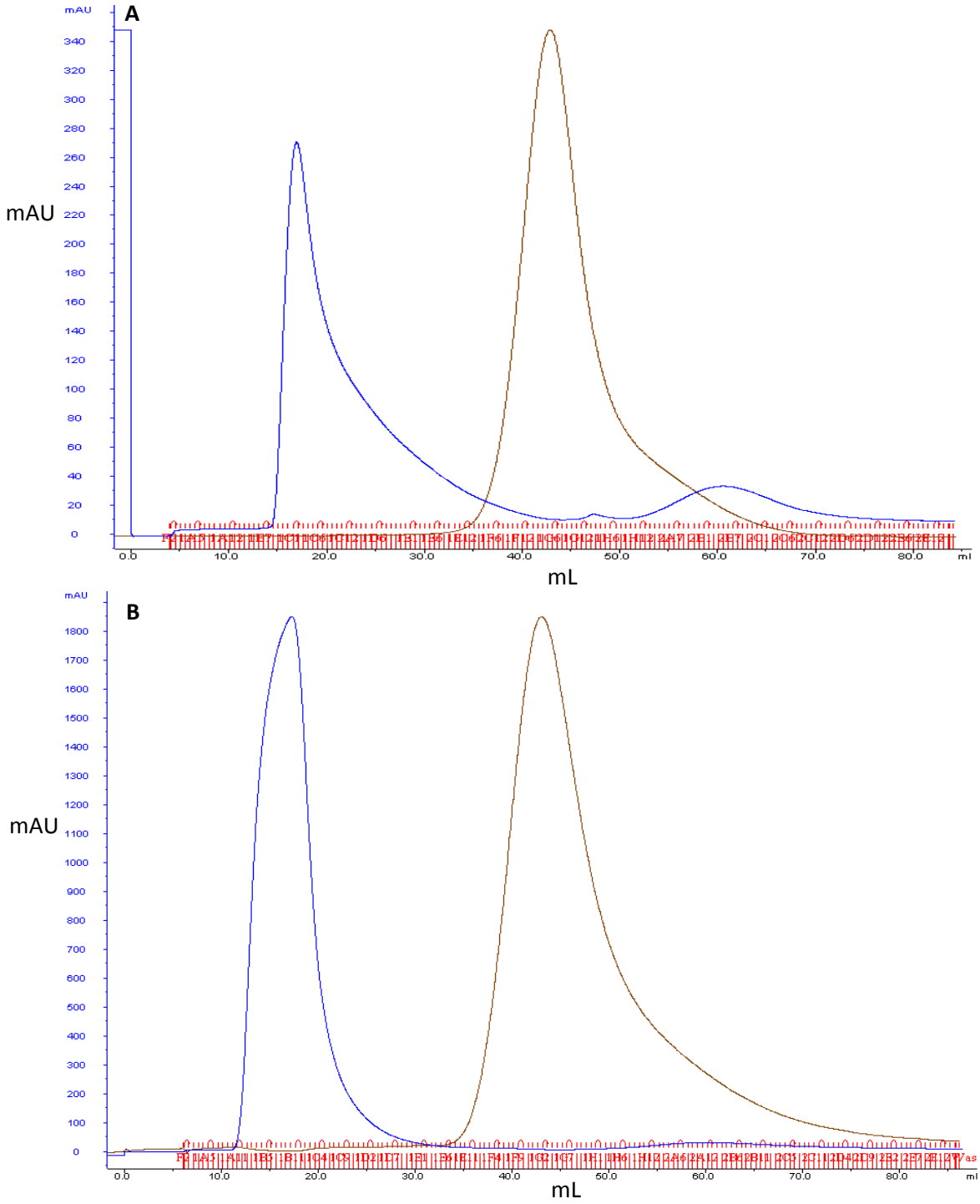


Fig 3.21: FPLC trace of CBP21 desalting and buffer exchange using the HiPrep 26/10 desalting column (GE Healthcare). A; CBP21 desalting into dH_2O, B; CBP21 – desalting and buffer exchange into PBS (see Section 2.3). FPLC trace of CBP21 desalting and buffer exchange into dH_2O and PBS respectively. The blue line indicates the absorbance at A_{280}, conductivity readings are denoted by the brown line, while the red markers signify the fractionation of the flow-through.

Desalting and buffer exchange of CBP21 using the HiPrep 26/10 desalting column from GE Healthcare proved successful. Using this technique it was possible to separate CBP21 from the imidazole salts present in the IMAC buffers. Desalting of CBP21 into dH_2O resulted in a sharp peak with a broad tail (see Fig 3.21a). When desalting protein into a water-equilibrated column peak broadening, such as this, can occur, resulting in higher sample dilution and poorer separation than when a buffered solution is used. Buffer exchange of CBP21 into PBS was more favourable, CBP21 resolved in a sharp peak with no broadening tail (see Fig 3.21b).

3.5 Total yield of recombinant CBP21

The expression conditions for the production of recombinant C-terminally $(His)_6$ tagged CBP21 from pCBP21_60 were optimised. A purification strategy and buffer exchange strategy were also developed that allowed for the preparation of high yields of highly homologous recombinant CBP21. The average yield of cell paste following expression of CBP21 in 500 mL LB broth was 3.2 g. The average yield of CBP21 per g/cell paste, following IMAC purification and buffer exchange into PBS was 3.1 mg/g cell paste.

3.6 Determination of protein size using size exclusion chromatography

C-terminally $(His)_6$ tagged CBP21 has been cloned into and expressed in *E. coli*. It is necessary to establish the protein size to determine if the N-terminal signal sequence is processed by *E. coli*. Previously, it has been shown that recombinant wild type CBP21 exists as a monomer, but crystallised as a dimer (Vaaje-Kolstad, *et al.* 2005b). The effect of the affinity tag incorporation into the protein structures are less well understood and can have deleterious effects on protein activity as well as adverse effects on protein conformation (Fitzpatrick, *et al.* 2004, Mohanty and Wiener 2004, Bucher, *et al.*, 2002, Goel, *et al.* 2000). To validate further the tertiary structure of recombinant C-terminally $(His)_6$ tagged CBP21 size exclusion chromatography (SEC) was employed.

3.6.1 Validation of CBP21 conformation using size exclusion chromatography

Gel filtration separates proteins based on differences in size as they pass through a packed column. The larger proteins will pass through the matrix more hastily than smaller proteins, due to the retardation of lower relative molecular mass proteins within the porous beads of the gel filtration medium. This technique can be used to characterise proteins through comparison with a set of standards used to create a standard curve. Equation 3.1 was used to create the standard curves, the standards used included alcohol dehydrogenase (ADH), bovine serum albumin (BSA), bovine carbonic anhydrase, ovalbumin, myoglobin, soybean trypsin inhibitor and cytochrome C.

Equation 3.1

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

In equation 3.1 V_t represents the total fill volume, V_o the void volume and V_e the elution volume of the protein standard. The Kav values were plotted against the log of the molecular weight (MW) to create a standard curve (see Fig 3.25 and 3.31). The elution of each protein from the column can be seen in Fig 3.24 and 3.30.

3.6.2 Size exclusion chromatography using Superdex 75 column

The Superdex 75 gel filtration matrix (GE Healthcare) has a special composite matrix of dextran and agarose (see Fig 3.22), which links the excellent gel filtration properties of cross-linked dextran with the physical and chemical stability of highly cross-linked agarose.

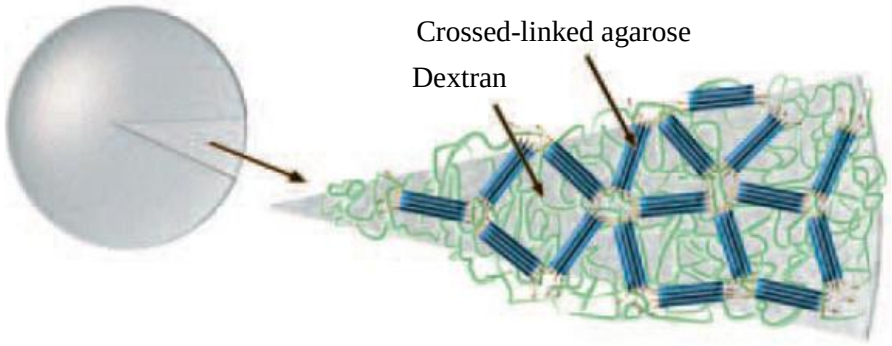


Fig 3.22: Schematic view through a section of a bead of superdex. The average particle size is 13 µm. (Image taken from Amersham Ltd.)

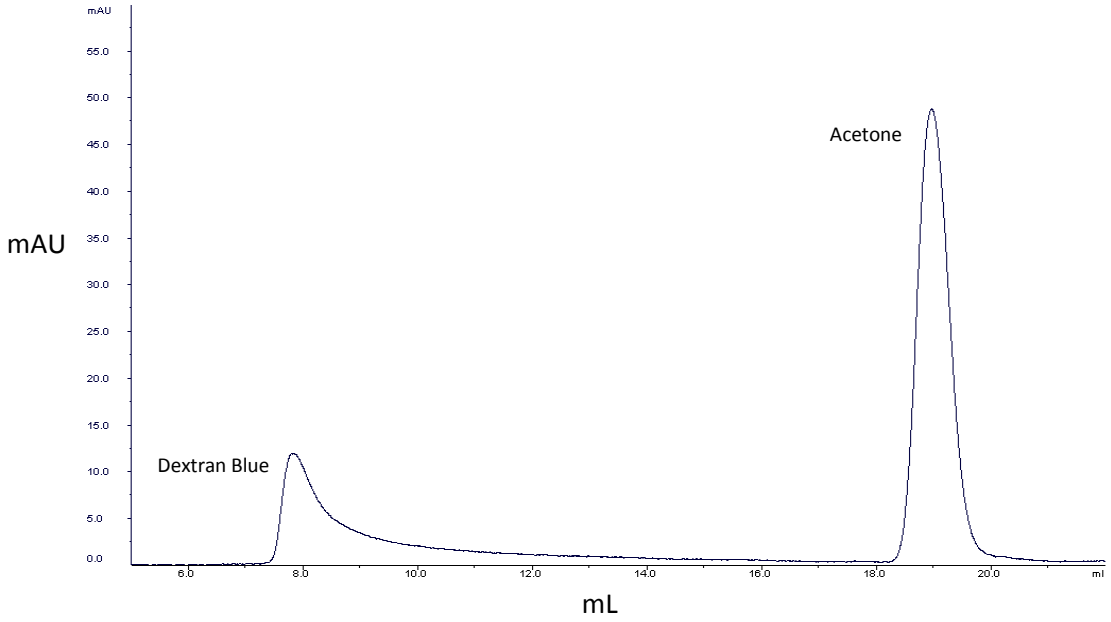


Fig 3.23: Determination of Superdex 75 total column and void volumes. The total column volume and void volumes for the Superdex 75 column were determined using the elution volumes of blue dextran and acetone. A void volume (V_o) of 7.73 mL was determined from the elution volume of blue dextran, and total volume (V_t) was 19.9 mL, determined with acetone.

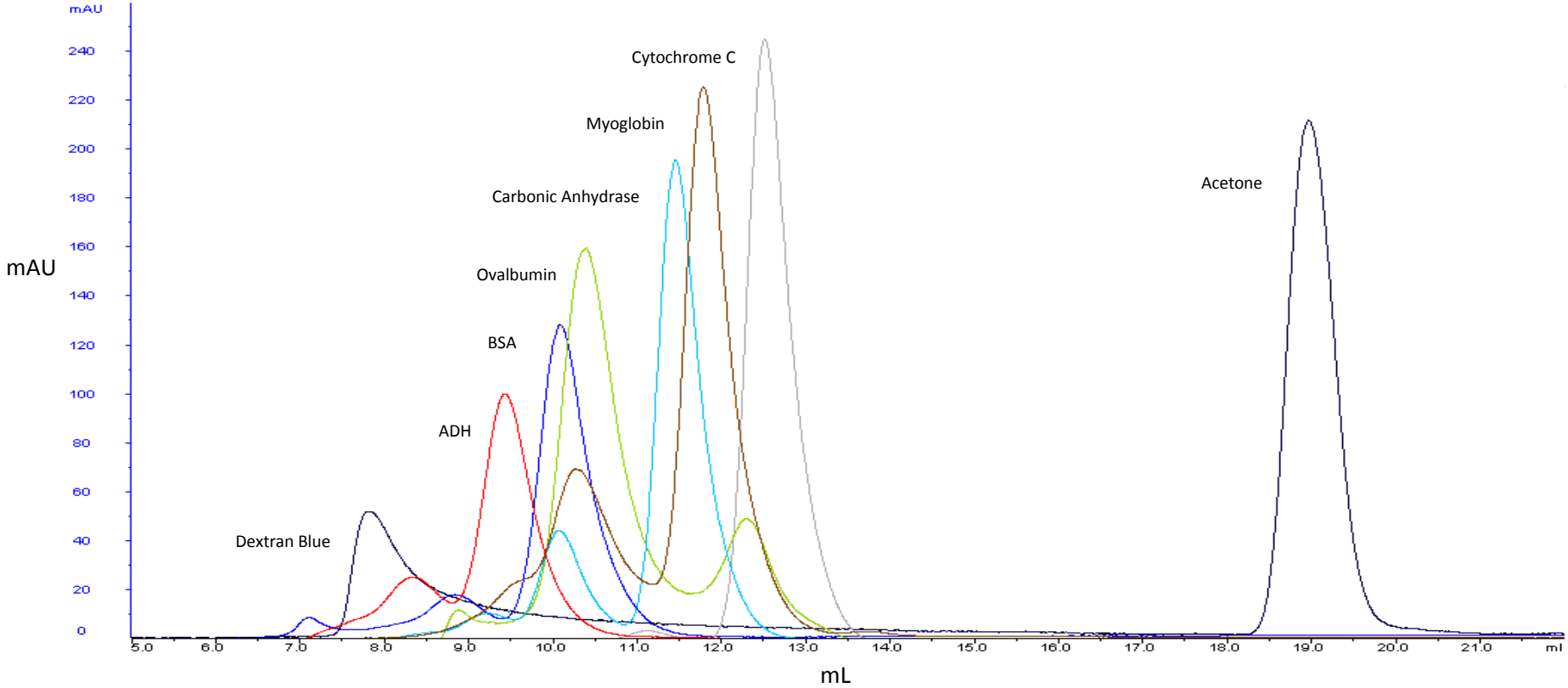


Fig 3.24: Elution of protein standards from the Superdex 75 size exclusion chromatography column. The elution profile of protein standards using size exclusion chromatography with the Superdex 75 column. The optical densities at OD280 of the eluted standard proteins are plotted against the elution volumes. The elution volume of each protein was used in the construction of the standard curve (see Fig 3.25).

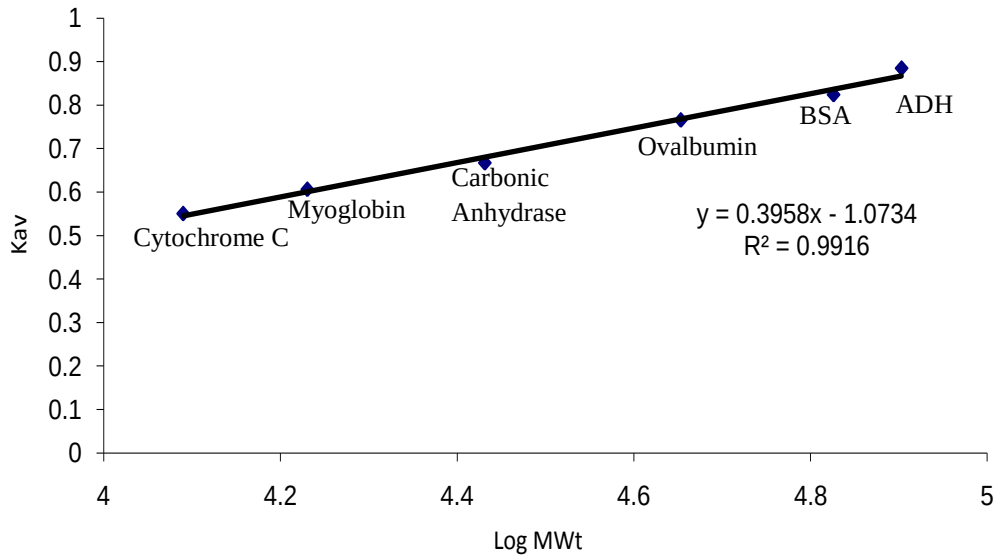


Fig 3.25: Development of size exclusion chromatography standard curve using the Superdex 75 SEC column. The protein molecular weight standard curve was created using SEC with the Superdex 75 column (see Section 2.22). The elution volumes of protein standards were as follows; ADH; 9.13 mL, BSA; 9.87 mL, Ovalbumin; 10.58 mL, Carbonic Anhydrase; 11.78 mL, Myoglobin; 12.52 mL, Cytochrome C; 13.2 mL. These values were used to construct a plot of Kav versus log MW. This could be in turn used to calculate the relative MW of all protein samples that were eluted in this range**.**

Table 3.1: Construction of a protein molecular weight standard curve for the Superdex 75 SEC column.

Standard	Elution Volume	Expected MW (Da)	log MW	Kav	Calculated Mw (Da)	% Error
ADH	9.13	80000	4.90	0.88	97240	21.55
BSA	9.87	67000	4.82	0.82	66828	0.26
Ovalbumin	10.58	45000	4.65	0.76	46631	3.63
Carbonic Anhydrase	11.78	27000	4.43	0.66	25383	5.99
Myoglobin	12.52	17000	4.23	0.60	17444	2.61
Cytochrome C	13.20	12300	4.08	0.55	12359	0.48

3.6.3 Investigation of protein compatibility with Superdex 75 SEC column

CBP21 was passed over the Superdex 75 SEC column as outlined in section 2.25.1. Initial runs suggested that the molecular weight of CBP21 as calculated from the elution volumes were too low (Table 3.2). The concentration of NaCl in the running buffer (PBS) was increased to 0.5 M, in an attempt to decrease the retention time. The expected size of CB21 without cleavage of the N-terminal signal sequence is 22709.3 Da, with the predicted mass with cleavage of the signal sequence 19860 Da. It was observed from table 3.2 that CBP21 had a similarly retarded elution volume in both 0.5 M NaCl replete and deplete PBS conditions in the Superdex 75 column (see Fig 3.26).

Due to the high carbohydrate content of the matrix it is possible that passage through the packed medium may be further retarded through interaction with the sugar component. To investigate if any binding was occurring between CBP21 and the column matrix CBP21 was applied to the column in the presence of competing sugars, as outlined in section 2.25.1. Any difference in elution volumes between sugar replete and deplete conditions could then be accounted for by matrix-protein interaction. It was found that the addition of competing sugars did not cause the protein to elute from the matrix any sooner (see Fig 3.27), again resulting in lower than expected molecular weight calculations (Table 3.3). This could indicate either a non sugar-specific interaction with the column matrix or that the appropriate competing sugar was not used. This column is therefore not suitable for CBP21 characterisation.

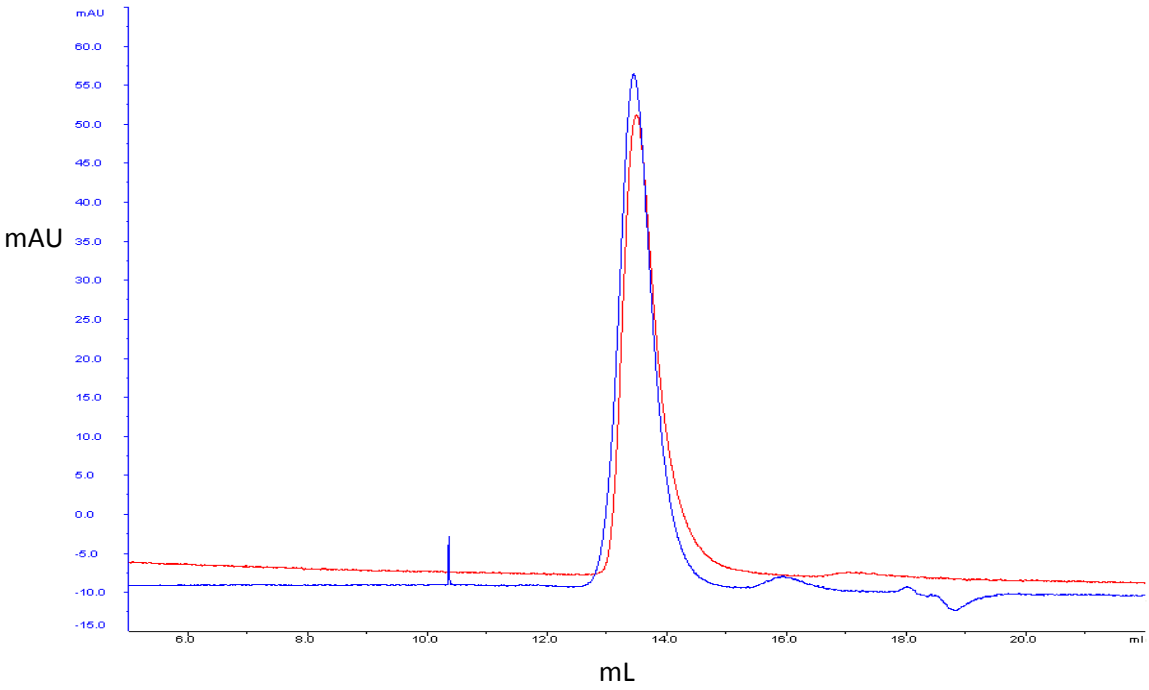


Fig 3.26: The adsorption A_{280nm} of CBP21 through the Superdex 75 column in the presence of PBS and PBS with 0.5 M NaCl. The elution profile of CBP21 in PBS is indicated with a blue line, with the elution profile in PBS with 0.5 M NaCl highlighted in red. Peak maxima are at 13.67 mL and 13.43 mL respectively.

Table 3.2: Molecular mass calculation using the Superdex 75 SEC column.

Running Buffer	Elution Volume	Kav	Log MW	Calculated MW
PBS	13.67	0.51	3.99	9739
PBS (500 mM NaCl)	13.43	0.53	4.04	10999

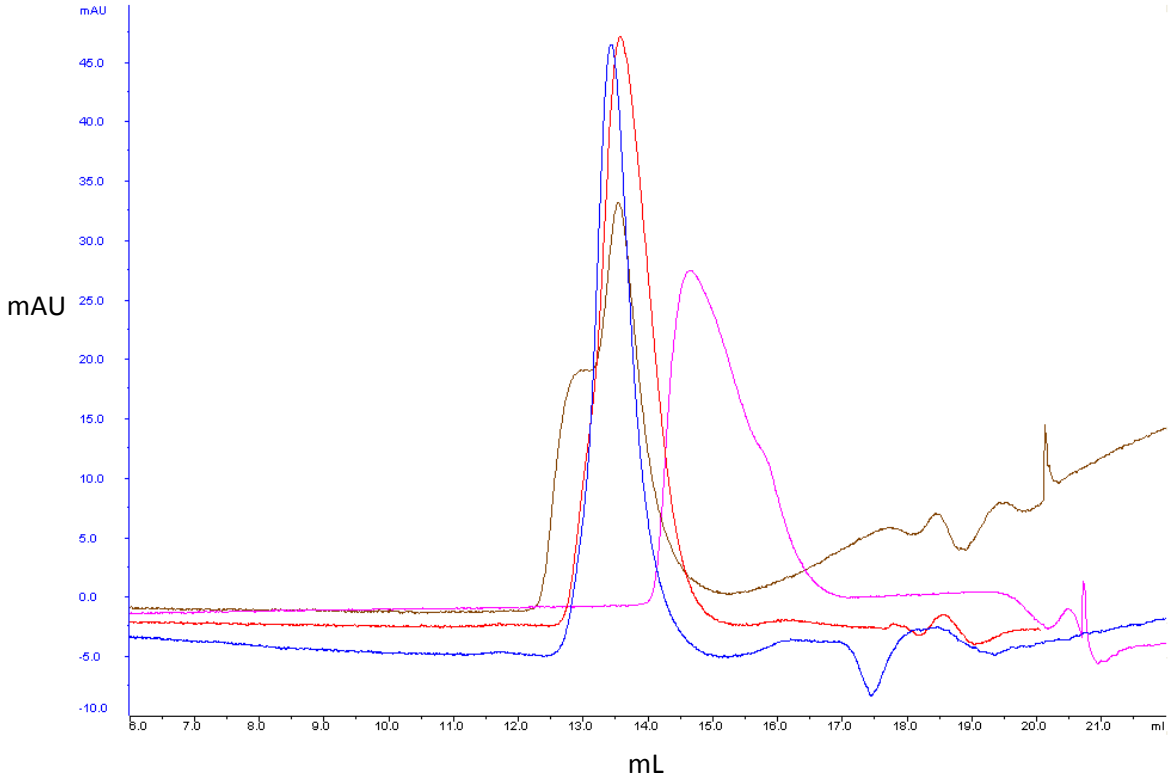


Fig 3.27: The adsorption A_{280nm} of CBP21 through the Superdex 75 column in the presence of galactose, glucose and GlcNAc. The elution profile of CBP21 in PBS is indicated in blue, galactose is indicated in red, glucose is indicated with a pink line, with the elution profile of CBP21 in GlcNAc indicated in brown. Peak maxima are at 13.43 mL, 13.57 mL, 14.65 mL and 13.53 mL respectively.

Table 3.3: CBP21 molecular mass calculations under sugar replete conditions, using the Superdex 75 SEC column.

Running Buffer	Elution Volume	Kav	Log MW	Calculated MW
PBS + 0.2 M GlcNAc	13.53	0.52	4.01	10245
PBS + 0.2 M Glucose	14.65	0.43	3.77	5927
PBS + 0.2 M Galactose	13.57	0.52	4.02	10455

3.6.4 Investigation of protein compatibility with Toyopearl HW-55S

As CBP21 interacted with the matrix in the Superdex 75 gel filtration column, it was decided to use another gel filtration matrix to examine protein size and to investigate if any non-specific interactions were causing the unexplained protein size. The Toyopearl SuperButyl-550 (HW-55S base bead) (Tosoh Bioscences, Germany) column matrix is composed of methyl-methacrylate, the compounds structure is displayed in fig 3.28, and has no carbohydrate component. As a result, it could be said that this matrix would be more suitable for the examination of carbohydrate binding molecules.

The HW-55S column had to be poured manually and as a result an extra column validation step had to be carried out. Column performance was checked at regular intervals, as manually poured columns can be susceptible to compaction over time. The number of theoretical plates (N), the figure used to validate column efficiency, was determined using the following equation.

Equation 3.2

$$N = 5.54 \left(\frac{Ve}{V_{1/2}}\right)^2 x \frac{1000}{L}$$

In this equation, Ve corresponds to the retention volume; $V_{1/2}$ is the peak width at half the peak height; and L the peak height (mm). Depending on the matrix type, typical values for column performance will vary. Values are calculated by running 0.5 mL of 1% (v/v) acetone over the column at the flowrate to be used for running samples. The retention time measured is at OD_{280nm}. The resulting elution profile can be seen in fig 3.29. Values for L, V_e and $V_{1/2}$ are measured from this profile and imputed into equation 3.2, to obtain N. This is then compared to manufacturers' operating parameters to determine if the column is operational.

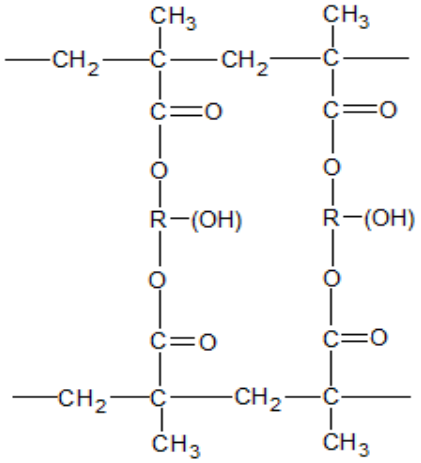


Fig 3.28: Polymeric structure of methyl methracylate. The hydroxylated polymeric structure that forms the filtration matrix in the SEC column HW-55S. R is a hydroxylated aliphatic group. Image produced using Chemsketch.

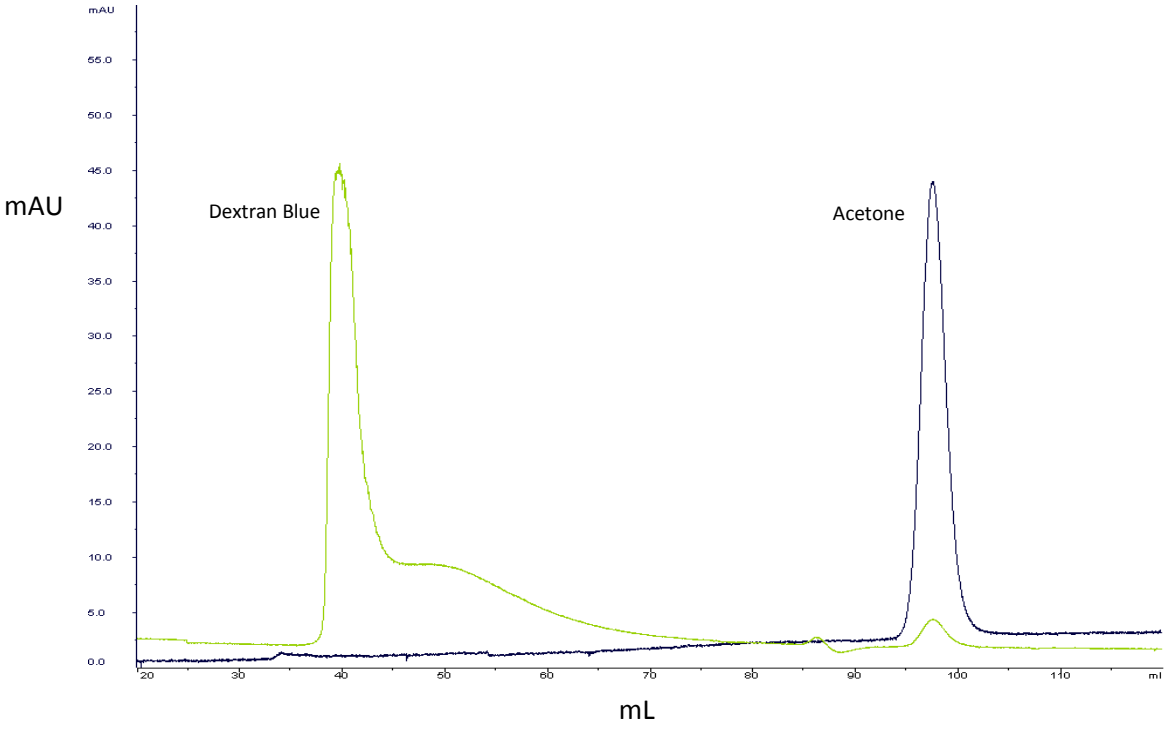


Fig 3.29: Determination of Toyopearl HW-55S column efficiency. The elution profile of blue dextran and acetone (1% v/v) using SEC with the Toyopearl HW-55S column. The optical density at OD_{280nm} of eluted blue dextran and acetone are plotted against the elution volumes. The void volume (elution volume for blue dextran was determined to be 39.67 mL. The total volume (elution volume for acetone) was determined to be 97.45 mL.

Prior to any investigation using the Toyopearl HW-55S column its efficiency was determined using equation 3.2 and the values obtained from Fig 3.29. The theoretical plate count for this column was calculated to be 8,284 which is within the operable range (3,500 – 10,000) for this matrix. Before the unknown relative molecular mass of a protein could be determined a molecular weight standard curve was constructed as before for the Superdex 75 column (see Section 3.6.2). The standards used were in the range of the expected size of CBP21, with the R^2 value of the standard curve 0.98 (see Fig 3.31).

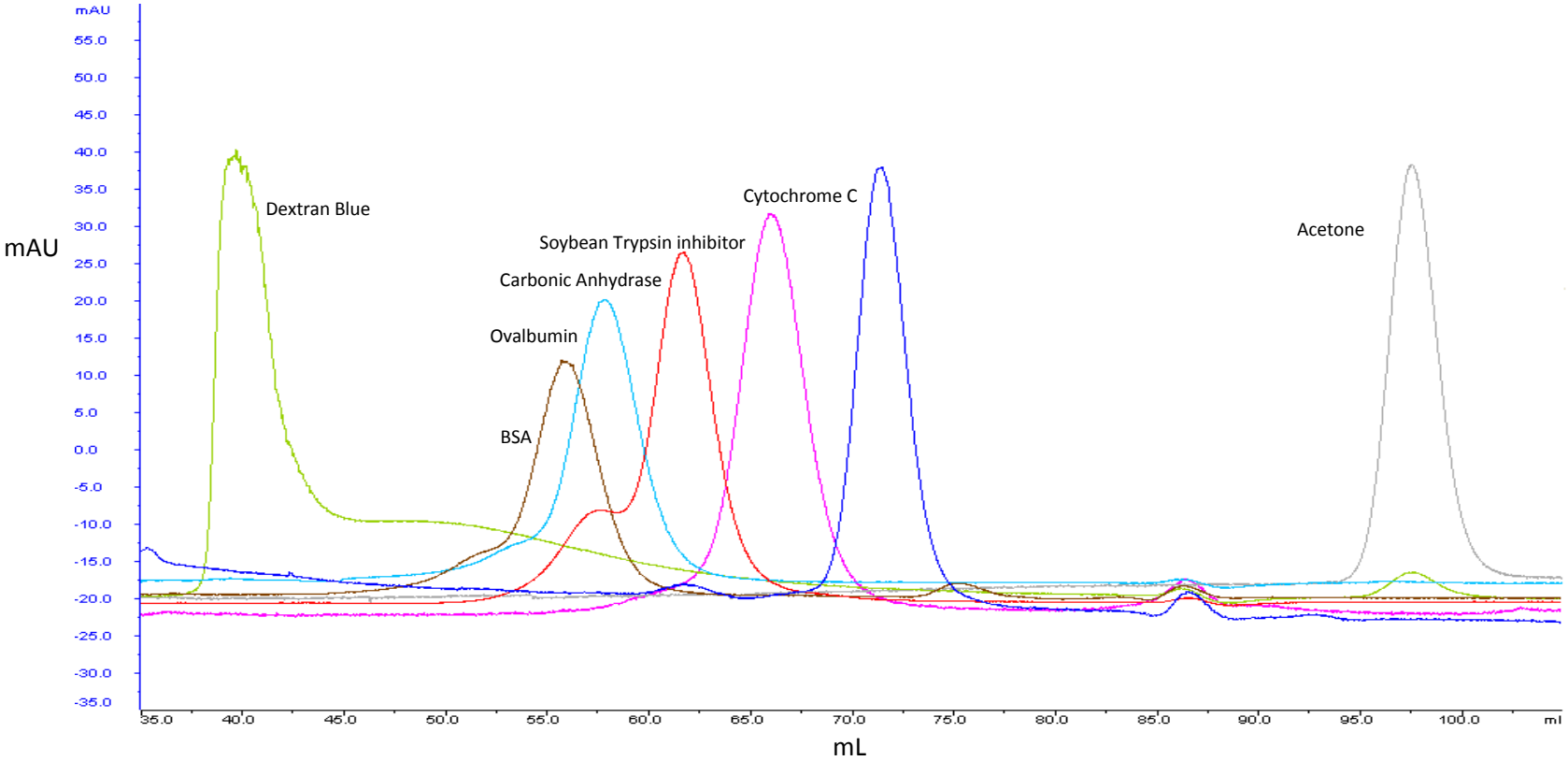


Fig 3.30: Elution of protein standards from the Tyopearl HW-55S size exclusion chromatography column. The elution profile of protein standards using size exclusion chromatography with the Tyopearl HW-55S column. The optical density at OD_{280} of eluted standard proteins are plotted against the elution volume. The elution volume of each protein was used in the construction of the standard curve (see Fig 3.31).

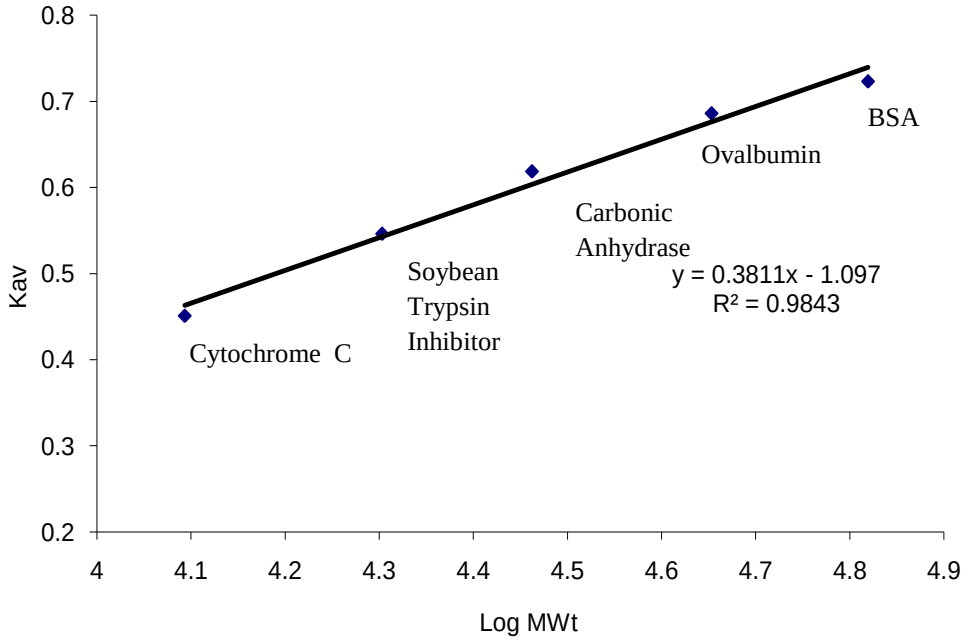


Fig 3.31: Development of size exclusion chromatography standard curve using the Toyopearl HW-55S SEC column. The protein molecular weight standard curve was created using SEC with the Tyopearl HW-55S SEC column (see Section 2.25). Void volume (Vo) of 39.67 mL was determined from the elution volume of blue dextran, and total volume (Vt) was 97.45 mL, determined with acetone. The elution volumes of protein standards were as follows; BSA; 55.66 mL, Ovalbumin; 57.80 mL, Carbonic Anhydrase; 61.70 mL, Soybean trypsin inhibitor; 65.89 mL, Cytochrome C; 71.40 mL. These values were used to construct a plot of Kav versus log MW. This could be in turn used to calculate the relative MW of all protein samples that were eluted in this range**.**

Table 3.4: Construction of a protein molecular weight standard curve for the Tyopearl HW-55S SEC column.

Standard	Elution Volume	Expected MW (Da)	log MW	Kav	Calculated Mw (Da)	% Error
BSA	55.66	66000	4.82	0.72	59778	9.43
Ovalbumin	57.80	45000	4.65	0.69	47769	6.15
Carbonic Anhydrase	61.70	29000	4.46	0.62	31771	9.56
Soybean trypsin inhibitor	65.89	20100	4.30	0.55	20500	1.99
Cytochrome C	71.40	12300	4.09	0.45	11521	7.08

3.6.5 Investigation of CBP21 compatibility with Toyopearl HW-55S column

Having successfully created a standard curve for the Toyopearl SEC column the relative molecular mass of CBP21 was estimated using this standard curve (see Fig 3.31). CBP21 was applied to the Toyopearl HW-55S column as outlined in section 2.25.2, and was run three times with a representative elution profile shown in Fig 3.32. The predicted mass of CBP21 without cleavage of the signal sequence is 22709.3 Da, with the predicted mass with cleavage of the signal sequence 19860 Da. The average elution volume for CBP21 was 66.46 mL (see Fig 3.32) which correlates to a molecular mass of 19314 Da, indicating that CBP21 is a monomer, which is processed within an *E. coli* cell with cleavage of the signal sequence.

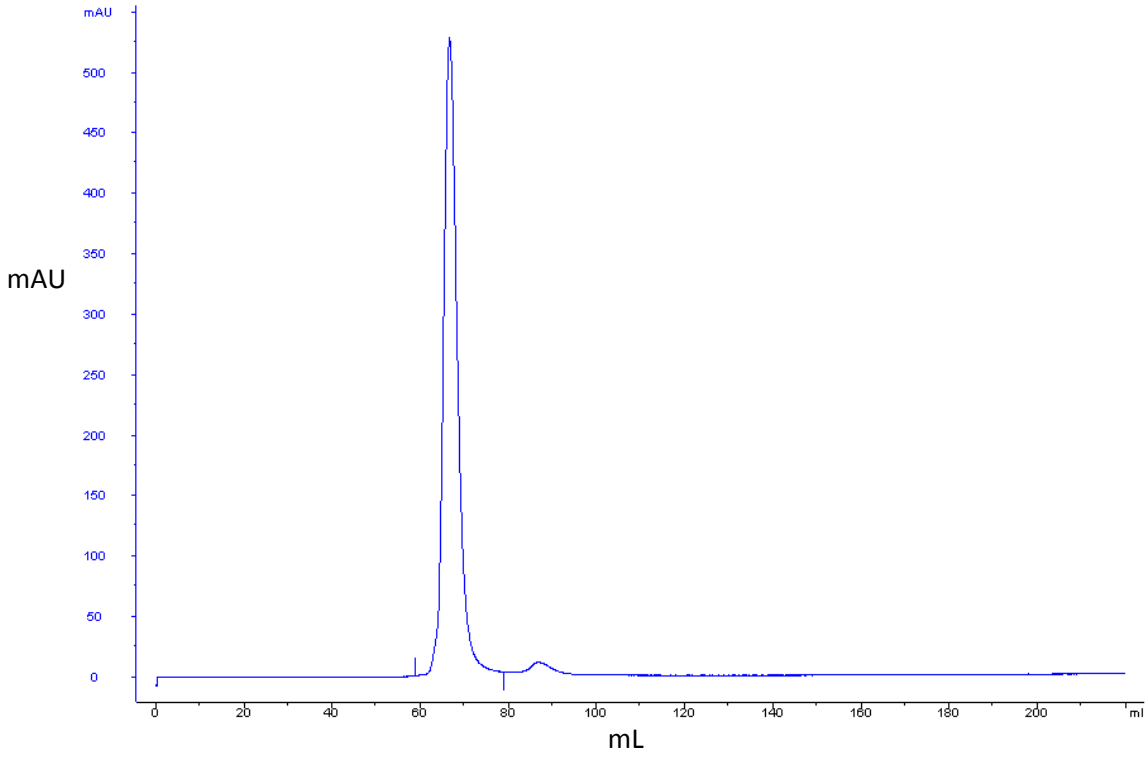


Fig 3.32: Elution profile of CBP21 in PBS pH 7.2 from the Toyopearl HW-55S SEC matrix. The elution profile of CBP21 using SEC with the Toyopearl HW-55S column. The optical density at OD280nm was plotted against the elution volume. The peak maximum was at 66.46 mL.

3.7 Mass Spectrometry

Mass spectrometry is an analytical technique used to determine the mass of particles, through measurement of the mass-to-charge ratio of charged particles. The molecule is first loaded into the MS instrument where it undergoes vaporisation, followed by ionization. The charged ions are subsequently separated according to their mass-to-charge ratio, with the ion signal processed into a mass spectrum (see Fig 3.33).

In order to confirm the size of the CBP21 protein expressed from pCBP21_60 electrospray ionization mass spectrometry coupled with a nanoflow Z-spray source was carried out (see Section 2.28). A mass spectrometry method named collision-induced dissociation (CID) was used to break up ions in the gas phase. It dissociates the ions, in this case charged proteins, by accelerating them under high electric potential in the vacuum of the mass spectrometer, before collision with a neutral gas. The transfer of kinetic energy to internal energy within the analyte results in bond breakage. Results obtained could then be compared with those obtained from the SEC as described in section 3.6.

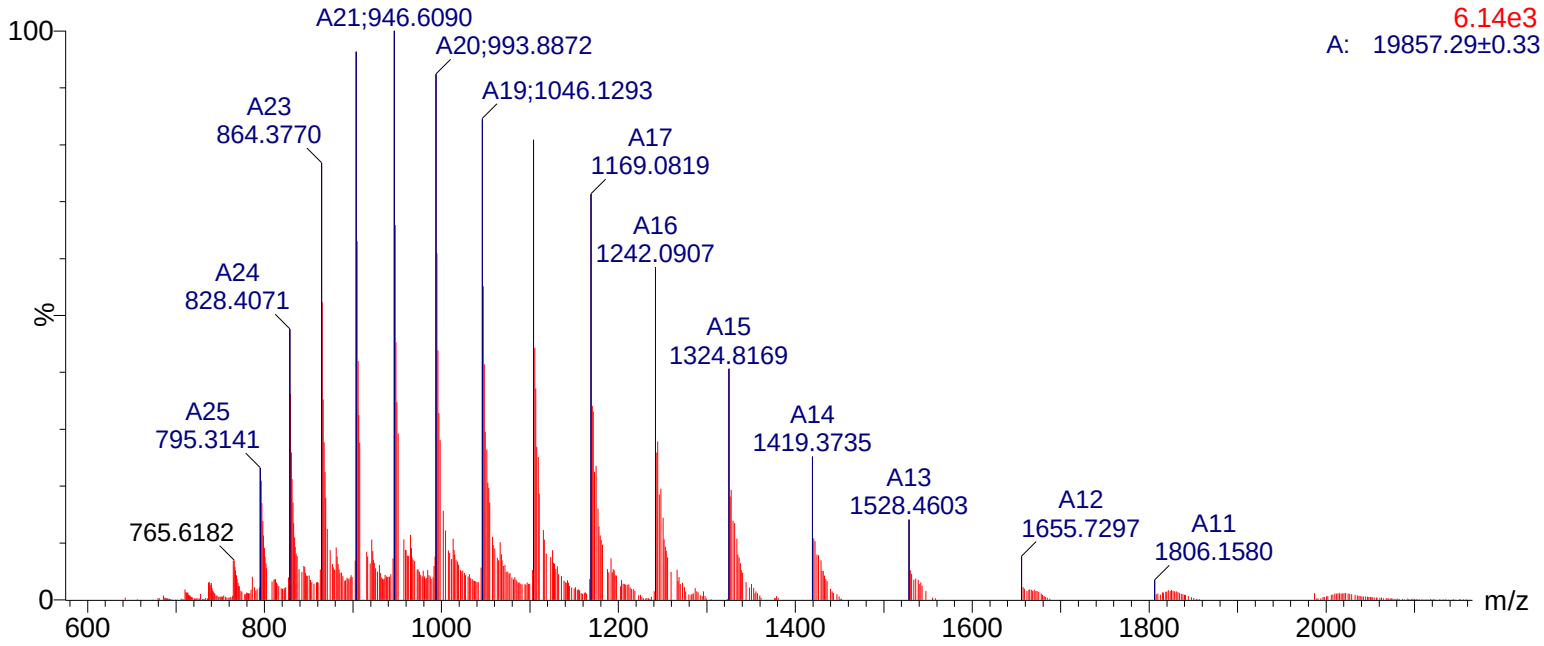


Fig 3.33: Manually de-convoluted mass spectrum of CBP21 in 50% ACN, 50% H_2O and 0.2% formic acid, using the nano flow Z-spray source. CBP21 was analysed using the nano flow Z-spray source and 50% ACN, 50% H_2O and 0.2% formic acid as a solvent. The calculated molecular mass of CBP21 was 19857.29 ± 0.33 Da.

Fig 3.33 shows that the molecular mass of CBP21 to be 19857.29 ± 0.33 Da. The predicted mass of CBP21 without cleavage of the signal sequence is 22709.3 Da, with the predicted mass with cleavage of the signal sequence 19860 Da. This indicates that *E. coli* is capable of recognising the N-terminal native CBP21 signal sequence, transporting CBP21 to the periplasm where cleavage of the signal sequence occurs.

3.8 Insoluble substrate assays

It has been established that it is possible to express soluble CBP21 protein from pCBP21_60 (see Section 3.2), and purify this protein to homogeneity (see Section 3.4). In order to determine if the $(His)_6$ tagged protein is active, solid substrate activity assays were undertaken. Wild type untagged CBP21 has previously been shown to bind to β-chitin, and to avicel (crystalline cellulose) to a lesser extent (Vaaje-Kolstad, *et al.* 2005b, Vaaje-Kolstad, *et al.* 2005a). These polymeric substrates are insoluble, unless dissolved in an acidic solution, conditions unfavourable for a protein activity assay. Many chitin binding proteins and chitin binding protein domains also appear to bind the chitinous polymers so tightly that conventional methods for eluting proteins from a matrix are ineffective (Howard, *et al.* 2003). However, SDS-PAGE loading dye can be used to elute bound proteins for gel analysis. Assessment of CBP21 interaction with the insoluble substrates chitin from crustacean shells (α-chitin), chitosan from crustacean shells (α-chitosan), chitin from squid bone (β-chitin), chitosan from carapacea skin and avicel was typically carried out as described in section 2.26, with mixing for 16 hrs (overnight).

The specificity of CBP21 for various insoluble substrates is shown in Fig 3.34 and 3.35. CBP21 appeared to bind equally well to α-chitin from crustacean shells and β-chitin from squid pen. Crystalline cellulose appeared to be the next best substrate in terms of CBP21 affinity. A little interaction between CBP21 and the chitosan from carapacea skin is observed in Fig 3.35b, with no interaction recorded with α-chitosan from crustacean shells.

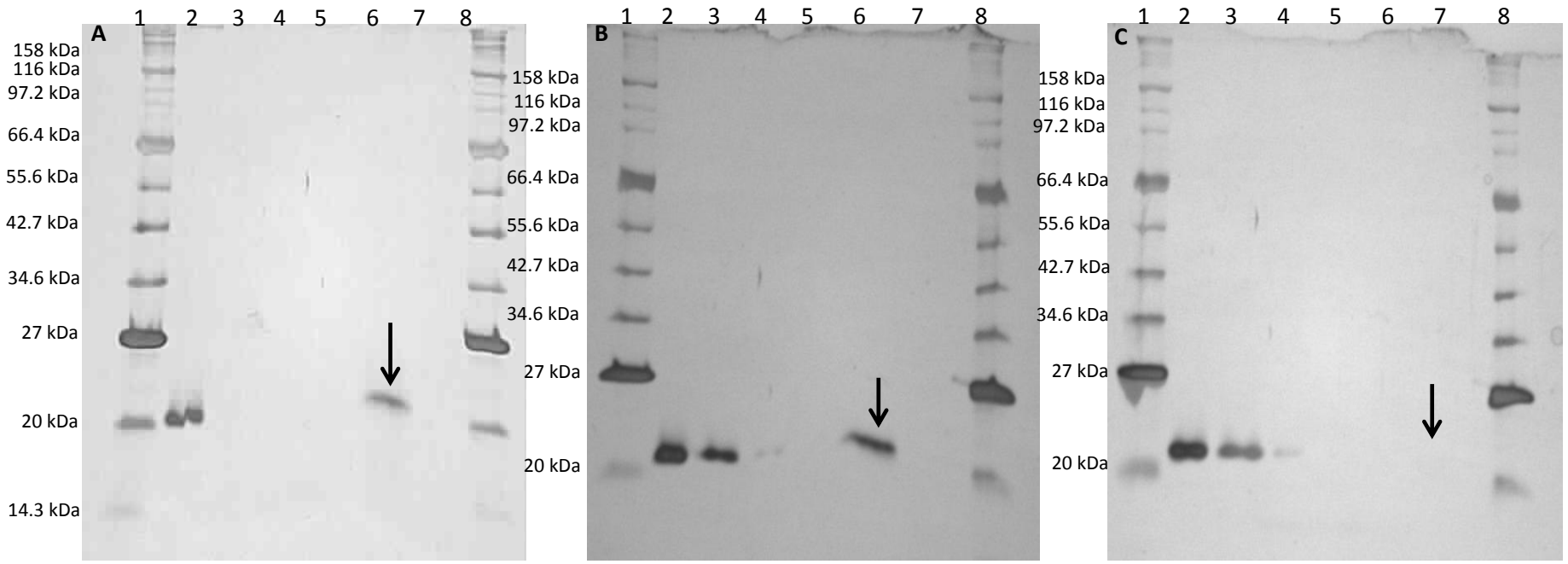


Fig 3.34: Efficiency of binding of CBP21 to various insoluble substrates. A; Crustacean shell chitin, B; Avicel, C; Chitosan from crustacean shells. 12% SDS-PAGE analysis of CBP21 insoluble substrate assay. Lane 1; Broad range protein ladder (see Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; column flowthrough, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21-substrate sample, Lane 7; Boiled PBS - substrate sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

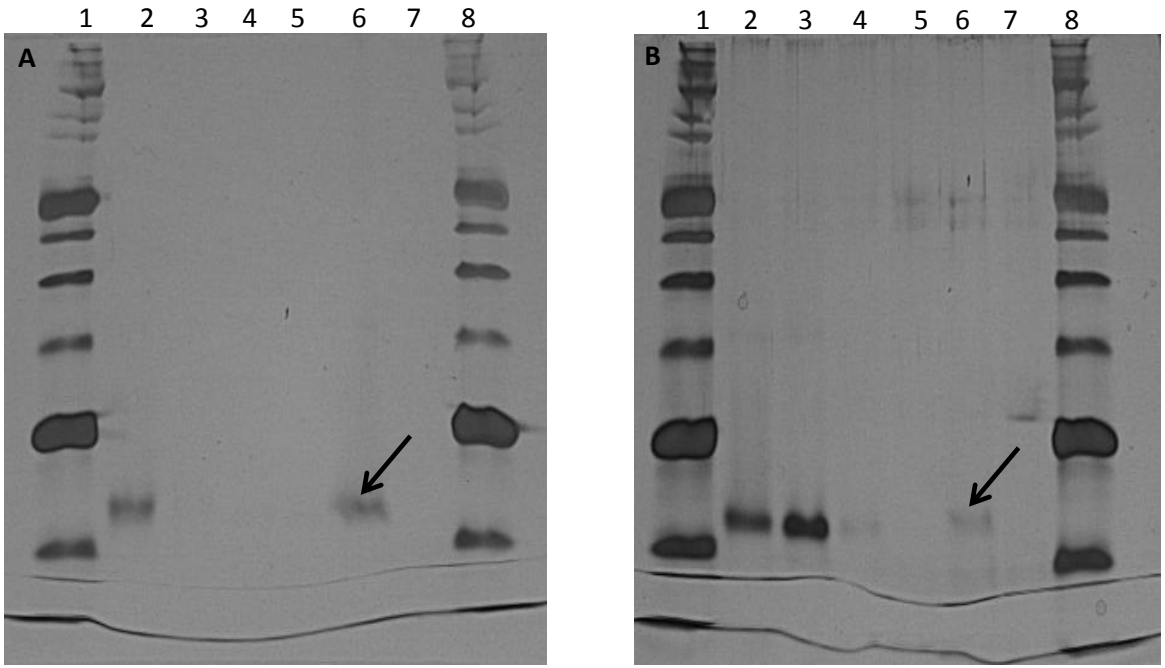


Fig 3.35: Efficiency of binding of CBP21 to various insoluble substrates, A; squid pen chitin, B; chitosan from carapacea skin. 12% SDS-PAGE analysis of CBP21 insoluble substrate assay. Lane 1; Broad range protein ladder (see Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; column flowthrough, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21-substrate sample, Lane 7; Boiled PBS-substrate sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

3.9 Discussion

The chitin binding protein CBP21 was successfully cloned, expressed and characterised, with regard to chitin binding. Extensive optimisation of its expression was essential due to the ample amount of homologous protein required for characterisation, such as size exclusion chromatography, mass spectrometry and activity assay development. The ability to produce large amounts of recombinant protein is also a pre-requisite if this molecule was to be incorporated onto an affinity matrix or into array format.

The *cbp21* gene was successfully cloned from *S. marcescens* koln genomic DNA (see Section 3.2). As outlined in sections 3.2.1 and 3.2.3 a number of constructs were examined, in order to study the effect of the position of the $(His)_6$ tag, and N-terminal signal sequence. The position of the $(His)_6$ tag and/or presence of the signal sequence were found to influence, greatly, the solubility of the target protein. The incorporation of a $(His)_6$ tag at either end of the protein was necessary for downstream protein purification using IMAC. Expression of N-tagged $(His)_6$ whole protein expression from pCBP21_30 resulted in protein localization to the insoluble fraction, with expression of C-tagged $(His)_6$ protein from pCBP21_60 resulting in soluble protein expression.

There are a number of reasons for insoluble protein production; including, incorrect protein folding, incorrect formation of disulphide bonds and the presence of inherent hydrophobic regions. As soluble CBP21 has previously been produced recombinantly, using *E. coli*, it was known that it is not an inherently insoluble protein. What was known, however, was that in *S. marcescens* the CBP21 protein is known to be involved in the degradation of chitin substrates outside the cell (Vaaje-Kolstad, *et al.* 2005a, Suzuki, *et al.* 1998, Suzuki, *et al.* 2002). In order for extracellular interactions to occur it is required that the protein is secreted from the cell, where it can gain access to the chitin polymers. Secretion of the protein from the cell requires that it must first be recognised by intracellular mechanisms. To facilitate this recognition and secretion CBPs have signal sequences at the N-terminus of the protein (~25aa). The signal peptide is

cleaved by a periplasmic signal peptidase when the exported protein reaches the periplasm.

In the case of pCBP21_30 the signal sequence and $(His)_6$ tag are both situated at the N-terminus, it was therefore possible that the position of the $(His)_6$ was interfering with the recognition of the signal sequence by *E. coli*. The cells inability to recognise the signal may have been preventing the protein from being transported to the correct cellular location, where correct protein folding could occur. It was also possible that the $(His)_6$ tag was not interfering with the recognition of the signal sequence and was simply preventing correct protein folding due to its position at the N-terminus. Although expression from pCBP21_60 gave rise to soluble CBP21 it was unclear at this point if *E. coli* was equipped to recognise the N-terminal signal sequence and perform the necessary cleavage.

Through sub-cloning the expression plasmids pCBP2130_N and pCBP2160_N were generated from pCBP21_30 and pCBP21_60 respectively. These plasmids expressed N- or C-tagged $(His)_6$ protein without the native leader sequence. While expression from pCBP2130_N resulted in no target protein over expression, CBP21 expressed from pCBP2160_N accumulated with the insoluble proteinaceous material, thus removal of the signal sequence had an adverse effect on protein solubility. This result suggested that the N-terminal signal sequence is recognised by *E. coli,* and is required for soluble protein expression.

The position of the N-terminal sequence is such that it could lie over the theoretical protein active site (see Fig 3.36). To confirm that *E. coli* was capable of recognising the native N-terminal signal sequence and transporting CBP21 across the inner cell membrane to the periplasmic space, cell compartmental analysis was undertaken (see Section 3.2.4). Having ascertained that *E. coli* was transporting CBP21 to the periplasm it was still unclear if cleavage of the signal sequence was occurring.

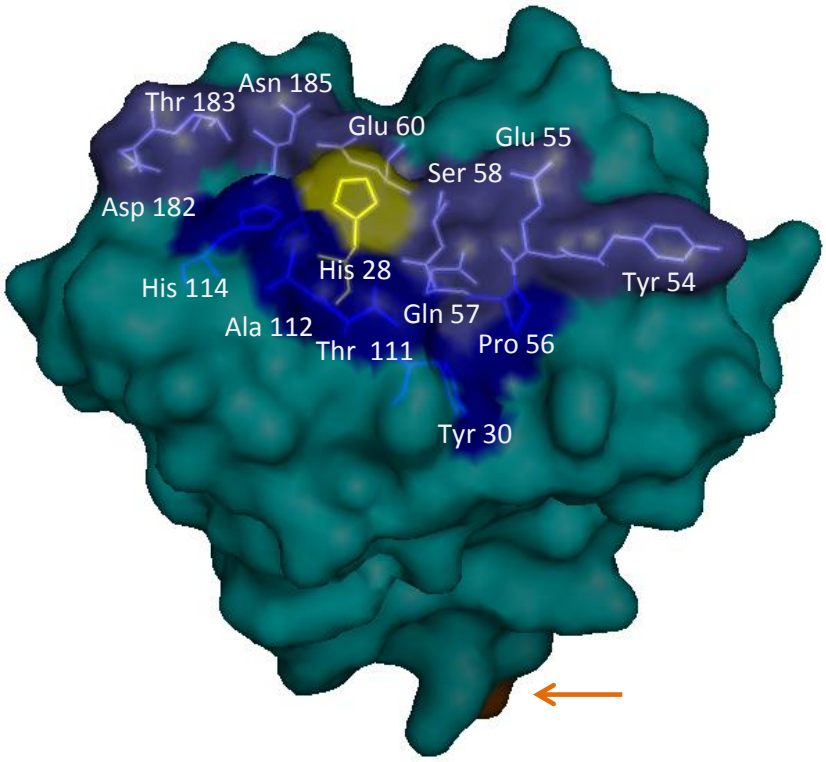


Fig 3.36: A 3-D structure of the theorized active site of CBP21. Image of the 3-D structure of CBP21. The residues thought to the responsible for chitin binding are highlighted in blue. The first N-terminal residue following cleavage of the mature protein is highlighted in yellow. The final C-terminal residue is indicated with orange.

Following the successful expression of soluble CBP21 from pCBP21_60, expression optimisation was carried out (see Section 3.3). Colony blot analysis was undertaken to determine the optimal genotype for expression of recombinant CBP21. Four *E. coli* strains, XL-10 Gold, KRX, JM109 and BL21, were transformed with pCBP21_60, to express C-terminally $(His)_6$ tagged CBP21 (see Section 3.3.1). The most affluent clone from each of the four strains was submitted to expression analysis by SDS-PAGE, with KRX selected as the most efficient clone for recombinant CBP21 expression. IPTG induction levels, incubation temperatures and harvest times were subsequently examined (see Section 3.3.2).

As IPTG concentrations and incubation temperature can influence cell growth rate, cultures may not be directly comparable, due to different cell densities. To combat this, all cultures were inoculated from the same starter culture and samples were also diluted to the same OD prior to cell harvest and analysis by SDS-PAGE. The effect of induction with varying levels of IPTG was investigated by looking at the percentage of soluble and insoluble product produced at certain stages of fermentation. It was concluded that increasing the IPTG concentration did result in increased levels of recombinant protein expression to the soluble protein fraction (see Fig 3.14), and that dramatically increasing the IPTG concentration did not lead to product accumulation in the insoluble fraction and did not adversely affect the cells growth rate (see Fig 3.15). From whole cell lysate analysis it was also determined that the optimal incubation temperature was 37°C with a harvest time at 6 hours post induction (see Fig 3.16). A harvest time of 6 hours post induction would necessitate immediate processing or overnight protein storage of the cell lysate. Although KRX is deficient in ompT and ompP proteases it has been observed (Creavin 2010) that prolonged protein exposure to other intra/extracellular proteases can cause protein degradation. Incubation at 30°C with overnight incubation would lend itself to higher cell densities compared to T = 6 with an increased amount of protein per cell. A morning harvest time would also lend itself to immediate downstream processing, thus reducing any prolonged protein exposure to proteases.

Having evaluated the conditions for optimal expression of CBP21 from *E. coli* a purification strategy was investigated. Due to the presence of the $(His)_6$ tag on the proteins C-terminus an IMAC purification strategy could be employed. The imidazole elution profile (see Fig 3.18) was used to establish suitable strategy for IMAC purification of $(His)_6$ tagged CBP21. The standard elution profile consisted of using a concentration of 20mM Imidazole in the binding buffer (wash buffer 1), followed by an intermediate wash step using 70 mM Imidazole (wash buffer 2), with a concentration of 200 mM Imidazole used for the elution of CBP21 from the matrix, free from any contaminants. These optimised conditions were transferred to the automated FPLC system, to aid in the large-

scale purification of recombinant proteins. The high level of purity obtained using IMAC can be seen in Fig 3.20. PBS was determined to be the most appropriate buffer for buffer exchange (see Fig 3.21). Using these standard expression and purification procedures, $(His)_6$ tagged CBP21 concentrations of up to 3.1 mg per g of cells were achieved.

A general feature of lectin binding sites is that they consist of a primary binding site that is capable of binding an individual monosaccharide unit, usually with a low affinity. This weak affinity is most often compensated for by the presence of multiple subunits (Loris, 2002). Consequently, any disruption to the lectins quaternary structure may have a greater impact on the lectins ability to interact with its target molecule. The crystal structure of CBP21 was determined by Vaaje-Kolstad *et al.,* (2005b), where it crystalised as a dimer. Their biochemical studies however suggested that CBP21 was a monomer in solution. In order to corroborate CBP21 conformation size exclusion chromatography was undertaken.

Initial gel filtration studies were undertaken using the Superdex-75 column from GE Healthcare. It was observed that the carbohydrate matrix impeded protein migration through the column. This migration retardation was not reverse by the addition of increased concentrations of NaCl (0.5M) to the running buffer nor excess free sugars GlcNAc, glucose, galactose (table 3.2). As the matrix consists of cross-linked dextran (polymeric glucose) and agarose (polymeric galactose), it is conceivable that while CBP21 does not appear to bind to the free monosaccharide sugar units in this instance, it may potentially interact with their polymeric counterparts. An alternative carbohydrate-free matrix, Toyopearl HW-55S, was subsequently employed for size exclusion studies. The relative molecular mass predicted for CBP21 using the Toyopearl-HW55S matrix was 19213 Da. The actual relative molecular mass for CBP21 without cleavage of the signal sequence is 22709, with the predicted mass with cleavage of the signal sequence 19860 Da. This result would indicate that the passage of CBP21 through the Toyopearl-HW55S matrix is not impeded, and that CBP21 exists in solution as a monomer.

The molecular mass of CBP21 as calculated from gel filtration also suggests that it exists in the periplasmic space in its processed form. Differences in the theoretical size as compared with the size calculated from gel filtration may be explained by the fact that the majority of standards used for the creation of the standard curve are globular proteins, thus it serves only as a rough estimate as to the size of non-globular shaped molecules. The Toyopearl-HW55S column was also poured manually, as a result the matrix will not give the same resolution as commercial columns, such as the Superdex 75 column. Therefore mass spectrometry was employed to determine the molecular mass of CBP21. The results obtained from mass spectrometry largely agreed with those from the size exclusion chromatography studies. The molecular mass as calculated using the nano-spray Z-spray source was 19857.29 ± 0.33 Da. This result confirms previous results that CBP21 exists in solution as a monomer, and that the protein is processed in *E. coli*, with the removal of the N-terminal signal sequence, thus theoretically displaying an accessible active site. The slight difference between the actual and calculated molecular mass of CBP21 of 2.71 ± 0.33 Da can be explained by slightly inaccurate calibration of the instrument with mass spec standards.

To confirm that the purified $(His)_6$ tagged CBP21 was active insoluble substrate assays were undertaken using a number of insoluble chitin and polymer derived substrates (see Fig 3.34 and 3.35). These assays confirmed that the recombinant CBP21 was active, and capable of binding to a number of different polymeric substrates, chitin in the α and β formation, chitosan and crystalline cellulose.

4.0 CBP21 Characterisation

4.1 Investigation of recombinant CBP21 affinity for glycoproteins

Following the preliminary characterisation of CBP21 with regards to its specificity for various insoluble chitin substrates (Section 3.8), a number of experiments were undertaken to examine the potential of CBP21 for use as a novel carbohydrate-binding molecule. This was carried out by investigating the specificity, if any of CBP21 for protein-linked glycans.

4.1.1 ELLA analysis of CBP21

The principle of the ELLA, similar to an ELISA has been outlined in section 1.5.5. In this study the primary antibody present in an ELISA has been replaced by a carbohydrate-binding protein, with the protein target a glycoprotein. As stated previously a $(His)_6$ affinity tag has been engineered into the recombinant protein, CBP21, which will allow for detection of CBP21–glycoprotein interaction, through the use of an anti-His murine antibody. The potential affinities of recombinant carbohydrate-binding proteins can be studied through the comparison of the binding profiles of biotinylated plant lectins. All lectin, glycoprotein and antibody concentrations are outlined in section 2.27. Calcium, magnesium and manganese are known to be required for the function of some plant lectins, due to the location of the ions in the sugar binding pocket. The dependence of CBP21 on metal ions is thus far unknown.

The binding of CBP21 to an array of glycoproteins was investigated, with the molecule showing no preference for any of the glycoproteins tested (see Fig 4.1). The glycoproteins used in this study were; bovine fetuin, bovine asialo fetuin, ovalbumin from chicken egg white, glucose oxidase from *Aspergillus niger*, invertase from *Saccharomyces cerevisiae*, bovine hyaluronidase and porcine thyroglobulin.

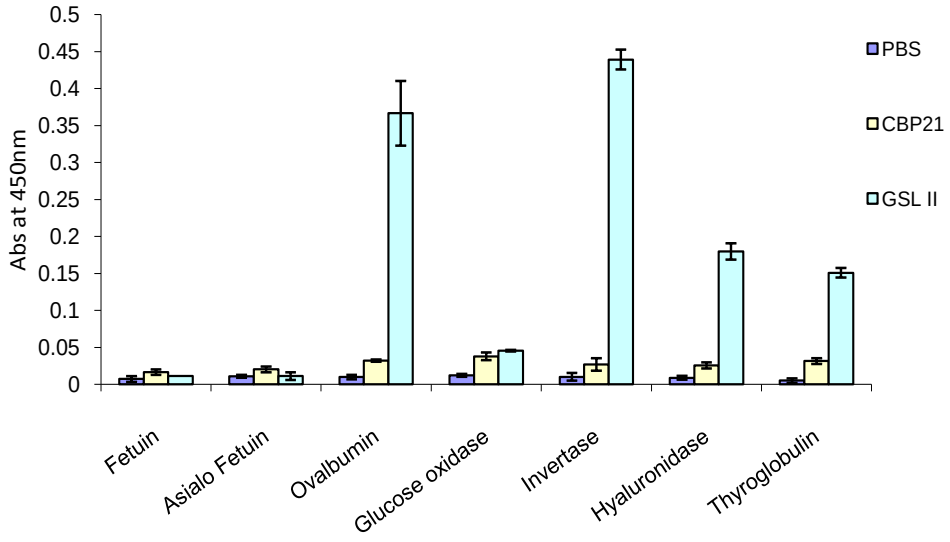


Fig 4.1: Comparison of CBP21 and GSLII ELLA binding profiles. Quantititive detection of carbohydrate-binding proteins bound to various glycoproteins, assayed by the activity of a HRP-linked antibody (OD_{450nm}). The binding profiles of CBP21 and GSLII, a comercially available plant lectin that has a binding preference for terminal GlcNAc, were compared. CBP21 did not have a similar glycoprotein binding profile to that of GSLII, and did not appear to interact with any of the glycoproteins tested to any extent.

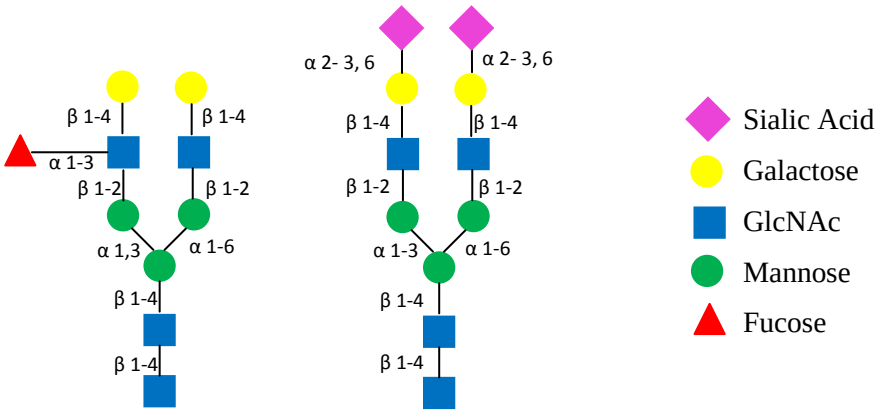
As stated in section 1.17.2 the know binding characteristics of CBP21 suggest that it may be capable of interacting with the terminal GlcNAc or core chitobiose structures of *N*-linked glycans. As outlined in section 1.4.1 mammalian lectins often contain terminal sialic acid or galactose, with terminal mannose residues most common among glycoproteins from yeasts and other microorganisms. Naturally occuring glycans with exposed terminal GlcNAc are more difficult to come across, while due to the nature of glycan synthesis it would prove impossible to find a glycoprotein displaying a simple core chitobiose structure. Glycan 'customisation' is therefore necessary to achieve the glycan structures required for CBP21 analysis.

Custom chemical synthesis of complex glycans, perhaps displaying terminal GlcNAc structures is a very expensive and time consuming process. The small quantities that could be produced in this manner would be inadequate for comprehensive carbohydrate-binding protein characterisation. A more straightforward way of creating a 'customised' glycan is through glycosidase treatment of pre-existing structures.

4.1.2 Glycoprotein treatments

The use of well defined glycosyl-hydrolases for the sequential hydrolysis of oligosaccharides is a technique commonlly used for the sequencing of glycoproteins. A wide variety of exo- and endoglycosidases, mostly of bacterial origin (Wongmadden and Landry 1995), are available commercially. In this instance a neuriminidase enzyme from *Clostridium perfringens*, capable of the hydrolysis of α 2-3, α 2-6 and α 2-8 linkages was used to cleave terminal sialic acid sturctures, while a β-galactosidase from *Bacteroides fragilis*, with a specificity for β 1-4 linkages was used for the hydrolysis of terminal galactose. The parameters of deglycosylation can vary with each glycoprotein, therefore they must be determined empirically (Tarentino and Plummer Jr. 1994). Two glycoproteins were chosen for glycosidase treatment and ELLA analysis, namely human transferrin and porcine thyroglobulin.

Transferrins are monomeric glycoproteins that function in iron binding. Human transferrin has two N-glycosylation sites at Asn^{413} and Asn^{611}, occupied by either bi-, tri- or tetraantennary glycans (Spik, *et al.*, 1975, Dorland, *et al.*, 1977, Spik, *et al.*, 1985). Due to the natural heterogeneity of the glycan stuructures whole serum transferrin includes a pool of glycovarients, 9 different transferrins can be separated electrophoretically due to differences in sialic acid content, with the two most abundant structures displayed in Fig 4.2 (De Jong and Van Eijk, 1989).



4.2 Structures of the two most anundant bi-antennary *N*-linked glycans present in human transferrin. Schematic representation of the structure of the most abundant glycan structure present on human transferrin (De Jong and Van Eijk 1989).

Thyroglobulin is a high molecular weight glycoprotein, which is the polypeptide precursor of the thyroid hormones. Analysis of thyroglobulin from several species has shown that two main carbohydrate species occur, unit-A type (oligomannose) and unit-B type (N-acetyllactossamine type). Porcine thyroglobulin is known to contain approximately 300 monosaccharides per molecule (Spiro 1965), of which unit-B type (see Fig 4.3) is the most abundant (Kamerling, *et al.*, 1988).

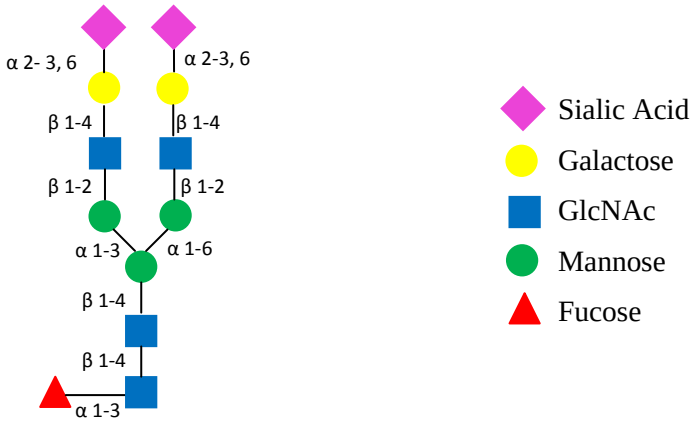


Fig 4.3: Structure of unit-B type glycan structure present on porcine thyroglobulin. Cartoon structure of the most abundant glycan structure present on porcine thyroglobulin (Kamerling, *et al.*, 1988).

Sequential treatment of these glycoproteins with neuriminidase and β 1-4 galactosidase should result in structures presenting terminal GlcNAc. Fig 4.4 displays a simplified version of the glycan structures expected following each glycosidase treatment.

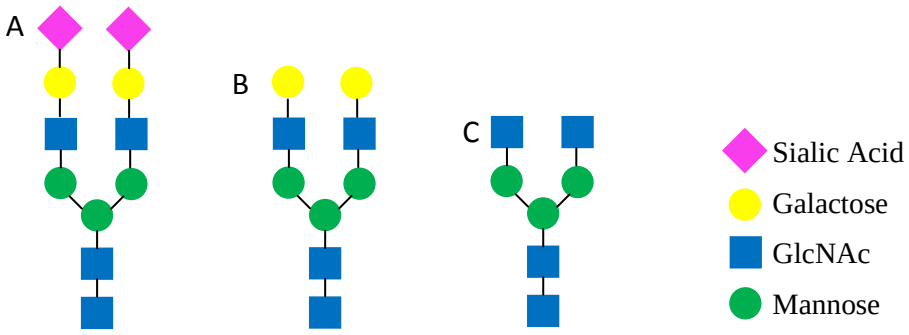


Fig 4.4: Glycoprotein terminal structures following glycosidase treatments, A; untreated glycoprotein, B; Asialo glycoprotein, C; Asialo Agalacto glycoprotein. Cartoon structure of the expected glycan patterns on human transferrin and porcine thyroglobuling following sugar release by glycosidases. Glycosidase treatment of glycans will results in the exposure of different terminal sugars. Neuriminidase enzymes will cleave all terminal sialic acid structures, while β 1-4 galactosidases will remove any terminally exposed β 1-4 linked galactose, exposing terminal GlcNAc.

ELLA analysis was carried out using untreated, neuraminidase and β 1-4 galactosidase treated transferrin and thyroglobulin. Glycoprotein samples were also probed with a number of biotinylated plant lectins, to compare binding profiles, and to gauge the efficiency of deglycosylation (see Fig 4.5). From these results it was evident that while the degylcosylation was successful, CBP21 was unable to bind to terminal GlcNAc or core chitobiose in this assay format.

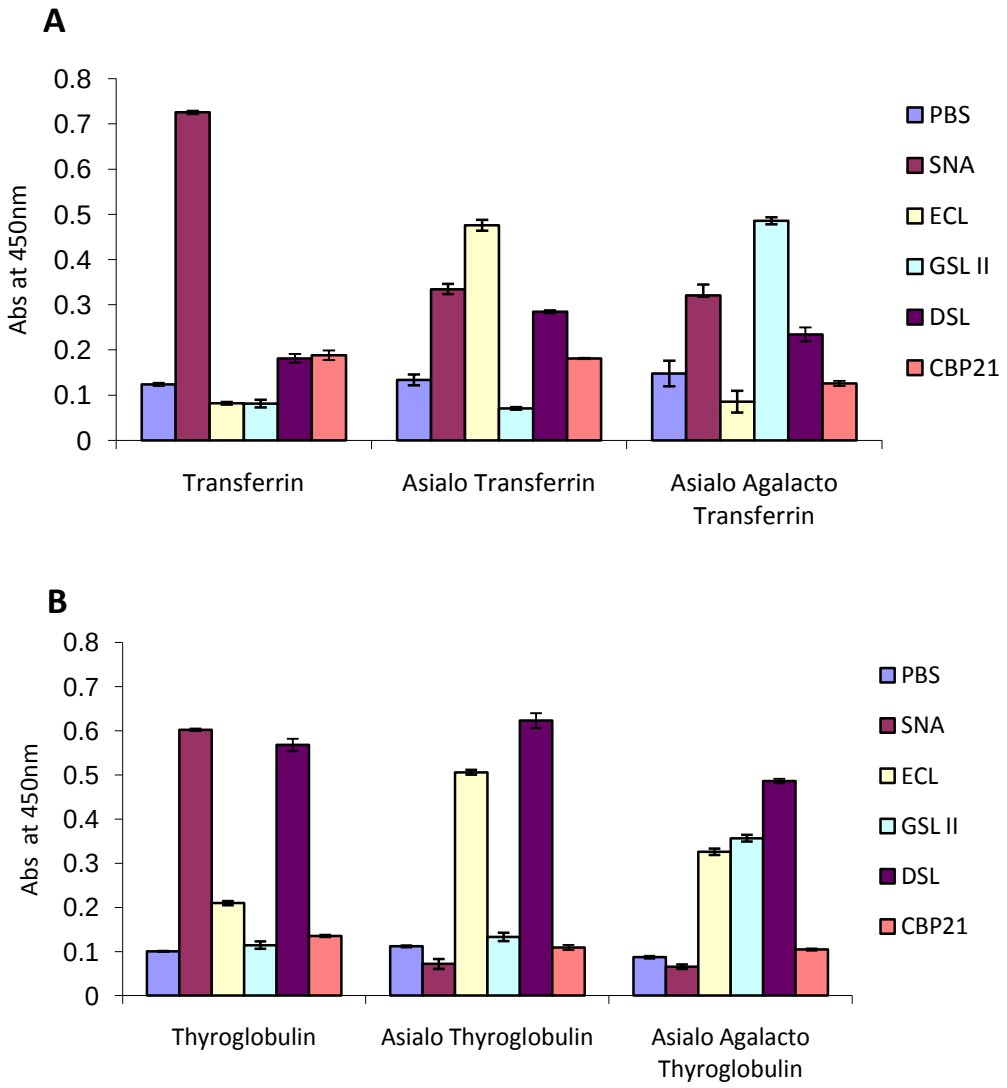


Fig 4.5: Quantitative detection of carbohydrate-binding protein interaction with A: Transferrin and B: Thyroglobulin. ELLA analysis of CBP21 and glycosidase treated transferrin and thyroglobulin. SNA is a terminal sialic binding lectin, ECL has a binding preference for terminal galactose, GSL II is a terminal GlcNAc binding lectin, DSL has been characterised as a chitin binding lectin, that appears to bind to core chitobiose.

While a neuriminidase and a β-galactosidase can be used to cleave terminal sialic acid and galactose structures, exposing the underlying galactose and GlcNAc residues, on both transferrin and thyroglobulin glycans, it is a more ardous task to cleave the underlying core mannose residues in order to expose the chitobiose core.

Invertase is an enzyme that has the ability to breakdown sucrose to fructose and glucose. Invertase also has a high carbohydrate component, which assists in correct protein folding, with all glycans of the high mannose type. Removal of the high mannose core structure was attempted using an N-acetylhexosaminidase and a α1-2,3 and α1-6 mannosidase from *Xanthomonas manihotis*. 100% cleavage of the core mannose structure to expose core chitobiose proved unachievable (see Fig 4.6).

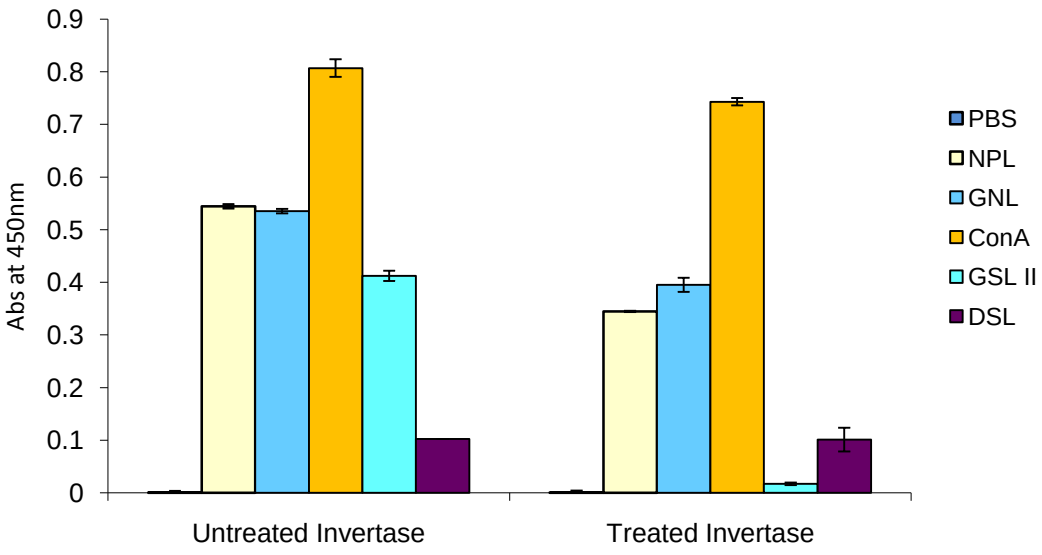


Fig 4.6: β-N-acetylglucosaminidase and mannosidase treatment of invertase. Quantatitive detection of lectin binding to untreated and β-N-acetylglucosaminidase and mannosidase treated invertase from *Saccharomyces cerevisiae.* NPL, GNL, ConA, GSLII and DSL are biotinylated plant lectins that have binding preferences for α1-6 (poly) mannose, α1-3 mannose, α-linked mannose/core mannose, α- or β- linked terminal GlcNAc, and GlcNAc oligomers respectively.

CBP21 did not appear to interact with any of the sugar species, sialic acid, galactose, GlcNAc, mannose or chitobiose, terminal or otherwise, in ELLA analysis (see Fig 4.5). The format of the ELLA may not be suitable for the characterisation of CBP21 in terms of glycan binding. An alternative method to determine if CBP21 can interact with the glycans attached to proteins is glycoprotein affinity chromatography.

4.1.3 Glycoprotein affinity chromatography

While the use of lectin affinity chromatography for the isolation of glycan species is prevalent (Section 1.5.5), the use of glycoproteins for the profiling of lectins is less commonplace. The latter can be achieved quite easily using a variety of surface chemistries. Cyanogen bromide-activated resins are one such surface that is commercially available and user friendly (Kohn and Wilchek 1982). The procedure for the coupling of proteins to cyanogen bromide activated surfaces is outlined in Fig 4.7.

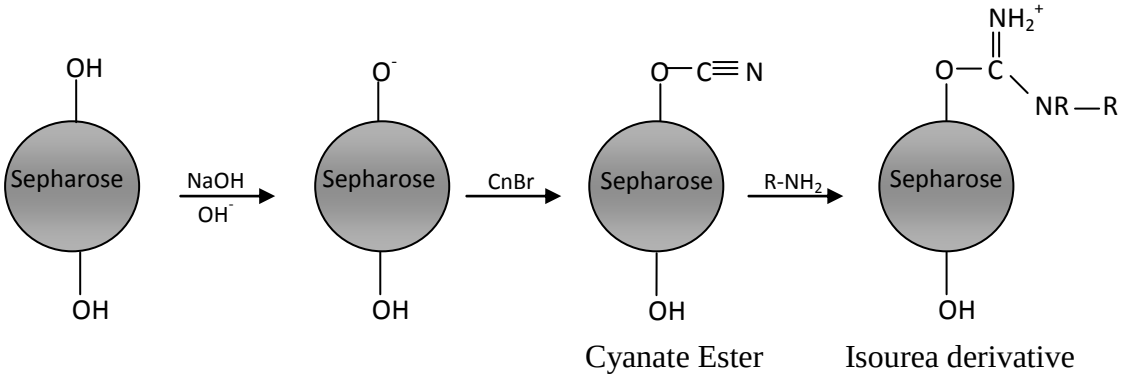


Fig 4.7: Reaction activation scheme of sepharose by cyanogen bromide and protein coupling through primary amines. Cyanogen bromide (CnBr) under basic conditions reacts with the –OH groups on agarose, forming reactive cyanate ester groups. These groups react with primary amines on the protein surface under very mild conditions, resulting in a covalent coupling of the ligand to the sepharose matrix

4.1.4 Invertase affinity chromatography

While terminal GlcNAc residues can be exposed through hydrolase treatment of oligosaccharides the enzymes are expensive, limiting the amount of glycoprotein that can be treated. Milligram quantities of glycoprotein are required for coupling to cyanogen bromide activated sepharose, therefore coupling of glycosidase treated glycans is not an option in this case. The glycans displayed on the surface of invertase are of the high mannose type. If CBP21 is capable of binding to the core chitobiose structure, the absence of galactose, sialic acid and/or bisecting fucose structures means that invertase oligosaccharides are smaller structures when compared with the more complex glycans of transferrin or thyroglobulin, theoretically resulting in less steric hindrance, giving the greatest chance of observing any interaction by CBP21. Previously GSLII, a plant lectin with an affinity for terminal GlcNAc has the shown the highest affinity for invertase compared to other glycoproteins (see Fig 4.1).

Invertase was coupled to the cyanogen bromide activated sepharose-4B, bound by the exposed primary amines on the molecules surface, as outlined in section 2.29. Table 4.1 summarises the extent of invertase immobilisation. The lectin capture ability of the invertase resin was validated using the mannose binding lectin NPL. The binding profile of NPL was then compared to that of CBP21 using OD_{280nm} readings (see Fig 4.8).

Fig 4.8 shows the attempted capture of CBP21 using invertase resin. If CBP21 was capable of binding invertase one would expect the unbound sample flow-through to contain less CBP21 than was loaded onto the column. 50 µg of CBP21 was loaded on to the resin, with an OD280nm reading of 0.072, CBP21 flow-through (OD_{280nm} 0.0426) and wash samples 1-3 (OD_{280nm} 0.028, 0.008, 0.004) were estimated to contain the entire loaded sample, indicating that no CBP21 was captured on the affinity resin.

Table 4.1: Immobilisation of invertase to cyanogen bromide activated sepharose 4B.

Starting Invertase	Dry weight resin	Unbound Invertase	Bound Invertase	mg Invertase/ mL resin
20 mg	0.66 g	5 mg	15 mg	7.5

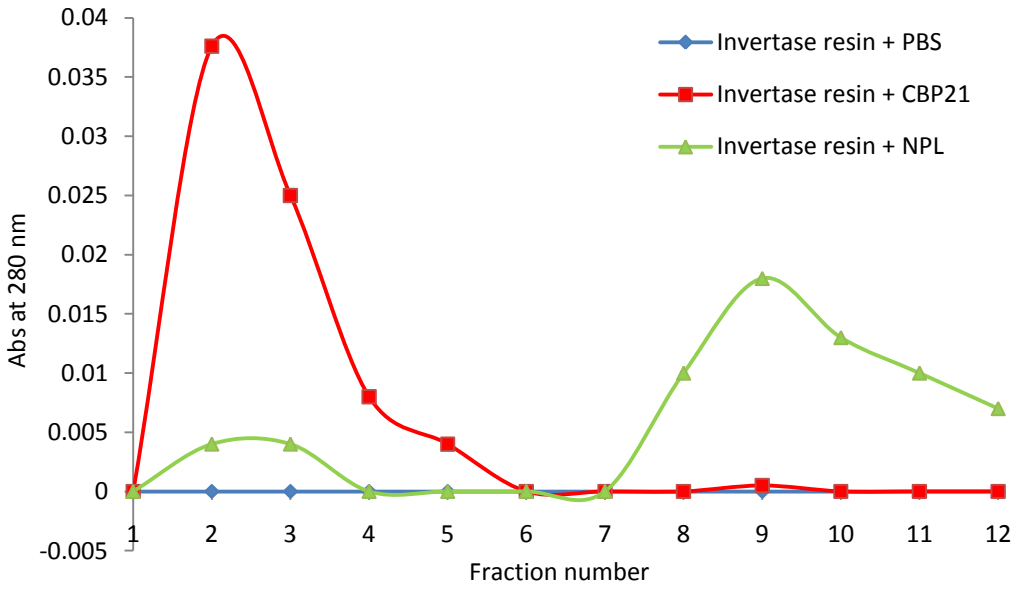


Fig 4.8: Application of carbohydrate-binding proteins to sepharose bound invertase. OD_{280nm} readings of invertase affinity chromatography using NPL and C BP21. The resin was pre-equilibrated with binding buffer until OD_{280nm} reached 0.000 (fraction 1). 1 ml of 50 µg/mL carbohydrate-binding proteins were then added to and mixed with 500 µL of invertase resin for 1 hr at room temperature. The column flow-through was collected (fraction 2), and the resin was washed with 5 x 1 mL of binding buffer (fractions 3-7), until the readings at OD_{280nm} returned to 0.000. Bound protein was eluted by the addition of 0.2 M mannose (fractions 8-12).

4.2 Establishment of a CBP21 activity assay

Although it has been show in section 3.8 that recombinant CBP21 is active and capable of binding to large chitin and chitin related polymers, it has not yet been established if CBP21 can bind to protein attached glycan structures. If a potential exists for CBP21 to interact with the chitobiose core of a glycan it is necessary to show that CBP21 can bind to smaller chitin units, such as chitobiose ($(GlcNAc)_2$) or chitotriose ($(GlcNAc)_3$). The most straightforward assaying technique would be in line with an ELLA, in 96-well MaxiSorp plate assay format. The MaxiSorp surface is a highly charged polystyrene surface with a high affinity to molecules with polar or hydrophilic groups. For direct immobilisation of a sugar molecule it would need to be attached to a support polymer, in the form of a glycoconjugate.

The use of neo-glycoconjugates for carbohydrate-binding protein profiling has escalated over the past two decades, in diverse fields, from xenotransplantation (Buonomano *et al.*, 1999) to affinity electrophoresis (Rye and Bovin, 1998). They have been exploited for use in the study of carbohydrate active enzymes (Bovin, 1998, Dupuy *et al.*, 1999), in the development of binding assays for cell adhesion molecules, including selectins and siglecs (Weitz-Schmidt *et al.*, 1996), as a tool for the study of lectins expressed on tumour cells (Kurmyshkina *et al.*, 2010), and most recently in the development of an expression system that permits genome-wide screening of proteins for interactions with glycans and proteins (Otto *et al.*, 2011).

Both uni- and multivalent glycoconjugate polymers are available from several companies including Glycotech (http://www.glycotech.com/index.htm) and Lectinity (http://www.lectinity.com/index.php). Polymers may be purchased in multiple formats included unlabelled, biotinylated and fluorescent, as well as sepharose bound probes. Polyacrylamide (PAA) is used as a high molecular weight carrier for the sugar molecules. It has a low non-specific sorption and it is stable to chemical and proteolytic action. The affinity of the probes is normally estimated at 10^2-10^5 higher than that of the corresponding free sugars. The spacer-arm for saccharides is normally $-(CH_2)_3-$, while the spacer-arm for Biotin

is -$(CH_2)_6$ (see Fig 4.9)-. The flexible polymer chain behaves as an additional spacer, with an approximate molecular weight of 30 kDa. PAA-linked biotinylated chitobiose and chitotriose were purchased from Glycotech.

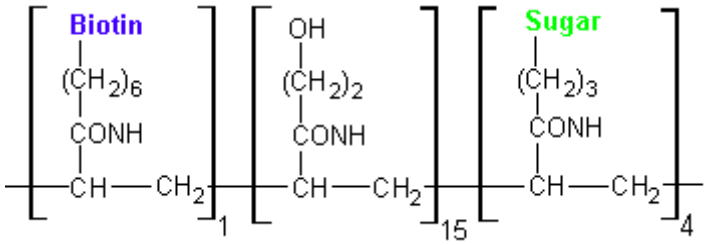


Fig 4.9: The general structure of PAA-linked multivalent biotinylated polymers. PAA is poly[N-(2-hydroxyethyl)acrylamide]. Image produced using chemsketch.

4.2.1 Determining the limit of detection for assay substrates

The first step in assay development was to ensure that both the CBP21 protein and the PAA-linked chitobiose and chitotriose oligos could be immobilised on the surface of a maxiSorp 96-well plate. The limit of detection would also need to be determined. Gradient concentrations (50 µL total volume) of protein and PAA-linked oligos were diluted in PBS buffer and immobilised on a 96-well plate, by passive adsorption, with incubation overnight at 4°C. The plates were blocked with blocking solution at 25°C for one hour and probed with anti-His or anti-Biotin antibody as appropriate, for one hour at 25°C. The peroxidase labelled antibody was subsequently detected using TMB solution (Section 2.3) and the absorbance read at OD_{450nm}.

It was determined that CBP21, chitobiose and chitotriose could be immobilised on the surface of a maxiSorp plate. It was observed that CBP21 could be detected at a concentration as low as 1 µg/mL, with no discernable difference between sample concentrations above 25 µg/mL (see Fig 4.10). Chitobiose was detectable at concentrations as low as 0.03 µg/mL (30 ng/mL), with an upper detection limit upwards of 20 µg/mL (see Fig 4.11). Chitotriose was detectable in the range of 0.01 µg/mL (10 ng/mL) to 10 µg/mL, with loss of linear detection at concentrations greater than 10 µg/mL (see Fig 4.12).

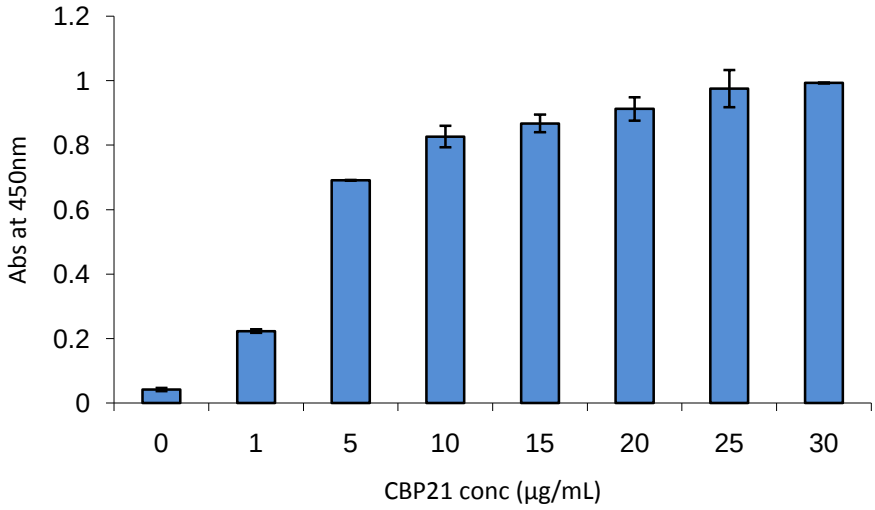


Fig 4.10: Determination of the limit of detection for CBP21. Quantitative detection of CBP21 using a peroxidase linked anti-$(His)_6$ antibody. 2.5% BSA was used as a blocking agent.

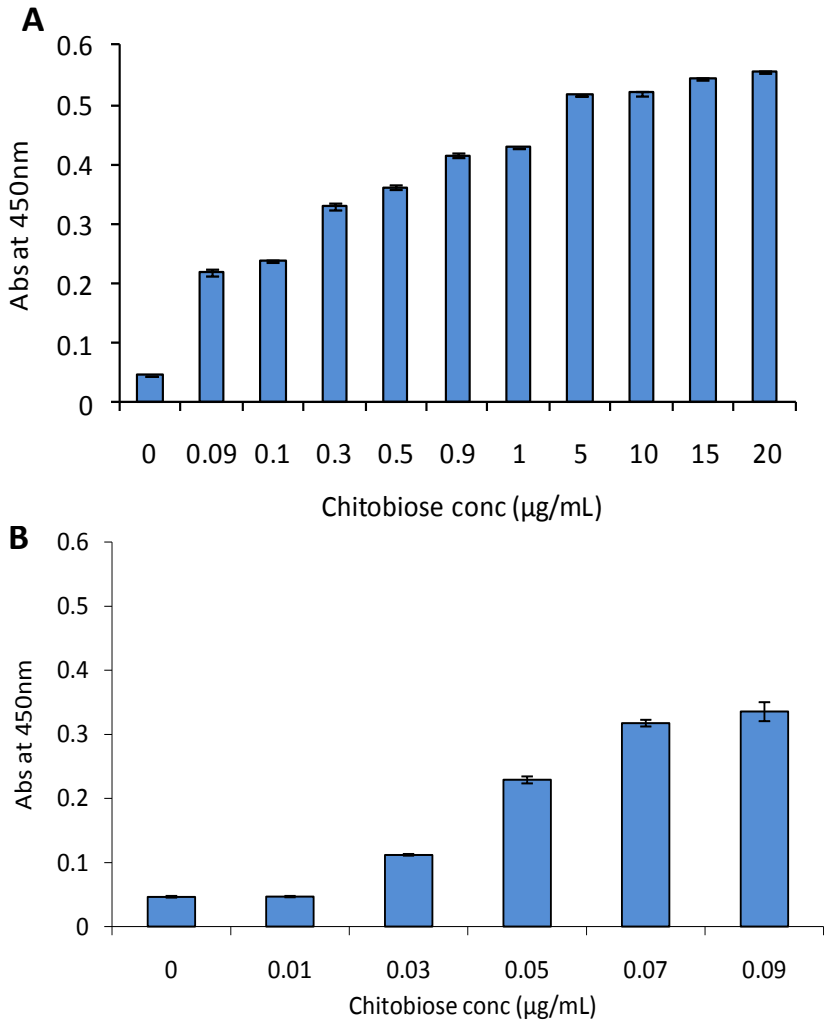


Fig 4.11: Calculating the limit of detection for chitobiose, A; chitobiose concentrations in the range 0.09 – 20 µg/mL, B; chitobiose concentrations in the range 0.01 – 0.09 µg/mL. Calculating the limit of detection of chitobiose using anti-Biotin antibody. 1% PVA (30,000-70,000 average mol wt, Sigma-Aldrich) was used as the blocking agent.

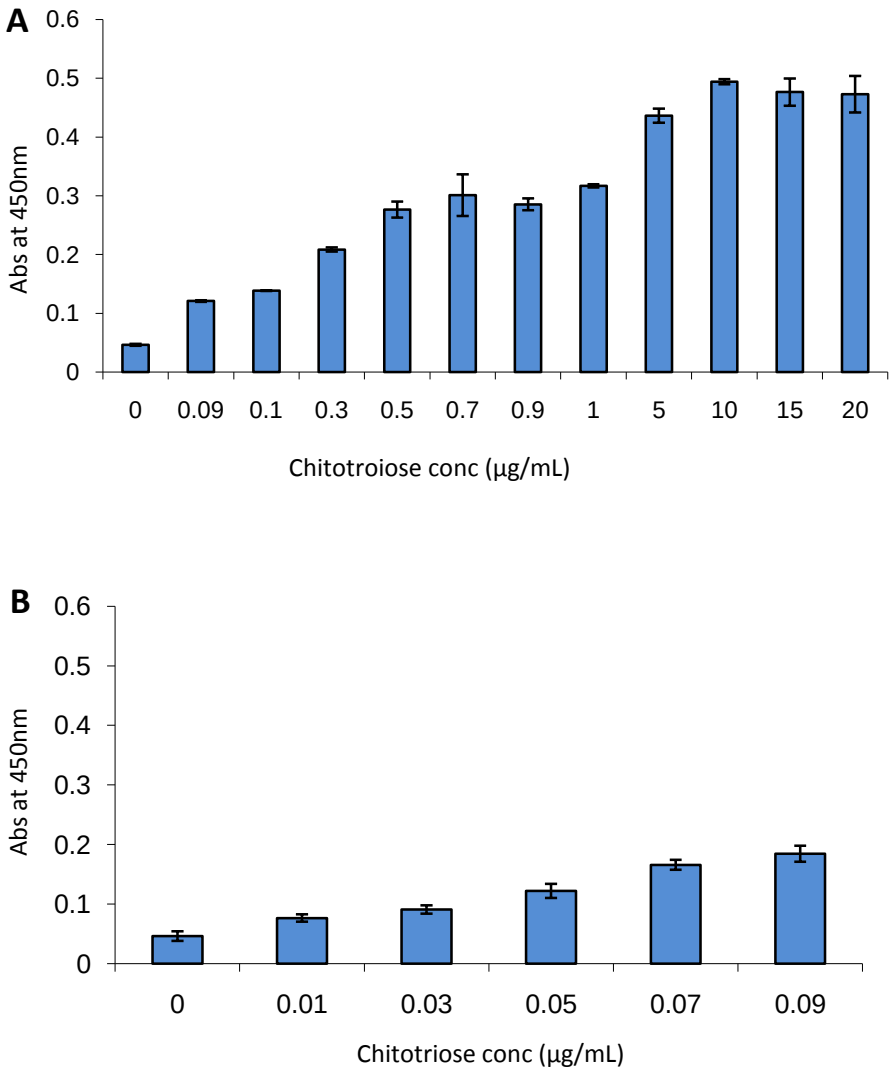


Fig 4.12: Calculating the limit of detection for chitotriose, A; chitotriose concentrations in the range 0.09 – 20 µg/mL, B; chitotriose concentrations in the range 0.01 – 0.09 µg/mL. Calculating the limit of detection of biotinylated chitobiose, using a peroxidase linked anti-Biotin antibody. 1% PVA (30,000-70,000 average mol wt, Sigma-Aldrich) was used as the blocking agent.

4.2.2 Assay format development

Once it had been established that CBP21 and the PAA-linked GlcNAc oligos could be immobilised on a maxiSorp 96-well plate the most appropriate assay format was explored. There were two potential assay formats, as outlined in Fig 4.13. Assay format A was carried out with immobilisation of 20 µg/mL of CBP21 in PBS on a 96-well maxiSorp plate overnight at 4°C. The plate was blocked for one hour at 25°C with 2.5% BSA in PBST. The blocked wells were subsequently probed with 50 µL of 10 µg/mL chitobiose in PBST for one hour at 25°C. The wells were washed four times with 200 µL of PBST prior to the addition of 50 µL 1/10,000 anti-Biotin antibody in PBST, which was subsequently incubated for one hour at 25°C. The wash step was repeated four times with PBST, the peroxidase linked antibody was detected by the addition of 100 µL TMB solution, the reaction stopped by the addition of 50 µL 10% H_2SO_4. The absorbance was read at 450nm (see Fig 4.14a).

Assay format B was carried out with immobilisation of 10 µg/mL of chitobiose in PBS on a 96-well maxiSorp plate overnight at 4°C. The plate was blocked for one hour at 25°C with 2.5% BSA in PBST. The blocked wells were subsequently probed with 50 µL of 20 µg/mL CBP21 in PBST for one hour at 25°C. The wells were washed four times with 200 µL of PBST prior to the addition of 50 µL 1/10,000 anti-His antibody in PBST, which was subsequently incubated for one hour at 25°C. The wash step was repeated with PBST. The peroxidase linked antibody was detected by the addition of 100 µL TMB solution. The reaction was stopped by the addition of 50 µL 10% H_2SO_4, with the absorbance read at 450nm (see Fig 4.14b). The results from assay format A indicated that the chitobiose PAA-linked oligo was interacting with the blocking agent (2.5% BSA), although some interaction between CBP21 and the GlcNAc polymer was detectable. This result highlighted the need for an alternative blocking agent. Non-specific binding of the PAA-linked oligos to BSA would result in high background signals that could lead to false positive results. The absorbance readings from assay format B seemed less promising. Although the results suggested that there was no interaction between CBP21 and the chitobiose polymer.

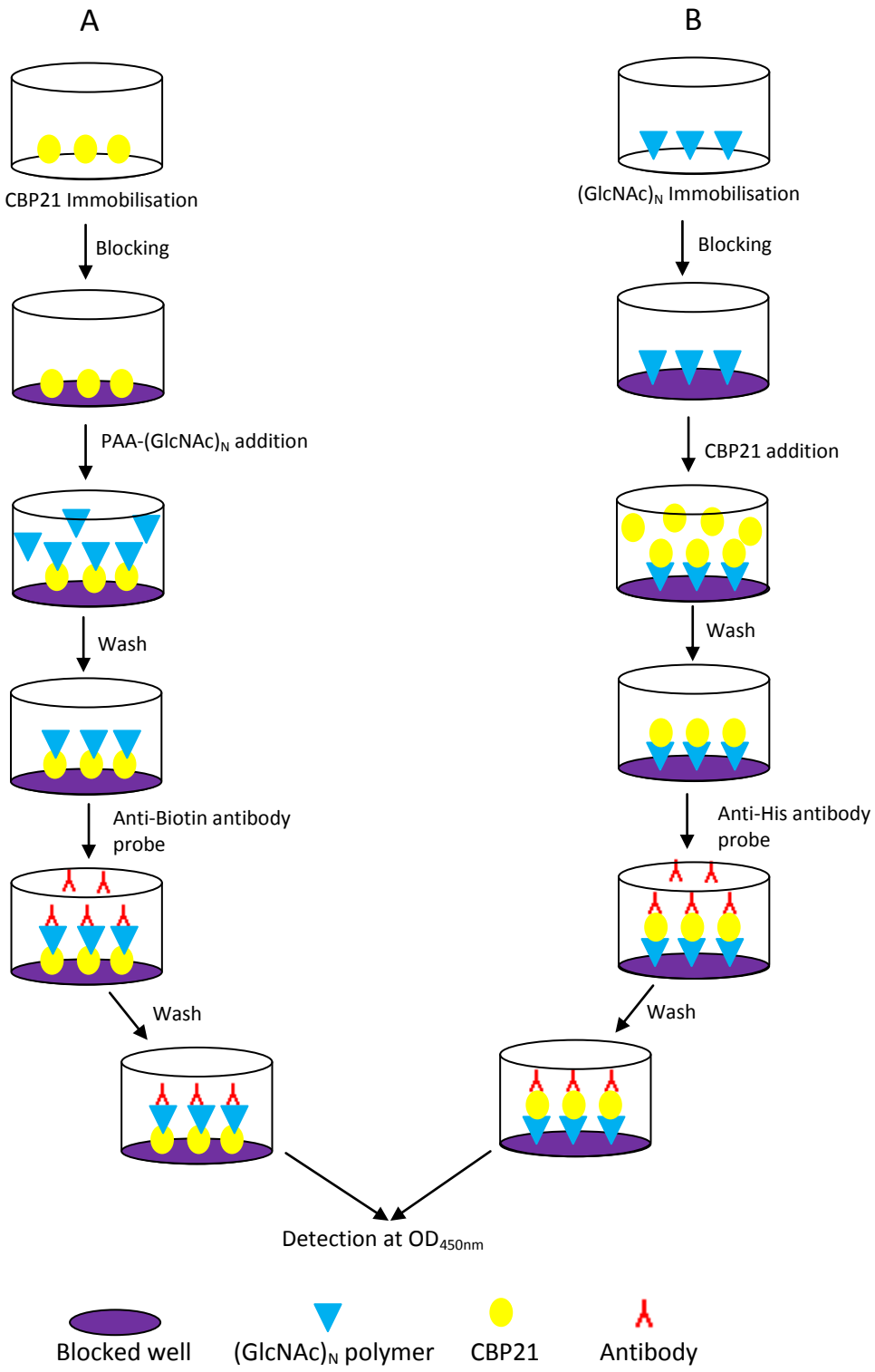


Fig 4.13: Potential assay formats for the detection of CBP21 interaction with PAA-linked $(GlcNAc)_N$ oligos. A; Assay format A, immobilisation of CBP21, blocking, probing with Biotin labelled $(GlcNAc)_N$, with detection by peroxidase linked ant-Biotin antibody. B; Assay format B, immobilisation of PAA-linked $GlcNAc)_N$, blocking, probing with $(His)_6$ labelled CBP21, with detection by peroxidase linked anti-His antibody.

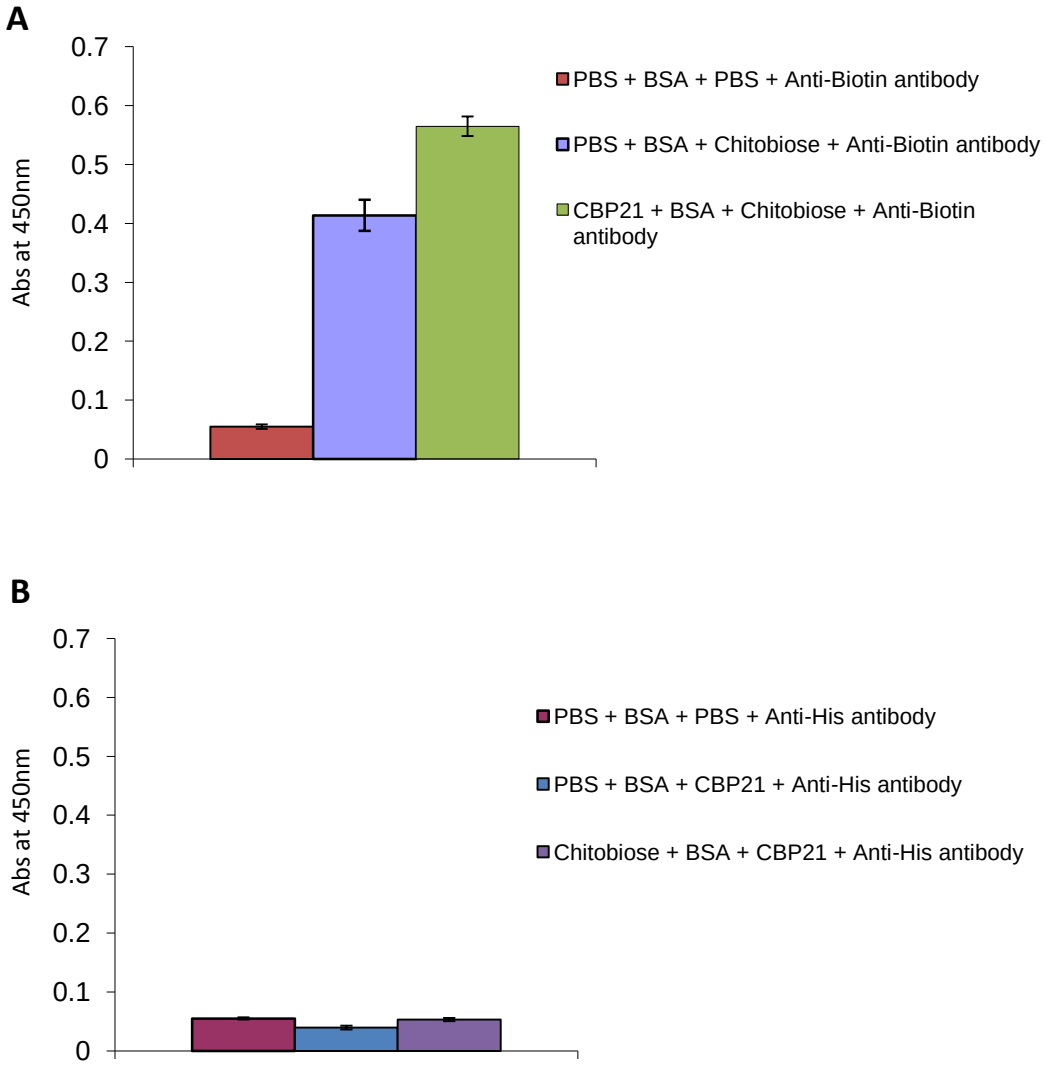


Fig 4.14: Development of CBP21 – $(GlcNAc)_N$ Binding assay. Two alternate assay formats were explored. $(His)_6$ labelled CBP21 was immobilised in assay format A, with probing by biotin labelled PAA-linked chitobiose. Biotinylated PAA-linked chitobiose was immobilised in assay format B, with the subsequent addition of $(His)_6$ labelled CBP21. The blocking agent used was 2.5% BSA. The absorbance was read at OD_{450nm}.

4.2.3 Blocking agent optimisation

Agents are used in ELISA type assays for blocking possible excess solid surface following protein immobilisation, avoiding non-specific immobilisation of succeeding reactants. Typical blocking reagents are neutral macromolecules, large enough to establish a stable attachment to the surface, yet small enough to find its way between reactants. While BSA (67,000 Da) is probably the most commonly used blocking agent the optimal blocking reagent for any particular assay should be determined empirically (Vogt Jr. *et al.*, 1987). The results from section 4.2.3 suggested that assay format A was the most appropriate. To use this assay format however the use of an alternative blocking agent would be required. While BSA is perhaps the most common blocking agent used in ELISAs there are many different types of blocking buffers available commercially. These include protein blocking buffers, protein-free blocking buffers, detergent blockers and SuperBlock blocking buffers (Rosas *et al.*, 2000).

The protein-free blocking buffers explored here were polyvinyl alcohol (PVA) (Rodda and Yamazaki, 1994, Studentsov *et al.*, 2002), polyvinylpyrrolidone (PVP) (Studentsov *et al.*, 2002, Mallorqui *et al.*, 2010, Bouchez-Mahiout *et al.*, 2010) and Synblock (AbD Serotec) (Afrough *et al.*, 2007). Protein based blocking buffers Carbo-Free blocking solution (vector labs) and casein (Kaur *et al.*, 2002) were also compared. The blocking buffers were evaluated using direct detection of CBP21 with peroxidase linked anti-His antibody. Varying concentrations of CBP21 were immobilised on the surface of maxiSorp plate, overnight at 4°C. The wells were subsequently blocked with 200 µL of blocking agent diluted in PBS, for one hour at 25°C. CBP21 was then probed with 1/10,000 dilution of anti-His antibody diluted in PBST, for one hour at 25°C. The wells were washed four times with PBST before development with TMB solution, the reaction stopped with 10% H_2SO_4 and the absorbance read at OD_{450nm}.

The ensuing results suggested that while some of the blocking agents did not block as well as others, some seemed to block any signal from developing (see

Fig 4.15). The tryptone solution (casein digests) was the least successful blocking agent, the signal at OD_{450nm} was the same across all CBP21 concentrations suggesting non-specific binding of the anti-His antibody to the plate surface. Wells blocked with Synblock yielded a higher than average signal on blank wells, with quenching of signal on CBP21 immobilised wells. Both PVA and PVP blocked wells displayed high degrees of signal quenching across all CBP21 concentrations, with very little discernable increase in signal with increasing CBP21 concentration. Carbo-free blocking solution from Vector labs was the most impressive agent next to BSA. While some degree of signal quenching was obvious when compared with BSA, there was a steady increase in signal in accordance with immobilised CBP21 concentrations, with very little difference between triplicate values. While Synblock is a protein based blocking solution PVA and PVP are synthetic blockers. These synthetic blockers consist of large polymers, with an average molecular weight of 30-70 kDa for PVA and 360 kDa for PVP, much larger than CBP21. In order to determine if the size of the synthetic polymers was influencing this phenomenon the experiment was repeated using PVA with an average molecular weight of 9-10 kDa (see Fig 4.16). The same result was observed using this lower molecular weight PVA polymer.

To establish if this problem of signal quenching was unique to CBP21 and/or the anti-His antibody the same experiment was undertaken using NPL (see Fig 4.17). NPL is a biotinylated plant lectin that exists as a dimer in solution, each monomer approximately 13 kDa, giving an overall mass of 26 kDa. Upon signal development it because apparent that while PVP didn't act as a sufficient blocker in this experiment (see Fig 4.17a) signal quenching of PVA blocked wells was again evident when compared to carbo-free or BSA blocked wells (see Fig 4.17b). This result suggested that the blocking agent was influencing the ability of the antibody to recognise the immobilised proteins. Although some signal quenching was observed when carbo-free blocking solution was used to block the plate wells it was determined that it was the best option going forward next to the unsuitable BSA.

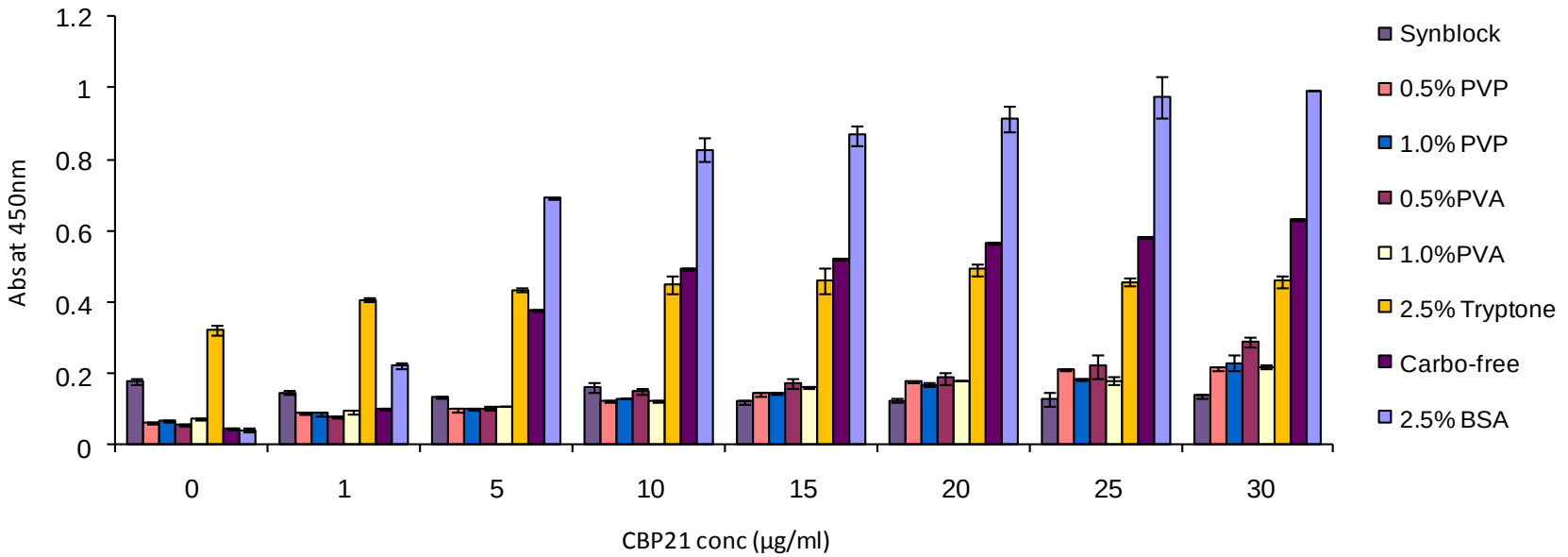


Fig 4.15: Comparison of blocking reagents. Quantitative detection of CBP21 using a peroxidase linked anti-$(His)_6$ antibody in differently blocked wells of a Nunc ELISA plate. The blocking agents investigated were Synblock (AbD Serotec), 1% PVA (30,000-70,000 average mol wt, Sigma-Aldrich), 0.5% PVA (30,000-70,000 average mol wt, Sigma-Aldrich), 1% PVP (360,000 average mol wt, Sigma-Aldrich), 0.5% PVP (360,000 average mol wt, Sigma-Aldrich), 2.5% Tryptone (LabM), Carbo-free blocking solution (Vector labs), 2.5% BSA (Sigma-Aldrich).

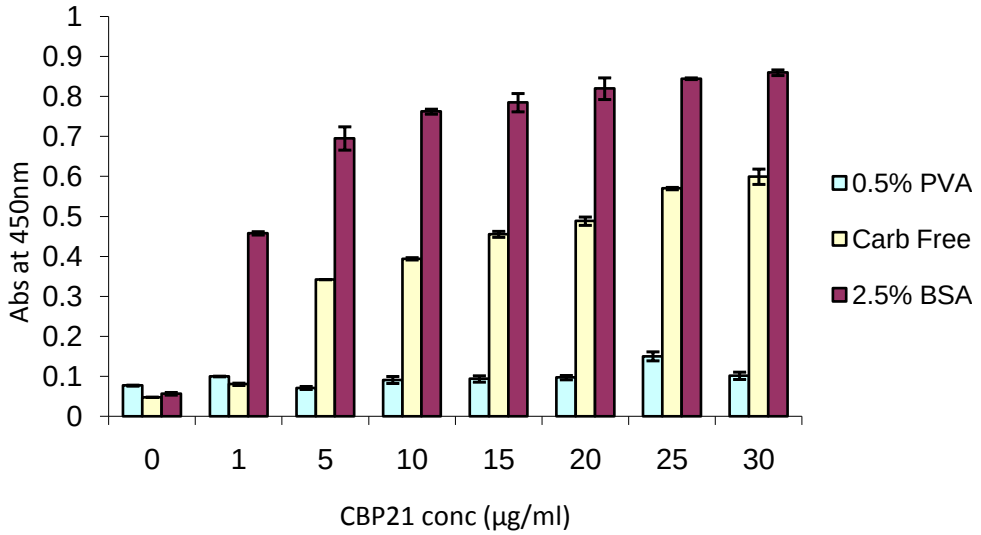


Fig 4.16: Comparison of blocking reagents. Quantitative detection of immobilised CBP21 using a peroxidase linked mouse anti-His antibody in differently blocked wells of a Nunc maxiSorp plate. The blocking agents investigated were 0.5% PVA (9,000-10,000 average mol wt, Sigma-Aldrich), 2.5% BSA (Sigma-Aldrich), and Carbo-free blocking buffer (Vector labs).

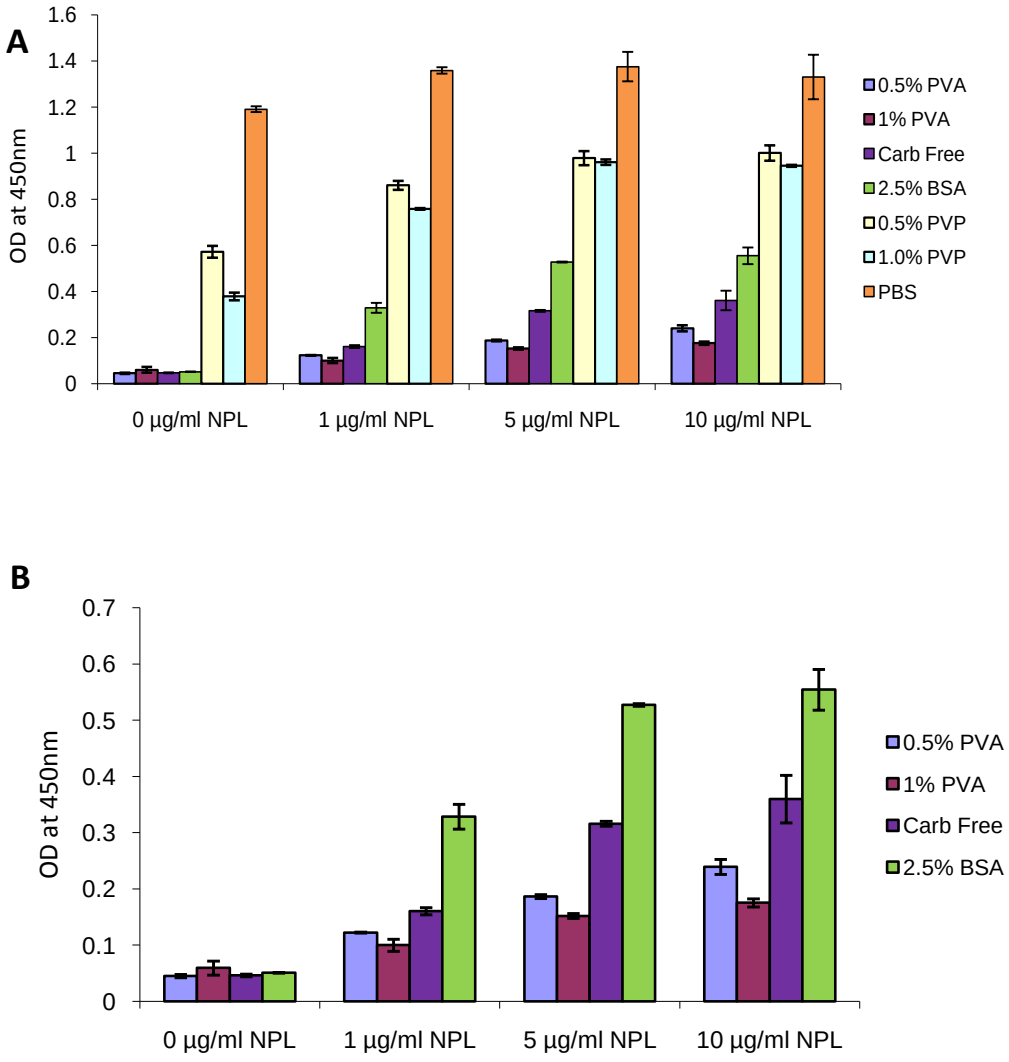


Fig 4.17: Comparison of blocking reagents using an immobilised biotin labelled lectin. A; Comparison of blocking agents PVA, PVP, BSA, Carbo-free and PBS, B; Highlighted comparison of PVA, BSA and Carbo-free blocking solutions. Quantitative detection of NPL using a HRP linked mouse anti-Biotin antibody. The blocking agents investigated were 0.5% PVA (30,000-70,000 average mol wt, Sigma-Aldrich), 1% PVA (30,000-70,000 average mol wt, Sigma-Aldrich), 0.5% PVP (360,000 average mol wt, Sigma-Aldrich), 1% PVP (360,000 average mol wt, Sigma-Aldrich), Carbo-free blocking buffer (Vector labs), 2.5% BSA (Sigma-Aldrich) and PBS.

4.2.4 CBP21 characterisation using the $(GlcNAc)_N$-binding assay

Following the selection of a blocking solution, assays of format A and B (see Fig 4.13) were repeated to determine if CBP21 could bind to chitotriose and chitobiose. Assay format A, as outlined in Fig 4.13a, was used to determine if immobilised CBP21 could bind chitotriose and chitobiose PAA-linked oligos (see Fig 4.18), while assay format B, (see Fig 4.13b) was used to determine if CBP21 could bind to immobilised chitotriose and chitobiose (see Fig 4.19).

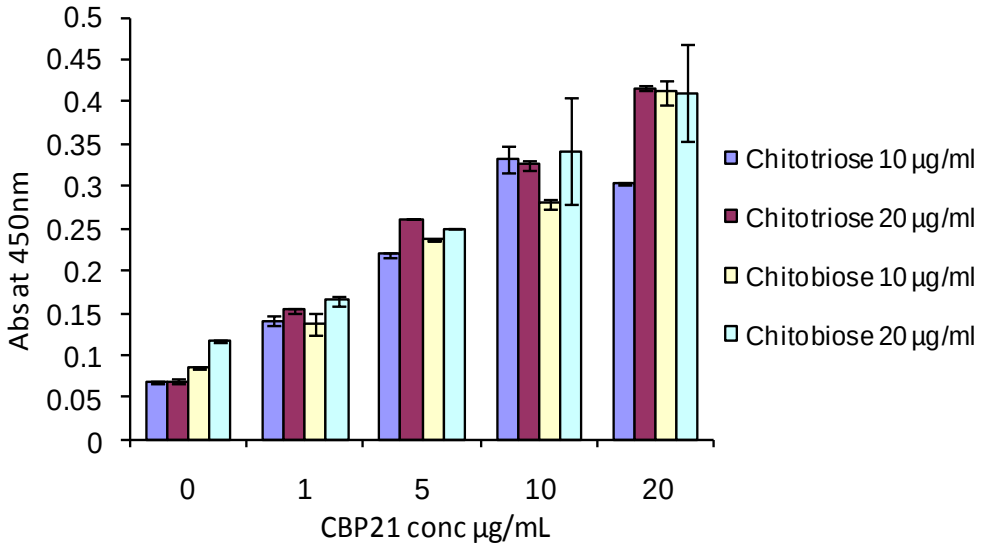


Fig 4.18: Quantitative detection of biotinylated PAA-linked $(GlcNAc)_N$ oligos bound to immobilised CBP21, using assay format A. Detection at OD_{450nm} of chitobiose and chitotriose PAA-linked olgios bound to immobilised C-terminally $(His)_6$ tagged CBP21, assayed by the activity of a HRP linked antibody. A gradient concentration of CBP21 (1–20 µg/mL) was immobilised on a NUNC maxiSorp plate overnight at 4°C. The plate was blocked with carbo-free blocking solution for one hour at 25°C, and probed with 50 µL 10 µg/mL and 20 µg/L chitobiose and chitotriose for one hour at 25°C. The biotinylated polymers were detected by probing with a 1/10,000 dilution of anti-Biotin antibody in PBST, with development using TMB solution. The absorbance was read at OD_{450nm}. Assay development time was 20 minutes.

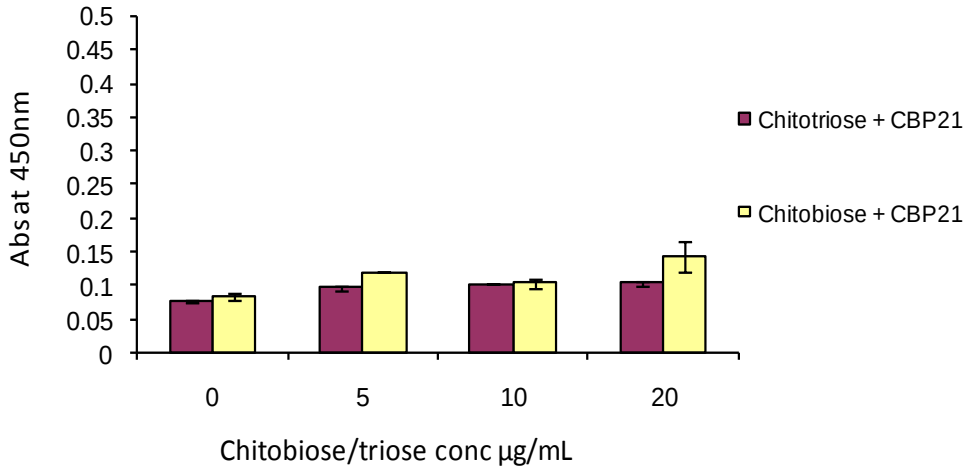


Fig 4.19: Detection of binding of $(His)_6$ tagged CBP21 to immobilised PAA-linked $(GlcNAc)_N$ oligos, using assay format B. Quantitative detection of CBP21 bound to immobilised PAA-linked Biotinlyated chitobiose and chitotriose oligos, assayed by the activity of HRP linked anti-His antibody (OD_{450mn}). A gradient concentration of PAA-linked chitobiose and chitotriose (5–20 µg/mL) was immobilised on a NUNC maxiSorp plate overnight at 4°C. The plate was blocked with carbo-free blocking solution for one hour at 25°C, and probed with 50 µL 20 µg/mL CBP21 for one hour at 25°C. Bound CBP21 was detected by probing with anti-His antibody, with development using TMB solution. The absorbance was read at OD_{450nm}. Assay development time was 30 minutes.

Repeat assays demonstrated that CBP21 could bind both chitobiose and chitotriose PAA-linked biotinylated polymers, in a concentration dependent manner, using assay format A. Optimum assay development time was 20 minutes, after which time high background signals developed. The biotinylated polymers did not appear to interact with the carbo-free blocking solution. It is unclear however if any signal quenching occurred. It does not appear that CBP21 is able to bind to the $(GlcNAc)_N$ oligos in assay format B. While assay development times of up to 30 minutes are acceptable, after this time false positives may result. There exist unsubstantial differences in signal at OD_{450nm} between different concentrations of PAA-linked oligos.

4.2.5 Validation of the chitin assay using wheat germ agglutinin

Wheat germ agglutinin (WGA) is a well characterised plant lectin. Like most lectins it is a multivalent molecule consisting of two identical subunits, with an overall molecular weight of 36 kDa, the crystal structure of which is shown in Fig 1.17. The receptor sugar for WGA is GlcNAc, with preferential binding to dimers and trimers of this sugar. WGA is commercially available in both biotinylated and un-conjugated forms, making it an ideal molecule for assay validation.

Un-conjugated WGA was used to validate format A of the CBP21 – $(GlcNAc)_N$ binding assay (see Fig 4.20). The assay was carried out as described in Fig 4.13a. It was observed that WGA bound the PAA-linked chitobiose and chitotriose polymers with greater affinity than CBP21. The development time for this assay was 3 minutes, compared to 20 minutes for CBP21, and had higher OD_{450nm} readings. It is evident from the results that WGA had a slightly higher affinity for chitobiose over chitotriose, this was also observed in the CBP21 assays. In terms of assay validation, the use of WGA has proven that this assay format is sufficient to profile chitin-binding proteins.

In order to validate format B an un-conjugated PAA-linked chitobiose was purchased from lectinity. The assay was carried out as described in Fig 4.13b, with immobilisation of the unlabelled chitobiose, and probing with biotinylated WGA. Although CBP21 was unable to bind to the PAA-linked chitobiose in this format, it was observed that WGA had a high affinity for the polymer (see Fig 4.21). The time required for adequate development of a detection signal in this assay was the same as in assay format A (3 minutes).

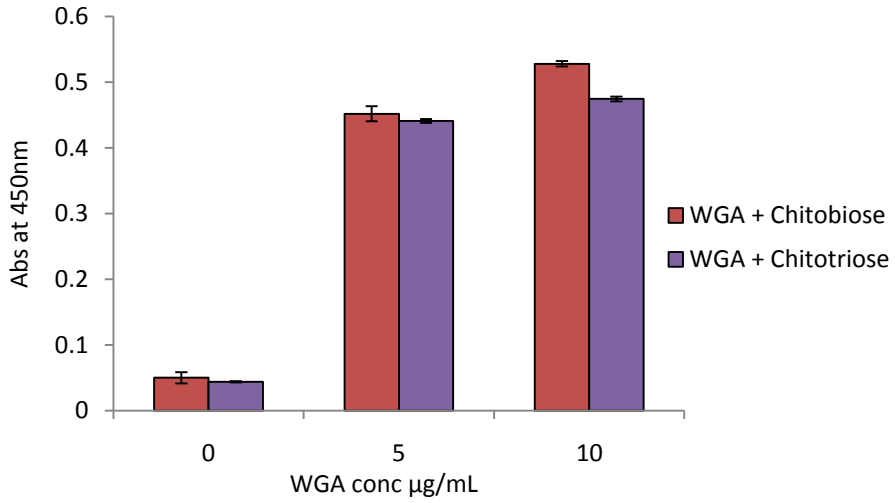


Fig 4.20: Binding of immobilised unlabelled wheat germ agglutinin to PAA-linked Biotin labelled chitobiose and chitotriose PAA-linked oligos. Quantitative detection of chitobiose and chitotriose bound to immobilised wheat germ agglutinin (WGA), assayed by the activity of an anti-Biotin HRP linked antibody (OD_{450nm}). Assay development time was 3 minutes.

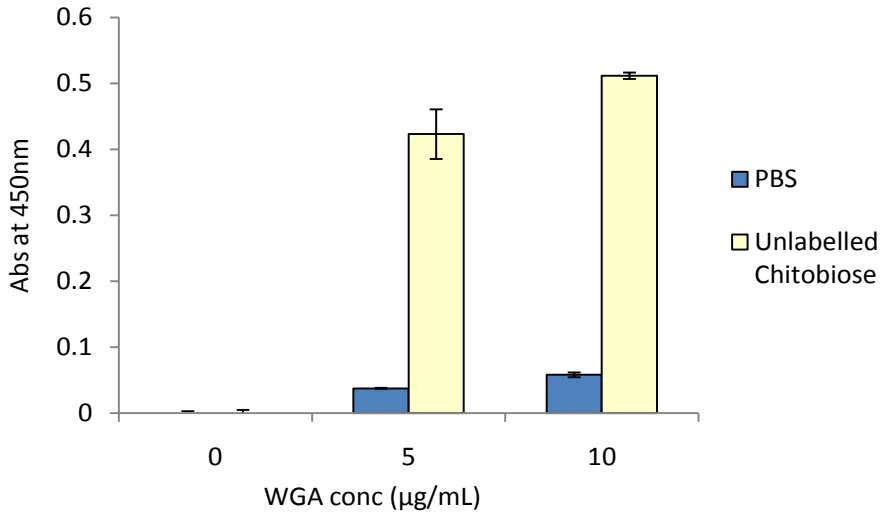


Fig 4.21: Binding of biotinylated wheat germ agglutinin to immobilised PAA-linked un-conjugated chitobiose. Quantitative detection of biotinylated WGA bound to immobilised un-conjugated PAA-linked chitobiose, assayed by the activity of an anti-Biotin HRP linked antibody (OD_{450nm}). Assay development time was 3 minutes.

Table 4.2: Final procedure for PAA-linked $(GlcNAc)_N$ assay

Step	Procedure
1	Immobilise 50 µL CBP21 in PBS on a maxiSorp plate, incubate overnight at 4°C
2	Block wells with 200 µL of Carbo-free blocker diluted in PBS, incubate for 1 hour at 25°C.
3	Probe wells with 50 µL $(GlcNAc)_N$ in PBS for one hour at 25°C
4	Wash wells with 4 x 200 µL of PBST
5	Probe with 1/10,000 HRP linked anti-Biotin antibody diluted in PBST, for one hour at 25°C
6	Wash wells with 4 x 200 µL of PBST
7	Develop wells by the addition of 100 µL of TMB solution
8	Stop reaction after 20 minutes by the addition of 50 µL 10% H_2SO_4
9	Read absorbance at 450nm

4.3 CBP21 affinity chromatography

It has been shown, through ELLA and glycoprotein affinity chromatography that CBP21 is incapable of binding to terminal GlcNAc and core chitobiose of glycans attached to glycoproteins. Through the development of a $(GlcNAc)_N$ assay it is known that CBP21 can bind the GlcNAc polymers chitobiose and chitotriose. It therefore may be possible that the secondary and tertiary structures of the glycoproteins may have prevented CBP21 from reaching the chitobiose core. CBP21 would be an extremely advantageous tool in the field of glycobiology if it could bind to the chitobiose core structure of protein released glycans.

A pool of released glycans could be generated using PNGase F. PNGase F is an endoglycosidase that cleaves N-linked glycans between the innermost GlcNAc and asparagines residue of high mannose, hybrid and complex oligosaccharides. It deaminates the asparagines to aspartic acid, but leaves the oligosaccharide intact. Released oligosaccharides could be separated from the protein mixture through use of C-18 reverse phase chromatography (see Fig 4.22).

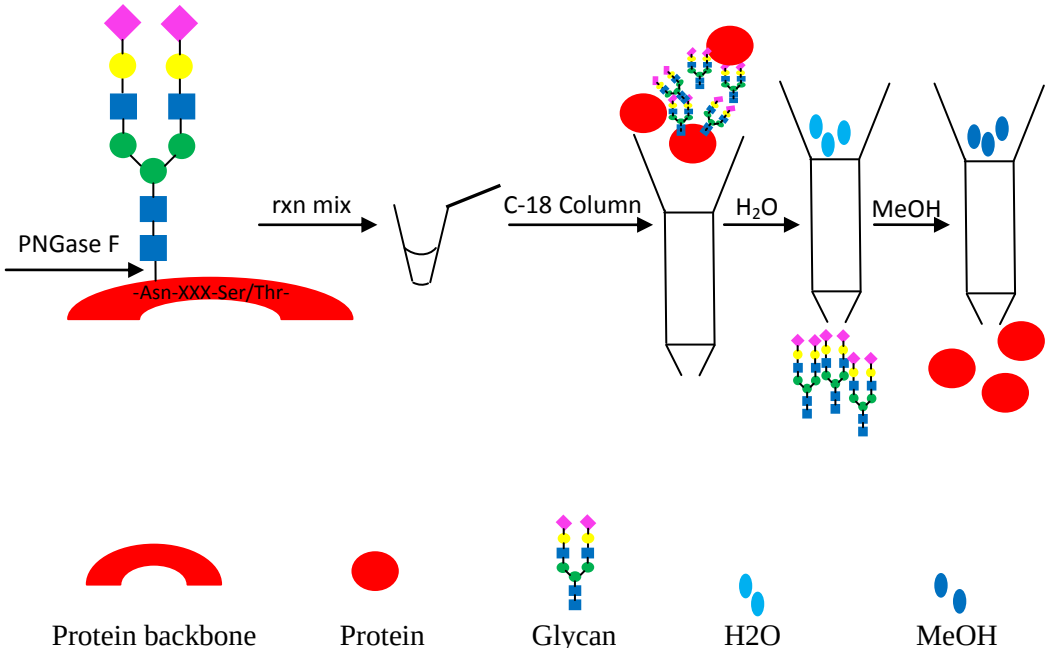


Fig 4.22: Release and isolation of PNGnase F liberated glycans using a C-18 column. PNGnase F acts to release *N*-linked oligosaccharide chains between the inner most GlcNAc and asparagine residue. Released glycans can be separated from the protein mix through using C-18 reverse phase column chromatography. The polar residues elute first and the non-polar residues eluting last.

As the protein released glycans could not be immobilised on a maxiSorp plate, and would not be conjugated to a His or Biotin label, an assay of the ELISA format, outlined in Fig 4.13, would be unsuitable. The most appropriate format would follow that of the glycoprotein affinity chromatography as discussed previously (Section 4.1.3). CBP21 could be immobilised on to cyanogen bromide activated sepharose, with addition of the protein released glycans. An assay for reducing sugars could be used to evaluate the success of the capture process.

CBP21 was immobilised onto cyanogen bromide activated sepharose 4B as outlined in section 2.29 (Table 4.3). The potential of the assay was evaluated through the use of non PAA-linked, un-conjugated chitobiose (Megazyme). The limit of detection of chitobiose was determined using the DNS assay (see Fig 4.24) as outlined in section 2.21. It was observed that chitobiose concentrations as low as 0.01 mg/mL could be detected using this assay, with linear detection between concentrations of 0.2-1 mg/mL.

300 µL chiotobiose (0.25 mg/mL) was passed over the CBP21 affinity resin with collection of 1 mL fractions. The resin was pre-equilibrated with PBS. 100 µL of 0.25 mg/mL chitobiose was mixed with the resin overnight at room temperature. Unbound sample was collected, and the resin was washed four times with 1 mL of PBS (fractions 3-7). The column was subsequently washed 5 times with 1 mL of sodium acetate, pH 4.0 (fractions 8-12). Elution fractions were assayed using the DNS assay as outlined in section 2.21 (see Fig 4.25).

The total amount of unbound chitobiose from the CBP21 affinity column was estimated as a percentage of the total unbound chitobiose from the blank column. It was estimated that an average of 87.2% of the chitobiose eluted from the CBP21 affinity column in the elution samples. Subtracting this value from 100% gave an average of 12.8% of chitobiose bound to the column. As 0.075 mg of chiotobiose was applied to the column this 12.8% equates to a total of 0.01 mg of bound chitobiose. Sodium acetate buffer, pH 4.0, was used to wash the resin in an attempt to elute any bound chitobiose. No chitobiose was detected in the eluted fractions. A sample of each resin was also mixed with DNS reagent and boiled in an attempt to discern further difference between CBP21-affinity and control resins (see Fig 4.26).

Subsequently PNGase F from *Flavobacterium meningosepticum* was used to digest bovine asialo fetuin. The released glycans were separated from the protein mix using a C-18 column (Alltech). The amount of released glycan in the purified sample was below the limit of detection of the DNS assay. A modified ELLA was also used in an attempt to detect any of the protein released glycans. CBP21 and un-conjugated WGA were immobilised on a maxiSorp plate, the plate was blocked and probed with the protein released glycans. The glycans were then probed with biotinylated ECL. No differences were distinguishable between glycan deplete and replete samples (see Fig 4.27).

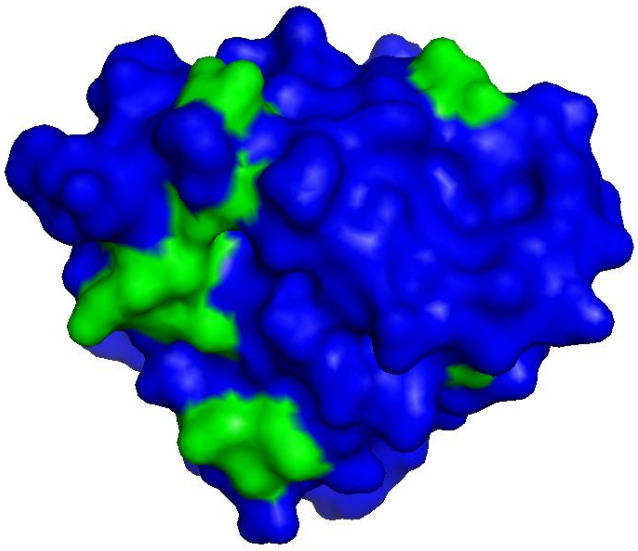


Fig 4.23: Location of surface exposed lysine residues on the CBP21 molecule. Representation of the location of the amino acid lysine on the surface of CBP21 molecule. Proteins are immobilised onto cyanogen bromide active sepharose through free amine groups. Lysine residues are highlighted in green. Image generated using PyMol.

Table 4.3: Immobilisation of CBP21 onto cyanogen bromide activated sepharose 4B.

Starting CBP21	**Resin volume**	**Unbound CBP21**	**Bound CBP21**	**mg CBP21/ mL resin**
8.9 mg	1 mL	0.1 mg	8.8 mg	8.8

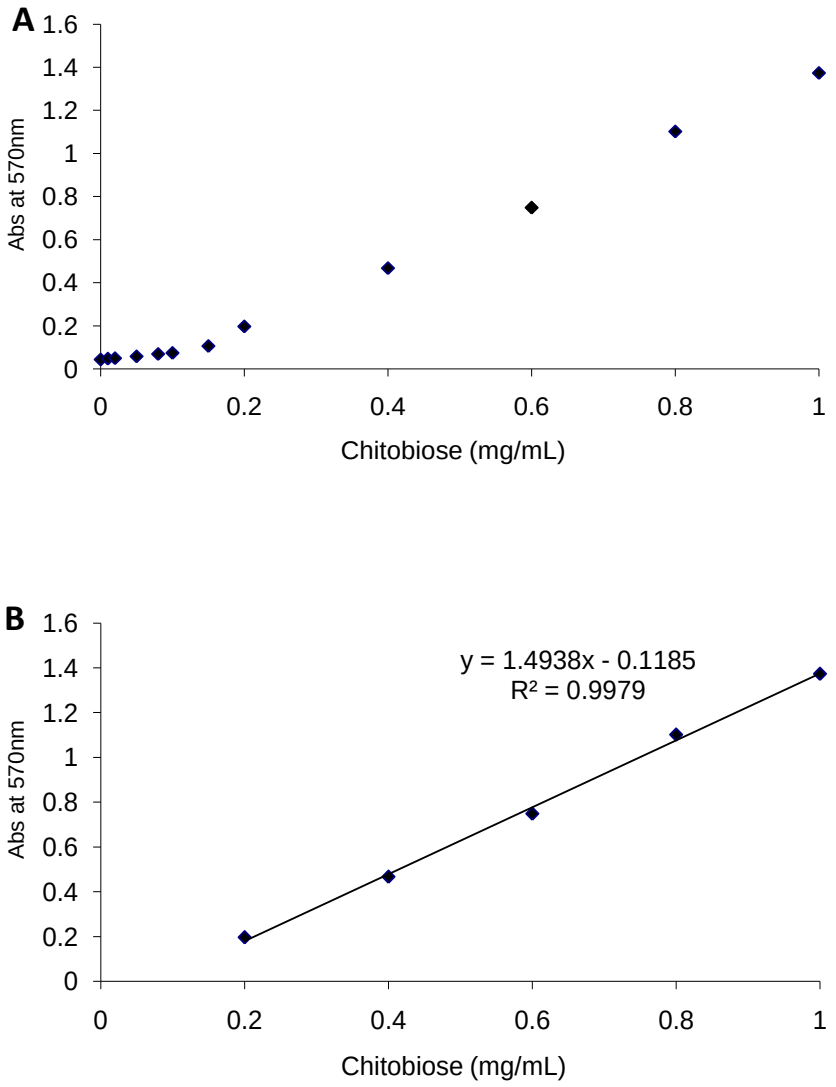


Fig 4.24: Determining the limit of detection of un-conjugated chitobiose using the DNS assay for reducing sugars, A; Detection of chitobiose between 0.01 – 1 mg/mL, B; Linear regression analysis of chitobiose between 0.2 – 1 mg/mL. The DNS assay was used to determine the limit of detection for un-conjugated chitobiose (Megazyme). The absorbance readings at OD_{570nm} were plotted against the chitobiose concentration (mg/mL).

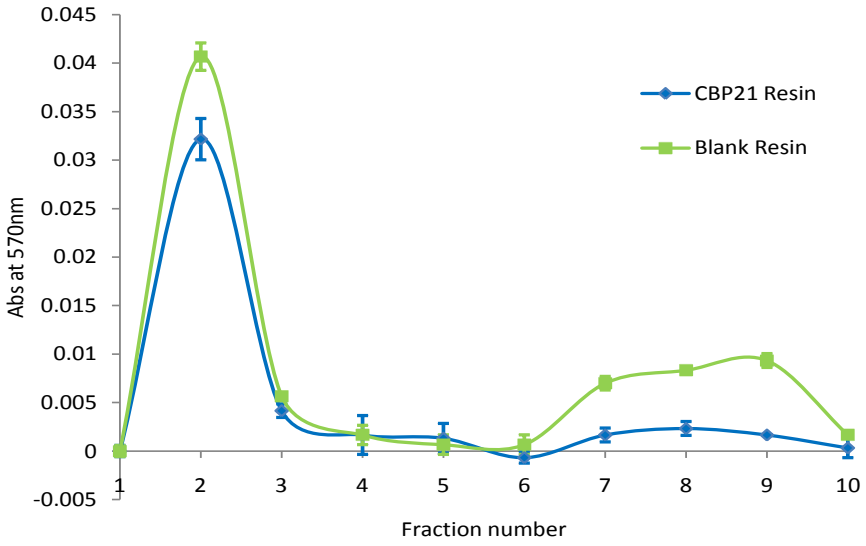


Fig 4.25: Analysis of chitobiose bound to sepharose immobilised CBP21. Quantitative analysis, by DNS assay of un-conjugated chitobiose (megazyme) bound to sepharose 4B immobilised CBP21. The resin was pre-equilibrated with PBS (fraction 1). 100 µL of 0.25 mg/mL chitobiose was mixed with the resin overnight at room temperature. Unbound sample was collected (fraction 2), and the resin was washed four times with 1 mL PBS (fractions 3-6). The column was subsequently washed four times with 1 mL sodium acetate, pH 4.0 (fractions 7-10). Eluted fractions were assayed using the DNS assay as outlined in section 2.21.

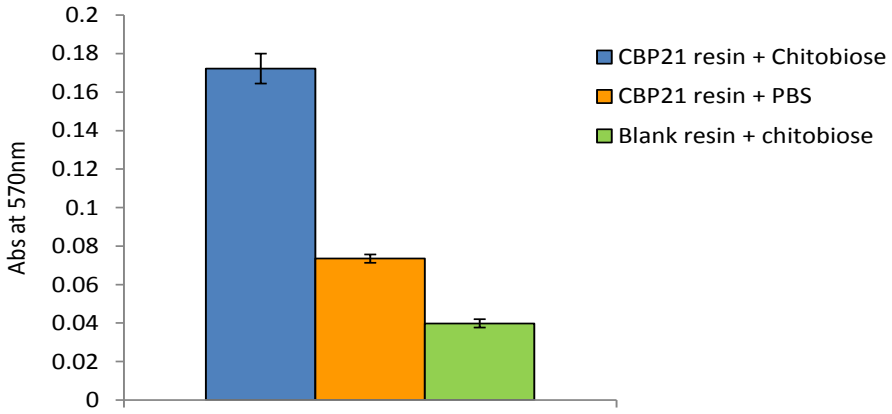


Fig 4.26: Analysis of CBP21 affinity column bound chitobiose by boiling and DNS assay. 300 µL of sample resin was added to 300 µL of PBS and mixed with 600 µL of DNS reagent. Samples were boiled for 10 mins, allowed to cool for 5 min, and the absorbance was read at OD_{570nm}.

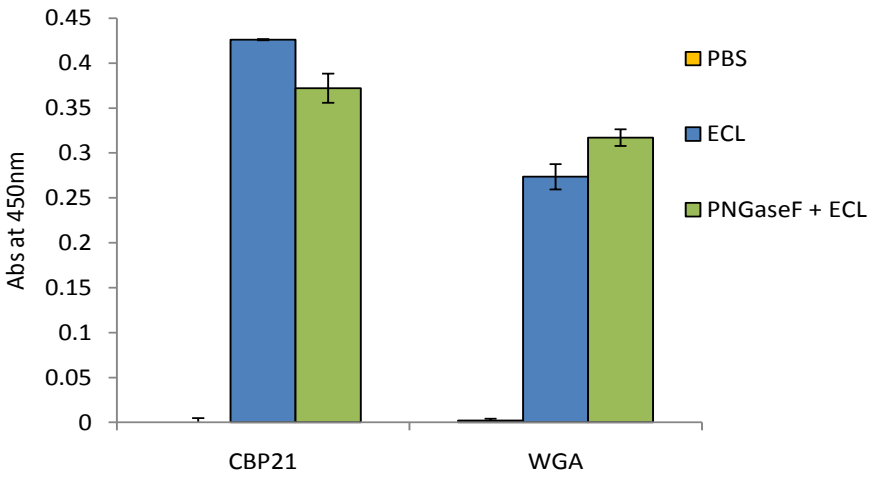


Fig 4.27: Analysis of PNGase F released glycans by modified ELLA. Quantitative detection of PNGase F released glycans bound to carbohydrate-binding molecules CBP21 and WGA by biotinylated ECL. ECL is a plant lectin known to have an affinity for terminal galactose.

4.4 Sugar Inhibition studies

Although CBP21 has not shown an affinity for any of the glycans tested (Section 4.1), it has shown an ability to bind to both un-conjugated chitobiose and PAA-linked chitobiose and chitotriose (Sections 4.2 and 4.3). In its natural environment CBP21 binds to long chain chitin polymers, and has been shown to bind to both chitosan and cellulose linked polymers (Section 3.8). CBP21 therefore while not able to bind single sugar units may have the ability to bind to other polysaccharides.

Sugar inhibition studies were carried out using the labelled $(GlcNAc)_N$ assay as developed in section 4.2 (see Fig 4.28), and the data validated using insoluble substrate assays coupled with a Bradford assay (see Fig 4.29). Sugar concentrations used were glucose – 3.4%, galactose – 3.4%, galactan – 6%, mannose – 3.4%, mannan from *Saccharomyces cerevisiae* – 6%, GlcNAc – 4.4%, sucrose – 6.8%, lactose – 6.8%, fucose – 3.2% and chitobiose – 2%, unless otherwise stated.

Sugar inhibition studies using the labelled chitobiose assay (see Fig 4.28) suggested that interaction between CBP21 and chitobiose was not affected by the presence of glucose, galactose, GlcNAc, lactose or the un-conjugated chitobiose, while galactan, mannose, sucrose and fucose appeared to slightly alter the affinity of CBP21 for the PAA-linked chitobiose. Mannan showed the highest level of disruption to chitobiose binding compared to the control sample. Both α- and β-chitin sugar inhibition studies (see Fig 4.29) revealed conflicting results. These assays indicated that CBP21 was least likely to bind to chitin in the presence of galactan, and that mannan had the next highest inhibitory effect on α–chitin binding (see Fig 4.29a), but showed little effect in the study using β–chitin from squid pen (see Fig 4.29b).

A saline solution of sepharose linked mannan, from *Saccharomyces cerevisiae*, was purchased from Sigma and used as an affinity matrix in an attempt to capture CBP21 (see Fig 4.30). Three 200 µL volumes of mannan agarose was aliquoted into three separate columns and equilibrated with binding buffer. A 200 µL sample of blank cyanogen bromide activated sepharose was added to a fourth column. The mannan agarose was mixed with A; 1 mL NPL (50 µg/mL), B; 1 mL CBP21 (50 µg/mL) and C; 1 mL PBS. The blank agarose was also mixed with 1 mL CBP21 (50 µg/mL), overnight at room temperature. Eluted sample was collected and the column was washed with 5 x 1 mL of binding buffer, prior to washing with 5 x 1 mL of 0.2 M mannose. The absorbance of each eluted sample was read at OD_{280nm} and was plotted against each fraction number. It was evident from the mannan affinity chromatography that CBP21 did not bind to the mannan agarose.

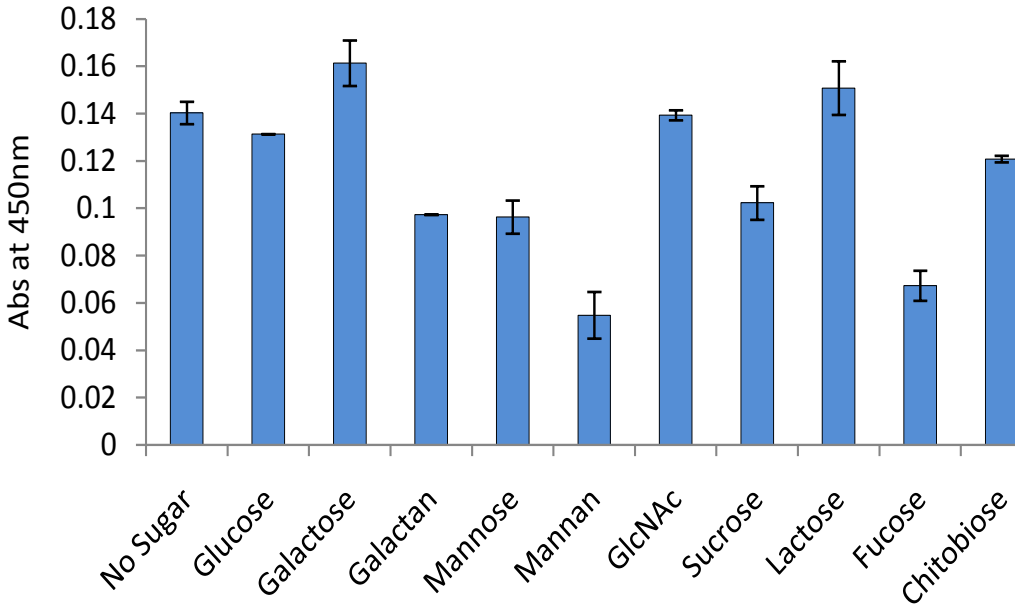


Fig 4.28: Sugar inhibition study of CBP21 using biotin labelled PAA-linked chitobiose. Quantatitive detection of biotin labelled PAA-linked chitobiose bound to sugar inhibited CBP21. Immobilised CBP21 was incubated with 50 µL of each sugar prior to incubation with 50 µL 10 µg/mL PAA-linked chitobiose prepared in sugar solution.

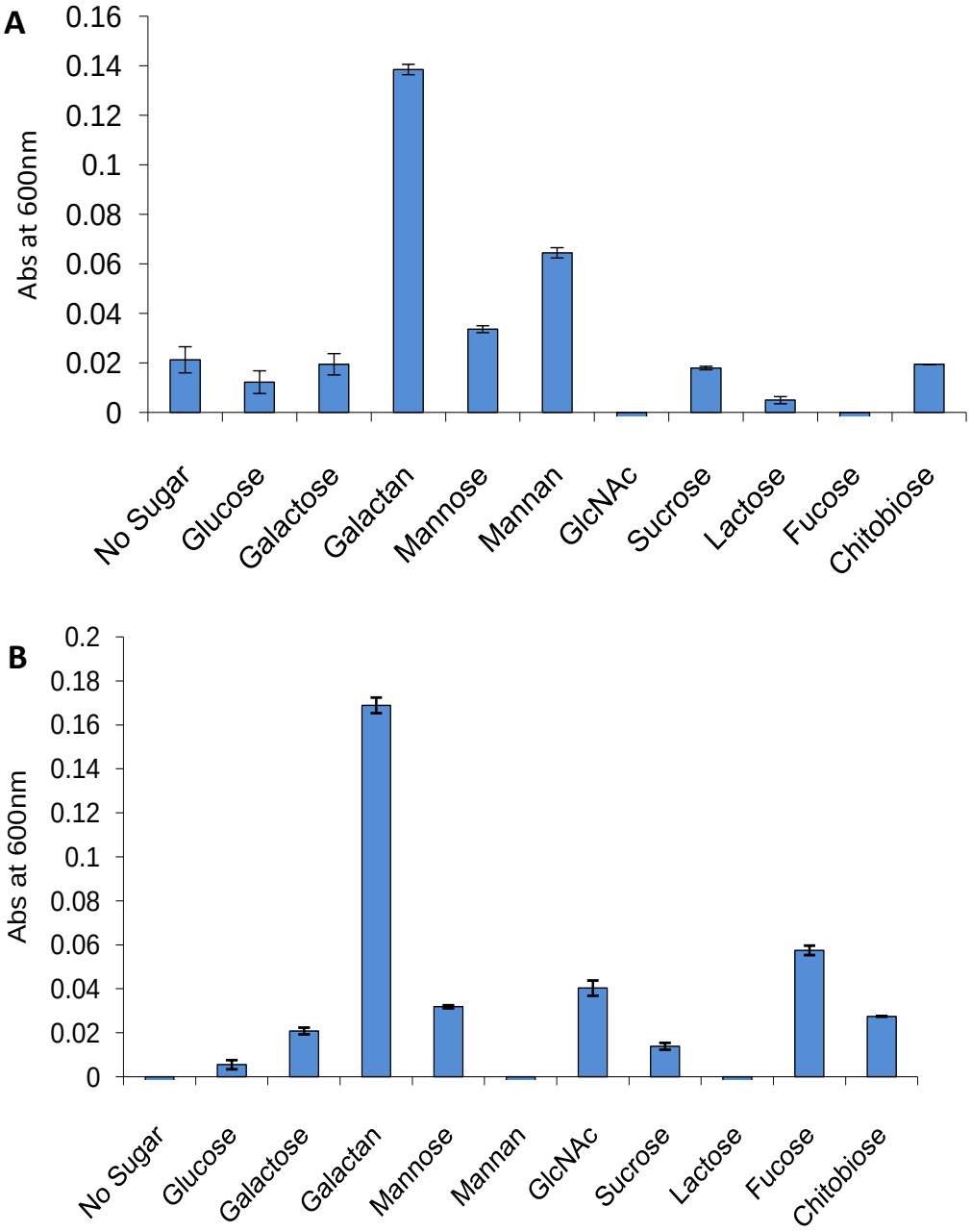


Fig 4.29: Sugar inhibition study of CBP21 using A; α-chitin and B; β-chitin. Quantitative detection of unbound CBP21 following sugar inhibition with various sugars. 1 mL of CBP21 (30 µg/mL) was incubated with each sugar for one hour prior to incubation with 0.05 g of α-chitin from crustacean shells or β-chitin from squid pen, overnight at room temperature. Unbound protein was detected using the Bradford Ultra assay (Section 2.20.1).

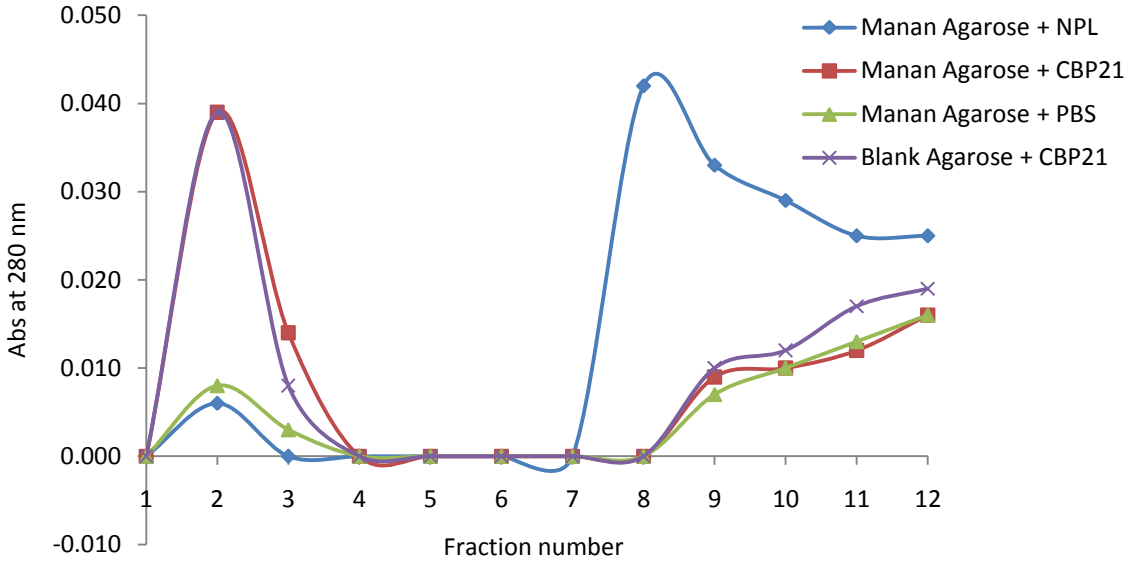


Fig 4.30: Application of CBP21 to mannan affinity agarose. OD_{280nm} readings taken during mannose affinity chromatography using CBP21. The resin was pre-equilibrated with binding buffer until OD_{280nm} reached 0.000 (fraction 1). 1 ml of 50 µg/mL carbohydrate-binding proteins were added to and mixed with 200 µL of resin, overnight at room temperature. The column flow-through was collected (fraction 2), and the resin was washed with 5 x 1 mL of binding buffer (fractions 3-7), until the readings at OD_{280nm} returned to 0.000. Bound protein was eluted by the addition of 0.2 M mannose (fractions 8-12). NPL is plant lectin with a known affinity for mannose.

4.5 CBP21 Mutagenesis

Protein–protein interactions are central to most biological processes, and detection of specific amino acid residues that contribute to the specificity and strength of protein interactions is a problem of the utmost importance (Moreira, *et al.*, 2007). Crystallographic studies of protein structures, especially those co-crystallised with various ligands can help to identify the specific residues in contact with a ligand. The crystal structure of CBP21 was eludicated by Vaaje-Kolstad *et al* in 2005. Although attempts were made to co-crystallise the protein with GlcNAc, they proved fruitless. Since the publication of the crystal structure Vaaje-Kolstad *et al.*, have attempted to pin point the active site residues through mutagenesis (Vaaje-Kolstad *et al.*, 2005a). While

they succeeded in decreasing the chitin-binding activity of a number of mutants (Table 1.4), complete disruption of the active site has not been achieved. Identification of the most influential residues in the active site will highlight good targets for site direct random mutagenesis, which has the potential to alter the sugar specificity of this protein (McMahon, 2010). The CBP21 residues thought to be involved in chitin binding are shown in Fig 4.31.

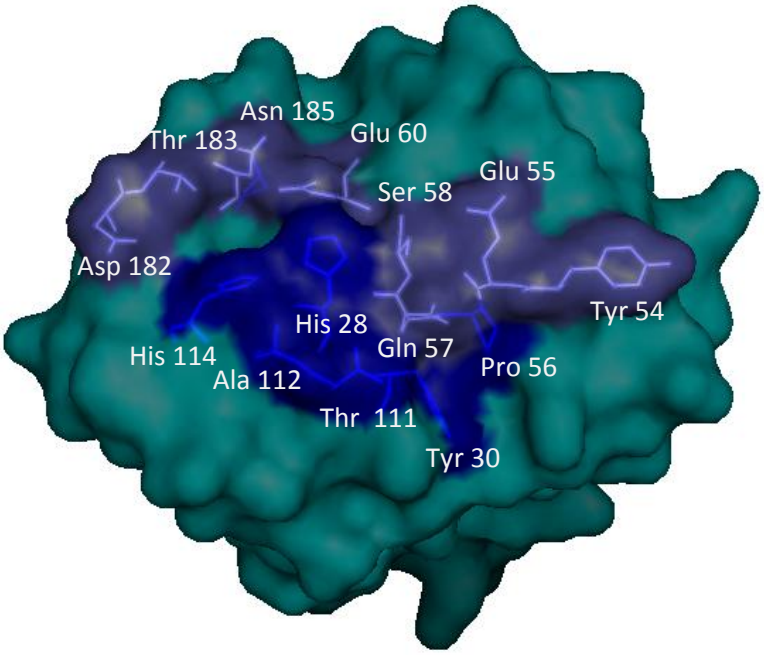


Fig 4.31: Structure of CBP21 with residues thought to be involved in chitin binding. The putative active site residues are highlighted in blue. Image generated using PyMol.

4.5.1 Site directed mutagenesis of CBP21

Alanine scanning is the most commonly used technique to map functional epitopes (Moreira *et al.*, 2007, Cunningham and Wells, 1989, Morrison and Weiss, 2001, Wells, 1990). Alanine substitution nullifies the amino acid side chain without altering the main-chain conformation, nor does it impost extreme electrostatic or steric effects (Cunningham and Wells, 1989). Amino acid replacement with proline or glycine would have similar effects, but could also introduce additional conformational freedom into the protein backbone. Alanine residues were introduced to CBP21 via site directed mutagenesis, using Phusion™ directed whole vector amplification (see Fig 4.32). Residues Y54, E55, P56, Q57, S58, E60, T111, H114 and D182 were targeted for mutagenesis. The resulting genes were named *cbp21*$_{Y54A}$, *cbp21*$_{E55A}$, *cbp21*$_{P56A}$, *cbp21*$_{Q57A}$, *cbp21*$_{S58A}$, *cbp21*$_{E60A}$, *cbp21*$_{T111A}$, *cbp21*$_{H114A}$ *and* *cbp21*$_{D182A}$ and the proteins named, CBP21$_{Y54A,}$ CBP21$_{E55A,}$ CBP21$_{P56A,}$ CBP21$_{Q57A,}$ CBP21$_{S58A,}$ CBP21$_{E60A}$, CBP21$_{T111A}$, CBP21$_{H114A}$ and CBP21$_{D182A}$ respectively. The construction of the CBP21 mutants is outlined in Fig 4.33 – 4.41. The predicted impact to the protein structure of the mutant proteins is show in Fig 4.42.

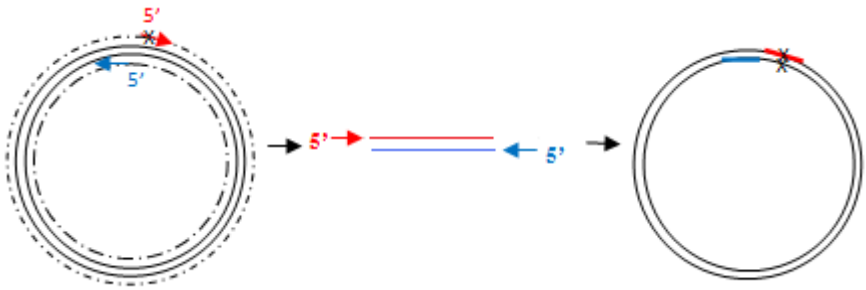


Fig 4.32: Schematic of Phusion™ site directed mutagenesis. Two primers are designed to amplify the entire expression plasmid, one of which contains the desired mutation. Following PCR (Section 2.10.1) a DpnI restriction digest can be carried out to restrict any of the methylated template DNA. The phosphorylated ends are then ligated. The ligation is subsequently transformed directly into an *E. coli* expression strain (Section 2.9.3). For the mutation of cbp21 the template plasmid was pCBP21_60.

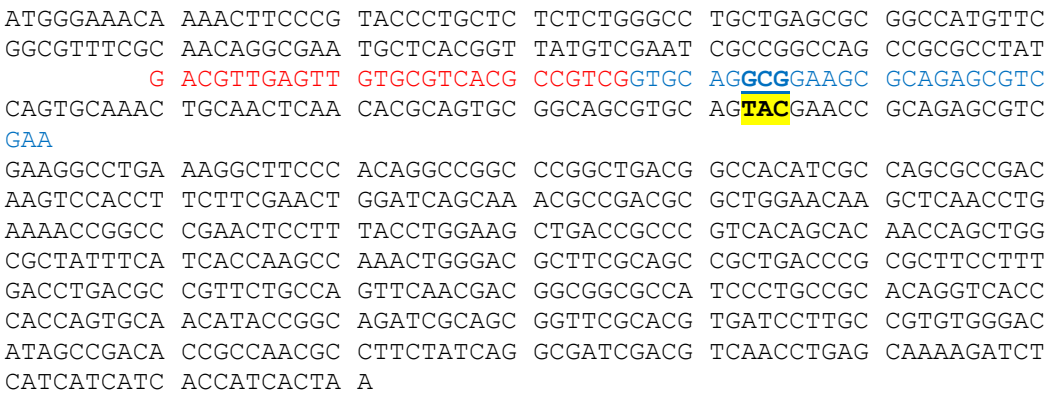


Fig 4.33: Outline of the construction of CBP21$_{Y54A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{Y54A}$, Y54A_F (blue), 55-60A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation Y54A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

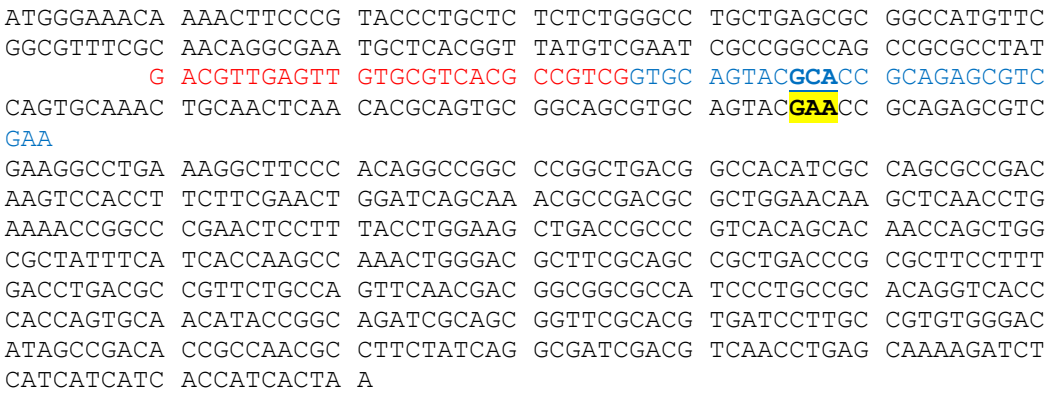


Fig 4.34: Outline of the construction of CBP21$_{E55A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{E55A}$, E55A_F (blue), 55-60A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation E55A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
G ACGTTGAGTT GTGCGTCACG CCGTCGGTGC AGTACGAA**GC** **G**CAGAGCGTC
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAA**CC** **G**CAGAGCGTC
GAA
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.35: Outline of the construction of CBP21$_{P56A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{P56A}$, P56A_F (blue), 55-60A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation P56A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on the forward primer in underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
G ACGTTGAGTT GTGCGTCACG CCGTCGGTGC AGTACGAACC G**GCG**AGCGTC
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC G**CAG**AGCGTC
GAA
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.36: Outline of the construction of CBP21$_{Q57A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{Q57A}$, Q57A_F (blue), 55-60A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation Q57A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
G ACGTTGAGTT GTGCGTCACG CCGTCGGTGC AGTACGAACC GCAGGCGGTC
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GAA
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.37: Outline of the construction of CBP21$_{S58A}$. Representation of the position of the primers used for the mutation of *cbp21$_{S58A}$*, S58A_F (blue), 55-60A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation S58A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
G ACGTTGAGTT GTGCGTCACG CCGTCGGTGC AGTACGAACC GGCGAGCGTC
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GCAGGCCTG
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.38: Outline of the construction of CBP21$_{E60A}$. Representation of the position of the primers used for the mutation of *cbp21$_{E60A}$*, E60A_F (blue), 55-60A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation E60A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
TT CGAGTTGGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
TTTTGGCCGG GCTTGAGGTT TACCTGGAAG CTGGCCGCCC GTCAC
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.39: Outline of the construction of CBP21$_{T111A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{T111A}$, T111A_F (blue), T111A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation T111A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
TT CGAGTTGGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
TTTTGGCCGG GCTTGAGGTT TACCTGGAAG CTGACCGCCC GTGCCAGCAC AACCAGCTGG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGC
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.40: Outline of the construction of CBP21$_{H114A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{H114A}$, H114A_F (blue), T111A_R (red) (Table 2.3) within the cbp21 sequence which creates the mutation H114A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CTAGCGTCG CCAAGCGTGC ACTAGGAACG GGTGTGGGAC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGCCA CCGCCAACGC CTTCTAT
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.41: Outline of the construction of CBP21$_{D182A}$. Representation of the position of the primers used for the mutation of *cbp21*$_{D182A}$, D 182A_F (blue), D182A R (red) (Table 2.3) within the cbp21 sequence which creates the mutation D182A. The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

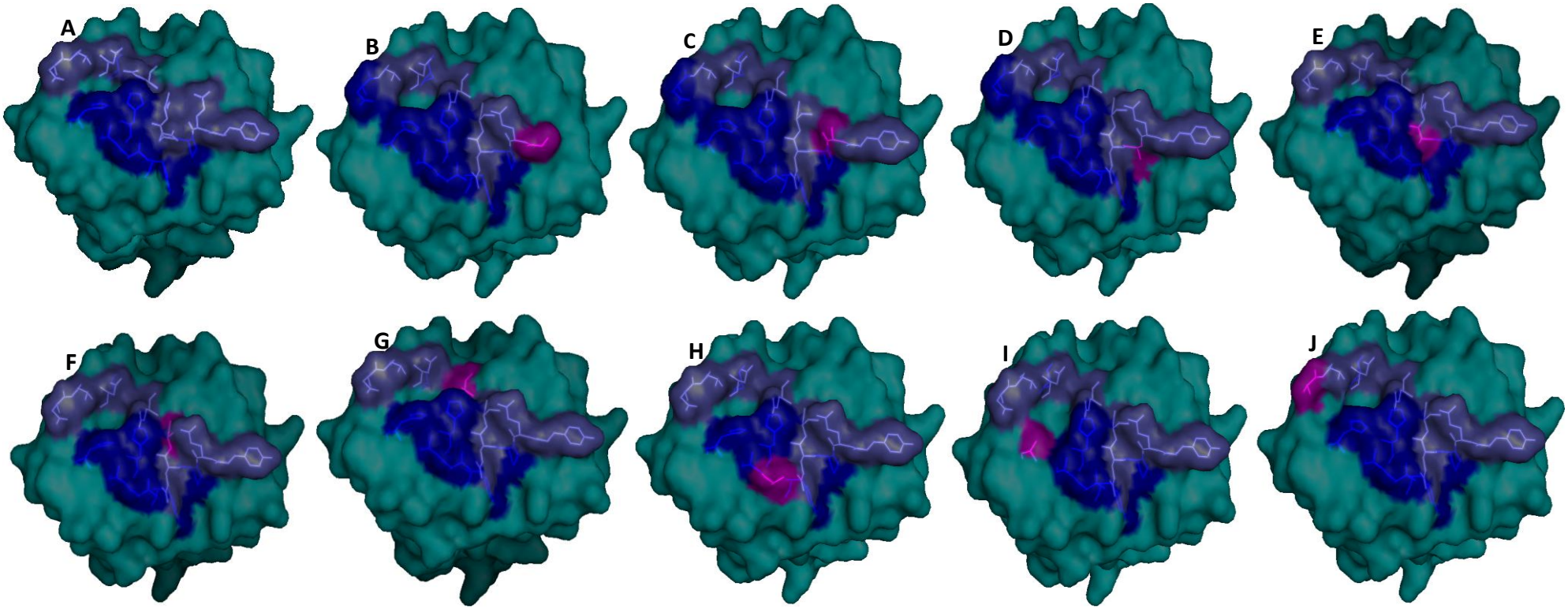


Fig 4.42: Structure of CBP21, A; $CBP21_{WT}$, B; $CBP21_{Y54A}$, C; $CBP21_{E55A}$, D; $CBP21_{P56A}$, E; $CBP21_{Q57A}$, F; $CBP21_{S58A}$, G; $CBP21_{E60A}$, H; $CBP21_{T111A}$, I; $CBP21_{H114A}$ and J; $CBP21_{D182A}$. Putative carbohydrate binding site residues are highlighted in blue, mutated residues are highlighted in purple. Image generated using PyMol.

4.5.2 Expression and purification of CBP21 mutants

Following successful construction of the CBP21 mutants, each plasmid was verified by sequencing and transformed into the *E. coli* KRX expression strain. Protein expression and purification was then undertaken. The optimised expression (Section 3.3.2) and purification (Section 3.4.2) strategies for pCBP21_60 were used for all derivatives of CBP21. It was observed that each mutant was successfully purified to a high degree of homogeneity using the FPLC and the purification strategy outlined in section 3.4.2 (see Fig 4.43 – 4.45). The average yield for each CBP21 variant per g/cell paste, following expression and purification, is given in table 4.4. The average yield is taken after 3 separate fermentations. The g/cell paste value is the weight of cells harvested from a 500ml culture. It was found that not all mutants expressed to the same level as the wild type molecule. Some variants gave rise to overall lower cell densities following overnight incubation, with much lower protein yields as a result.

Table 4.4: Average yield of recombinant CBP21 mutant variants from a 500 mL expression in LB medium.

Protein	Mass of cells (g)	mg protein/ g cell paste
$CBP21_{Y54A}$	5.96	3.13
$CBP21_{E55A}$	4.4	2.54
$CBP21_{P56A}$	2.72	0.79
$CBP21_{Q57A}$	5.18	1.68
$CBP21_{S58A}$	4.68	3.79
$CBP21_{E60A}$	3.21	2.15
$CBP21_{T111A}$	3.78	2.48
$CBP21_{H114A}$	3.92	2.37
$CBP21_{D182A}$	3.44	0.65

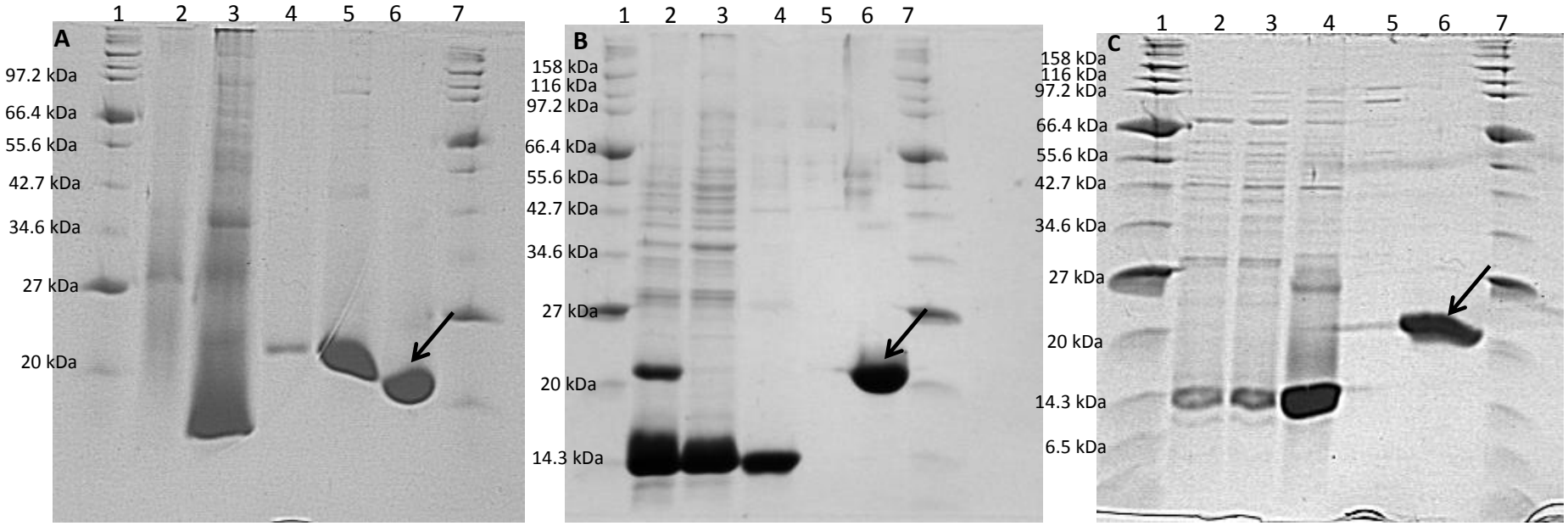


Fig 4.43: Purification of C-terminally tagged CBP21 mutants by IMAC, A; $CBP21_{Y54A}$, B; $CBP21_{E55A}$, C; $CBP21_{P56A}$. Analysis by 12.5% SDS-PAGE and silver staining of CBP21 purification by IMAC using Ni^{2+} charged sepharose. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude lysate, Lane 3; Column flow-through, Lane 4, 20mM Imidazole wash, Lane 5; 70mM Imidazole wash, Lane 6; 200mM Imidazole elution, Lane 7; Broad range protein ladder. Purified protein bands are indicated with an arrow.

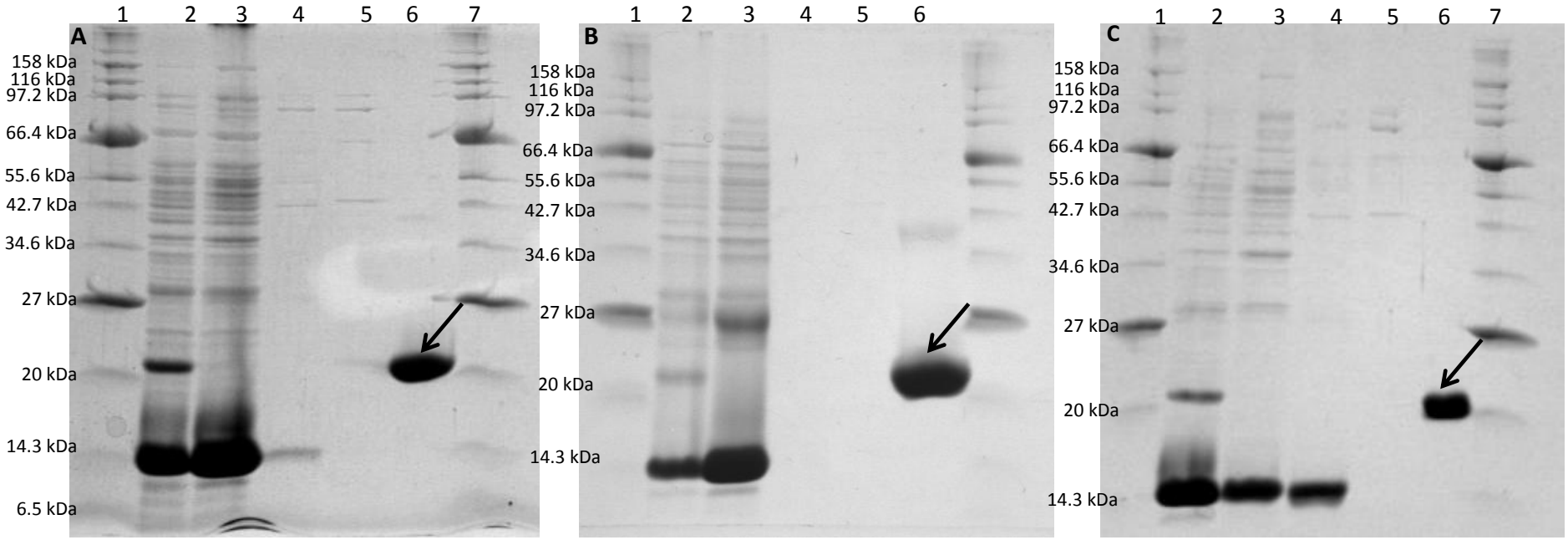


Fig 4.44: Purification of C-terminally tagged CBP21 mutants by IMAC, A; $CBP21_{Q57A}$, B; $CBP21_{S58A}$, C; $CBP21_{E60A}$. Analysis by 12.5% SDS-PAGE and silver staining of CBP21 purification by IMAC using Ni^{2+} charged sepharose. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude lysate, Lane 3; Column flow-through, Lane 4, 20mM Imidazole wash, Lane 5; 70mM Imidazole wash, Lane 6; 200mM Imidazole elution, Lane 7; Broad range protein ladder. Purified protein bands are indicated with an arrow.

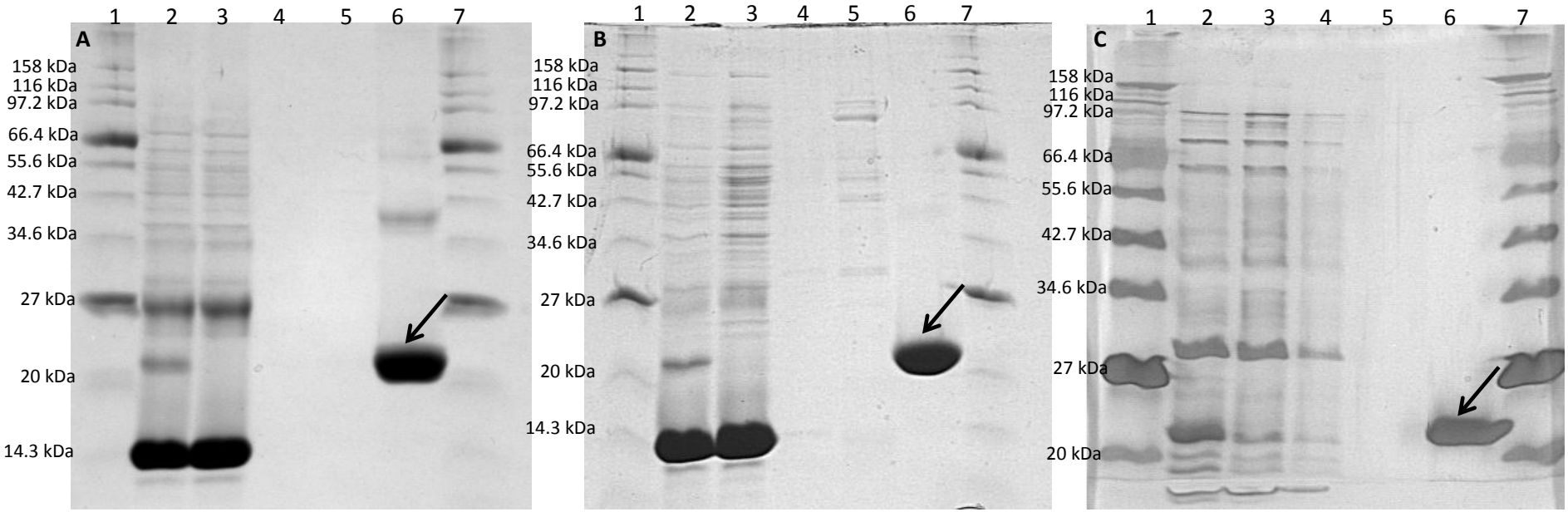


Fig 4.45: Purification of *C*-terminally tagged CBP21mutants by IMAC, **A; $CBP21_{T111A}$, B; $CBP21_{H114A}$, C; $CBP21_{D182AA}$.** Analysis by 12.5% SDS-PAGE and silver staining of CBP21 purification by IMAC using Ni^{2+} charged sepharose. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude lysate, Lane 3; Column flow-through, Lane 4, 20mM Imidazole wash, Lane 5; 70mM Imidazole wash, Lane 6; 200mM Imidazole elution, Lane 7; Broad range protein ladder. Purified protein bands are indicated with an arrow.

4.5.3 CBP21 mutant characterisation

Insoluble substrate assays were carried out to compare the strength of binding of the CBP21 mutants compared to the wild type protein (see Fig 4.46–4.50) Assays were carried out as outlined in Section 2.26, with definite differences observed in the affinity of the CBP21 mutants for the insoluble substrates compared to the wild type molecule (Table 4.5). The insoluble substrates used were β-chitin from squid pen, α-chitin from crustacean shells, crystalline cellulose, chitosan from carapacea skin and α-chitosan from crustacean shells.

Studies of WT CBP21 using β-chitin established that all of the protein bound to the resin with none present in the elution fractions. The affinity of the mutants to β-chitin, compared with that of the WT, appeared to be diminished, with some displaying a very low binding capacity (see Fig 4.48). The most profound effect of the alanine scanning was observed with $CBP21_{H114A}$, which appeared to have no affinity for β-chitin. $CBP21_{T111A}$ and $CBP21_{Y54A}$ also appeared to have a very low affinity for the insoluble substrate. While the WT protein appeared to bind to α- and β-chitin with similar affinities, the same was not true for all of the mutants. In the case of α-chitin all of the CBP21 mutants appeared to retain an ability to bind to the insoluble substrate (see Fig 4.49). Alanine scanning appeared to have little or no effect on $CBP21_{E55A}$, $CBP21_{P56A}$ and $CBP21_{Q57A}$, $CBP21_{S58A}$, $CBP21_{E60A}$, $CBP21_{T111A}$ and $CBP21_{D182A}$. Interestingly similarly to the results of the β-chitin study the binding of $CBP21_{Y54A}$ and $CBP21_{H114A}$ to α-chitin was diminished, with more protein present in elution fractions compared to the other mutants.

The WT protein had a low affinity for the chitosan from carpacea skin, with the majority of the loaded protein eluting in the first elution fraction (see Fig 4.50b). None of the mutants appeared to have an increased capacity to bind to this substrate (see Fig 4.50), with $CBP21_{S58A}$, $CBP21_{T111A}$ and $CBP21_{H114A}$ exhibiting extremely diminished capacities, with either notably faint or no protein bands visible in the boiled chitosan sample lane (Lane 6). The WT protein did not display a high affinity for chitosan from crustacean shells, and the same remained

true for the majority of the mutants (see Fig 4.51). $CBP21_{Q57A}$ however appeared to have an increased affinity for α-chitosan, while the $CBP21_{E60A}$ and $CBP21_{T111A}$ mutants displayed slight increases in band intensity. Significant differences between CBP21 mutants and the parent protein were again observed when crystalline cellulose was used as a substrate (see Fig 4.50). Mutants $CBP21_{Y54A}$, $CBP21_{E60A}$, $CBP21_{H114A}$ had little or no ability to bind to crystalline cellulose compared to the WT, while a marked increase was observed with $CBP21_{T111A}$ and $CBP21_{D182A}$ mutants.

Differences in the CBP21 mutants ability to bind to glycoproteins was explored using an ELLA (Section 2.27). Each mutant was subjected to analysis using untreated, nuriminidase and β-galactosidase treated thyroglobulin as the substrate (see Fig 4.51). Unfortunately none of the mutants displayed an increased affinity for the oligos tested.

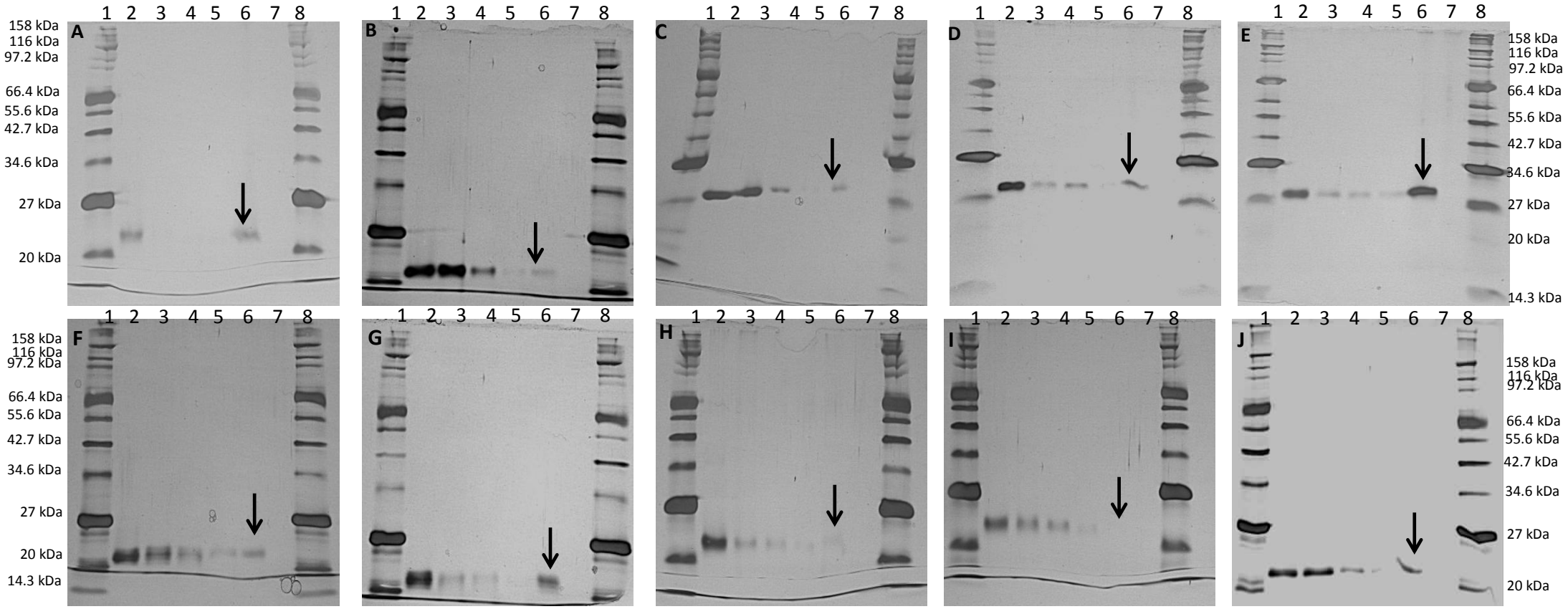


Fig 4.46: CBP21 solid substrate assay using β-chitin from squid pen. A; CBP21 WT, B; $CBP21_{Y54A}$, C; $CBP21_{E55A}$, D; $CBP21_{P56A}$, E; $CBP21_{Q57A}$, F; $CBP21_{S58A}$, G; $CBP21_{E60A}$, H; $CBP21_{T111A}$, I; $CBP21_{H114A}$ and J; $CBP21_{D182A}$ Lane 1; Broad range protein ladder (Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; Eluted sample, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21 β-chitin sample, Lane 7; Boiled PBS β-chitin sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

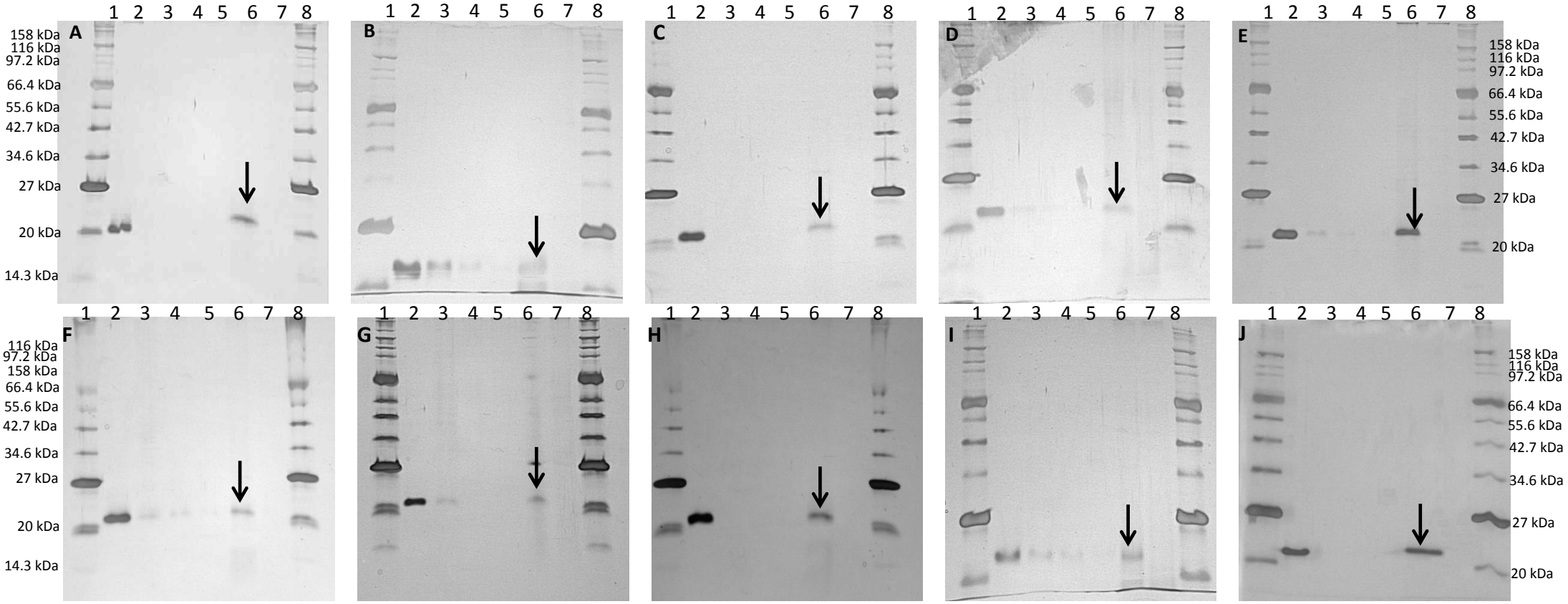


Fig 4.47: CBP21 solid substrate assay using α-chitin from crustacean shell. A; CBP21 WT, B; CBP21$_{Y54A}$, C; CBP21$_{E55A}$, D; CBP21$_{P56A}$, E; CBP21$_{Q57A}$, F; CBP21$_{S58A}$, G; CBP21$_{E60A}$, H; CBP21$_{T111A}$, I; CBP21$_{H114A}$ and J; CBP21$_{D182A}$. Lane 1; Broad range protein ladder (Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; Eluted sample, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21 α-chitin sample, Lane 7; Boiled PBS α-chitin sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

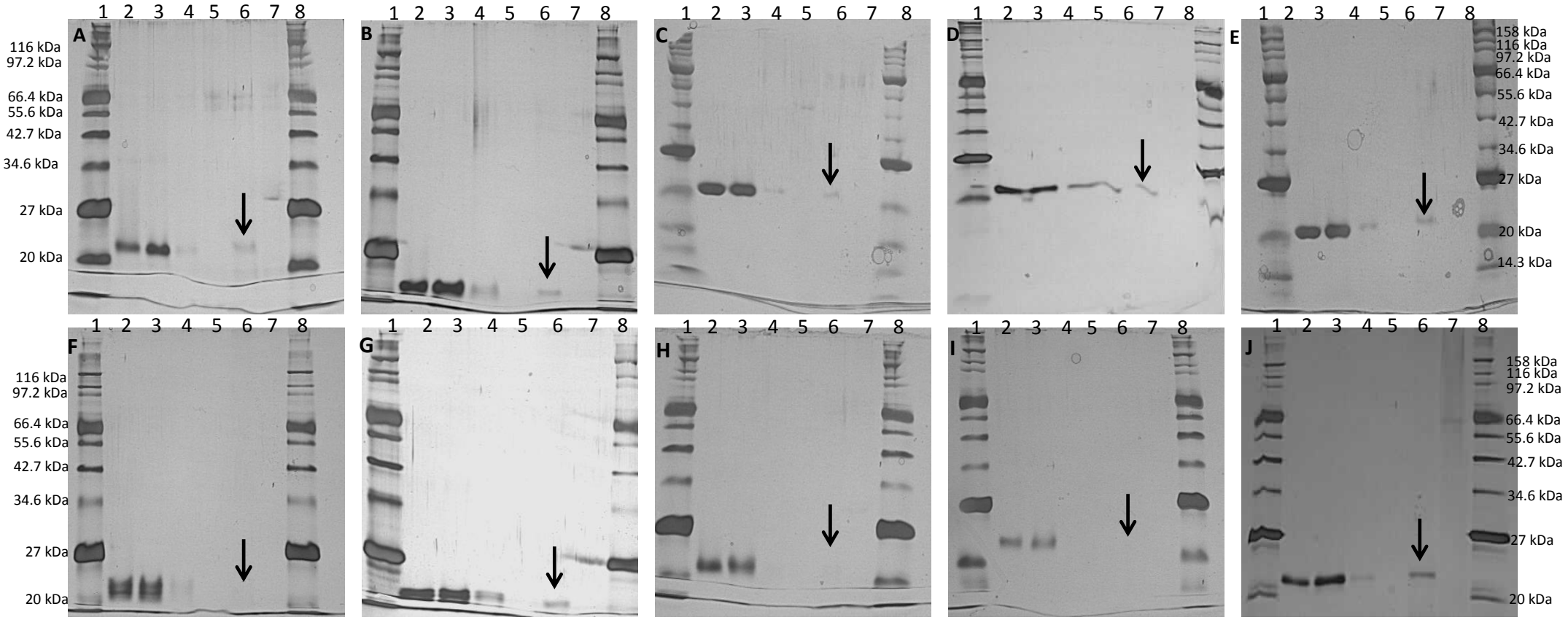


Fig 4.48: CBP21 solid substrate assay using chitosan from carapacea skin, A; CBP21 WT, B; CBP21$_{Y54A}$, C; CBP21$_{E55A}$, D; CBP21$_{P56A}$, E; CBP21$_{Q57A}$, F; CBP21$_{S58A}$, G; CBP21$_{E60A}$, H; CBP21$_{T111A}$, I; CBP21$_{H114A}$ and J; CBP21$_{D182A}$. Lane 1; Broad range protein ladder (Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; Eluted sample, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21 β-chitosan sample, Lane 7; Boiled PBS β-chitosan sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

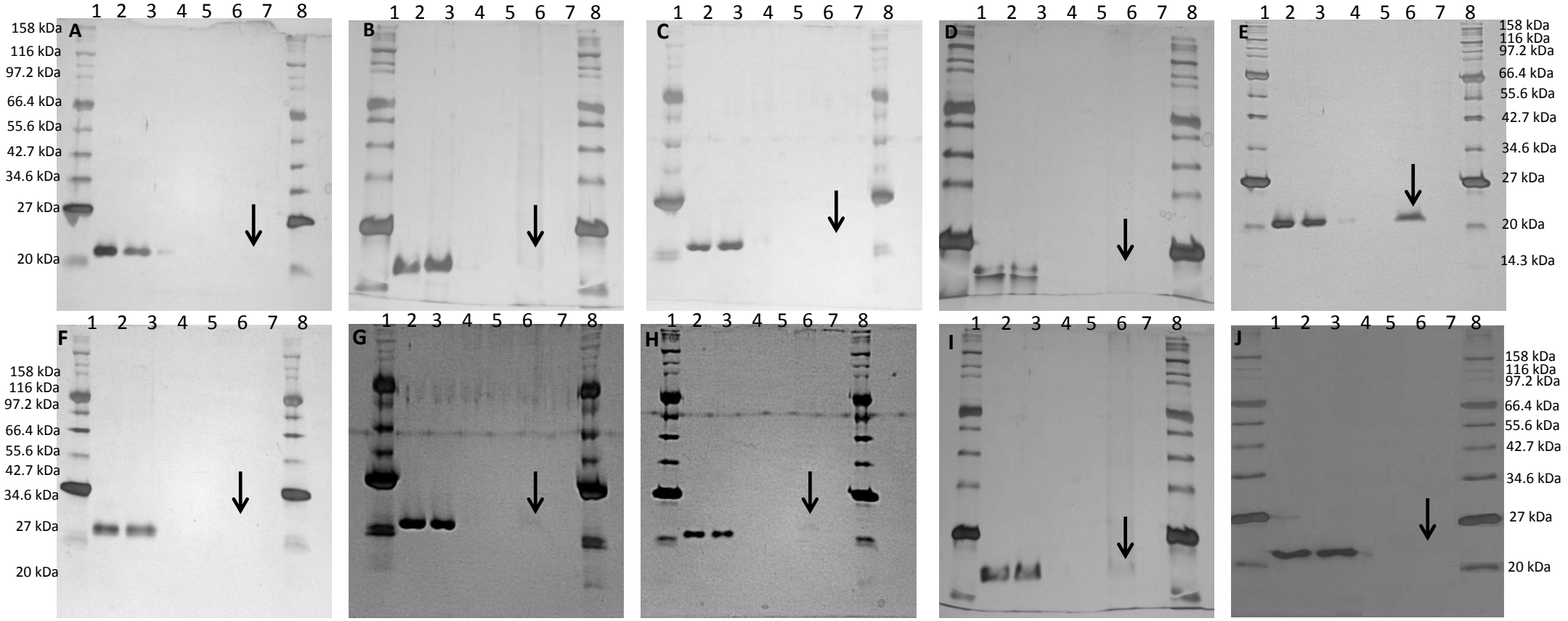


Fig 4.49: CBP21 solid substrate assay using α-chitosan from crustacean shell, A; CBP21 WT, B; $CBP21_{Y54A}$, C; $CBP21_{E55A}$, D; $CBP21_{P56A}$, E; $CBP21_{Q57A}$, F; $CBP21_{S58A}$, G; $CBP21_{E60A}$, H; $CBP21_{T111A}$, I; $CBP21_{H114A}$ and J; $CBP21_{D182A}$. Lane 1; Broad range protein ladder (Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; Eluted sample, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21 α-chitosan sample, Lane 7; Boiled PBS α-chitosan sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

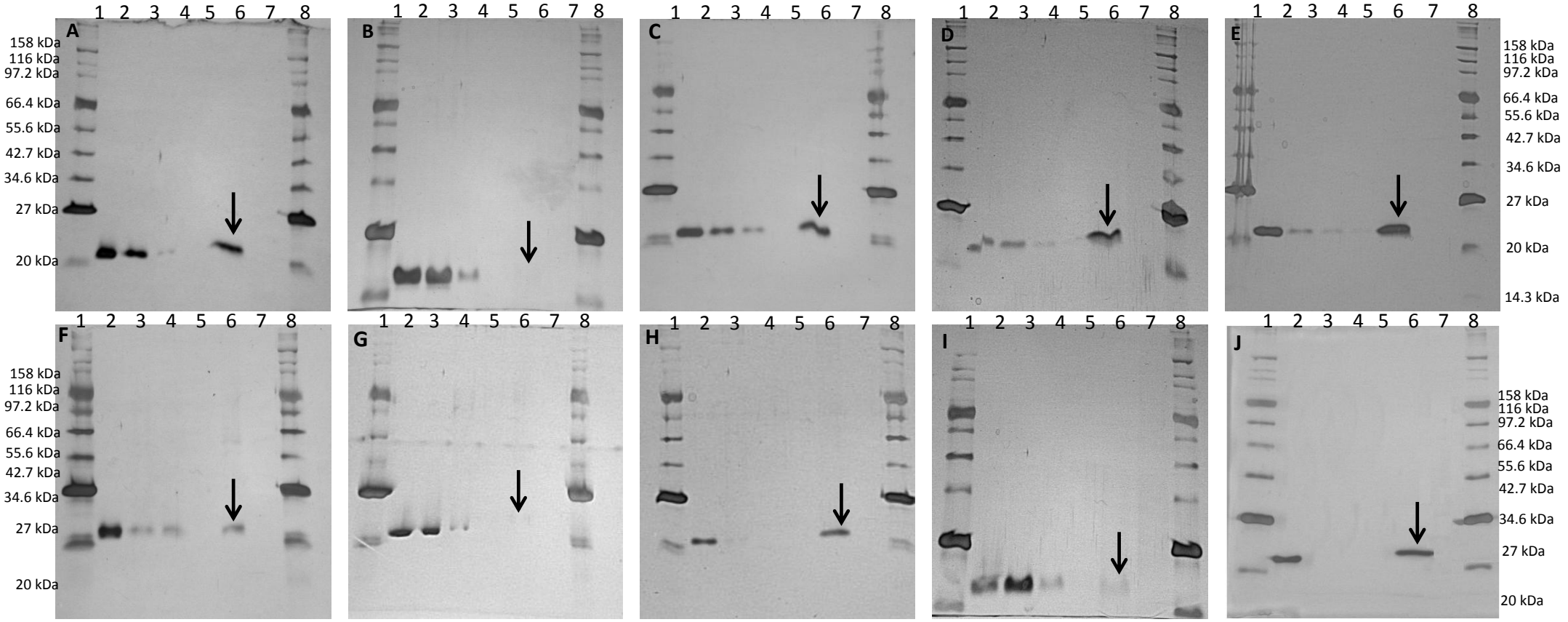


Fig 4.50: CBP21 solid substrate assay using crystalline cellulose, A; CBP21 WT, B; CBP21$_{Y54A}$, C; CBP21$_{E55A}$, D; CBP21$_{P56A}$, E; CBP21$_{Q57A}$, F; CBP21$_{S58A}$, G; CBP21$_{E60A}$, H; CBP21$_{T111A}$, I; CBP21$_{H114A}$ and J; CBP21$_{D182A}$. Lane 1; Broad range protein ladder (Section 2.23.2), Lane 2; Crude CBP21 sample, Lane 3; Eluted sample, Lane 4; Wash 1, Lane 5; Wash 2, Lane 6; Boiled CBP21 α-chitosan sample, Lane 7; Boiled PBS α-chitosan sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

Table 4.5: Overview of the marked changes observed between the CBP21 wild type protein and its mutants.

Mutation	β-chitin	α-chitin	Chitosan (carpacea skin)	Chitosan (crustacean shell)	Crystalline cellulose
WT	√	√	√	√	√
Y54A	- -	- -	√	√	X
E55A	- -	√	√	√	√
P56A	-	-	-	√	√
Q57A	- -	-	√	+ +	√
S58A	- -	-	X	√	√
E60A	-	-	√	√	- -
T111A	- -	√	X	+	+ +
H114A	X	- -	X	+	-
D182A	- -	√	√	√	+ +

√ Binding comparable to WT, slight increase in affinity compared to WT (+), larger increase in affinity compared to WT (+ +), slightly decrease in affinity compared to WT (-), larger decrease in affinity compared to WT (- -), no binding (X).

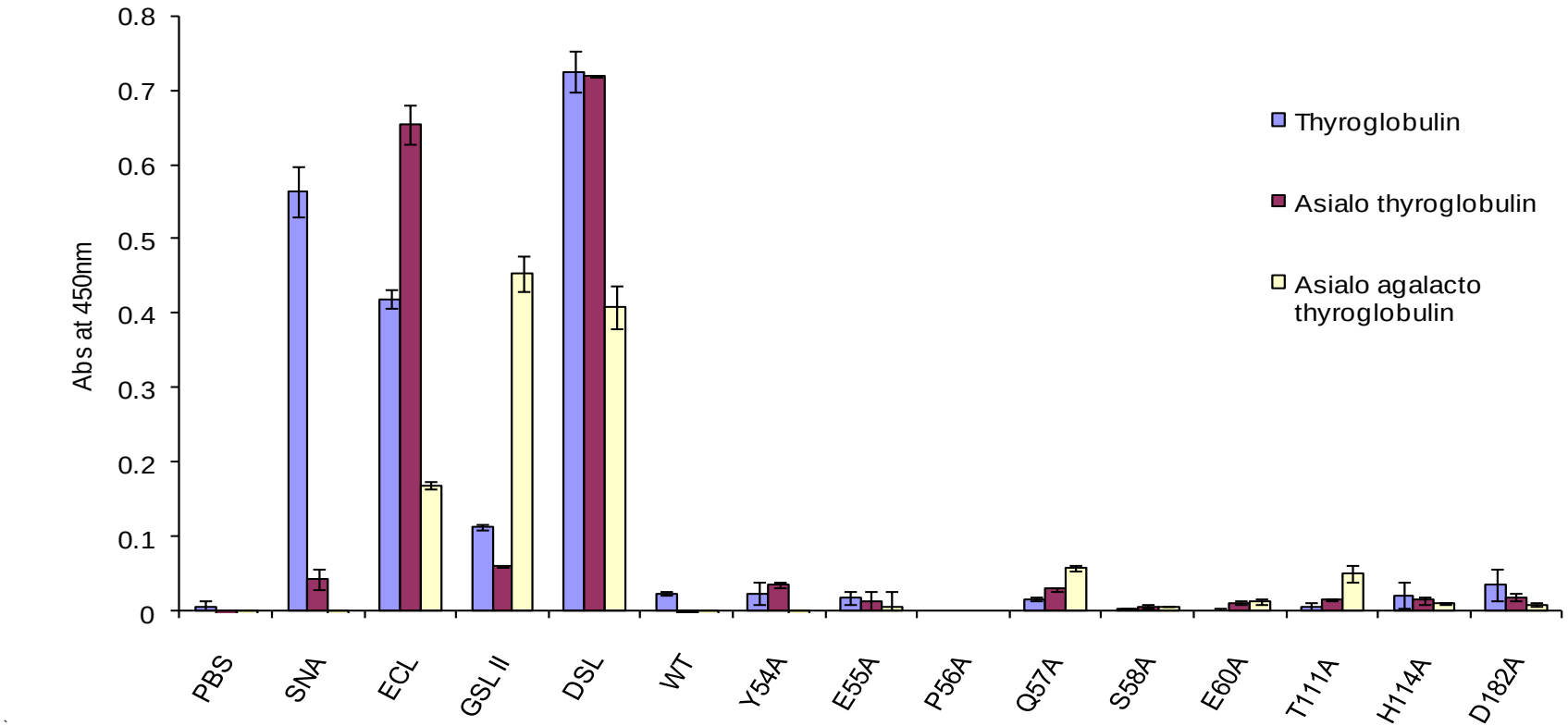


Fig 4.51: ELLA analysis of CBP21 mutants using glycosidase treated thyroglobulin as a glycoprotein source. SNA, ECL, GSLII and DSL are commercially available plant lectins that show a binding preference for sialic acid, terminal galactose, terminal GlcNAc and chitobiose respectively.

4.5.4 Random mutagenesis of CBP21

While it has been show in section 4.4.3 that alanine scanning of various CBP21 residues impacted on the proteins affinity for various insoluble substrates, the replacement of amino acids with alanine is more likely to decrease a proteins affinity for a given substrate rather than alter it. Random mutagenesis is a technique better suited for inducing amino acid changes that may produce proteins with altered specificities. Alanine scanning of these sites had previously been shown to have the greatest impact in terms of βmax on the affinity for β-chitin (Table 1.4) (Vaaje-Kolstad *et al.*, 2005a). Site directed random mutagenesis was carried out on two positions within the CBP21 protein, E55 and H114, as outlined in Fig 4.32. The positions of the residues to be mutated and the locations of primer annealing sites are shown in Fig 4.52 and 4.53.

Following PCR the reaction was cleaned up (Section 2.8) and subjected to DpnI digestion to remove any of the template DNA. The plasmids were then ligated (Section 2.10.2) and transformed (Section 2.9.3) directly into *E. coli* KRX. Following overnight incubation transformed colonies were subcultured in deep well 96-well plates, with transfer to fresh medium on Day 2. On Day 3 10 µl of cell culture was transferred to a fresh 96-well plate containing LB medium, with ampicillin and IPTG. The plate was incubated with shaking at 800-1000 rpm overnight, at 37°C. The plate was spun at 4,000 rpm to pellet the cells, and the medium removed. Cells were lysed as described in section 2.15.2. Asialo agalacto thyroglobulin was immobilised on a 96 well Maxisorp plate overnight at 4°C. The plate was blocked with 200 µL carbo-free blocking solution for one hour at 25°C. 50 µL of cell lysate was transferred to this blocked maxiSorp plate for ELLA analysis (see Fig 4.54 and 4.56). Plates were developed over 30 minutes, with lectin control samples stopped after 5 minutes.

Once the OD_{450nm} readings were normalised by the subtraction of background signal, few promising $CBP21_{E55}$ random clones remained. The analysis was repeated using fresh cultures and lysate and the same clones surfaced as having potential for glycoprotein interaction. Clones B1, B2, G3 and G10 were then

selected for further analysis, they were grown up in 500 mL culture volumes (Section 2.14), purified using IMAC, buffer exchanged into PBS and re-examined using ELLA analysis (see Fig 4.55). None of these selected clones showed an affinity for the glycoprotein derivatives tested in this assay format. Similarly for $CBP21_{H114}$ random mutants, no clones stood out as having a potential affinity for the glycoprotein tested, following the initial ELLA analysis (see Fig 4.56).

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
G ACGTTGAGTT GTGCGTCACG CCGTCGGTGC AGTACGAANN NNCAGAGCGTC
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GAA
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.52: Outline of the construction of $CBP21_{E55}$ random mutants. Representation of the position of the primers used for the mutation of CBP21 at position E55. The position of the primer E55A_R_F is represented in blue, with 55-60A_R represented in red (Table 2.3). The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

ATGGGAAACA AAACTTCCCG TACCCTGCTC TCTCTGGGCC TGCTGAGCGC GGCCATGTTC
GGCGTTTCGC AACAGGCGAA TGCTCACGGT TATGTCGAAT CGCCGGCCAG CCGCGCCTAT
CAGTGCAAAC TGCAACTCAA CACGCAGTGC GGCAGCGTGC AGTACGAACC GCAGAGCGTC
GAAGGCCTGA AAGGCTTCCC ACAGGCCGGC CCGGCTGACG GCCACATCGC CAGCGCCGAC
TT CGAGTTGGAC
AAGTCCACCT TCTTCGAACT GGATCAGCAA ACGCCGACGC GCTGGAACAA GCTCAACCTG
TTTTGGCCGG GCTTGAGGTT TACCTGGAAG CTGACCGCCC GTNNNAGCAC AACC
AAAACCGGCC CGAACTCCTT TACCTGGAAG CTGACCGCCC GTCACAGCAC AACCAGCTGG
CGCTATTTCA TCACCAAGCC AAACTGGGAC GCTTCGCAGC CGCTGACCCG CGCTTCCTTT
GACCTGACGC CGTTCTGCCA GTTCAACGAC GGCGGCGCCA TCCCTGCCGC ACAGGTCACC
CACCAGTGCA ACATACCGGC AGATCGCAGC GGTTCGCACG TGATCCTTGC CGTGTGGGAC
ATAGCCGACA CCGCCAACGC CTTCTATCAG GCGATCGACG TCAACCTGAG CAAAAGATCT
CATCATCATC ACCATCACTA A

Fig 4.53: Outline of the construction of CBP21$_{H114}$ random mutants. Representation of the position of the primers used for the random mutation of CBP21 at position H114. The position of the primer H114_R_F is indicated with blue, with T111A R represented in red (Table 2.3). The nucleotides to be mutated are highlighted in yellow, with the nucleotides to be introduced on each primer underlined bold.

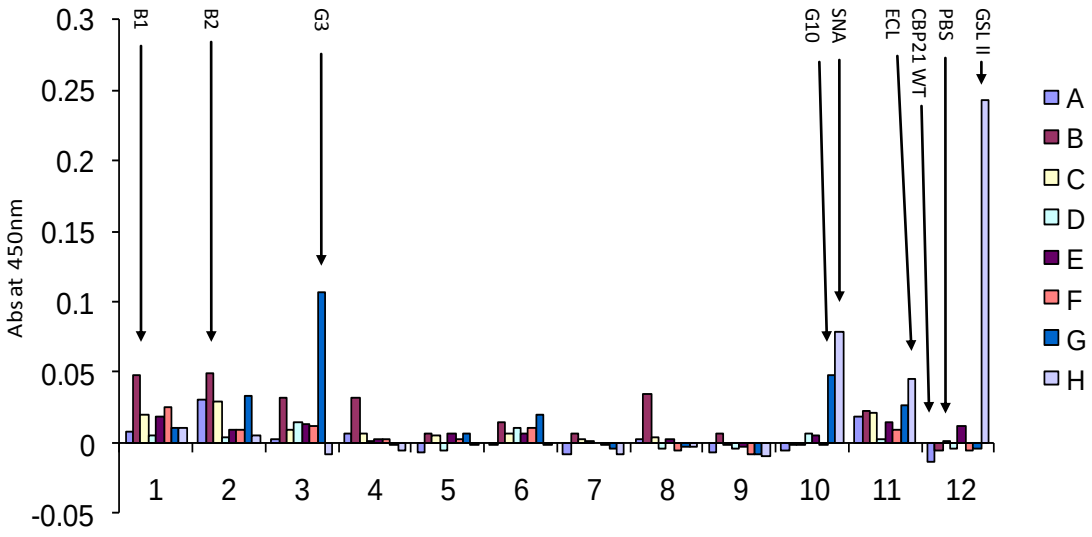


Fig 4.54: ELLA analysis of CBP21$_{E55}$ random mutants bound to asialo agalacto thyroglobulin. Quantitative detection of CBP21$_{E55}$ random mutants bound to asialo agalacto thyroglobulin. Binding of the mutants to the glycoprotein was detected through the use of a peroxidase linked antibody at OD$_{450nm}$. SNA is a terminal sialic binding lectin, ECL has a binding preference for terminal galactose, GSL II is a terminal GlcNAc binding lectin

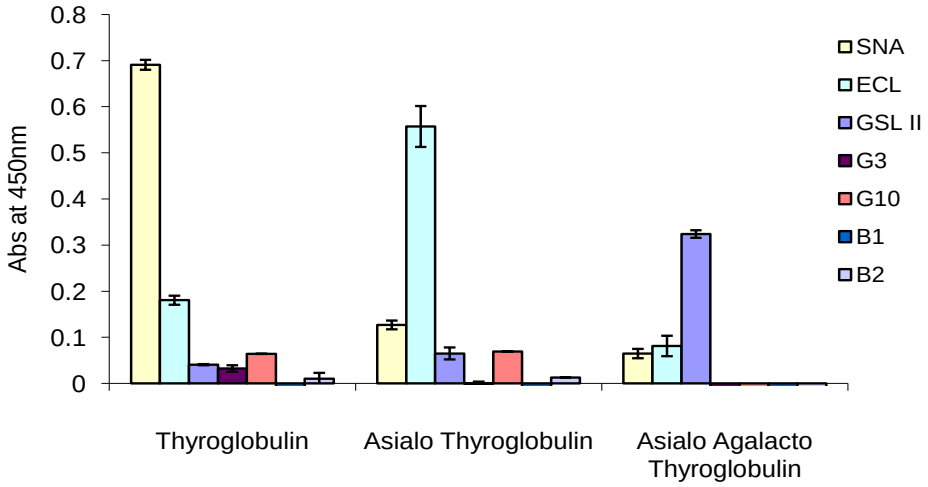


Fig 4.55: ELLA analysis of CBP21$_{E55}$ random mutants G3, G10, B1 and B2. Quantitative detection of CBP21$_{E55}$ random mutants bound to untreated, neuraminidase and β-galactosidase treated thyroglobulin through the use of a peroxidase linked antibody at OD_{450nm}. SNA is a terminal sialic binding lectin, ECL has a binding preference for terminal galactose, GSL II is a terminal GlcNAc binding lectin

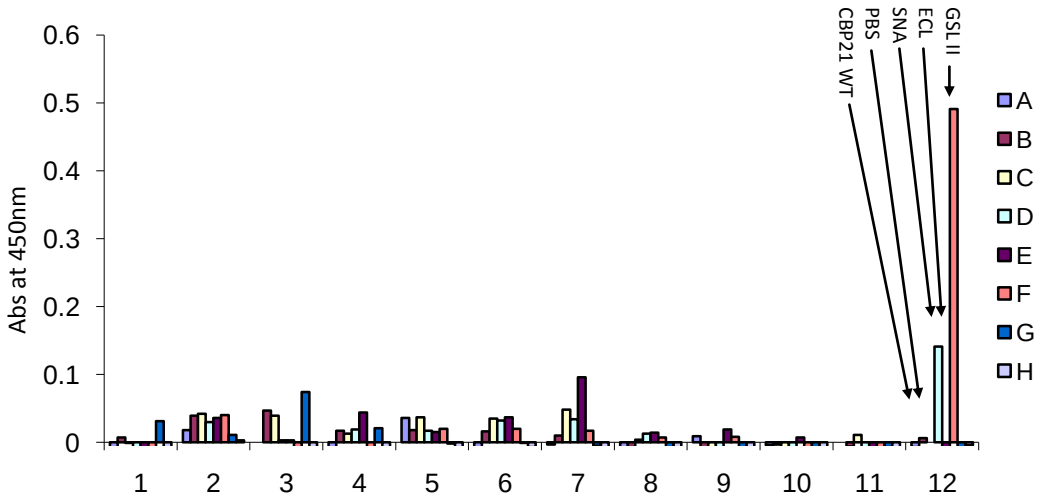


Fig 4.56: ELLA analysis of CBP21$_{H114}$ random mutagenesis using asialo agalacto thyroglobulin. Quantitative detection of CBP21$_{H114}$ random mutants bound to asialo agalacto thyroglobulin. Binding of the mutants to the glycoprotein was detected through the use of a peroxidase linked antibody at OD_{450nm}. SNA is a terminal sialic binding lectin, ECL has a binding preference for terminal galactose, GSL II is a terminal GlcNAc binding lectin

4.6 Discussion

Preliminary ELLA analysis of CBP21 revealed that it did not interact with the protein linked oligosaccharides tested. This result was perhaps unsurprising. As mentioned previously the purpose of this project was to evaluate if CBP21, a chitin-binding protein, has an affinity for GlcNAc sugars located within a protein linked glycan structure. It is known that all N-linked glycans share a common core structure, consisting of a chitobiose core, linked to 9 mannose residues (see Fig 1.2). Depending on the origin of the glycoprotein these structures are further processed, with trimming, reducing the mannose content, and further elongation with sugars such as GlcNAc, galactose, fucose and sialic acid (Section 1.2.1). GlcNAc is typically found in the chitobiose core, and as the first residue linked to the core mannose structure of N-linked glycans of glycoproteins.

The binding profile of CBP21 was compared to that of GSLII, a commercial plant lectin, with a known affinity for terminal β-linked GlcNAc. While some of the glycans tested displayed sialic acid (fetuin and thyroglobulin) or galactose (asialo fetuin) as their terminal residues, others are high mannose structures, while for avian ovalbumin most of the terminal sites of the antennary structures of the N-glycans are GlcNAc (Kim *et al.*, 2007). Although GSLII demonstrated an affinity for the glycoproteins ovalbumin, invertase and hyaluronidase, the apparent lack of affinity for these residues by CBP21 could be as a result of a number of reasons; CBP21 may not be able to bind to single GlcNAc residues, meaning it won't bind to terminal GlcNAc, or perhaps the format of this assay might not suit CBP21. The glycoprotein structure and nature of the glycosidic linkages might introduce some steric hindrance, or one may need a larger concentration of glycan structures on the protein to detect binding.

Glycosidase treatment of the glycoproteins transferrin and thyroglobulin was used to expose terminal GlcNAc. Neuriminidase and β-galactosidase treatment of these glycoproteins was sufficient in removing the unwanted sialic acid and galactose residues. While both glycoproteins have similar glycan structures porcine thyroglobulin has a much higher concentration of monosaccharides per

molecule (approx 300) (Spiro, 1965). Although the terminal moiety on display was GlcNAc, CBP21 did not show any affinity for the molecules. In 2005(b) Vaaje-Kolstad *et al*, succeeded in crystallising the CBP21 molecule. They also attempted to crystallise the protein in complex with GlcNAc, but did not succeed. This is probably because CBP21 cannot bind to a single GlcNAc sugar residue. From this experiment it was also evident that DSL, a plant lectin described as a chitin binding protein, showed affinity for all forms of transferrin and thyroglobulin, suggesting that it was interacting with the chitobiose core. It was observed that absorbance readings for DSL bound to thyroglobulin were significantly higher than those for transferrin bound DSL, this is probably as a result of the increased concentration of glycans displayed on the surface of thyroglobulin. As CBP21 did not display any binding pattern one could conclude that CBP21 is not capable of binding to the chitobiose core in a similar manner to DSL. An attempt was made to trim back a glycoprotein structure to completely expose the chitobiose core, but efforts to eliminate the core mannose structure using glycosidases proved fruitless. An alternative technique was then employed to test if the use of a different assay format would improve the understanding the situation. Glycoprotein affinity chromatography was carried out using invertase immobilised to cyanogen bromide activated sepharose. No interaction between CBP21 and the invertase resin was observed. From the cumulative results it was evident that CBP21 was unable to interact with terminal GlcNAc, nor the chitobiose core structure of undigested protein-linked glycans.

Through the use of the chemically synthesised PAA-linked oligosaccharides a $(GlcNAc)_N$ assay was developed. The aim of this assay was to determine if CBP21 had an affinity for protein-free $(GlcNAc)_N$ oligosaccharides. Preliminary assay results revealed problems with the blocking agent BSA. There appeared to be some interaction between the PAA-linked oligos and the blocking solution, which gave rise to high background readings. This occurrence deemed BSA unsuitable as a blocking agent as its use could lead to false positive results. An alternative blocking reagent was identified by the examination of a number of protein and non-protein based blockers.

There are numerous blocking agents on the market today, with a notable increase in the availability of non-protein based blocking solutions for sale over the last decade. Unlike BSA or other sera, casein, and other common protein-containing blocking agents, these products are free of glycoproteins, making them ideal for applications in which glycoprotein contamination could generate false positive results. The suitability of the blocking solutions was tested, with some interesting results (see Fig 4.15). The quenching of signal at OD_{450nm} from the direct detection of CBP21 was observed with the use of PVA, PVP and synblock. These three blocking agents are non-protein based. While the exact content of synblock is proprietary, it is known that PVA and PVP are large water soluble synthetic polymers. The most likely explanation for this phenomenon is that these long chain polymers are inhibiting the anti-His antibody from reaching the antigen binding site on the protein surface. CBP21 is a small molecule (19.8 kDa), with the antibody approximately 7.5 times larger (~150 kDa). Unlike protein blockers the synthetic agents won't necessarily immobilise in a horizontal position on the plate, but will be free to move into any number of positions, possibly branching over CBP21 blocking the antigen binding site (see Fig 4.57).

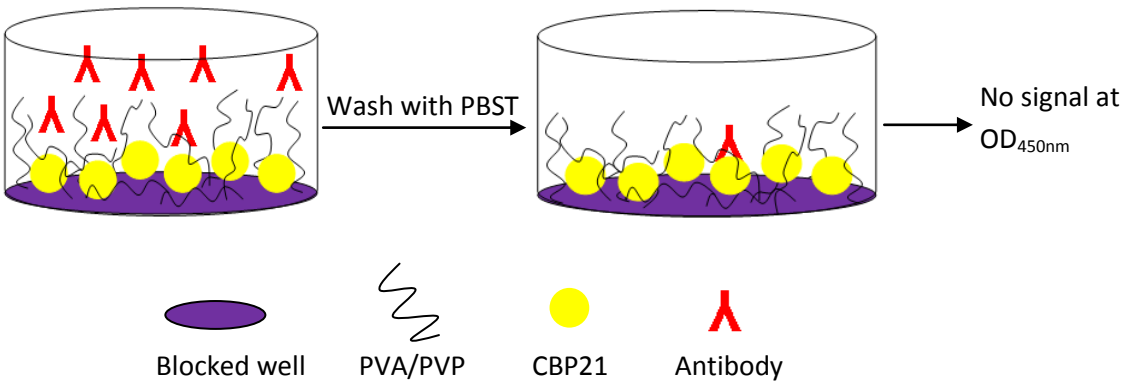


Fig 4.57: Reduced detection of CBP21 by anti-His antibody caused by polymer based blocking agents PVA and PVP. Blocking with large polymeric blocking agents obstructs the detection of the $(His)_6$ tagged protein by inhibiting the anti-His antibody from reaching the antigen binding site.

This theory was verified by repetition of the experiment using the biotinylated lectin NPL, with a reduced level of quenching observed (see Fig 4.17). This reduced level of signal loss is probably due to the steric availability of multiple biotins on each NPL monomer versus the single His tag present on CBP21

Immobilisation of the same concentration of a larger molecule on a plate surface will reduce the amount of blocking agent used. The space between the polymer and the antigen binding site of a larger molecule will also be increased, theoretically leaving more room for the antibody-antigen interaction. Carbo-free blocker was determined to be the best alternative blocking agent and was selected for further experiments. Although this is a protein-based blocking solution this product is certified as being essentially free of glycoproteins.

The optimal format for the $(GlcNAc)_N$ binding assay was then determined. Assay format A revealed that surface immobilised CBP21 could bind to PAA-linked chitobiose and chitotriose oligosaccharides (see Fig 4.18). This interaction was concentration dependent with slightly higher affinities for chitobiose over chitotriose. As CBP21 could bind to the oligos when free in solution its inability to do so in assay format B was unexpected. The inability of CBP21 to bind to the surface immobilised oligos may be due to the nature of immobilisation of the PAA-linked oligos. Although the sugar units are projected away from the PAA backbone by means of a spacer arm CBP21 may be unable to orientate itself for binding so close to the plate surface.

The $(GlcNAc)_N$-binding assay was validated using WGA, with WGA displaying an affinity for both free and immobilised PAA-linked chitobiose and chitotriose in assay formats A and B respectively. WGA also displayed higher affinity for the oligosaccharides than CBP21 in assay format A. Compared to the monomeric CBP21 WGA is known to exist as a dimer, made up of two identical polypeptide subunits (18 kDa each). When assembled as a dimer the molecule forms a horseshoe-shaped complex where the sub-unit interface contains eight putative independent carbohydrate-binding sites (Lienemann 2009, Wright 1992), whereas CBP21 is theorized to have just one carbohydrate-binding site.

Therefore WGA should theoretically have an increased affinity for the substrate compared to the same unit concentration of CBP21.

With just one putative carbohydrate-binding site the immobilisation of CBP21 on the maxiSorp plate is an important experimental variable. The molecule adsorbs by passive adsorption, without the ability to control the orientation. Adsorption of CBP21 in multiple orientations will increase the variability of the assay, with the putative binding site displayed in multiple directions, meaning than not all CBP21 molecules will be functional i.e. capable of binding the oligosaccharides.

This variability could be eliminated and the assay sensitivity improved by the use of Immobilizer™ 96-well plates from NUNC. This technology provides a simple one-step procedure for covalently coupling bio-molecules in an orientation specific way. This technology exploits the use of a binding module covalently bound to a linker arm, which is coupled to the plate surface. The linker distances the bound molecule away from the plate surface optimizing accessibility of the immobilised molecule. A number of configurations are available, including streptavidin module, glutathione and Ni-chelate plates.

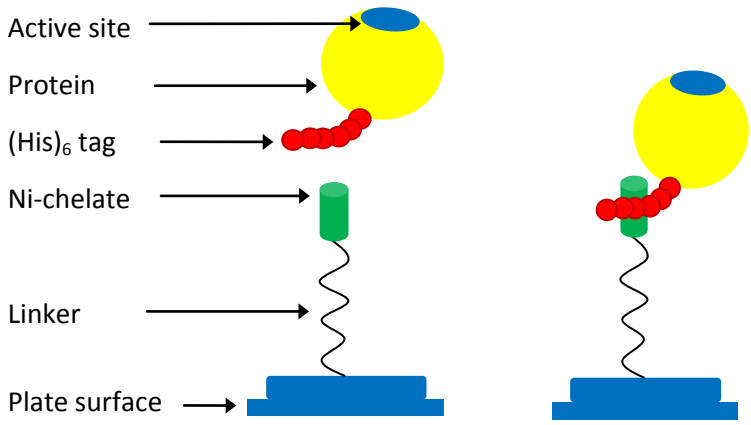


Fig 4.58: Immobilisation of a $(His)_6$ tagged protein onto a NUNC Immobilizer™ Ni-chelate plate. Specific orientation of a (His)6 tagged protein could be achieved by covalent coupling to the NUNC Immobilizer™ Ni-chelate plate.

As discussed in section 1.5 the enrichment of protein released glycans for structural analysis is an important rate limiting step. Affinity chromatography using plant lectins is often used, but due to the cost of these affinity matrices can be quite expensive process. As CBP21 can bind to free chitobiose structures there existed the possibility that it may display an affinity for protein released glycans. The potential ability of CBP21 to bind to the core structure of protein release glycans would be exceptionally beneficial in the field of glycobiology. The ability of CBP21 to bind to the chitobiose core, a structure common to all N-linked glycans would mean that the same molecule could be used for to capture all glycan species with the facility for recombinant production of the molecule reducing the overall cost.

Through the use of affinity chromatography it was shown that CBP21 could be used to capture free chitobiose oligosaccharides from solution (see Fig 4.25). Similarly to the 96-well plate assays discussed above the sensitivity of this affinity chromatography could be improved by using alternative immobilisation matrices/chemistries. As proteins are immobilised on to the cyanogen bromide activated sepharose by the presence of free amine groups, specific orientation of the molecule cannot be achieved. The use of a matrix such as streptavidin agarose (Updyke, 1984) or a similar Ni-NTA agarose would allow projection of the molecule in an orientation specific manner which could increase the availability of the binding site, thus improving the capture of the target molecule. Alternatively a lysine tag could be incorporated into the protein recombinantly with the option of an additional spacer arm.

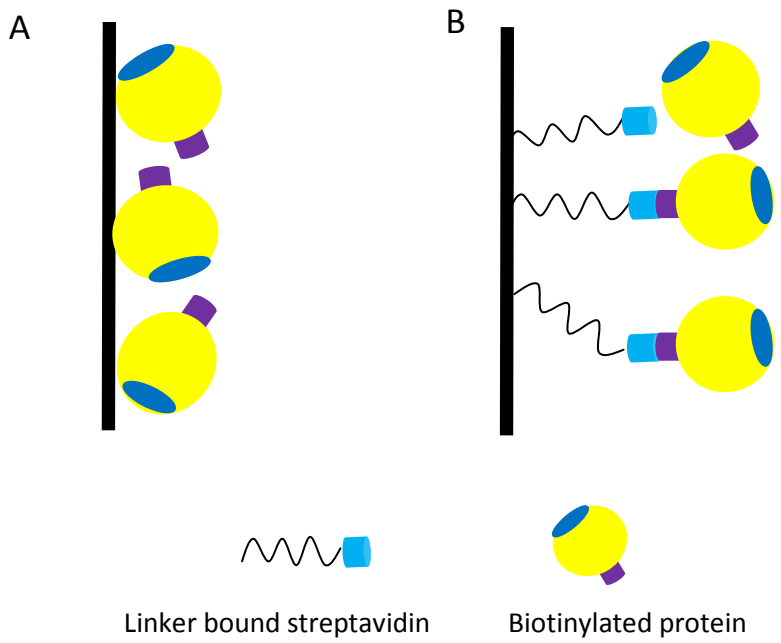


Fig 4.59: Comparison of A; Cyanogen bromide and B; Covalent coupling of protein molecules to an agarose support columns. Specific orientation of a biotin linked protein could be achieved by covalent coupling to streptavidin bio support matrix.

Asialo fetuin was selected for N-glycan cleavage. The selection of a glycoprotein is important step as PNGase F will not remove oligosaccharide containing α1-3 linked core fucose. The secondary and tertiary structure of proteins can also prevent endoglycosidases from reaching their substrate, thus making denaturation a crucial step in efficient cleavage. Unfortunately the molar concentration of CBP21 in the released glycan pool was too low to be quantified using a DNS assay. Significant improvements in assay sensitivity, to detect the released glycans could be introduced by using an online system such as LC-MS.

Sugar inhibition studies (Section 4.4) revealed that the binding of CBP21 to chitobiose was inhibited by mannan and galactan polymers. Mannan affinity chromatography was undertaken, but CBP21 did not bind to the mannan agarose. Similar studies could be repeated using a range of sugar resins and/or PAA-

linked oligosaccharides with a range of repeating sugar units, but would be cost prohibitive.

The structures of chitinases, chitin binding domains, cellulases and cellulose binding domains show that the common carbohydrate binding architecture consists of a surface groove, cleft, or tunnel lined with aromatic residues that are necessary for binding. The CBP21 monomer shows neither a groove/tunnel nor stretches of solvent-exposed aromatic residues (Vaaje-Kolstad *et al.*, 2005b). The structural and sequence analysis of the CBP21 molecule by Vaaje-Kolstad *et al.*, (2005a) suggested that the binding of CBP21 to chitin was primarily governed by polar interactions, involving residues in a polar surface patch. Alanine scanning of the residues proposed to be involved in chitin-binding was undertaken. Analysis of the mutants using the insoluble substrate assays, revealed a number of differences between mutant and WT proteins (Table 4.5).

Interestingly the site directed mutagenesis appeared to have a greater impact on the interaction with β-chitin compared to α-chitin. The main difference between α β-chitin concerns the spatial ordering of groups that may engage in hydrogen bonds with a carbohydrate-binding protein (see Fig 1.13 and 1.14). It is conceivable that as different chitin-binding proteins have different affinities for the various forms of chitin, that mutations can have effects on affinity for certain types of chitin only (Vaaje-Kolstad *et al.*, 2005b). Although the affinities of CBP21 mutants E55A and H114A for β-chitin were significantly impaired site directed random mutagenesis of these molecules revealed no increase in affinity for single GlcNAc residues attached to glycoproteins. Analysis of 90+ clones for a given mutant should cover all possible 20 amino acid combinations.

5.0 Cloning, Expression, Purification and characterisation of CbpD, CbpA and CbpL

5.1 Overview

This chapter describes the cloning of the CBP21 homologues, CbpD from *Psuedomonas aeruginosa*, CbpA from *Photorhabdus asymbiotica* and CbpL from *Photorhabdus luminescens*, the development of purification strategies and preliminary physiochemical characterisation.

5.2 Initial cloning and small scale expression of the CBP21 chitin-binding protein homologues CbpD, CbpA and CbpL

It has been demonstrated that the CBP21 was unable to bind to GlcNAc structures on protein linked glycans (Chapter 4). Although all chitin-binding proteins share a similar level of homology most show different characteristics in terms of their preference for the different forms of chitin (α, β or γ). The proteins CbpD, CbpA and CbpL were identified from BLAST analysis of CBP21 (Section 2.13). CpbD has previously been reported in the literature (Section 1.16.3), while CbpL from *P. luminescens* has been identified in a BLAST alignment (Vaaje-Kolstad *et al.*, 2005a). CbpA has not yet been identified in the literature, and none of the proteins have been produced recombinantly. As stated in section 1.16.3, unlike CbpA, CbpL and CBP21, CbpD appears to have a second C-terminal binding domain (see Fig 5.1). It is unknown if this second domain is also a chitin-binding domain. BLAST analysis of this protein revealed that both domains clustered well with other putative GlcNAc-binding proteins and chitinases (see Fig 1.24).

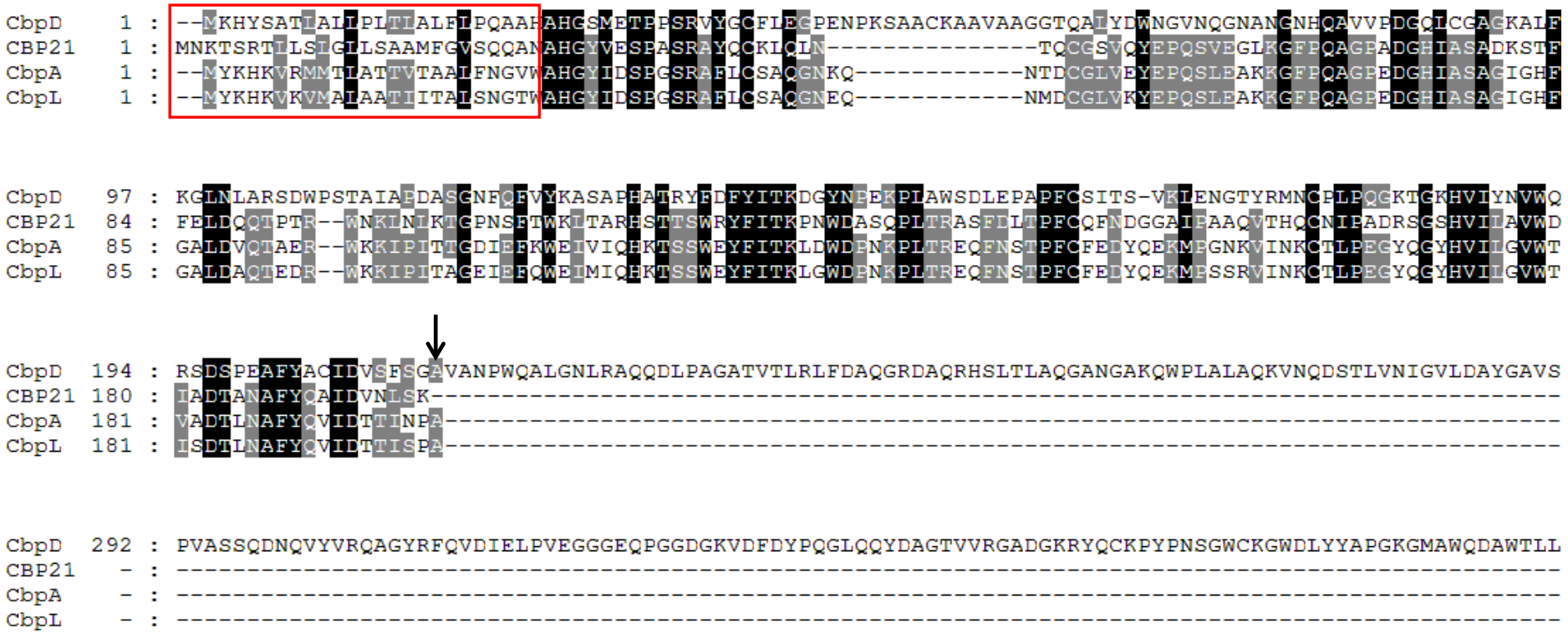


Fig 5.1: Sequence alignment of CBP21 with homologues CbpD, CbpA and CbpL. Amino acid sequence alignment of CBP21, CbpD, CbpA and CbpL from *S. marcescens*, *P. aeruginosa*, *P. asymbiotica* and *P. luminescens* respectively. The N-terminal signal sequences are highlighted with red. The beginning of the putative C-terminal binding domain of CbpD is indicated with an arrow. Alignment was performed using ClustalW and the image generated using GeneDoc 2.6 (Section 2.13).

5.2.1 Initial cloning of CbpD, CpbA and CbpL

As discussed in section 3.2.3, the chitin-binding proteins have N-terminal signal sequences that promote extracellular localisation (see Fig 5.1). To accommodate downstream purification and detection an affinity tag, in the form of 6 histidine residues, can be engineered into the proteins termini. The tagging of CBP21 at the N-terminus resulted in the expression of insoluble CBP21, while engineering of the $(His)_6$ tag at the C-terminus gave rise to a soluble form of CBP21, that could be isolated from the periplasmic space. The genes encoding CbpD, CbpA and CbpL were cloned into the pQE60 vector from Qiagen, facilitating the incorporation of the $(His)_6$ tag at the C-terminus.

The *P. aeruginosa* strain PAO1 was obtained from Prof. Keith Poole, and the sequence for the organism obtained from the NCBI data bank (Accession No. NC002516). The *P. luminescens* strain was provided by Dr David Clarke, University College Cork. Primers were designed using the TT01 sequence, available through the NCBI (Accession No. NC005126). *P. asymbiotica* genomic DNA was also provided by Dr David Clarke (UCC), and the sequence for the ATCC43949 strain was available from the NCBI data bank (Accession No. NC012962).

Genomic DNA was prepared according to the method outlined in section 2.6.1. The genes were amplified using PCR (Section 2.10.1). The primers used for amplification of *C*-terminally tagged CBPD were Psuedo-F3 and Pseudo-R2, for CbpA, Asymb-F3 and Asymb-R2 and Lumin-F3 and Lumin R-2 were used for CbpL. The PCR products were analysed by agarose gel electrophoresis and a band corresponding to the expected sizes were present. The PCR products were subjected to restriction analysis with *Nco*I and *Bgl*II, then ligated (Section 2.10.2) into an *Nco*I *Bgl*II digested pQE60 vector (see Fig 5.5). *E. coli* XL10 Gold cells were transformed with the ligation mixture (Section 2.9.3). The plasmid DNA obtained from some of the resulting colonies was screened for inserts using gel electrophoresis and restriction analysis. The presence of the correct inserts was further confirmed by DNA sequencing (Section 2.12).

ATGAAACACT ACTCAGCCAC CCTGGCACTC CTGCCACTCA CCCTCGCCCT GTTCCTGCCC
CAGGCAGCCC ATGCCCACGG CTCGATGGAA ACGCCGCCCA GTCGGGTCTA CGGCTGCTTC
CTCGAAGGTC CGGAGAATCC CAAGTCGGCC GCCTGCAAGG CCGCCGTCGC CGCCGGCGGC
ACCCAGGCAC TGTACGACTG GAATGGCGTC AACCAGGGCA ACGCCAACGG CAACCACCAG
GCGGTGGTCC CCGACGGCCA GCTCTGCGGC GCCGGCAAGG CACTGTTCAA GGGCCTGAAC
CTGGCTCGCA GCGACTGGCC CAGCACTGCC ATCGCGCCGG ACGCCAGCGG CAACTTCCAG
TTCGTCTACA AGGCCAGCGC GCCGCACGCG ACCCGCTACT TCGACTTCTA CATCACCAAG
GACGGCTATA ACCCCGAGAA GCCGCTGGCC TGGAGCGACC TGGAACCCGC GCCGTTCTGC
TCGATCACCA GCGTCAAGCT GGAGAACGGC ACCTACCGGA TGAACTGCCC GCTGCCCCAG
GGCAAGACCG GCAAGCATGT GATCTATAAC GTCTGGCAGC GCTCGGACAG CCCGGAAGCC
TTCTACGCCT GCATCGACGT GAGCTTCAGC GGCGCCGTCG CCAACCCCTG GCAAGCGCTG
GGCAACCTGC GCGCGCAGCA GGACCTGCCA GCCGGTGCTA CCGTCACCCT GCGTCTGTTC
GATGCCCAGG GCCGCGACGC CCAGCGTCAC AGCCTGACCC TGGCCCAGGG CGCCAACGGT
GCCAAGCAAT GGCCGCTGGC GCTGGCGCAG AAGGTCAACC AGGACTCCAC CCTGGTCAAC
ATCGGCGTGC TGGATGCCTA CGGGGCGGTC AGCCCGGTGG CCAGCTCGCA GGACAACCAG
GTCTACGTGC GCCAGGCCGG CTACCGCTTC CAGGTCGACA TCGAACTGCC GGTCGAGGGC
GGCGGCGAGC AACCGGGCGG CGACGGCAAG GTCGACTTCG ACTATCCGCA AGGCCTGCAG
CAATACGACG CCGGGACCGT AGTGCGCGGT GCCGATGGCA AGCGCTACCA GTGCAAGCCC
TACCCGAACT CCGGCTGGTG CAAGGGCTGG GACCTCTACT ACGCCCCGGG CAAGGGCATG
GCCTGGCA**GG ACGCCTGGAC CCTGCTG**TAA

Fig 5.2: *P. aeruginosa cbpd* coding sequence. The nucleotide sequence of *cbpD*. The annealing sites of primers Asymb-F3 and Asymb -R2 are indicated with red. There are no *Nco*I or *Bgl*II restriction sites within the gene.

ATGTACAAAC ATAAAGTGAG AATGATGACT TTAGCCACCA CGGTAACCGC AGCTTTATTT
AATGGTGTAT GGGCTCATGG TTATATTGAT AGTCCAGGAA GTCGAGCATT CCTTTGTTCT
GCACAGGGAA ATAAACAAAA TACCGATTGC GGGTTGGTGG AATATGAGCC TCAATCTCTT
GAAGCTAAGA AAGGCTTCCC GCAAGCTGGT CCAGAAGATG GTCATATTGC CAGTGCTGGT
ATTGGTCACT TCGGCGCGTT GGATGTTCAA ACAGCAGAGC GCTGGAAGAA AATTCCCATT
ACTACCGGTG ATATTGAGTT CAAATGGGAA ATTGTAATCC AACACAAGAC TTCAAGTTGG
GAATATTTTA TTACTAAATT AGACTGGGAC CCCAATAAGC CTTTAACCAG AGAGCAATTT
AATAGTACAC CTTTCTGTTT TGAGGATTAC CAGGAAAAAA TGCCGGGGAA TAAAGTTATT
AATAAATGTA CTTTGCCAGA AGGTTATCAA GGCTACCACG TTATATTAGG TGTCTGGACA
GTTGCAGATA CCCTGAATGC ATTTTACCAA GTGATTGAT**A CAACCATTAAC CCAGCC**TGA

Fig 5.3: *P. asymbiotica cbpA* coding sequence. The nucleotide sequence of *cbpA*. The annealing sites of primers Lumin-F3 and Lumin -R2 are indicated with red. There are no *Nco*I or *Bgl*II restriction sites within the gene.

ATGTATAAAC ATAAAGTGAA AGTGATGGCT TTAGCCGCTA CGATAATTAC AGCTTTATCT
AATGGTACAT GGGCTCATGG TTATATTGAT AGCCCAGGAA GTCGTGCATT CCTTTGTTCC
GCACAAGGAA ATGAACAAAA TATGGATTGT GGGTTGGTTA AGTATGAGCC TCAATCTCTT
GAAGCTAAGA AAGGCTTTCC ACAAGCTGGC CCGGAAGATG GTCATATTGC CAGCGCTGGC
ATTGGTCATT TTGGCGCATT GGATGCTCAA ACCGAAGATC GTTGGAAAAA AATCCCTATT
ACTGCCGGTG AGATTGAGTT CCAGTGGGAA ATTATGATTC AACACAAAAC CTCAAGTTGG
GAATATTTTA TTACCAAACT TGGTTGGGAT CCCAATAAAC CTTTAACCAG AGAGCAATTT
AATAGTACAC CTTTCTGCTT TGAGGATTAT CAGGAAAAAA TGCCGAGCAG CAGAGTTATT
AATAAATGTA CCTTACCAGA GGGTTATCAA GGCTATCACG TTATACTGGG TGTTTGGACA
ATCTCAGATA CCCTAAATGC ATTTTACCAA GTAATT**GACA CAACAATTAG CCCTGCT**TGA

Fig 5.4: *P. luminescens cbpL* coding sequence. The nucleotide sequence of *cbpL*. The annealing sites of primers Psuedo-F3 and Pseudo-R3 are indicated with red. There are no *Nco*I or *Bgl*II restriction sites within the gene.

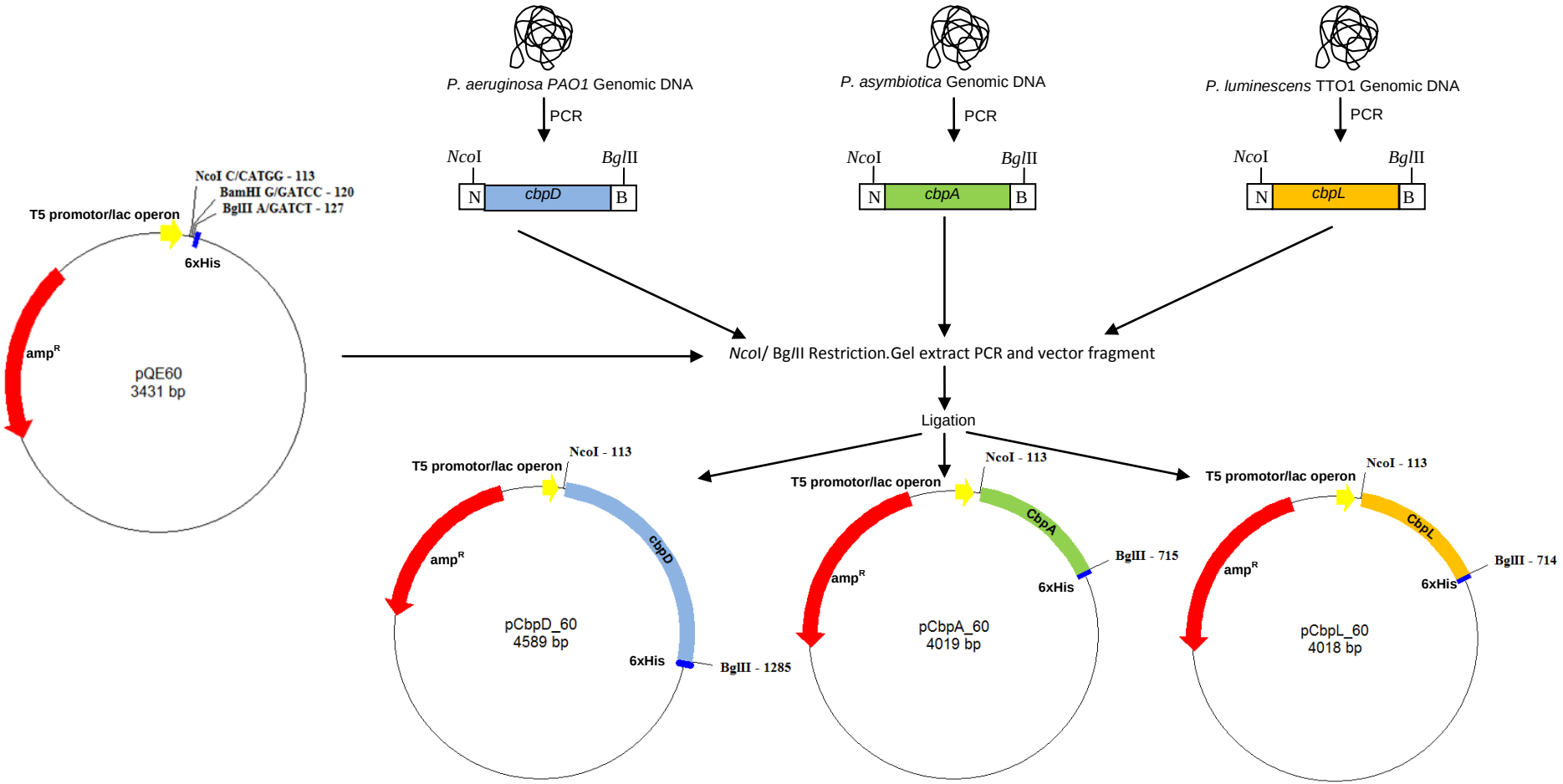


Fig 5.5: Outline of the cloning strategy for the genes encoding CbpD, CbpA and CbpL. *Cbpd*, *cbpa* and *cbpL* were cloned into the pQE60 vector from Qiagen as *Nco*I *Bgl*II fragments to yield the vector pCbpD_60, pCbpA_60 and pCbpL_60 respectively.

5.2.2 Small scale expression of CbpD, CbpA and CbpL

The proteins were expressed from their respective vectors in 100 mL cultures (Section 2.14), to ensure the recombinant proteins were soluble in *E. coli.* It was observed that unlike CBP21 all three proteins associated with the insoluble protein fraction (see Fig 5.6 - 5.9). Optimisation of the expression conditions for CbpD was undertaken in an effort to improve protein solubility. CbpD was expressed in 100 mL *E. coli* cultures at either 37 or 30°C, with varying levels of IPTG (see Fig 5.10). Unfortunately, no improvement in protein solubility was observed.

Table 5.1: Relative molecular masses of chitin-binding proteins CbpD, CbpA and CbpL

Protein	MW including signal sequence	MW without signal sequence
CbpD	42739 Da	40065 Da
CbpA	23192 Da	19545.7 Da
CbpL	23108 Da	20406.7 Da

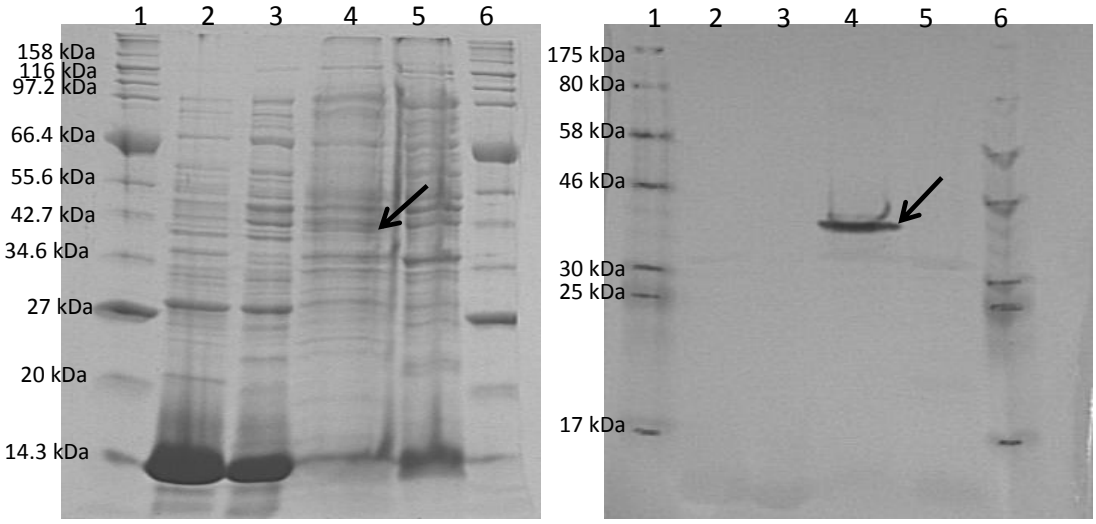


Fig 5.6: Expression of C-terminally $(His)_6$ tagged CbpD in *E. coli* XL10 Gold. SDS-PAGE and corresponding western blot analysis of C-tagged CbpD expressed in *E. coli*. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; pCbpD_60 soluble fraction, Lane 3; empty vector soluble fraction Lane 4; pCbpD_60 insoluble fraction, Lane 5; empty vector insoluble fraction, Lane 6; Broad Range protein ladder. The over-expressed band corresponding to CbpD is indicated with an arrow.

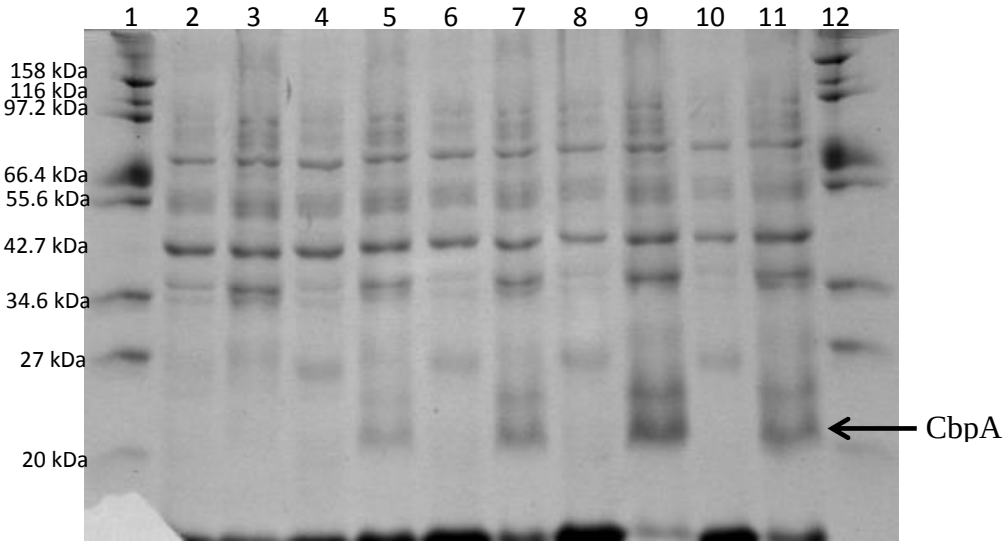


Fig 5.7: Time course expression analysis of C-terminally $(His)_6$ tagged CbpA in *E. coli* XL-10 Gold. Time course analysis by 12.5% SDS-PAGE of C-tagged CbpA expressed in *E. coli*. Broad range protein ladder (Section 2.23.3), Lane 2; Soluble fraction 0 hours, Lane 3; Insoluble fraction 0 hours, Lane 4; Soluble 1 hour, Lane 5; Insoluble 1 hour, Lane 6; Soluble 2 hours, Lane 7; Insoluble 2 hours, Lane 8; Soluble 3 hours, Lane 9; Insoluble 3 hours, Lane 10 Soluble 4 hours, Lane 11; Insoluble 4 hours Lane 12; Broad range protein marker.

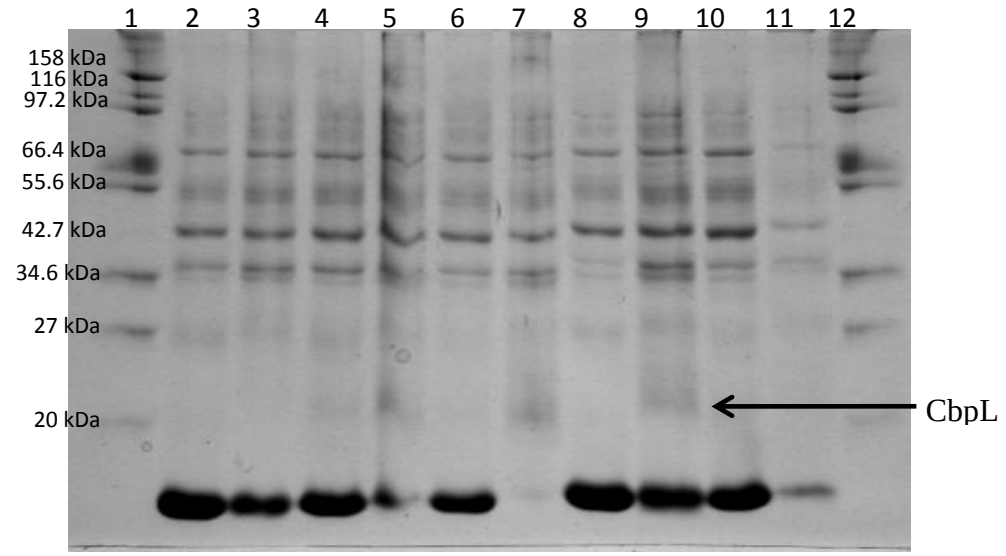


Fig 5.8: Time course expression analysis of C-terminally $(His)_6$ tagged CbpL in *E. coli* XL10 Gold. 12.5% SDS-PAGE time course analysis of C-tagged CbpL expressed in *E. coli*. Broad range protein ladder (Section 2.23.3), Lane 2; Soluble fraction 0 hours, Lane 3; Insoluble fraction 0 hours, Lane 4; Soluble 1 hour, Lane 5; Insoluble 1 hour, Lane 6; Soluble 2 hours, Lane 7; Insoluble 2 hours, Lane 8; Soluble 3 hours, Lane 9; Insoluble 3 hours, Lane 10 Soluble 4 hours, Lane 11; Insoluble 4 hours Lane 12; Broad range protein marker.

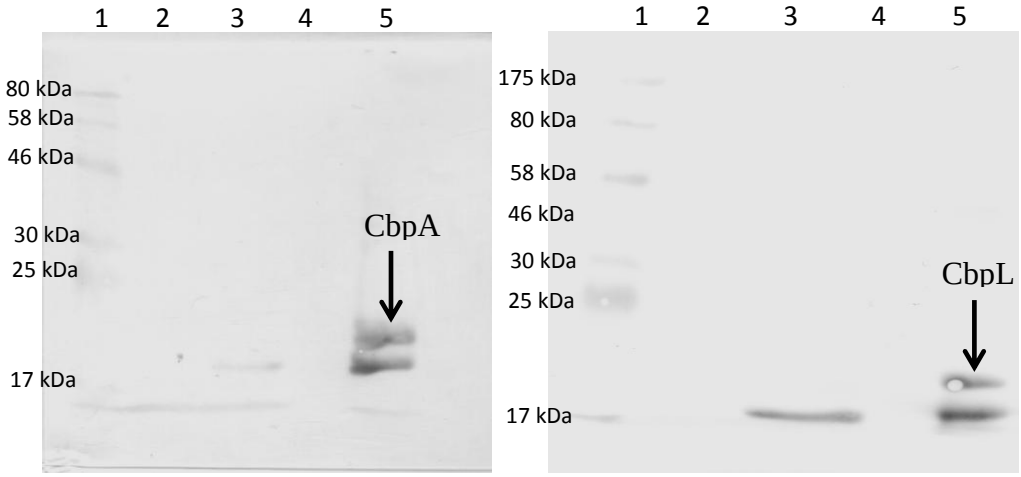


Fig 5.9: Western blot analysis of C-terminally $(His)_6$ tagged A; CbpA and B; CbpL. Analysis of the soluble and insoluble fractions from the expression of CbpA and CbpL from the pQE60 vector in *E. coli* XL10 Gold. Lane 1; Pre-stained protein ladder (Section 2.23.3), Lane 2; pQE60 empty vector soluble fraction, Lane 3; pQE60 expression vector soluble fraction, Lane 4; pQE60 empty vector insoluble fraction, Lane 5; pQE60 expression vector insoluble fraction. CbpA and CbpL are indicated with an arrow

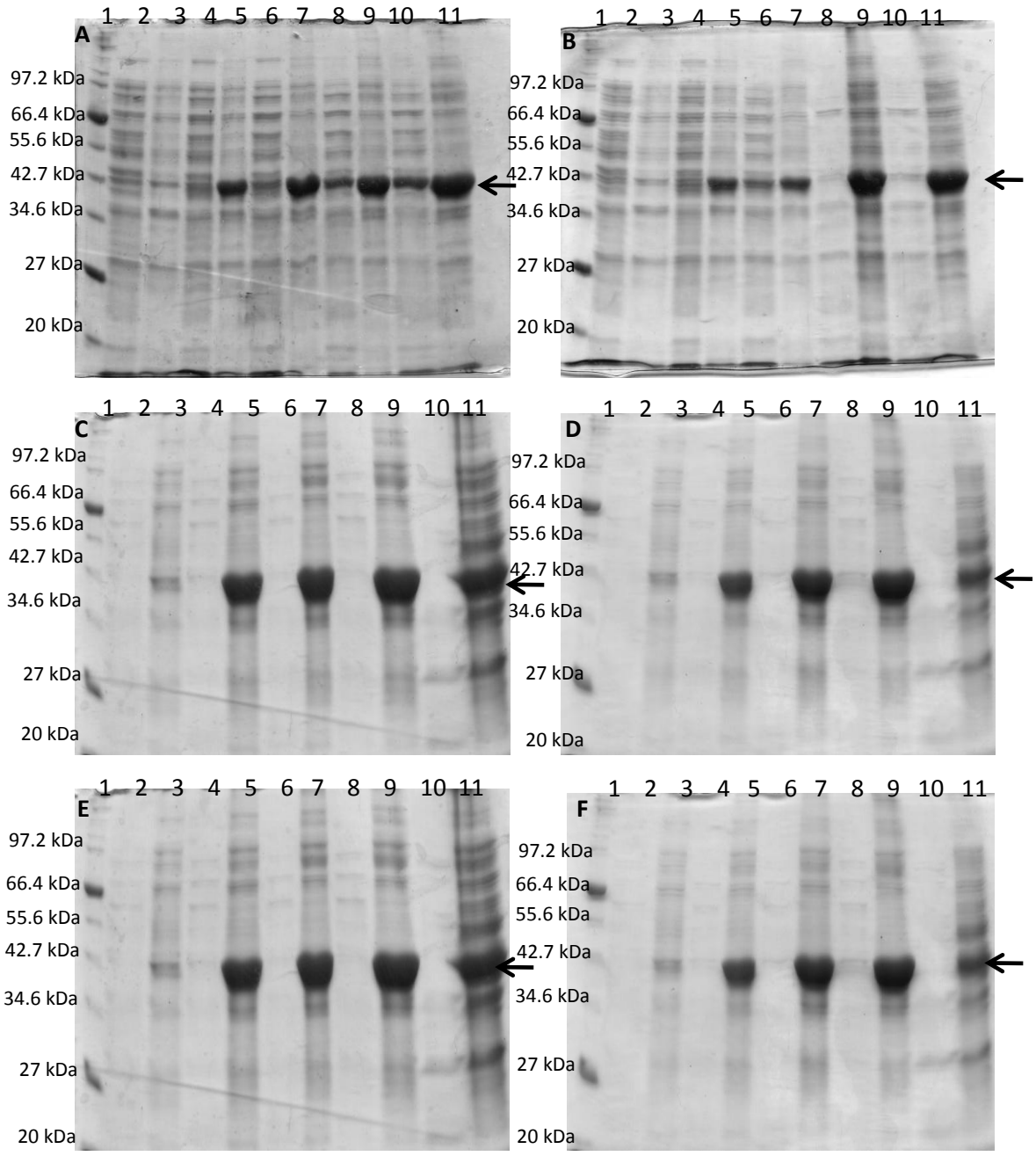


Fig 5.10: Analysis of the expression of CbpD under varying temperature and induction conditions. 12.5% SDS-PAGE analysis of the expression of CbpD from pCbpD_60, **A;** 37°C incubation, 50 mM IPTG, **B**; 30°C incubation, 50mM IPTG, **C;** 37°C incubation, 25mM IPTG, **D;** 37°C incubation, 10 mM IPTG, **E;** 30°C incubation, 25mM IPTG, **F;** 30°C incubation, 10 mM IPTG Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; soluble 0 hours, Lane 3; insoluble 0 hours, Lane 4; soluble 2 hours, Lane 5; insoluble 2 hours, Lane 6; soluble 4 hours, Lane 7; insoluble 4 hours, Lane 8, soluble 6 hours, Lane 9; insoluble 6 hours, Lane 10; soluble o/n, Lane 11; insoluble o/n.

5.3 Sub-cloning of CbpD

5.3.1 Cloning of untagged CbpD

As the position of additional tags can effect protein solubility CbpD was sub-cloned without the C-terminal $(His)_6$ tag to observe the effect, if any of the affinity tag. To create untagged CBPD the *cbpd* gene was amplified using Phusion directed whole vector amplification (see Fig 3.7), pCbpD_60 was used as the template. The primers used for this amplification were pQE60_F1 and Pseudo-R6 (Table 2.3). A stop codon was included in the reverse primer. Once incorporated into the gene this stop codon would stop transcription before the $(His)_6$ tag located at the 3' end of the gene (see Fig 5.11). When expressed in *E. coli* untagged CbpD still associated with the insoluble protein fraction (see Fig 5.12).

ATAAATAGAT TCAATTGTGA GCGGATAACA ATTTCACACA GAATTCATTA AAGAGGAGAA

ATTAA**C**CATG GGAAAACACT ACTCAGCCAC CCTGGCACTC CTGCCACTCA CCCTCGCCCT

GTTCCTGCCC CAGGCAGCCC ATGCCCACGG CTCGATGGAA ACGCCGCCCA GTCGGGTCTA

CGGCTGCTTC CTCGAAGGTC CGGAGAATCC CAAGTCGGCC GCCTGCAAGG CCGCCGTCGC

CGCCGGCGGC ACCCAGGCAC TGTACGACTG GAATGGCGTC AACCAGGGCA ACGCCAACGG

CAACCACCAG GCGGTGGTCC CCGACGGCCA GCTCTGCGGC GCCGGCAAGG CACTGTTCAA

GGGCCTGAAC CTGGCTCGCA GCGACTGGCC CAGCACTGCC ATCGCGCCGG ACGCCAGCGG

CAACTTCCAG TTCGTCTACA AGGCCAGCGC GCCGCACGCG ACCCGCTACT TCGACTTCTA

CATCACCAAG GACGGCTATA ACCCCGAGAA GCCGCTGGCC TGGAGCGACC TGGAACCCGC

GCCGTTCTGC TCGATCACCA GCGTCAAGCT GGAGAACGGC ACCTACCGGA TGAACTGCCC

GCTGCCCCAG GGCAAGACCG GCAAGCATGT GATCTATAAC GTCTGGCAGC GCTCGGACAG

CCCGGAAGCC TTCTACGCCT GCATCGACGT GAGCTTCAGC GGCGCCGTCG CCAACCCCTG

GCAAGCGCTG GGCAACCTGC GCGCGCAGCA GGACCTGCCA GCCGGTGCTA CCGTCACCCT

GCGTCTGTTC GATGCCCAGG GCCGCGACGC CCAGCGTCAC AGCCTGACCC TGGCCCAGGG

CGCCAACGGT GCCAAGCAAT GGCCGCTGGC GCTGGCGCAG AAGGTCAACC AGGACTCCAC

CCTGGTCAAC ATCGGCGTGC TGGATGCCTA CGGGGCGGTC AGCCCGGTGG CCAGCTCGCA

GGACAACCAG GTCTACGTGC GCCAGGCCGG CTACCGCTTC CAGGTCGACA TCGAACTGCC

GGTCGAGGGC GGCGGCGAGC AACCGGGCGG CGACGGCAAG GTCGACTTCG ACTATCCGCA

AGGCCTGCAG CAATACGACG CCGGGACCGT AGTGCGCGGT GCCGATGGCA AGCGCTACCA

GTGCAAGCCC TACCCGAACT CCGGCTGGTG CAAGGGCTGG GACCTCTACT ACGCCCCGGG

ACCGTCC TGCGGACCTG GGACGACTAA AGATCTCAT CACCATCACC

CAAGGGCATG GCCTGGCAGG ACGCCTGGAC CCTGCTG AGATCTCAT CACCATCACC

AT

ATCACTAAGC TTAATTAGTG AGCTTGGACT CCTGTTGATA GATCCAGTAA TTAGTGAGCT

Fig 5.11: Cloning of untagged CbpD. Amplification of untagged CbpD from pCBPD_60. The forward primer, pQE60_F1, is shown in red, and the reverse primer, Pseudo R6, is shown in blue. The nucleotides encoding the poly histidine tag are underlined in purple, with the nucleotides (stop codon) to be introduced on the forward primer underlined in red.

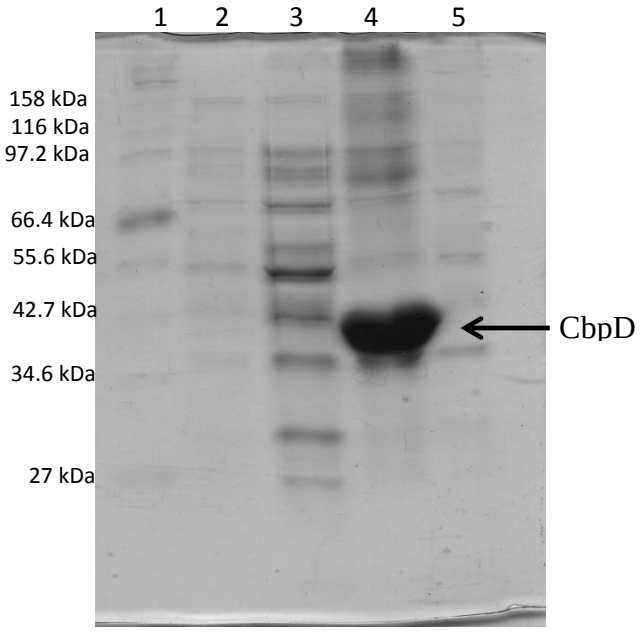


Fig 5.12: Analysis of the expression of untagged CbpD. 12.5% SDS-PAGE analysis of the expression of CbpD from pCbpD_UN in *E. coli* KRX. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; pCbpD_UN soluble fraction, Lane 3; pQE60 empty soluble fraction, Lane 4; pCbpD_UN insoluble fraction, Lane 5; pQE60 empty insoluble fraction.

5.3.2 Cloning of CbpD with alternative N-terminal signal sequences

One of the overall goals in the production of recombinant proteins is to simultaneously reach a high specific recombinant protein production rate and a high product quality. Large scale protein recovery from insoluble inclusion bodies is not favourable. It is both costly and time consuming and the isolation and re-folding of insoluble proteins has been shown to impact on a proteins activity (Sun, *et al.*, 2011). While reasons for the formation of protein inclusion bodies are plentiful it can sometimes be due to the malformation of disulfide bonds. Many cytoplasmic proteins have free cysteines that cannot form disulfide bonds in the reducing environment of the cytoplasm (Pugsley 1993). A strategy that can be employed to avoid inclusion body formation is to target a protein to an outer compartment of the host cell, where a less reducing environment exists (Shokri, *et al.*, 2003).

There are three major pathways for the translocation of polypeptides across the cytoplasmic membrane in gram negative bacteria; the Sec pathway (Pugsley, 1993, Fekkes and Driessen, 1999, Stephenson, 2005), the signal recognition particle (SRP) pathway (Steiner *et al.*, 2006, Keenan, *et al.*, 2001), and the Tat pathway (Berks *et al.*, 2000). The Sec and SRP translocation systems are responsible for transmembrane transport of unfolded proteins whereas the Tat system is used by folded proteins. The transport of proteins through the inner membrane of *E. coli* can be either post-translational (Sec) or co-translational (SRP) (Fekkes and Driessen, 1999, Koch *et al.*, 2003) (see Fig 5.13).

Transport of polypeptides via these pathways is achieved by the use of N-terminal signal sequences. Whether a signal targets its protein to SRP or to SecA depends on the degree of hydrophobicity of the core region of the signal sequence; the highly hydrophobic inner membrane protein signals associate with SRP, while the less hydrophobic signals utilize SecA (Lee and Bernstein, 2001). Two of the most commonly used signal sequences are the PelBss from the pectate lyase protein PelB from *Erwinia carotovora* (PelBss) (Lei *et al.*, 1987, Khushoo *et al.*, 2004, Johnson, *et al.*, 1996, Shevchik *et al.*, 1997) and DsbAss from DsbA, a periplasmic enzyme involved in protein disulfide bond formation (Zhang and Donnenberg, 1996, Bardwell *et al.*, 1991, Jeong and Lee, 2000). PelBss directs proteins via the Sec based system while DsbAss directs proteins via the SRP system.

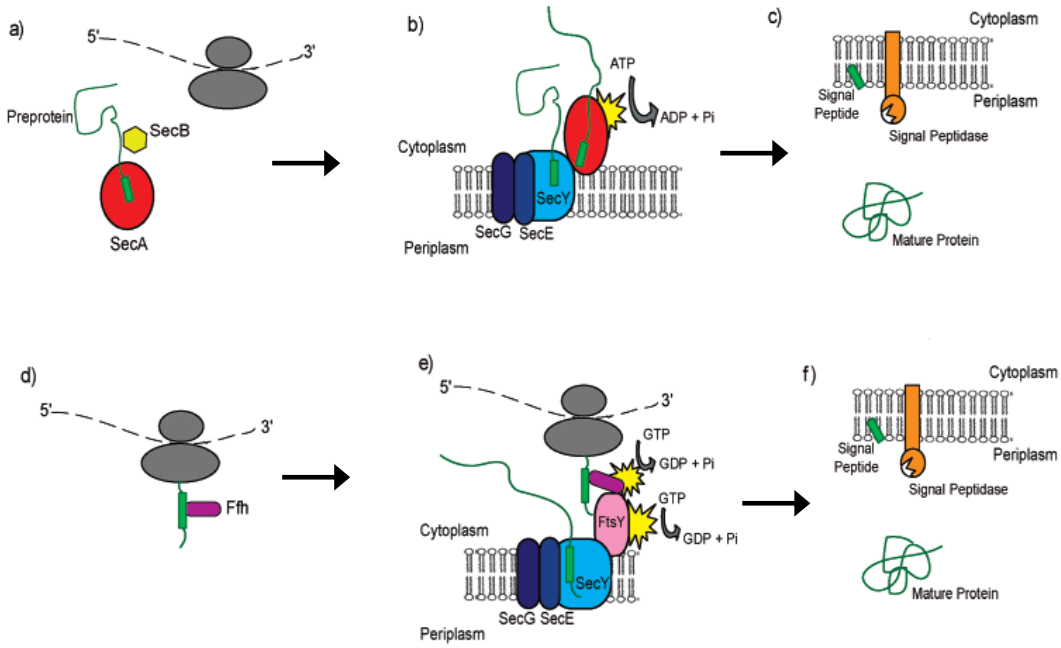


Fig 5.13: Schematic representation of single peptide interactions of the Sec system. Signal peptides of secretory proteins are bound by SecA (a) in the cytoplasm, delivered to, bound by, and translocated via SecY of the pore (b), and ultimately bind to signal peptidase for cleavage of the signal from the mature proteins (c). Inner membrane proteins signals are bound by SRP as they emerge from the ribosome (d) and targeted to the membrane. The signal anchor interacts with SecY (e) and translocation occurs via the membrane-embedded SecYEG channel followed by signal peptidase cleavage and folding of the mature protein (f). Image adapted from (Rusch and Kendall, 2007).

CbpD was amplified from pCbpD_60 using primers Pseudo-F7 and Psuedo-R2. The PCR products were analysed by agarose gel electrophoresis and a band corresponding to the expected size was present. The PCR product was subjected to restriction analysis with *Not*I and *Bgl*II, then ligated (Section 2.10.2) into *Not*I *Bgl*II restricted pPelB and pDsbA to create pPelB_CbpD and pDsbA_CbpD respectively (see Fig 5.14). The ligated vectors were transformed (Section 2.9.2) into *E. coli* XL10 Gold cells and the plasmid DNA obtained from some of the resulting colonies was screened for inserts using gel electrophoresis and restriction analysis. The presence of the correct inserts was further confirmed by DNA sequencing (Section 2.12). Time course analysis was undertaken to establish the solubility of the recombinant proteins in *E. coli* (see Fig 5.16 and

5.17). It was observed that CbpD expressed with PelBss and DsbAss associated with the insoluble protein fraction.

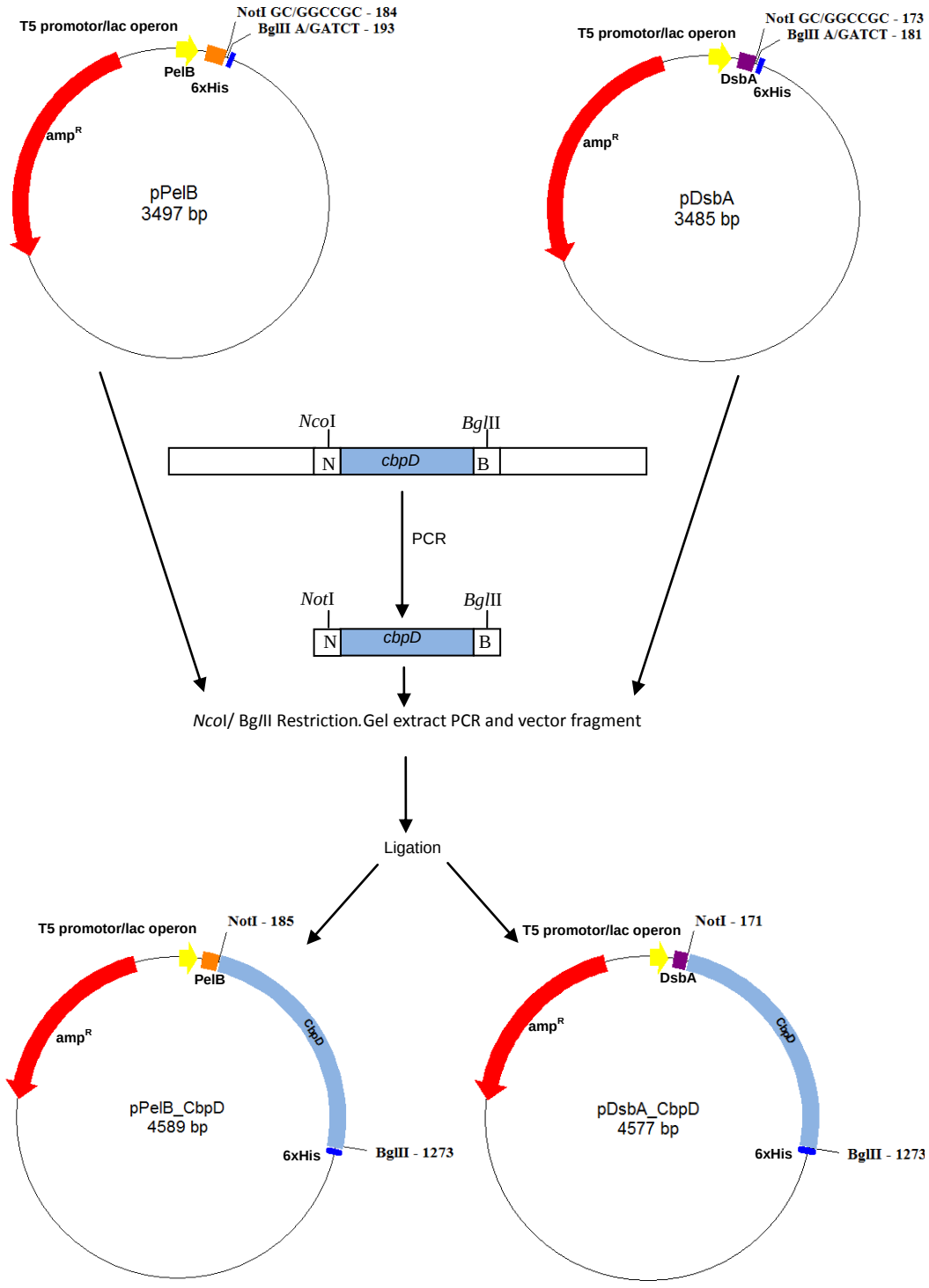


Fig 5.14: Schematic of the cloning of CbpD into pPelB and pDsbA. CbpD was cloned into the pDsbA and pPelB vectors as *Not*I *Bgl*II fragments to yield the vectors pDsbA_CbpD and pPelB_CbpD respectively.

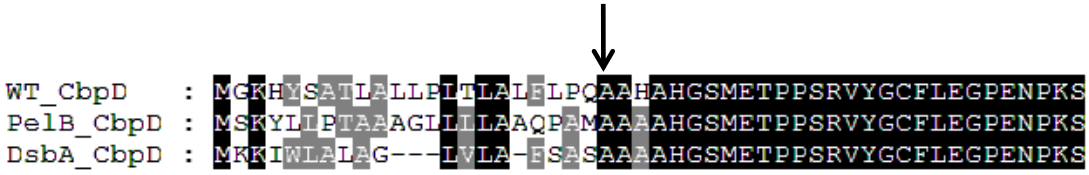


Fig 5.15: Amino acid alignment of the N-terminal domains of CbpD with WT, PelB and DsbA signal sequences. The signal sequences begin with the first residue in the alignment, the arrow indicates the first residue of the mature protein. This alignment was generated using ClustalW and Gendoc 2.6.

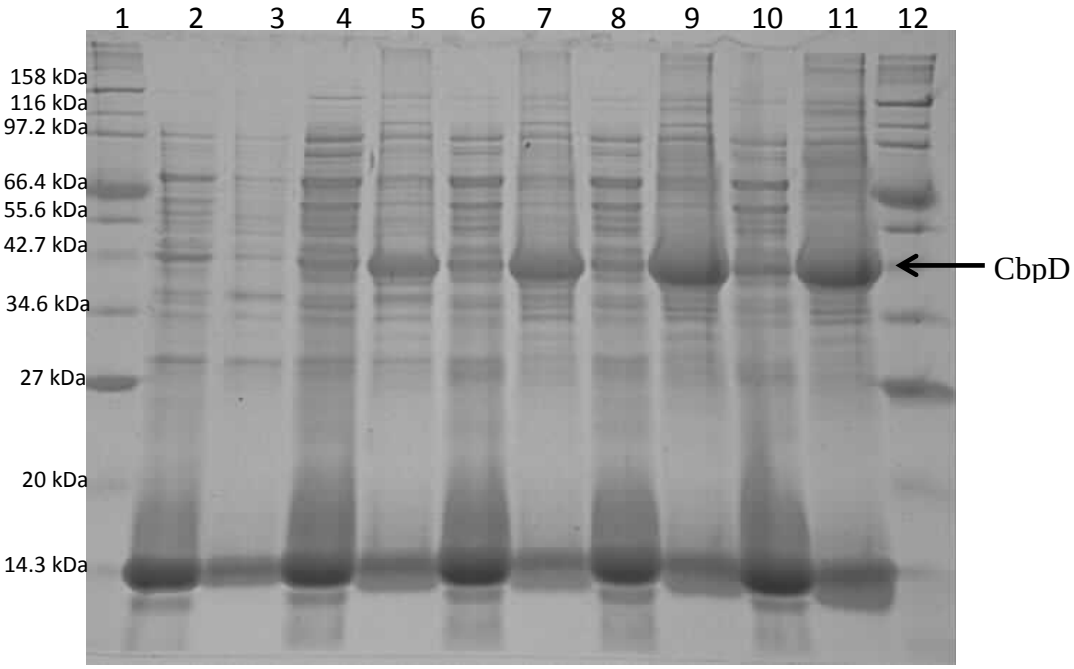


Fig 5.16: Time course expression analysis of CbpD expressed with PelBss. Analysis by 12.5% SDS-PAGE of the expression of CbpD expressed from pCbpD_PelB. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; soluble 0 hours, Lane 3, insoluble 0 hours, Lane 4; soluble 1 hour, Lane 5; insoluble 1 hour, Lane 6; soluble 2 hours, Lane 7; insoluble 2 hours, Lane 8; soluble 3 hours, Lane 9; insoluble 3 hours, Lane 10, soluble 4 hours, Lane 11; insoluble 4 hours, Lane 12; Broad range protein ladder.

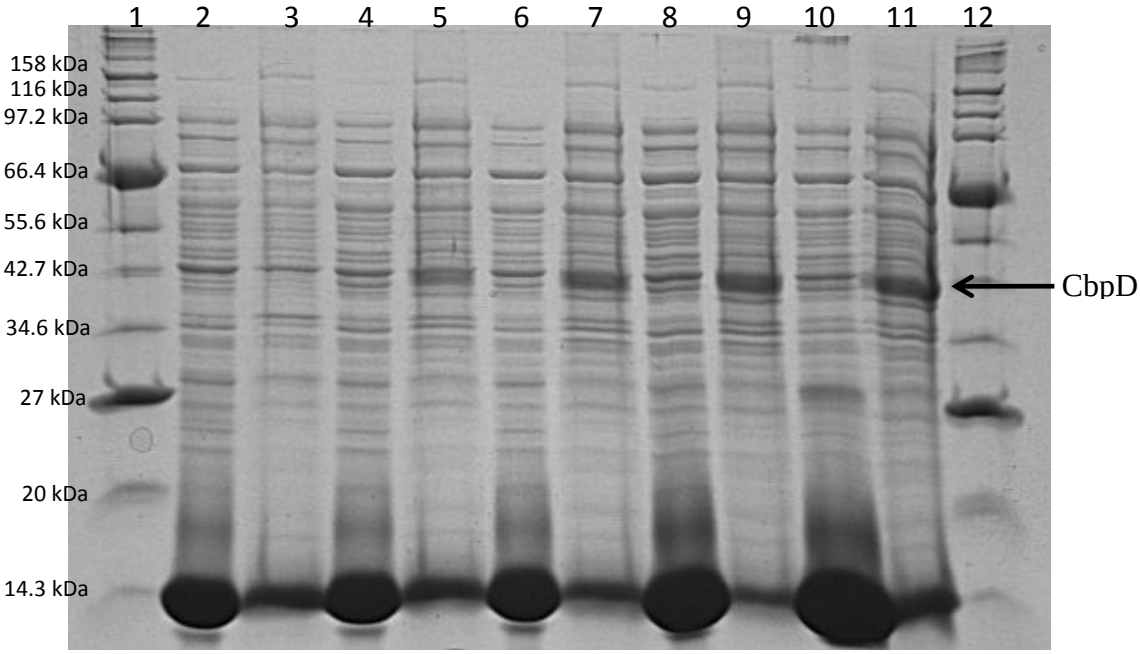


Fig 5.17: Time course expression analysis of CbpD expressed with DsbAss. Analysis by 12.5% SDS-PAGE of the expression of CbpD expressed from pCbpD_DsbA. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; soluble 0 hours, Lane 3, insoluble 0 hours, Lane 4; soluble 1 hour, Lane 5; insoluble 1 hour, Lane 6; soluble 2 hours, Lane 7; insoluble 2 hours, Lane 8; soluble 3 hours, Lane 9; insoluble 3 hours, Lane 10, soluble 4 hours, Lane 11; insoluble 4 hours, Lane 12; Broad range protein ladder.

5.3.3 Cloning of CbpD sub-domains

In an effort to improve protein solubility and to eludicate if both putative domains are chitin-binding domains a number of CbpD sub-clones were constructed. These sub-clones consisted of amino acid residues 1-212, 212-389 and 117-389 and were named CbpDa, CbpDb and CbpDc respectively. The cloning of these gene fragments is outlined in Fig 5.18; CbpDa was cloned using the primers Pseudo-F3 and Pseudo-R5, CbpDb was amplified using primers Psuedo-F8 and Pseudo-R2 and CbpDc was amplified using the primers Pseudo-F9 and Pseudo-R2. The PCR products were analysed by agarose gel electrophoresis and bands corresponding to the correct sizes were present. The PCR products were subjected to restriction with *Nco*I and *Bgl*II, then ligated (Section 2.10.2) into an *Nco*I *Bgl*II digested pQE60 vector to create pCbpDa_60, pCbpDb_60 and pCbpDc_60. *E. coli* XL10 Gold cells were transformed with the ligation mixture (Section 2.9.3). The plasmid DNA obtained from some of the resulting colonies was screened for inserts using gel electrophoresis and restriction analysis.

Once confirmed by sequencing each plasmid was transformed into *E. coli* KRX and expressed in a 100 mL culture (Section 2.14) to assess protein solubility. Clones CbpDa and CbpDc were found to express to the insoluble protein fraction, but a band corresponding to the size of CbpDb was visible in the soluble protein fraction following expression and western blot analysis (see Fig 5.20).

GTCAGTCCATGGGAAAACACTACTCAGCCACCC

ATGAAACACT ACTCAGCCAC CCTGGCACTC CTGCCACTCA CCCTCGCCCT
GTTCCTGCCC CAGGCAGCCC ATGCCCACGG CTCGATGGAA ACGCCGCCCA GTCGGGTCTA
CGGCTGCTTC CTCGAAGGTC CGGAGAATCC CAAGTCGGCC GCCTGCAAGG CCGCCGTCGC
CGCCGGCGGC ACCCAGGCAC TGTACGACTG GAATGGCGTC AACCAGGGCA ACGCCAACGG
CAACCACCAG GCGGTGGTCC CCGACGGCCA GCTCTGCGGC GCCGGCAAGG CACTGTTCAA

CAGTCCATGGGAAGCGG

GGGCCTGAAC CTGGCTCGCA GCGACTGGCC CAGCACTGCC ATCGCGCCGG ACGCCAGCGG
CAACTTCCAG TTCGTCTAC
CAACTTCCAG TTCGTCTACA AGGCCAGCGC GCCGCACGCG ACCCGCTACT TCGACTTCTA
CATCACCAAG GACGGCTATA ACCCCGAGAA GCCGCTGGCC TGGAGCGACC TGGAACCCGC
GCCGTTCTGC TCGATCACCA GCGTCAAGCT GGAGAACGGC ACCTACCGGA TGAACTGCCC
GCTGCCCCAG GGCAAGACCG GCAAGCATGT GATCTATAAC GTCTGGCAGC GCTCGGACAG

CTGCA CTCGAAGTCG CCGCGGTCTAGATGACTG

CAGTCCATGGGAGCCGTCGCCAACCCCTG

CCCGGAAGCC TTCTACGCCT GCATCGACGT GAGCTTCAGC GGCGCCGTCG CCAACCCCTG
G
GCAAGCGCTG GGCAACCTGC GCGCGCAGCA GGACCTGCCA GCCGGTGCTA CCGTCACCCT
GCGTCTGTTC GATGCCCAGG GCCGCGACGC CCAGCGTCAC AGCCTGACCC TGGCCCAGGG
CGCCAACGGT GCCAAGCAAT GGCCGCTGGC GCTGGCGCAG AAGGTCAACC AGGACTCCAC
CCTGGTCAAC ATCGGCGTGC TGGATGCCTA CGGGGCGGTC AGCCCGGTGG CCAGCTCGCA
GGACAACCAG GTCTACGTGC GCCAGGCCGG CTACCGCTTC CAGGTCGACA TCGAACTGCC
GGTCGAGGGC GGCGGCGAGC AACCGGGCGG CGACGGCAAG GTCGACTTCG ACTATCCGCA
AGGCCTGCAG CAATACGACG CCGGGACCGT AGTGCGCGGT GCCGATGGCA AGCGCTACCA
GTGCAAGCCC TACCCGAACT CCGGCTGGTG CAAGGGCTGG GACCTCTACT ACGCCCCGGG

CC TGCGGACCTG GGACGACTCTAGATGACTG

CAAGGGCATG GCCTGGCAGG ACGCCTGGAC CCTGCTGTAA

Fig 5.18: The cloning of CbpD segments CbpDa, CbpDb and CbpDc. CbpDa was cloned using the primers Pseudo-F3 and Pseudo-R5, CbpDb was amplified using primers Psuedo-F8 and Pseudo-R2 and Cbpc was amplified using the primers Pseudo-F9 and Pseudo-R2. Forward primers are highlighted in red with reverse primers highlighted in blue.

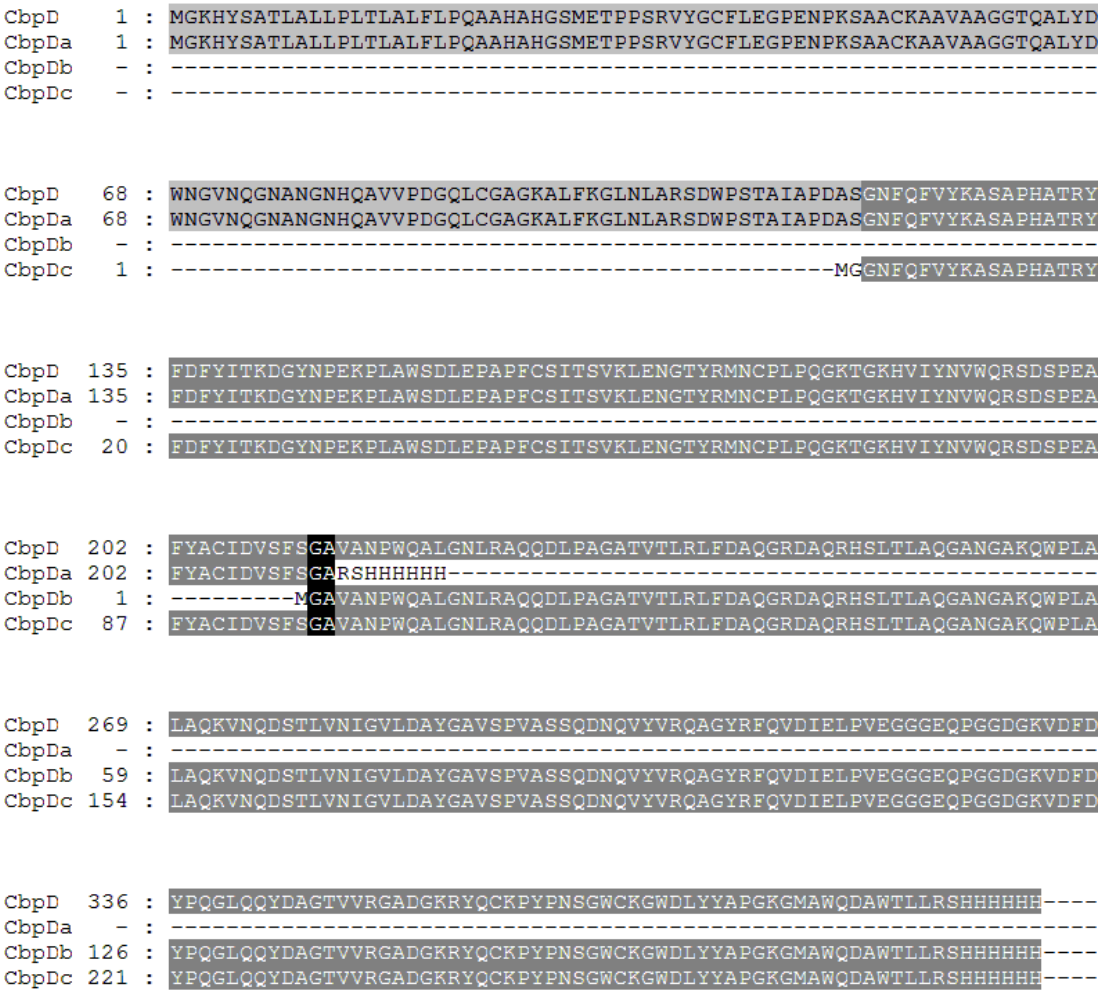


Fig 5.19: Amino acid alignment of CbpD and sub-clones CbpDa, CbpDb and CbpDc. This image was generated using ClustalW2 and Gendoc 2.7.

Table 5.2: Properties of CbpD sub-clones CbpDa, CbpDb and CbpDc

Protein	CbpD residues	Amino Acid residues	Expected protein size (Da)
CbpDa	1-212	221	23886
CbpDb	212-389	188	20496
CbpDc	117-389	283	31260

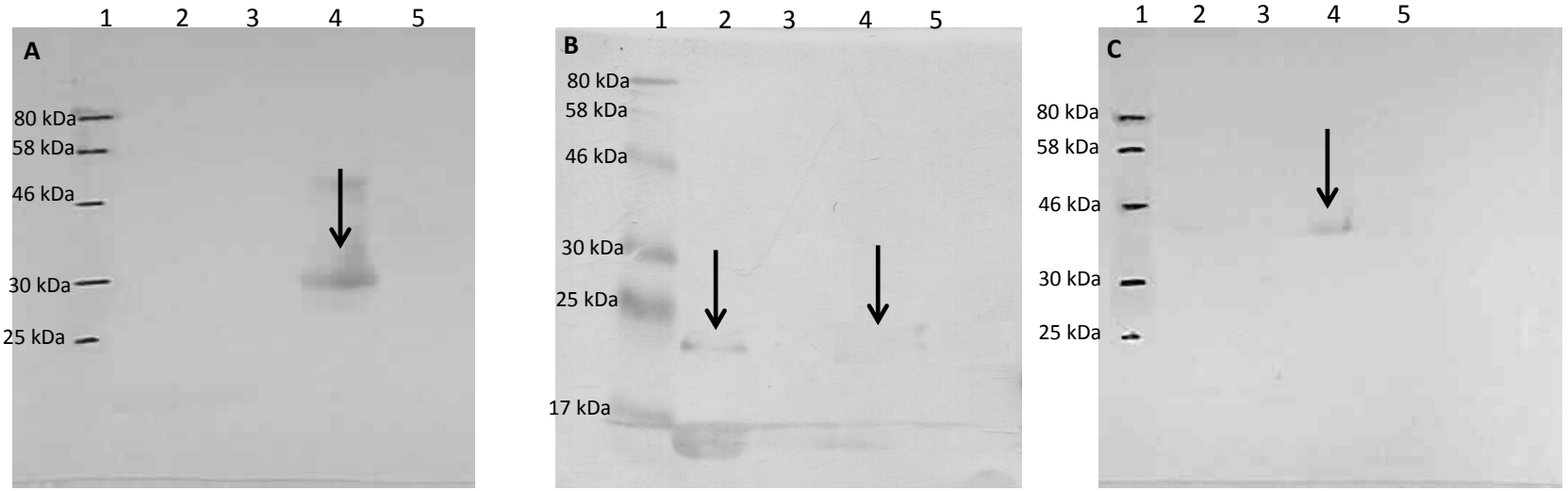


Fig 5.20: Western blot analysis of A; CbpDa, B; CbpDb and C; CbpDc expressed in *E. coli* XL-10 Gold. Lane 1; Pre-stained protein ladder (Section 2.23.3), Lane 2; expressed protein soluble fraction, Lane 3; empty vector soluble fraction, Lane 4; expressed protein insoluble fraction, Lane 5; empty vector insoluble fraction. $(His)_6$ positive bands are indicated with arrows.

5.4 Purification of CBP21 homologues

5.4.1 Solubilisation, refolding and purification of CbpD, CbpA and CbpL

Recombinant protein in the form of inclusion bodies lacks any biological activity. While the expression of a protein to inclusion bodies is not ideal protein isolation and solubilisation is possible. The isolation of proteins from inclusion bodies does not usually give full recovery of biological activity, making the preparation of inclusion bodies of significant importance. Detergents such as tween or triton are often used to break down the inclusion bodies, while common denaturants such as urea and guanidine hydrochloride are used to solubilise the protein inclusion bodies prior to protein refolding. It is important that the denaturing agents are removed before refolding to allow for the formation of the correct intramolecular associations. There are a number of methods that can be used for protein refolding, including dialysis, dilution, gel filtration and on-column re-folding.

Dialysis is one of the most common techniques to aid protein refolding (Misawa and Kumagai, 1999), however it takes several days and uses large volumes of buffer. Dilution is another simple technique, but again it is quite slow and cumbersome. It also necessitates extensive dilution which is not beneficial to protein recovery. Gel filtration is another useful technique that allows for the separation of high concentrations of native protein from the aggregated material. The aggregated material can however be difficult to remove from the often expensive columns and this method can be quite slow despite the use of small volumes. On-column refolding has come to the forefront over the past decade as the foremost techniques for protein folding. It merits less time, high concentrations of product, high biological activity and easy scale up (Sun *et al.*, 2011, Middelberg 2002). In this instance solubilised protein was immobilised on an affinity matrix, a urea gradient was used to gently refold the protein on the column, thus helping it to refold into its natural structure.

CbpD, CbpA and CbpL were isolated from inclusion bodies as outlined in section 2.15.3 and subjected to on-column refolding using the FPLC procedure described in section 2.18.3. Satisfactory solubilisation of the inclusion bodies was assessed by looking at the clarity of the pellet following the addition of denaturant (see Fig 5.21). On-line monitoring using the AKTA FPLC system revealed that the target protein was eluting from the column in the correct elution stage (see Fig 5.22, and 5.23), although the height of the elution peaks were quite low and subsequent quantification revealed low protein concentrations (Table 5.3). SDS-PAGE analysis was used to gauge the size of the refolded proteins and the success of the purification (see Fig 5.24).

Surprisingly analysis of the imidazole elution fractions following CbpD purification revealed that although there was a low but sharp protein elution peak upon initial application of the imidazole elution buffer the target protein continued to leach from the column beyond 74 mL of buffer application (see Fig 5.25). It was theorized that the proteins could be associating with the column material. The purification was repeated using an elution buffer of pH 4.0 and an elution buffer containing GlcNAC to see if the elution of CbpD could be improved, but did not appear successful (see Fig 5.26).

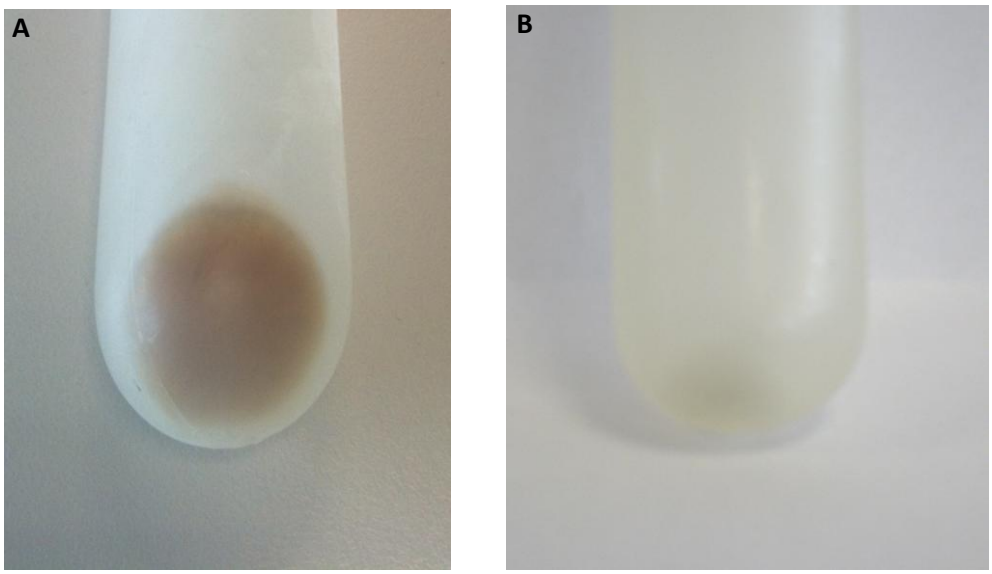


Fig 5.21: Solubilisation of inclusion bodies from a pellet of *E. coli* cells expressing CbpD. A; CbpD insoluble cell pellet, B; solubilised CbpD pellet. The success of protein solubilisation can be measured by the clarity of the cell pellet.

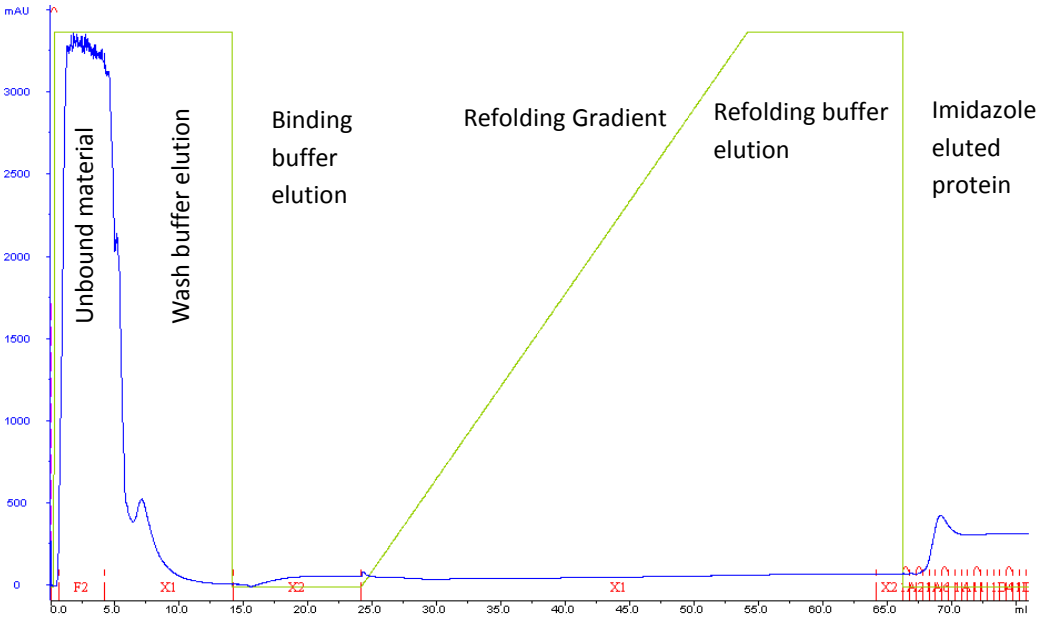


Fig 5.22: On-column refolding and purification of CbpD. Refolding and purification of resolubilised CbpD from a 400 mL *E. coli* expression culture on an Amersham HisTrap FF crude 1 mL column, using FPLC. The blue line indicates the absorbance at A_{280}, the green line indicates the percentage of buffer B in the running buffer, while the red markers signify the fractionation of the eluent.

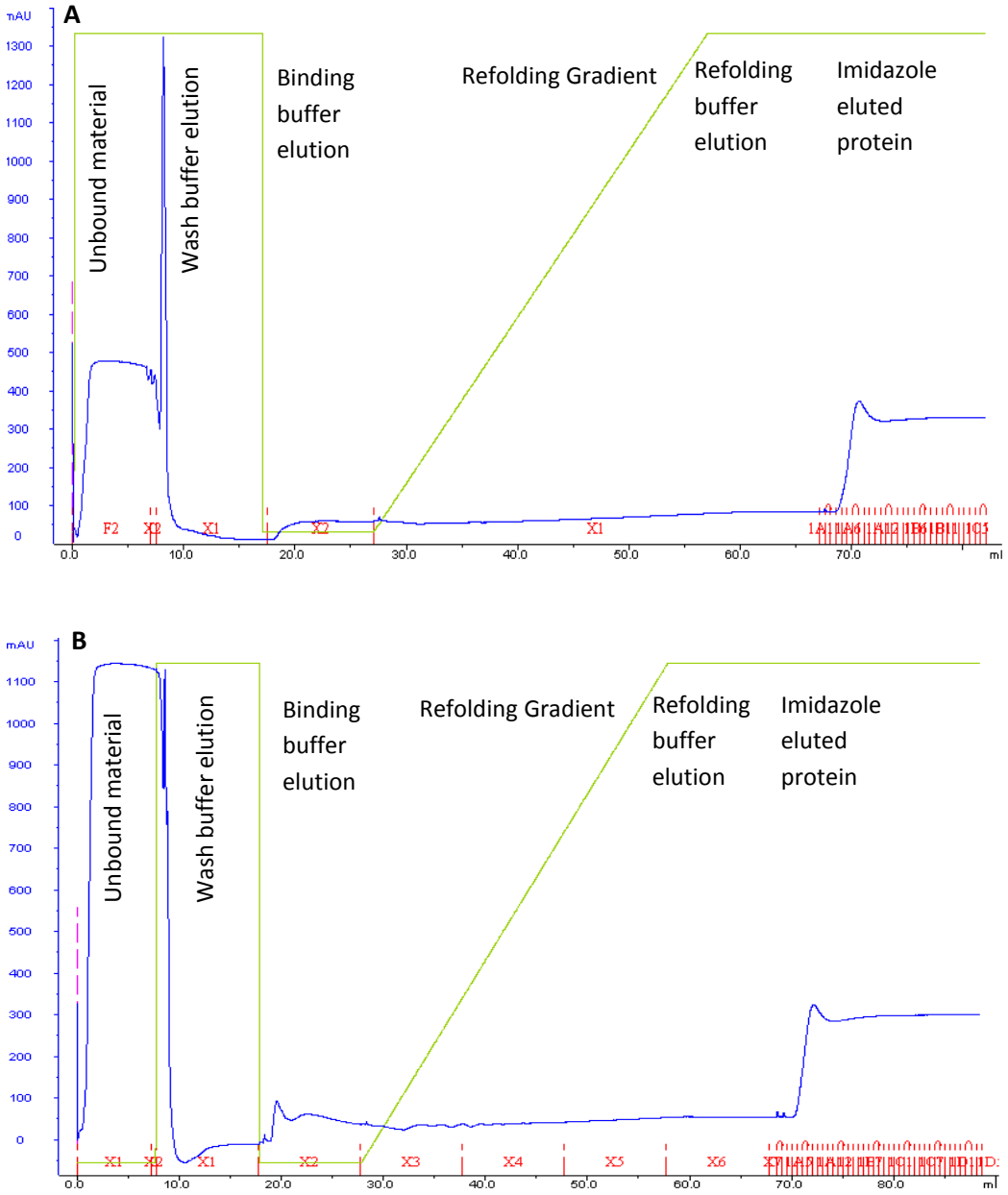


Fig 5.23: On-column refolding and purification of A; CbpA and B; CbpL. Refolding and purification of resolubilised CbpA and CbpL from a 400 mL *E. coli* expression culture on an Amersham HisTrap FF crude 1 mL column, using FPLC. The blue line indicates the absorbance at A_{280}, the green line indicates the percentage of buffer B in the running buffer, while the red markers signify the fractionation of the eluent.

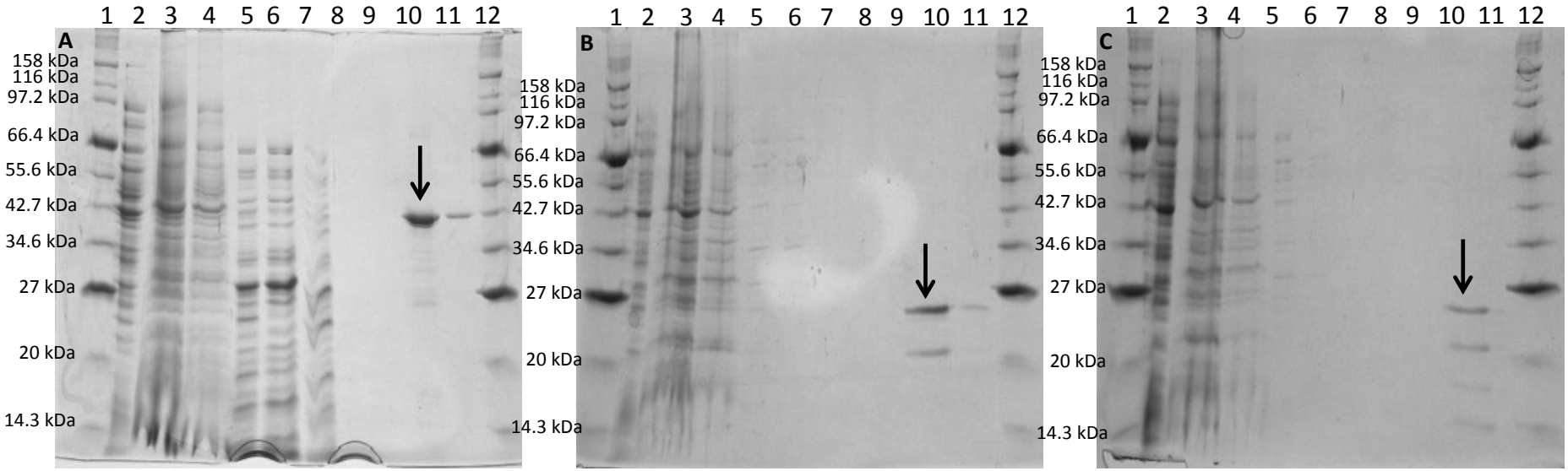


Fig 5.24: Analysis of fractions from the insoluble purification of A; CbpD, B; CbpA and C; CbpL. 12.5% SDS-PAGE analysis of the fractions resulting from the on column refolding and purification of CbpD, CbpA and CbpL using FPLC. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Spin 1 supernatant, Lane 3; Spin 2 supernatant, Lane 4; Spin 3 supernatant, Lane 5; Column flow-through, Lane 6; Binding buffer elution, Lane 7; Wash buffer elution , Lane 8; Refolding gradient elution, Lane 9; Refolding buffer elution, Lane 10; Pooled elutions 1A4 – 1A11 , Lane 11; 1B4 , Lane 12; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

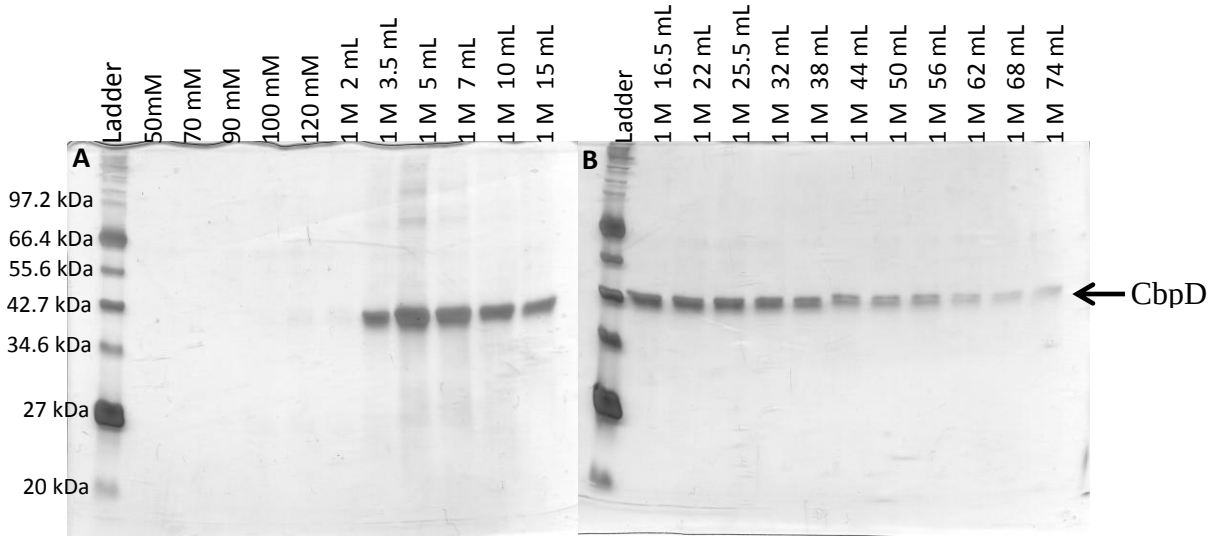


Fig 5.25: Analysis of the elution fractions from purification of CbpD from insoluble cell pellet. 12.5% SDS-PAGE and silver stain analysis of the elution fractions of CbpD from a 1 mL HisTrap FF column following on column refolding and purification of solubilised CbpD. The ladder refers to the broad range protein ladder (Section 2.23.3). Imidazole wash concentrations and elution volumes are indicated.

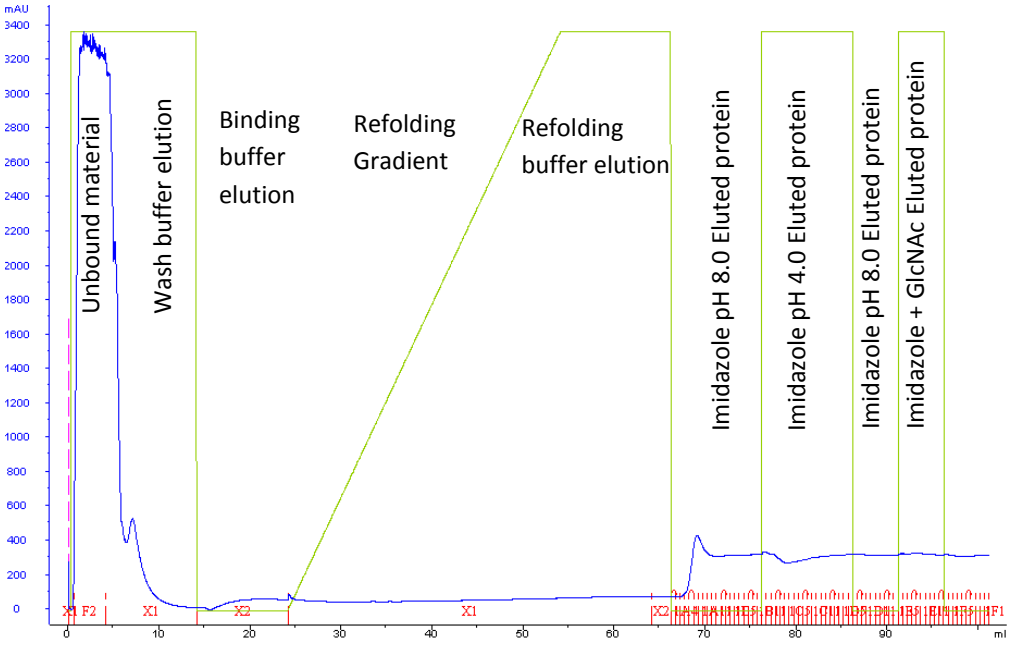


Fig 5.26: On-column refolding and purification of CbpD. Refolding and purification of resolubilised CbpD from a 400 mL *E. coli* expression culture on an Amersham HisTrap FF crude 1 mL column, using FPLC. The blue line indicates the absorbance at A_{280}, the green line indicates the percentage of buffer B in the running buffer, while the red markers signify the fractionation of the eluent.

Table 5.3: Average yield of recombinant proteins CbpD, CbpA and CbpL following expression from a 400 mL *E. coli* culture and on-column refolding

Protein	Mass of cells (g)	Total protein yield (mg)	mg protein/ g cell paste
CbpD	2.58	0.877	0. 34
CbpA	1. 34	0.185	0.138
CbpL	1 .42	0.372	0.262

5.4.2 IMAC purification of CbpDb

Previous western blot analysis of CbpDb containing lysate revealed that the protein was present in the soluble protein fraction. The protein was subsequently expressed as outlined in section 2.14, subjected to cell lysis (Section 2.15) and purified from the soluble cell lysate using IMAC as outlined in section 2.18.1. SDS-PAGE analysis of the elution fractions revealed that the protein began to elute from the column at an imidazole concentration of 200 mM and purified to homogeneity (see Fig 5.27). The average yield of CbpDd following expression and IMAC purification was 0.07 mg/g of cells.

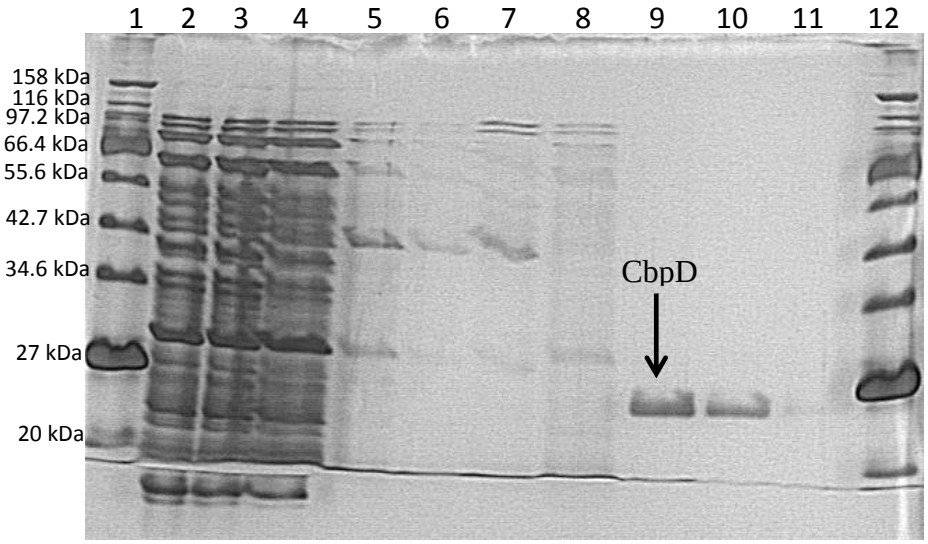


Fig 5.27: IMAC purification optimisation of CbpDb from soluble *E. coli* cell lysate. Analysis by 12.5% SDS-PAGE and silver staining of the elution fractions resulting from the IMAC purification CbpDb. Lane 1; Broad range protein ladder (2.23.3), Lane 2; Crude lysate, Lane 3; column flow-through, Lane 4; 20 mM imidazole elution 1, Lane 5; 20 mM imidazole elution 2, Lane 6; 20 mM imidazole elution 3, Lane 7; 70 mM imidazole elution 1, Lane 8; 70 mM imidazole elution 2, Lane 9; 200 mM imidazole elution, Lane 10; 200 mM imidazole elution 2, Lane 11; 200 mM imidazole elution 3, Lane 12; Broad range protein ladder.

5.5 Protein Characterisation

5.5.1 Insoluble substrate assays

The specificities of CbpD, CbpDb, CbpA and CbpL for various insoluble substrates are shown in Figs 5.28-5.31. CbpD bound α-chitin from crustacean shells, with no protein visible in flow-through and wash elution fractions, while in contrast to CBP21 equal amounts appeared to have bound to both chitosan and crystalline cellulose. CbpDb bound to α-chitin, with partial binding to crystalline cellulose, and no apparent binding to α-chitosan from crustacean shells. CbpA and CbpL had identical binding profiles. They bound to β- and α-chitin with similar affinities, both forms of chitosan, with faint protein bands visible in the unbound protein elution fraction, and to crystalline cellulose.

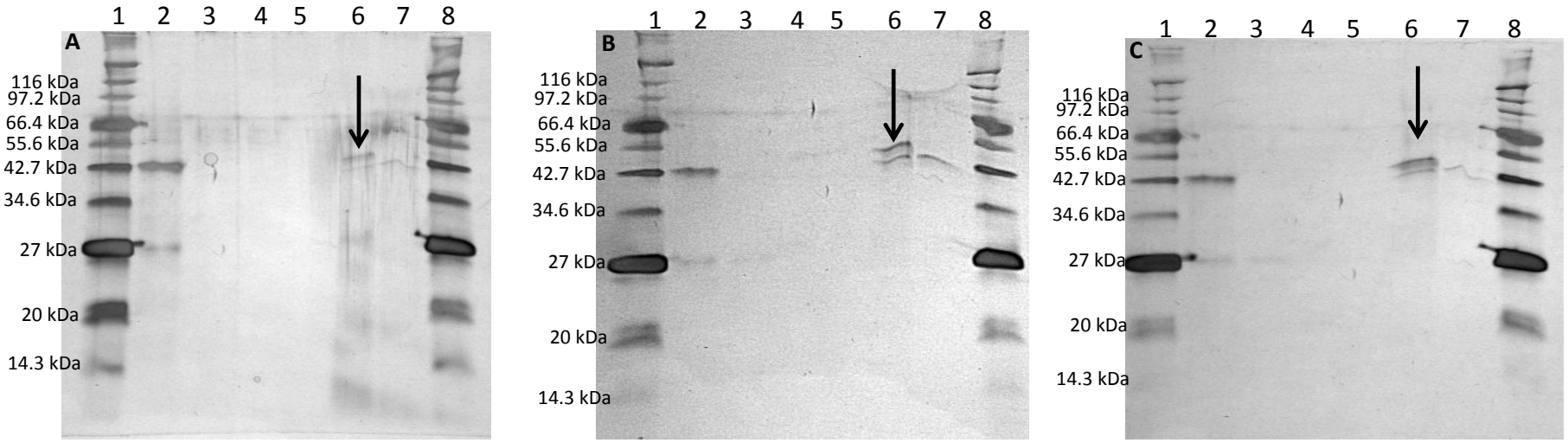


Fig 5.28: Analysis of CbpD bound to A; α-chitin, B; α-chitosan and C; Cellulose. 12.5% SDS-PAGE analysis of CbpD from *P. aeruginosa* bound to α-chitin and α-chitosin from crustacean shells and crystalline cellulose following incubation overnight, as outlined in section 2.26. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude CbpD sample, Lane 3; unbound CbpD, Lane 4; Wash 1 elution, Lane 5; Wash 2 elution, Lane 6; Boiled protein-substrate sample, Lane 7; Boiled PBS-substrate sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

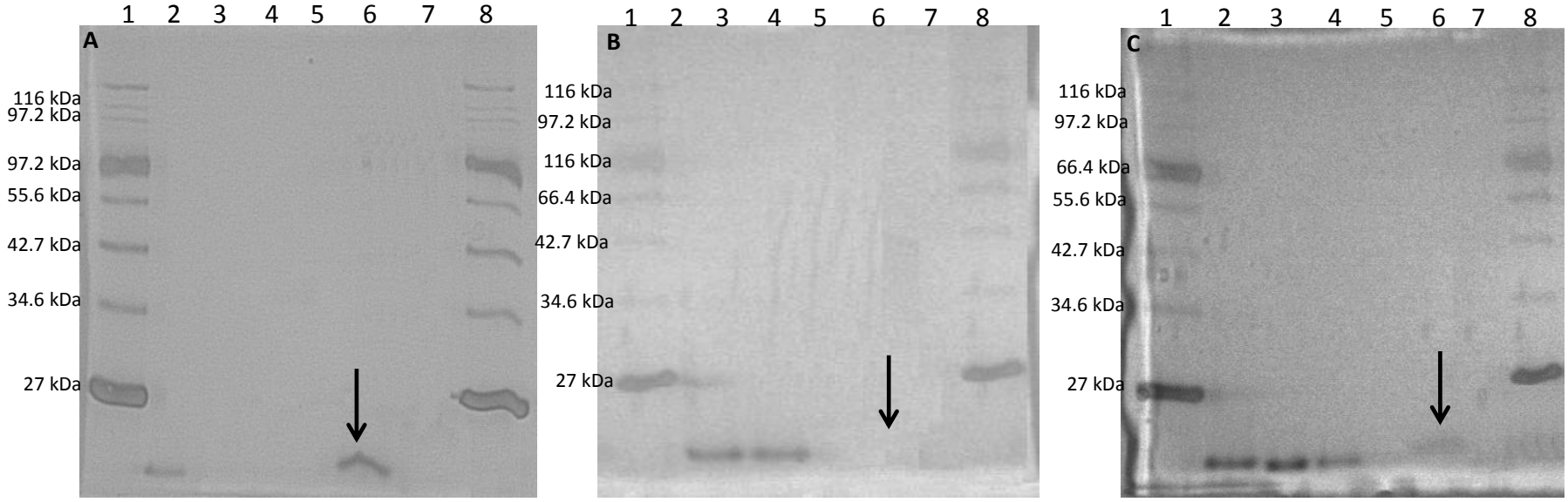


Fig 5.29: Analysis of CbpDb bound to A; α-chitin, B; α-chitosan and C; Cellulose. 12.5% SDS-PAGE analysis of CbpDb from *P. aeruginosa* bound to α-chitin and β-chitin from crustacean shells and crystalline cellulose following incubation overnight, as outlined in section 2.26. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude CbpD sample, Lane 3; unbound CbpD, Lane 4; Wash 1 elution, Lane 5; Wash 2 elution, Lane 6; Boiled protein-substrate sample, Lane 7; Boiled PBS-substrate sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

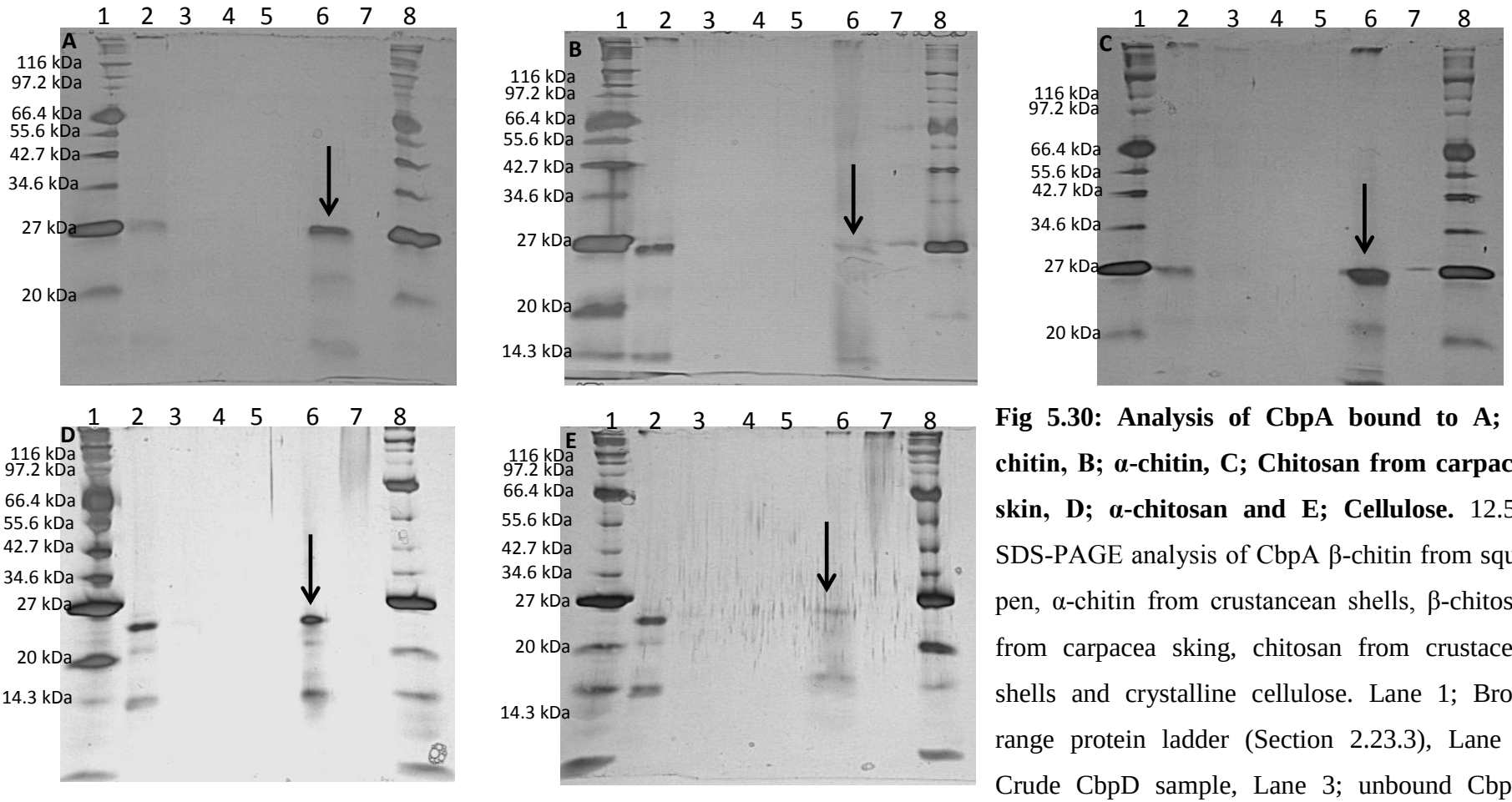


Fig 5.30: Analysis of CbpA bound to A; β-chitin, B; α-chitin, C; Chitosan from carpacea skin, D; α-chitosan and E; Cellulose. 12.5% SDS-PAGE analysis of CbpA β-chitin from squid pen, α-chitin from crustancean shells, β-chitosan from carpacea sking, chitosan from crustacean shells and crystalline cellulose. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude CbpD sample, Lane 3; unbound CbpD, Lane 4; Wash 1 elution, Lane 5; Wash 2 elution, Lane 6; Boiled protein-substrate sample, Lane 7; Boiled PBS-substrate sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

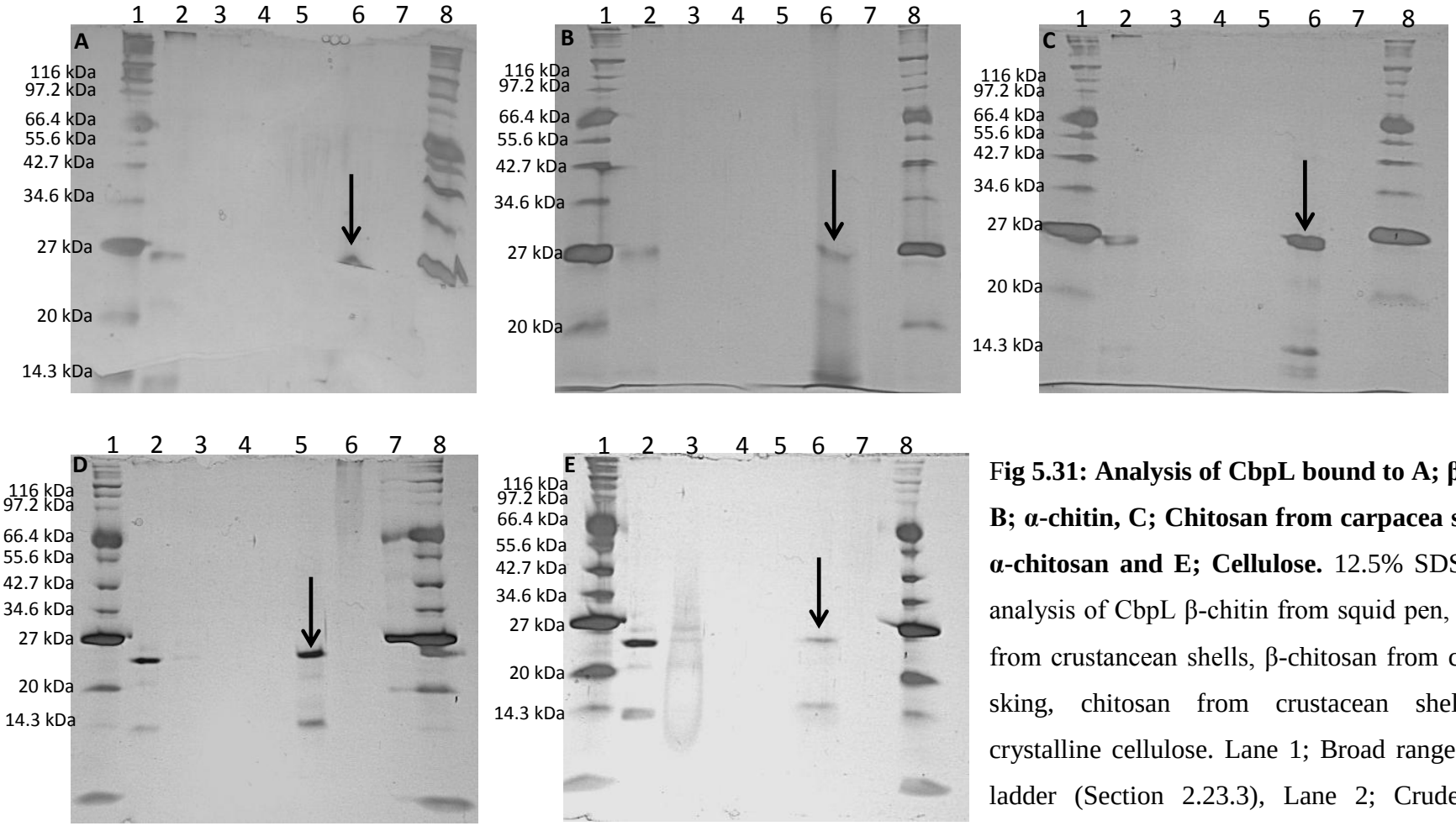


Fig 5.31: Analysis of CbpL bound to A; β-chitin, B; α-chitin, C; Chitosan from carpacea skin, D; α-chitosan and E; Cellulose. 12.5% SDS-PAGE analysis of CbpL β-chitin from squid pen, α-chitin from crustancean shells, β-chitosan from carpacea sking, chitosan from crustacean shells and crystalline cellulose. Lane 1; Broad range protein ladder (Section 2.23.3), Lane 2; Crude CbpD sample, Lane 3; unbound CbpD, Lane 4; Wash 1 elution, Lane 5; Wash 2 elution, Lane 6; Boiled protein-substrate sample, Lane 7; Boiled PBS-substrate sample, Lane 8; Broad range protein ladder. Substrate bound protein is indicated with an arrow.

5.5.2 ELLA analysis

ELLA analysis of CbpD, CbpA and CbpL using untreated, neuraminidase treated and β-galactosidase treated thyroglobulin revealed that none of the proteins were capable of binding to terminal or core GlcNAc in this format (see Fig 5.32).

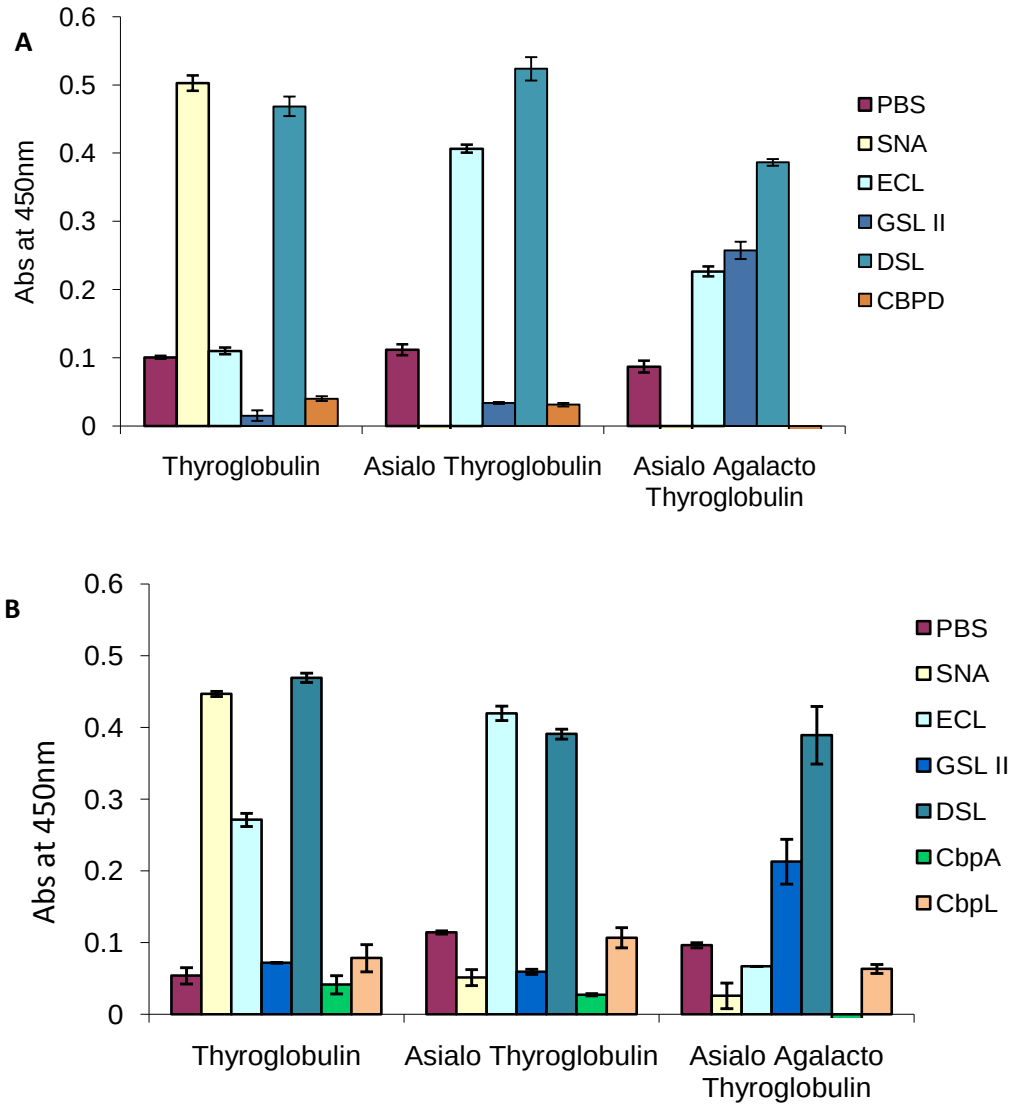


Fig 5.32: Quantitative detection of A; CbpD, B; CbpA and CbpL interaction with thyroglobulin. ELLA analysis of CbpA and CbpL and glycosidase treated thyroglobulin. SNA is a terminal sialic binding lectin, ECL has a binding preference for terminal galactose, GSL II is a terminal GlcNAc binding lectin, DSL has been characterised as a chitin binding lectin that appears to bind to core chitobiose.

5.6 Discussion

The CBP21 chitin-binding proteins homologues CbpD, CbpA and CbpL were successfully cloned, expressed and characterised, with regard to chitin binding. Although the proteins showed high levels of homology to CBP21, unlike CBP21 each protein associated with the insoluble protein fraction when expressed in *E. coli*. Solubilisation and refolding of the proteins was necessary for characterisation.

The genes *cbpd, cbpa* and *cbpl* were successfully amplified from *P. aeruginosa, P. asymbiotica and P. luminescens* genomic DNA respectively, and cloned into the commercial pQE vector from Qiagen (Section 5.2.1). The proteins were expressed from their respective constructs in *E. coli* with a C-terminal $(His)_6$ tag and over-expressed protein bands of the correct size were found to associate with the insoluble protein fraction. Examination of the CbpA and CbpL western blots revealed the presence of two $(His)_6$ positive bands in the insoluble fraction, around 21 kDa and 18 kDa in size, with the lower relative molecular mass band also visible in the soluble protein fraction. It is possible that both bands represent different forms of each protein. The size difference could potentially be due to proteolytic degradation but it is also possible that the lower molecular weight bands could represent CbpA and CbpL in which the N-terminal signal sequences have been proteolytically cleaved. Accumulation of over-expressed protein in this size range was not visible on SDS-PAGE analysis. Neither protein appeared to express particularly well in *E. coli* which suggested that the protein may have been slightly toxic, a commonly observed phenomenon in recombinant protein expression (Viaplana *et al.*, 1997). Protein expression could be improved through the use of an enriched medium (Khushoo *et al.*, 2004) or alternatively through the use of a strain with tight promoter control e.g. BL21 (DE3).

As stated previously there are a number of reasons for insoluble protein production; including, incorrect protein folding, incorrect formation of disulphide bonds and the presence of inherent hydrophobic regions. It was shown in section 3.2 that the N-terminal signal sequence of CBP21 was necessary for soluble protein expression. In the case of CbpD, CbpA and CbpL the most likely explanation for protein insolubility was that the presence of the native N-terminal leader sequences was not

promoting protein transport to the periplasmic space of the *E. coli* cell. From an evolutionary point of view it is plausible that *E. coli* is more closely related to *Serratia* spp. and therefore it is possible that *E. coli* incapable of recognising the N-terminal signal sequences of CbpD, CbpA and CbpL.

It was also indeterminate at this stage if the location of the poly-histidine tag, at the C-terminus, was affecting the ability of the proteins to fold properly. As CbpA and CbpL shared such high homology with CBP21 it was thought to be unlikely in their case. CbpD was cloned without the C-terminal $(His)_6$ tag and expressed in *E. coli*. It was observed that the protein still expressed to the insoluble protein fraction. This result ruled out the presence of the affinity tag as the cause of protein insolubility.

There are a number of strategies that can be employed to improve recombinant protein solubility, including reduction of the growth temperature, reduction of IPTG induction levels, secretion of the protein to the periplasm, the addition of fusion partners, use of alternative host strains (Zhang *et al.*, 2011, Hatahet *et al.*, 2010) and the expression of smaller soluble protein domains (Makrides, 1996). The expression of CbpD in *E. coli* was monitored at different growth temperatures at varying levels of IPTG induction. The effect of induction with varying levels of IPTG was investigated using SDS-PAGE, looking at the presence of soluble and insoluble product produced as certain stages of fermentation. It was observed that dramatically decreasing the IPTG concentration or lowering the growth temperature to 30°C did not result in increased levels of soluble recombinant protein production.

The four cysteine residues of CBP21 appear as two disulfide bridges (Vaaje-Kolstad *et al.*, 2005b). These cysteine residues are conserved in CbpA and CbpL whereas CbpD contains six cysteine residues. Cytoplasmic proteins do not generally contain structural disulfide bonds as they cannot form in the reducing environment causing the protein to misfold (Derman and Beckwith, 1991). This problem is often overcome through the addition of a signal sequence to the N-terminus of a protein, which directs the protein across the cytoplasmic membrane to an alternative cellular compartment. The oxidative periplasmic space is an ideal compartment for the formation of disulfide bonds. CbpD was cloned and amplified with two difference signal sequences, PelBss and DsbAss. The PelBss directs transcribed proteins to the

periplasm through use of the Sec system (post-translational process), while the DsbAss directs actively transcribing proteins through use of the SRP and Sec pathway components (co-translational process). Expression of CbpD with both signal sequences still resulted in insoluble protein expression. It is possible that replacement of the native signal sequences of CbpD, CbpA and CbpL with that from CBP21 might result in the transport of the proteins to the periplasm leading to soluble protein expression.

CbpD contains two putative chitin-binding domains. These domains were cloned separately; CbpDa and CbpDb, along with a third construct CbpDc, both in an attempt to improve protein solubility but also to eludicate if both domains are chitin binding domains. The presence of multiple protein binding domains is common. Typically proteins that display multiple domains also have multiple activities, evolved through the recruitment of the different domains (Jaffe *et al.*, 1996, Montanier *et al.*, 2009), although some proteins display multiple domains with identical activities (Guillen *et al.*, 2007, Guillen *et al.*, 2010). While CbpDa and CbpDc both expressed to the insoluble protein fraction in *E. coli* it was clear that there were traces of CbpDb present in the soluble fraction.

An insoluble protein purification protocol was adapted based on the method by Amersham. The protocol for the isolation and solubilisation of the inclusion bodies worked well, with complete solubilisation of the protein pellet apparent from the change in colour. Although final re-suspension of the cell lysate is in 5 mL per 100 mL of culture the cell debris becomes very viscous following solubilisation and was difficult to sediment even using high-speed centrifugation, thus reducing the amount of protein that is passed through the FPLC system. Each protein was successfully purified using the refolding and purification protocol outlined in section 2.18.3 with few contaminants evident in subsequent silver stained SDS-PAGE gels. Intricacies were observed with protein elution, specifically with CbpD which leached very slowly from the column upon the application of imidazole. The efficiency of elution showed no improvement with pH adjustments or by the addition of GlcNAc. CbpDb was isolated from *E. coli* soluble cell lysate through IMAC purification. CbpDb withstood washes up to 100 mM imidazole and eluted homogeneously upon the application of 200 mM imidazole.

Protein characterisation was undertaken using both the insoluble substrate assays and ELLA analysis developed in section 4.1.2. From the insoluble substrate assay analysis it was evident that all four proteins were active and capable of binding to chitin, following solubilisation and re-folding from inclusion bodies. Unlike CBP21 CbpD appeared to have similar affinities for α-chitin, α-chitosan from crustacean shells and crystalline cellulose, and very little protein, if any, was visible in the unbound protein elution fractions. CbpDb appeared to mirror the affinity of CBP21 somewhat more closely with complete binding to α-chitin, partial binding to crystalline cellulose and no apparent binding to α-chitosan from crustacean shells. This suggests that the N-terminal chitin-binding domain of CbpD is responsible for chitosan binding. CbpA and CbpL differ in just 11 amino acids and as predicted had extremely similar binding profiles. Both bound to β and α-chitin with no protein visible in the first elution fraction. They displayed an almost equal affinity for α-chitosan from crustacean shells, chitosan from carpacea skin, and crystalline cellulose with minute amounts of protein visible in the unbound protein elution fraction. Although multiple protein bands were present in CbpA and CbpL His-Trap purified protein samples, each of the bands bound to the insoluble substrates, suggesting that they are all smaller, proteolytically cleaved, forms of CbpA and CbpL. Analogous to the ELLA analysis of CBP21 CbpD, CbpA nor CbpL appeared to bind to either the treated or untreated forms of thyroglobulin.

6.0 Final Discussion and Recommendations

This research thesis described the cloning of the *cbp21* gene that encodes the chitin-binding protein CBP21 from *S. marcescens*, into a bacterial expression system, enabling recombinant expression and affinity purification. The physico-chemical characterisation of CBP21 protein was carried out (see Chapter 3) including the investigation of the affinity and specificity of the recombinant molecule using a series of both novel and pre-established assays (see Chapter 4). The ability of CBP21 to bind to both α- and β-chitin, chitosan and crystalline cellulose was established. Despite displaying an affinity for GlcNAc polymers, CBP21 was found to have no affinity for the GlcNAc sub-units of protein attached glycans.

A novel GlcNAc binding assay was developed and described in section 4.2. Using this assay CBP21 was shown to have an affinity for polyacrylamide-linked chitobiose and chitotriose. The possibility of incorporating the molecule onto an affinity chromatography platform was also investigated. It was observed that CBP21 immobilised successfully on a sepharose matrix, with successful capture of chitobiose. The capture of PNGaseF released glycans was also explored. However an alternative technique, with increased sensitivity would need to be employed for the detection of the captured/released glycans as the quantities released were too small for the reducing sugar assay used.

As discussed in section 1.5, glycan enrichment of protein released glycans, is an important step in the characterisation of novel glycoproteins. Currently lectin affinity chromatography is the most commonly used technique for this type of characterisation. However, the availability and cost of the affinity matrices is the most restrictive feature of this method. As the chitobiose core is a common structural feature of all N-linked glycans, if bound by CBP21 this molecule could be exploited as a broad range N-linked glycan binding protein, used to concentrate glycan mixes prior to glycan profiling. The ability of CBP21 to bind to the chitobiose core of N-linked glycans could also lead to the creation of an alternative immobilisation strategy for carbohydrate arrays, with each glycan species bound directly to a matrix using CBP21, ensuring optimal orientation of the sugar species on the immobilisation matrix.

Novel CBP21 mutants were created to explore the role of specific residues in the putative carbohydrate binding site. It was theorised that through mutagenesis of the residues involved in chitin binding, novel molecules with alternative binding affinities or specificities could be generated. However, it was shown in section 4.5 that although mutagenesis of these residues did impact on the molecules ability to bind to chitin, chitin sub-structures and cellulose, complete disruption of binding to all of the polymers tested was not achieved and the affect of binding to glycoprotein structures was negligible, suggesting that no one residue was responsible for recognition and binding.

The availability of a crystal structure of CBP21 in complex with a GlcNAc polymer would allow for further site-directed mutagenesis studies, which may allow for the development of alanine mutants with more pronounced differences in specificities. The incorporation of alternative amino acids, through site-directed random mutagenesis, might even open the possibility of creating mutants with affinities for other glycan residues such as mannose or galactose. As crystallisation requires large amounts of pure soluble protein the optimisation of expression of CBP21 carried out in section 3.3.2 is of key importance. Further improvements to protein yields could be attained through the introduction of alternative growth media. Although LB was used throughout this study a number of alternative enriched media exist, including TB medium, SOB broth, YT broth, as well as several proprietary media components. Alternatively the introduction of more stable antibiotic resistance cassettes, such as carbenicillin, could improve plasmid retention within the cell, thus increasing protein production. The use of vectors with stronger promoters could also be introduced.

The creation of a library of carbohydrate-binding proteins displaying a wide variety of specificities all with a common core molecule would be highly advantageous in the field of glycobiology and specifically in the manufacture protein/lectin microarrays. An array consisting of recombinant molecules of common origin would mean that each carbohydrate-binding protein would have the same requirements for activity, such as temperature, pH and buffer. The ability to produce these proteins recombinantly would not only ensure consistent production (not seen with plant lectins) but would also allow the incorporation of affinity tags into the protein,

allowing more direct immobilisation strategies, leading to a dramatic decrease in cost, and greater accessibility.

Finally this thesis included an initial study of CBP21 related molecules. Bioinformatic analysis revealed the existence of a number of closely related CBP21 homologues, which had not been the subject of any previous characterisation studies. CbpD from *Pseduomonas aeruginosa*, CbpA from *Photorhabdus asymbiotica* and CbpL from *Photorhabdus luminescens* were selected for further study and subsequently cloned into bacterial expression systems similar to that of CBP21, enabling recombinant protein expression. Significant difficulties were encountered with the production of these molecules as they were expressed to the insoluble protein fraction. Although a number of strategies were explored to improve protein solubility, none proved fully successful. Solubilisation of these recombinant molecules provided enough protein to be purified and isolated for basic characterisation. It was shown that all three molecules are in fact chitin-binding molecules, that can bind to both the α and β forms of chitin and that the second putative chitin-binding domain of CbpD is capable of binding to chitin. As with CBP21 these molecules did not show any affinity for glycans attached proteins when tested using ELLA analysis.

Chitinolytic microorganisms play an important role in the biosphere by recycling chitin, the second most abundant carbohydrate, next to cellulose. Like organisms that degrade cellulose chitinolytic organisms produce a battery of enzymes with different specificities that act together to break down chitin microfibrils. As with cellulases many chitinases also have simple binding domains, not involved in catalysis, thought to anchor the enzymes to the substrate surface (Beguin and Aubert 1994). However, independent proteins such as CBP21, CbpD, CbpA and CbpL have been shown to increase the ability of chitinases to breakdown chitin.

Although development of these molecules for glycan analysis was not successful there are many commercially viable fields where the availability of characterised molecules such as these would be of great advantage. As outlined in section 1.18 the extraction of chitin from source is invariably carried out using a chemical process, leading to a build up of waste chemicals which are expensive to dispose of

(Goycoolea *et al.*, 2000). The use of prokaryotic chitinases with the synergistic effect of chitin-binding proteins would potentially lead to a more environmentally friendly, efficient process. Similarly in the field of bio-control chitinases have been engineered into plant expression systems to act as bio-pesticides in an effort to stave off attack by insects and pathogenic fungi. Enhanced resistance to chitin containing plant pathogens could be achieved through the introduction of chitin-binding proteins, to increase the effectiveness of chitinase action (Dehestani, *et al.*, 2010). Chitinolytic enzymes also have wide ranging applications in the preparation of chitooligosaccharides, preparation of single cell protein, isolation of protoplasts from fungi and yeast, the treatment of chitinous waste and the control of malaria transmission, all of which could benefit from the collaborative effect of the small chitin-binding proteins.

7.0 References

Adams, E. W., Ratner, D. M., Bokesch, H. R. 2004. Oligosaccharide and glycoprotein Microarrays as tools in HIV glycobiology: Glycan-dependent gp120/protein interactions. *Chemistry & biology*. 11 (6), pp875-881.

Afrough, B., Dwek, M.V. and Greenwell, P. 2007. Identification and elimination of false-positives in an ELISA-based system for qualitative assessment of glycoconjugate binding using a selection of plant lectins. *BioTechniques*, 43(4), pp.458.

Altschul, S., Madden, T., Schaffer, A. 1998. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Faseb Journal*. 12 (8), pp102.

Ambrosi, M., Cameron, N. R. and Davis, B. G. 2005. Lectins: tools for the molecular understanding of the glycocode. *Organic & Biomolecular Chemistry*. 3 (9), pp1593-1608.

Andersen, O. A., Dixon, M. J., Eggleston, I. M. and van Aalten, D. M. F. 2005. Natural product family 18 chitinase inhibitors. *Natural product reports*. 22 (5), pp563-579.

Andrade, P. P., Schottelius, J. and Andrade, C. R. 1988. An Enzyme-Linked Lectin Assay for the Study of Lectin Receptors of Leishmania. *Brazilian Journal of Medical and Biological Research*. 21 (3), pp517-521.

Andrews, P. 1964. Estimation of Molecular Weights of Proteins by Sephadex Gel-Filtration. *Biochemical Journal*. 91 (2), pp222-&.

Apweiler, R., Hermjakob, H. and Sharon, N. 1999. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochimica Et Biophysica Acta-General Subjects*. 1473 (1), pp4-8.

Arnold, F., Bedouet, L., Batina, P. 1998. Biochemical and immunological analyses of the flagellin of Clostridium tyrobutyricum ATCC 25755. *Microbiology and immunology*. 42 (1), pp23-31.

Ashwell, G. and Harford, J. 1982. Carbohydrate-Specific Receptors of the Liver. *Annual Review of Biochemistry*. 51pp531-554.

Bardwell, J.C.A., Mcgovern, K. and Beckwith, J. 1991. Identification of a Protein Required for Disulfide Bond Formation Invivo. *Cell*, 67(3), pp.581-589.

Beck, A., Cochet, O. and Wurch, T. 2010. GlycoFi's technology to control the glycosylation of recombinant therapeutic proteins. *Expert Opinion on Drug Discovery*. 5 (1), pp95-111.

Becker, D. J. and Lowe, J. B. 2003. Fucose: biosynthesis and biological function in mammals. *Glycobiology*. 13 (7), pp41R-53R.

Beguin, P. and Aubert, J.P. 1994. The Biological Degradation of Cellulose. *FEMS Microbiology Reviews,* 13(1), pp.25-58.

Beintema, J. J. 1994. Structural Features of Plant Chitinases and Chitin-Binding Proteins. *Febs Letters*. 350 (2-3), pp159-163.

Berks, B.C., Sargent, F. and Palmer, T. 2000. The Tat protein export pathway. *Molecular Microbiology,* 35(2), pp.260-274.

Blum, H., Beier, H. and Gross, H. J. 1987. Improved Silver Staining of Plant-Proteins, Rna and Dna in Polyacrylamide Gels. *Electrophoresis*. 8 (2), pp93-99.

Boot, R. G., Blommaart, E. F. C., Swart, E. 2001. Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. *Journal of Biological Chemistry*. 276 (9), pp6770-6778.

Boot, R. G., Renkema, G. H., Strijland, A. 1995. Cloning of a Cdna-Encoding Chitotriosidase, a Human Chitinase Produced by Macrophages. *Journal of Biological Chemistry*. 270 (44), pp26252-26256.

Boot, R. G., Renkema, G. H., Verhoek, M. 1998. The human chitotriosidase gene - Nature of inherited enzyme deficiency. *Journal of Biological Chemistry*. 273 (40), pp25680-25685.

Bormann, C., Baier, D., Horr, I. 1999. Characterization of a novel, antifungal, chitin-binding protein from Streptomyces tendae Tu901 that interferes with growth polarity. *Journal of Bacteriology*. 181 (24), pp7421-7429.

Bovin, N.V. 1998. Polyacrylamide-based glycoconjugates as tools in glycobiology. *Glycoconjugate Journal*, 15(5), pp.431-446.

Bowen, A. R., Chenwu, J. L., Momany, M. 1992. Classification of Fungal Chitin Synthases. *Proceedings of the National Academy of Sciences of the United States of America*. 89 (2), pp519-523.

Boyd, W. C. and Shapleigh, E. 1954. Specific Precipitating Activity of Plant Agglutinins (Lectins). *Science*. 119 (3091), pp419-419.

Bradford, M. M. 1976. Rapid and Sensitive Method for Quantitation of Microgram Quantities of Protein Utilizing Principle of Protein-Dye Binding. *Analytical Biochemistry*. 72 (1-2), pp248-254.

Brameld, K. A. and Goddard, W. A. 1998. The role of enzyme distortion in the single displacement mechanism of family 19 chitinases. *Proceedings of the National Academy of Sciences of the United States of America*. 95 (8), pp4276-4281.

Brockhausen, I., Yang, J. M., Burchell, J. 1995. Mechanisms Underlying Aberrant Glycosylation of Muc1 Mucin in Breast-Cancer Cells. *European Journal of Biochemistry*. 233 (2), pp607-617.

Bryan, M. C., Plettenburg, O., Sears, P. 2002. Saccharide display on microtiter plates. *Chemistry & biology*. 9 (6), pp713-720.

Bucher, M.H., Evdokimov, A.G. and Waugh, D.S. 2002. Differential effects of short affinity tags on the crystallization of Pyrococcus furiosus maltodextrin-binding protein. *Acta Crystallographica Section D-Biological Crystallography*, 58pp.392-397.

Bouchez-Mahiout, I., Boulenc, E., Snegaroff, J., Choudat, D., Pecquet, C., Raison-Peyron, N., Vigan, M., Chabane, H., Dron-Gonzalves, M., Branlard, G., Tanis-Plant, S. and Lauriere, M. 2010. Immunoblotting analysis of wheat allergens: control of side reactions through wheat polypeptides naturally present in dried cow milk. *Food and Agricultural Immunology*, 21(3), pp.237-251.

Budnik, B. A., Lee, R. S. and Steen, J. A. J. 2006. Global methods for protein glycosylation analysis by mass spectrometry. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 1764 (12), pp1870-1880.

Buhi, W. C. 2002. Characterization and biological roles of oviduct-specific, oestrogen-dependent glycoprotein. *Reproduction*. 123 (3), pp355-362.

Buonomano, R., Tinguely, C., Rieben, R., Mohacsi, P.J. and Nydegger, U.E. 1999. Quantitation and characterization of anti-Gal alpha 1-3Gal antibodies in sera of 200 healthy persons. *Xenotransplantation*, 6(3), pp.173-180.

Burda, P. and Aebi, M. 1999. The dolichol pathway of N-linked glycosylation. *Biochimica Et Biophysica Acta-General Subjects*. 1426 (2), pp239-257.

Bussink, A. P., Speijer, D., Aerts, J. M. R. G. and Boot, R. G. 2007. Evolution of mammalian chitinase (-like) members of family 18 glycosyl hydrolases. *Genetics*. 177 (2), pp959-970.

Cambi, A., Koopman, M. and Figdor, C. G. 2005. How C-type lectins detect pathogens. *Cellular Microbiology*. 7 (4), pp481-488.

Campos-Gongora, E., Ebert, F., Willhoeft, U. 2004. Characterization of chitin synthases from Entamoeba. *Protist*. 155 (3), pp323-330.

Cantarel, B. L., Coutinho, P. M., Rancurel, C. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research*. 37ppD233-D238.

Castric, P., Cassels, F. J. and Carlson, R. W. 2001. Structural characterization of the Pseudomonas aeruginosa 1244 pilin glycan. *Journal of Biological Chemistry*. 276 (28), pp26479-26485.

Chang, W., Chen, M. and Wang, S. 2010. An antifungal chitinase produced by Bacillus subtilis using chitin waste as a carbon source. *World Journal of Microbiology & Biotechnology*. 26 (5), pp945-950.

Chaston, J. and Goodrich-Blair, H. 2010. Common trends in mutualism revealed by model associations between invertebrates and bacteria. *FEMS microbiology reviews*. 34 (1), pp41-58.

Chen, G., Dunphy, G.B. and Webster, J.M. 1994. Antifungal Activity of 2 Xenorhabdus Species and Photorhabdus-Luminescens, Bacteria Associated with the Nematodes Steinernema Species and Heterorhabditis-Megidis. *Biological Control*, 4(2), pp.157-162.

Chen, G., Zhang, Y., Li, J., Dunphy, G.B., Punja, Z.K. and Webster, J.M. 1996. Chitinase Activity ofXenorhabdusandPhotorhabdusSpecies, Bacterial Associates of Entomopathogenic Nematodes. *Journal of Invertebrate Pathology*, 68(2), pp.101-108.

Chong, S. R., Mersha, F. B., Comb, D. G. 1997. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene*. 192 (2), pp271-281.

Christlet, T. H. T. and Veluraja, K. 2001. Database analysis of O-glycosylation sites in proteins. *Biophysical journal*. 80 (2), pp952-960.

Cohen, E. 2001. Chitin synthesis and inhibition: a revisit. *Pest management science*. 57 (10), pp946-950.

Chu, H. H., Hoang, V., Hofemeister, J. and Schrempf, H. 2001. A Bacillus amyloliquefaciens ChbB protein binds beta- and alpha-chitin and has homologues in related strains. *Microbiology-Sgm*. 147pp1793-1803.

Creavin, A. 2010. *PA-IIL* as a model lectin for the structural and functional characterisation of related lectins. PhD thesis. Dublin City University

Cunningham, B.C. and Wells, J.A. 1989. High-Resolution Epitope Mapping of Hgh-Receptor Interactions by Alanine-Scanning Mutagenesis. *Science,* 244(4908), pp.1081-1085.

Dahms, N. M., Olson, L. J. and Kim, J. P. 2008. Strategies for carbohydrate recognition by the mannose 6-phosphate receptors. *Glycobiology*. 18 (9), pp664-678.

Dai, T., Tegos, G. P., Burkatovskaya, M. 2009. Chitosan Acetate Bandage as a Topical Antimicrobial Dressing for Infected Burns. *Antimicrobial Agents and Chemotherapy*. 53 (2), pp393-400.

Dalpathado, D. S. and Desaire, H. 2008. Glycopeptide analysis by mass spectrometry. *Analyst*. 133 (6), pp731-738.

Darling, R. J., Kuchibhotla, U., Glaesner, W. 2002. Glycosylation of erythropoietin affects receptor binding kinetics: Role of electrostatic interactions. *Biochemistry*. 41 (49), pp14524-14531.

Davies, G. and Henrissat, B. 1995. Structures and Mechanisms of Glycosyl Hydrolases. *Structure*. 3 (9), pp853-859.

Dehestani, A., Kazemitabar, K., Ahmadian, G. 2010. Chitinolytic and antifungal activity of a Bacillus pumilus chitinase expressed in Arabidopsis. *Biotechnology Letters*. 32 (4), pp539-546.

De Jong, G. and Van Eijk, H.G. 1989. Functional properties of the carbohydrate moiety of human transferrin. *International Journal of Biochemistry*, 21(3), pp.253-263.

Derman, A.I. and Beckwith, J. 1991. Escherichia-Coli Alkaline-Phosphatase Fails to Acquire Disulfide Bonds when Retained in the Cytoplasm. *Journal of Bacteriology*, 173(23), pp.7719-7722.

Dini, L., Autuori, F., Lentini, A. 1992. The Clearance of Apoptotic Cells in the Liver is Mediated by the Asialoglycoprotein Receptor. *FEBS letters*. 296 (2), pp174-178.

Doig, P., Kinsella, N., Guerry, P. and Trust, T. J. 1996. Characterization of a post-translational modification of Campylobacter flagellin: Identification of a sero-specific glycosyl moiety. *Molecular microbiology*. 19 (2), pp379-387.

Dorland, L., Haverkamp, J., Schut, B.L., Vliegenthart, J.F.G., Spik, G., Strecker, G., Fournet, B. and Montreuil, J. 1977. The structure of the asialo-carbohydrate units of human serotransferrin as proven by 360 MHz proton magnetic resonance spectroscopy. *FEBS Letters*, 77(1), pp.15-20.

Drickamer, K. and Taylor, M. E. 2002. Glycan arrays for functional glycomics. *Genome biology*. 3 (12), pp1034.1.

Dube, D. H. and Bertozzi, C. R. 2005. Glycans in cancer and inflammation. Potential for therapeutics and diagnostics. *Nature Reviews Drug Discovery*. 4 (6), pp477-488.

Duchaud, E., Rusniok, C., Frangeul, L. 2003. The genome sequence of the entomopathogenic bacterium Photorhabdus luminescens. *Nature Biotechnology*. 21 (11), pp1307-1313.

Duck, L. W., Walter, M. R., Novak, J. 2007. Isolation of flagellated bacteria implicated in Crohn's disease. *Inflammatory bowel diseases*. 13 (10), pp1191-1201.

Duo-Chuan, L. 2006. Review of fungal chitinases. *Mycopathologia,* 161(6), pp.345-360.

Dupuy, F., Petit, J.M., Mollicone, R., Oriol, R., Julien, R. and Maftah, A. 1999. A single amino acid in the hypervariable stem domain of vertebrate alpha 1,3/1,4-fucosyltransferases determines the type 1 type 2 transfer - Characterization of acceptor substrate specificity of the Lewis enzyme by site-directed mutagenesis. *Journal of Biological Chemistry,* 274(18), pp.12257-12262.

Dutta, P. K., Tripathi, S., Mehrotra, G. K. and Dutta, J. 2009. Perspectives for chitosan based antimicrobial films in food applications. *Food Chemistry*. 114 (4), pp1173-1182.

Eijsink, V. G. H., Vaaje-Kolstad, G., Varum, K. M. and Horn, S. J. 2008. Towards new enzymes for biofuels: lessons from chitinase research. *Trends in biotechnology*. 26 (5), pp228-235.

Elvin, C. M., Vuocolo, T., Pearson, R. D. 1996. Characterization of a major peritrophic membrane protein, peritrophin-44, from the larvae of Lucilia cuprina - cDNA and deduced amino acid sequences. *Journal of Biological Chemistry*. 271 (15), pp8925-8935.

Fekkes, P. and Driessen, A.J.M. 1999. Protein targeting to the bacterial cytoplasmic membrane. *Microbiology and Molecular Biology Reviews*, 63(1), pp.161-+.

Fisher, K. J. and Aronson, N. N. 1992. Cloning and Expression of the Cdna Sequence Encoding the Lysosomal Glycosidase Di-N-Acetylchitobiase. *Journal of Biological Chemistry*. 267 (27), pp19607-19616.

Fischer-Le Saux, M., Viallard, V., Brunel, B. 1999. Polyphasic classification of the genus Photorhabdus and proposal of new taxa: P-luminescens subsp luminescens subsp nov., P-luminescens subsp akhurstii subsp nov., P-luminescens subsp laumondii subsp nov., P.temperata sp nov., P-temperata subsp temperata subsp nov and P-asymbiotica sp nov. *International Journal of Systematic Bacteriology*. 49pp1645-1656.

Fitches, E., Woodhouse, S. D., Edwards, J. P. and Gatehouse, J. A. 2001. In vitro and in vivo binding of snowdrop (Galanthus nivalis agglutinin; GNA) and jackbean (Canavalia ensiformis; Con A) lectins within tomato moth (Lacanobia oleracea) larvae; mechanisms of insecticidal action. *Journal of insect physiology*. 47 (7), pp777-787.

Fitzpatrick, T.B., Auweter, S., Kitzing, K., Clausen, T., Amrhein, N. and Macheroux, P. 2004. Structural and functional impairment of an Old Yellow Enzyme homologue upon affinity tag incorporation. *Protein Expression and Purification*, 36(2), pp.280-291.

Folders, J., Tommassen, J., van Loon, L. C. and Bitter, W. 2000. Identification of a chitin-binding protein secreted by Pseudomonas aeruginosa. *Journal of Bacteriology*. 182 (5), pp1257-1263.

Freeze, H. H. 1998. Disorders in protein glycosylation and potential therapy: Tip of an iceberg? *Journal of Pediatrics*. 133 (5), pp593-600.

Friedman, B. A., Vaddi, K., Preston, C. 1999. A comparison of the pharmacological properties of carbohydrate remodeled recombinant and placental-derived beta-glucocerebrosidase: Implications for clinical efficacy in treatment of Gaucher disease. *Blood*. 93 (9), pp2807-2816.

Fuchs, R. L., Mcpherson, S. A. and Drahos, D. J. 1986. Cloning of a Serratia-Marcescens Gene Encoding Chitinase. *Applied and Environmental Microbiology*. 51 (3), pp504-509.

Fukui, S., Feizi, T., Galustian, C. 2002. Oligosaccharide microarrays for high-throughput detection and specificity assignments of carbohydrate-protein interactions. *Nature biotechnology*. 20 (10), pp1011-1017.

Fukushima, K., Satoh, T., Baba, S. and Yamashita, K. 2010. alpha 1,2-Fucosylated and beta-N-acetylgalactosaminylated prostate-specific antigen as an efficient marker of prostatic cancer. *Glycobiology*. 20 (4), pp452-460.

Gerrard, J. G., Joyce, S. A., Clarke, D. J. 2006. Nematode symbiont for Photorhabdus asymbiotica. *Emerging Infectious Diseases*. 12 (10), pp1562-1564.

Geyer, H. and Geyer, R. 2006. Strategies for analysis of glycoprotein glycosylation. *Biochimica Et Biophysica Acta-Proteins and Proteomics*. 1764 (12), pp1853-1869.
Gooday, G. W. 1990. The Ecology of Chitin Degradation. *Advances in Microbial Ecology*. 11pp387-430.

Goel, A., Colcher, D., Koo, J.S., Booth, B.J.M., Pavlinkova, G. and Batra, S.K. 2000. Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *Biochimica Et Biophysica Acta-General Subjects*, 1523(1), pp.13-20.

Goycoolea, F.M., Argüelles-Monal, W., Peniche, C. and Higuera-Ciapara, I.2000. Chitin and chitosan. *IN:* G. Doxastakis and V. Kiosseoglou (ed.) *Developments in Food Science.* Elsevier.

Graham, L. S. and Sticklen, M. B. 1994. Plant Chitinases. *Canadian Journal of Botany-Revue Canadienne De Botanique*. 72 (8), pp1057-1083.

Guan, S., Mok, Y., Koo, K. 2009. Chitinases: Biomarkers for Human Diseases. *Protein and Peptide Letters*. 16 (5), pp490-498.

Guerry, P. 2007. Campylobacter flagella: not just for motility. *Trends in microbiology*. 15 (10), pp456-461.

Guillen, D., Santiago, M., Linares, L., Perez, R., Morlon, J., Ruiz, B., Sanchez, S. and Rodriguez-Sanoja, R. 2007. Alpha-amylase starch binding domains: Cooperative effects of binding to starch granules of multiple tandemly arranged domains. *Applied and Environmental Microbiology*, 73(12), pp.3833-3837.

Guillen, D., Sanchez, S. and Rodriguez-Sanoja, R. 2010. Carbohydrate-binding domains: multiplicity of biological roles. *Applied Microbiology and Biotechnology*, 85(5), pp.1241-1249.

Gum, J. R. 1992. Mucin Genes and the Proteins they Encode - Structure, Diversity, and Regulation. *American Journal of Respiratory Cell and Molecular Biology*. 7 (6), pp557-564.

Hakala, B. E., White, C. and Recklies, A. D. 1993. Human Cartilage Gp-39, a Major Secretory Product of Articular Chondrocytes and Synovial-Cells, is a Mammalian Member of a Chitinase Protein Family. *Journal of Biological Chemistry*. 268 (34), pp25803-25810.

Hakomori, S. 1985. Aberrant Glycosylation in Cancer Cell-Membranes as Focused on Glycolipids - Overview and Perspectives. *Cancer research*. 45 (6), pp2405-2414.

Hakomori, S. 2002. Glycosylation defining cancer malignancy: New wine in an old bottle. *Proceedings of the National Academy of Sciences of the United States of America*. 99 (16), pp10231-10233.

Hakomori, S. 1996. Tumor malignancy defined by aberrant glycosylation and sphingo(glyco)lipid metabolism. *Cancer research*. 56 (23), pp5309-5318.

Hamilton, S. R., Davidson, R. C., Sethuraman, N. 2006. Humanization of yeast to produce complex terminally sialylated glycoproteins. *Science*. 313 (5792), pp1441-1443.

Hammond, C., Braakman, I. and Helenius, A. 1994. Role of N-Linked Oligosaccharide Recognition, Glucose Trimming, and Calnexin in Glycoprotein Folding and Quality-Control. *Proceedings of the National Academy of Sciences of the United States of America*. 91 (3), pp913-917.

Hardy, S. M., Roberts, C. J., Brown, P. R. and Russell, D. A. 2010. Glycoprotein microarray for the fluorescence detection of antibodies produced as a result of erythropoietin (EPO) abuse. *Analytical Methods*. 2 (1), pp17-23.

Hatahet, F., Nguyen, V.D., Salo, K.E.H. and Ruddock, L.W. 2010. Disruption of reducing pathways is not essential for efficient disulfide bond formation in the cytoplasm of E. coli. *Microbial Cell Factories,* 9pp.67.

Hejgaard, J., Jacobsen, S., Bjorn, S. E. and Kragh, K. M. 1992. Antifungal Activity of Chitin-Binding Pr-4 Type Proteins from Barley-Grain and Stressed Leaf. *FEBS letters*. 307 (3), pp389-392.

Helenius, A. and Aebi, M. 2004. Roles of N-linked glycans in the endoplasmic reticulum. *Annual Review of Biochemistry*. 73pp1019-1049.

Henrissat, B., Callebaut, I., Fabrega, S. 1995. Conserved Catalytic Machinery and the Prediction of a Common Fold for several Families of Glycosyl Hydrolases. *Proceedings of the National Academy of Sciences of the United States of America*. 92 (15), pp7090-7094.

Herscovics, A. and Orlean, P. 1993. Glycoprotein-Biosynthesis in Yeast. *Faseb Journal*. 7 (6), pp540-550.

Hirschberg, C. B. 2001. Golgi nucleotide sugar transport and leukocyte adhesion deficiency II. *Journal of Clinical Investigation*. 108 (1), pp3-6.

Hitchen, P. G. and Dell, A. 2006. Bacterial glycoproteomics. *Microbiology-Sgm*. 152pp1575-1580.

Hollak, C. E. M., Vanweely, S., Vanoers, M. H. J. and Aerts, J. M. F. G. 1994. Marked Elevation of Plasma Chitotriosidase Activity - a Novel Hallmark of Gaucher Disease. *Journal of Clinical Investigation*. 93 (3), pp1288-1292.

Horlacher, T. and Seeberger, P. H. 2008. Carbohydrate arrays as tools for research and diagnostics. *Chemical Society Reviews*. 37 (7), pp1414-1422.

Houseman, B. T. and Mrksich, M. 2002. Carbohydrate arrays for the evaluation of protein binding and enzymatic modification. *Chemistry & biology*. 9 (4), pp443-454.

Howard, M. B., Ekborg, N. A., Weiner, R. M. and Hutcheson, S. W. 2003. Detection and characterization of chitinases and other chitin-modifying enzymes. *Journal of Industrial Microbiology & Biotechnology*. 30 (11), pp627-635.

Hu, B., Trinh, K., Figueira, W. F. and Price, P. A. 1996. Isolation and sequence of a novel human chondrocyte protein related to mammalian members of the chitinase protein family. *Journal of Biological Chemistry*. 271 (32), pp19415-19420.

Huet, J., Rucktooa, P., Clantin, B., Azarkan, M., Looze, Y., Villeret, V. and Wintjens, R. 2008. X-ray structure of papaya chitinase reveals the substrate binding mode of glycosyl hydrolase family 19 chitinases. *Biochemistry*, 47(32), pp.8283-8291.

Imberty, A., Wimmerova, M., Mitchell, E. P. and Gilboa-Garber, N. 2004. Structures of the lectins from Pseudomonas aeruginosa: insights into the molecular basis for host glycan recognition. *Microbes and Infection*. 6 (2), pp221-228.

Ioffe, E. and Stanley, P. 1994. Mice Lacking N-Acetylglucosaminyltransferase-i Activity Die at Midgestation, Revealing an Essential Role for Complex Or Hybrid N-Linked Carbohydrates. *Proceedings of the National Academy of Sciences of the United States of America*. 91 (2), pp728-732.

Jaeken, J. and Matthijs, G. 2001. Congenital disorders of glycosylation. *Annual Review of Genomics and Human Genetics*. 2pp129-151.

Jaffe, J., NatansonYaron, S., Caparon, M.G. and Hanski, E. 1996. Protein F2, a novel fibronectin-binding protein from Streptococcus pyogenes, possesses two binding domains. *Molecular Microbiology*, 21(2), pp.373-384.

James, B. S. and Stacey, F. H. 1936. The Indetification of the Haemagglutinin of the Jack Beah with Concanavalin A. *J. Bacteriol.* 32 (2), pp227.

Janakiraman, M. N., White, C. L., Laver, W. G. 1994. Structure of Influenza-Virus Neuraminidase B/Lee/40 Complexed with Sialic-Acid and a Dehydro Analog at 1.8-Angstrom Resolution - Implications for the Catalytic Mechanism. *Biochemistry*. 33 (27), pp8172-8179.

Jang, M. K., Kong, B. G., Jeong, Y. I. 2004. Physicochemical characterization of alpha-chitin, beta-chitin, and gamma-chitin separated from natural resources. *Journal of Polymer Science Part A-Polymer Chemistry*. 42 (14), pp3423-3432.

Jayakumar, R., Prabaharan, M., Nair, S. V. 2010. Novel carboxymethyl derivatives of chitin and chitosan materials and their biomedical applications. *Progress in Materials Science*. 55 (7), pp675-709.

Jenkins, N., Parekh, R. B. and James, D. C. 1996. Getting the glycosylation right: Implications for the biotechnology industry. *Nature biotechnology*. 14 (8), pp975-981.

Jeong, K.J. and Lee, S.Y. 2000. Secretory production of human leptin in Escherichia coli. *Biotechnology and Bioengineering,* 67(4), pp.398-407.

Johnson, D.L., Middleton, S.A., McMahon, F., Barbone, F.P., Kroon, D., Tsao, E., Lee, W.H., Mulcahy, L.S. and Jolliffe, L.K. 1996. Refolding, purification, and characterization of human erythropoietin binding protein produced in Escherichia coli. *Protein Expression and Purification,* 7(1), pp.104-113.

Jones, A. J. S., Papac, D. I., Chin, E. H. 2007. Selective clearance of glycoforms of a complex glycoprotein pharmaceutical caused by terminal N-acetylglucosamine is similar in humans and cynomolgus monkeys. *Glycobiology*. 17 (5), pp529-540.

Joshi, M. C., Sharma, A., Kant, S. 2008. An insecticidal GroEL protein with chitin binding activity from Xenorhabdus nematophila. *Journal of Biological Chemistry*. 283 (42), pp28287-28296.

Ju, T., Aryal, R. P., Stowell, C. J. and Cummings, R. D. 2008. Regulation of protein O-glycosylation by the endoplasmic reticulum-localized molecular chaperone Cosmc. *Journal of Cell Biology*. 182 (3), pp531-542.

Kamerling, J.P., Rijkse, I., Maas, A.A.M., van Kuik, J.A. and Vliegenthart, J.F.G. 1988. Sulfated N-linked carbohydrate chains in porcine thyroglobulin. *FEBS Letters*, 241(1-2), pp.246-250.

Kanninen, K., Goldsteins, G., Auriola, S. 2004. Glycosylation changes in Alzheimer's disease as revealed by a proteomic approach. *Neuroscience letters*. 367 (2), pp235-240.

Karlsson, M. and Stenlid, J. 2009. Evolution of Family 18 Glycoside Hydrolases: Diversity, Domain Structures and Phylogenetic Relationships. *Journal of Molecular Microbiology and Biotechnology*. 16 (3-4), pp208-223.

Kasai, K. and Hirabayashi, J. 1996. Galectins: A family of animal lectins that decipher glycocodes. *Journal of Biochemistry*. 119 (1), pp1-8.

Kaur, R., Dikshit, K.L. and Raje, M. 2002. Optimization of immunogold labeling TEM: An ELISA-based method for evaluation of blocking agents for quantitative detection of antigen. *Journal of Histochemistry & Cytochemistry*, 50(6), pp.863-873.

Kawabata, S., Nagayama, R., Hirata, M. 1996. Tachycitin, a small granular component in horseshoe crab hemocytes, is an antimicrobial protein with chitin-binding activity. *Journal of Biochemistry*. 120 (6), pp1253-1260.

Keenan, R.J., Freymann, D.M., Stroud, R.M. and Walter, P. 2001. The signal recognition particle. *Annual Review of Biochemistry,* 70pp.755-775.

Keyhani, N. O. and Roseman, S. 1999. Physiological aspects of chitin catabolism in marine bacteria. *Biochimica Et Biophysica Acta-General Subjects*. 1473 (1), pp108-122.

Khoushab, F. and Yamabhai, M. 2010. Chitin Research Revisited. *Marine Drugs*. 8 (7), pp1988-2012.

Khushoo, A., Pal, Y., Singh, B.N. and Mukherjee, K.J. 2004. Extracellular expression and single step purification of recombinant Escherichia coli l-asparaginase II. *Protein Expression and Purification,* 38(1), pp.29-36.

Kim, Y., Jang, K., Joo, H., Kim, H., Lee, C. and Kim, B. 2007. Simultaneous profiling of N-glycans and proteins from human serum using a parallel-column system directly coupled to mass spectrometry. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 850(1-2), pp.109-119.

Kim, Y. J. and Varki, A. 1997. Perspectives on the significance of altered glycosylation of glycoproteins in cancer. *Glycoconjugate journal*. 14 (5), pp569-576.

Kobata, A. 1992. Structures and Functions of the Sugar Chains of Glycoproteins. *European Journal of Biochemistry*. 209 (2), pp483-501.

Koch, H.G., Moser, M. and Muller, M. 2003. Signal recognition particle-dependent protein targeting, universal to all kingdoms of life. *Reviews of Physiology, Biochemistry and Pharmacology, Vol 146,* 146pp.55-94.

Kohn, J. and Wilchek, M. 1982. A New Approach (Cyano-Transfer) for Cyanogen-Bromide Activation of Sepharose at Neutral pH, which Yields Activated Resins, Free of Interfering Nitrogen Derivatives. *Biochemical and Biophysical Research Communications,* 107(3), pp.878-884.

Kolbe, S., Fischer, S., Becirevic, A. 1998. The Streptomyces reticuli alpha-chitin-binding protein CHB2 and its gene. *Microbiology-Uk*. 144pp1291-1297.

Krajewska, B. 2004. Application of chitin- and chitosan-based materials for enzyme immobilizations: a review. *Enzyme and microbial technology*. 35 (2-3), pp126-139.

Kraus, A., Groenendyk, J., Bedard, K. 2010. Calnexin Deficiency Leads to Dysmyelination. *Journal of Biological Chemistry*. 285 (24), pp18928-18938.

Krist, P., Vannucci, L., Kuzma, M. 2004. Fluorescent labelled thiourea-bridged glycodendrons. *Chembiochem*. 5 (4), pp445-452.

Krykbaev, R., Fitz, L. J., Reddy, P. S. 2010. Evolutionary and biochemical differences between human and monkey acidic mammalian chitinases. *Gene*. 452 (2), pp63-71.

Kuranda, M. J. and Robbins, P. W. 1991. Chitinase is Required for Cell-Separation during Growth of Saccharomyces-Cerevisiae. *Journal of Biological Chemistry*. 266 (29), pp19758-19767.

Kurmyshkina, O., Rapoport, E., Moiseeva, E., Korchagina, E., Ovchinnikova, T., Pazynina, G., Belyanchikov, I. and Bovin, N. 2010. Glycoprobes as a tool for the study of lectins expressed on tumor cells. *Acta Histochemica,* 112(2), pp.118-126.

Kzhyshkowska, J., Mamidi, S., Gratchev, A. 2006. Novel stabilin-1 interacting chitinase-like protein (SI-CLP) is up-regulated in alternatively activated macrophages and secreted via lysosomal pathway. *Blood*. 107 (8), pp3221-3228.

Laemmli, U. K. 1970. Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-T4. *Nature*. 227 (5259), pp680.

Laine, R. A. 1997. Information capacity of the carbohydrate code. *Pure and Applied Chemistry*. 69 (9), pp1867-1873.

Lambre, C. R., Terzidis, H., Greffard, A. and Webster, R. G. 1991. An Enzyme-Linked Lectin Assay for Sialidase. *Clinica Chimica Acta*. 198 (3), pp183-194.

Lee, H.C. and Bernstein, H.D. 2001. The targeting pathway of Escherichia coli presecretory and integral membrane proteins is specified by the hydrophobicity of the targeting signal. *Proceedings of the National Academy of Sciences of the United States of America*, 98(6), pp.3471-3476.

Leffler, H., Carlsson, S., Hedlund, M. 2002. Introduction to galectins. *Glycoconjugate journal*. 19 (7-9), pp433-440.

Lei, S.P., Lin, H.C., Wang, S.S., Callaway, J. and Wilcox, G. 1987. Characterization of the Erwinia-Carotovora Pelb Gene and its Product Pectate Lyase. *Journal of Bacteriology*, 169(9), pp.4379-4383.

Lepenies, B. and Seeberger, P. H. 2010. The promise of glycomics, glycan arrays and carbohydrate-based vaccines. *Immunopharmacology and immunotoxicology*. 32 (2), pp196-207.

Lerouge, P., Cabanes-Macheteau, M., Rayon, C. 1998. N-glycoprotein biosynthesis in plants: recent developments and future trends. *Plant Molecular Biology*. 38 (1-2), pp31-48.

Liang, C. and Wu, C. 2009. Glycan array: a powerful tool for glycomics studies. *Expert Review of Proteomics*. 6 (6), pp631-645.

Lienemann, M. 2009. Characterization of the wheat germ agglutinin binding to self-assembled monolayers of neoglycoconjugates by AFM and SPR. *Glycobiology (Oxford)*, 19(6), pp.663.

Loris, R. 2002. Principles of structures of animal and plant lectins. *Biochimica Et Biophysica Acta-General Subjects*. 1572 (2-3), pp198-208.

Loris, R., Tielker, D., Jaeger, K. E. and Wyns, L. 2003. Structural basis of carbohydrate recognition by the lectin LecB from Pseudomonas aeruginosa. *Journal of Molecular Biology*. 331 (4), pp861-870.

Ludwig, J. A. and Weinstein, J. N. 2005. Biomarkers in cancer staging, prognosis and treatment selection. *Nature Reviews Cancer*. 5 (11), pp845-856.

Lyristis, M., Boynton, Z. L., Petersen, D. 2000. Cloning, sequencing, and characterization of the gene encoding flagellin, flaC, and the post-translational modification of flagellin, FlaC, from Clostridium acetobutylicum ATCC824. *Anaerobe*. 6 (2), pp69-79.

Maezaki, Y., Tsuji, K., Nakagawa, Y. 1993. Hypocholesterolemic Effect of Chitosan in Adult Males. *Bioscience Biotechnology and Biochemistry*. 57 (9), pp1439-1444.

Mahdavi, J., Sonden, B., Hurtig, M. 2002. Helicobacter pylori SabA adhesin in persistent infection and chronic inflammation. *Science*. 297 (5581), pp573-578.

Makrides, S.C. 1996. Strategies for achieving high-level expression of genes in Escherichia coli. *Microbiological Reviews,* 60(3), pp.512-&.

Middelberg, A.P.J. 2002. Preparative protein refolding. *Trends in Biotechnology*, 20(10), pp.437-443.

Mislovicova, D., Gemeiner, P., Kozarova, A. and Kozar, T. 2009. Lectinomics I. Relevance of exogenous plant lectins in biomedical diagnostics. *Biologia*, 64(1), pp.1-19.

Mallorqui, J., Llop, E., de Bolos, C., Gutierrez-Gallego, R., Segura, J. and Pascual, J.A. 2010. Purification of erythropoietin from human plasma samples using an immunoaffinity well plate. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences*, 878(23), pp.2117-2122.

Manos, J., Arthur, J., Rose, B. 2009. Gene expression characteristics of a cystic fibrosis epidemic strain of Pseudomonas aeruginosa during biofilm and planktonic growth. *FEMS microbiology letters*. 292 (1), pp107-114.

Brurberg, Synstad, Klemsdal, Van Aalten, Sundheim and Eijsink. 2000. *Chitinases from Serratia marcescens.*

Matsumoto, H., Shinzaki, S., Narisada, M. 2010. Clinical application of a lectin-antibody ELISA to measure fucosylated haptoglobin in sera of patients with pancreatic cancer. *Clinical Chemistry and Laboratory Medicine*. 48 (4), pp505-512.

Taylor and Drickamer. 2006. *Introduction to Glycobiology.* second ed. Oxford University Press.

Mazeau, K., Winter, W. T. and Chanzy, H. 1994. Molecular and Crystal-Structure of a High-Temperature Polymorph of Chitosan from Electron-Diffraction Data. *Macromolecules*. 27 (26), pp7606-7612.

Mccoy, J. P., Varani, J. and Goldstein, I. J. 1983. Enzyme-Linked-Lectin Assay (Ella) - Specific Enzyme Assay for Detection and Quantitation of Terminal Carbohydrate Groups. *Federation proceedings*. 42 (4), pp933-933.

McMahon, K. 2010. Structural and functional characterisation of lectins from PA-IL superfamily of lectins. PhD thesis. Dublin City University.

Meibom, K. L., Li, X. B. B., Nielsen, A. T. 2004. The Vibrio cholerae chitin utilization program. *Proceedings of the National Academy of Sciences of the United States of America*. 101 (8), pp2524-2529.

Mellado, E., Specht, C. A., Robbins, P. W. and Holden, D. W. 1996. Cloning and characterization of chsD, a chitin synthase-like gene of Aspergillus fumigatus. *FEMS microbiology letters*. 143 (1), pp69-76.

Merzendorfer, H. and Zimoch, L. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *Journal of Experimental Biology*. 206 (24), pp4393-4412.

Mescher, M. F., Stroming.Jl and Watson, S. W. 1974. Protein and Carbohydrate Composition of Cell-Envelope of Halobacterium-Salinarium. *Journal of Bacteriology*. 120 (2), pp945-954.

Mescher, M. F. and Strominger, J. L. 1976. Purification and Characterization of a Prokaryotic Glycoprotein from Cell-Envelope of Halobacterium-Salinarium. *Journal of Biological Chemistry*. 251 (7), pp2005-2014.

Millar, D. J., Allen, A. K., Smith, C. G. 1992. Chitin-Binding Proteins in Potato (Solanum-Tuberosum L) Tuber - Characterization, Immunolocalization and Effects of Wounding. *Biochemical Journal*. 283pp813-821.

Miller, G. L. 1959. Use of Dinitrosalicylic Acid Reagent for Determination of Reducing Sugar. *Analytical Chemistry*. 31 (3), pp426-428.

Misawa, S. and Kumagai, I. 1999. Refolding of therapeutic proteins produced in Escherichia coli as inclusion bodies. *Biopolymers*, 51(4), pp.297-307.

Mislovicova, D., Gemeiner, P., Kozarova, A. and Kozar, T. 2009. Lectinomics I. Relevance of exogenous plant lectins in biomedical diagnostics. *Biologia*. 64 (1), pp1-19.

Mitsuhashi, W., Kawakita, H., Murakami, R. 2007. Spindles of an entomopoxvirus facilitate its infection of the host insect by disrupting the peritrophic membrane. *Journal of virology*. 81 (8), pp4235-4243.

Miyoshi, E., Nakano, M., Okuyama, N. and Taniguchi, N. 2006. Fucosylated haptoglobin is a novel marker for pancreatic cancer; the detailed structure of oligosaccharide and a possible mechanism for fucosylation. *Molecular & Cellular Proteomics*. 5 (10), pp861.

Moens, S. and Vanderleyden, J. 1997. Glycoproteins in prokaryotes. *Archives of Microbiology*. 168 (3), pp169-175.

Mohanty, A.K. and Wiener, M.C. 2004. Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expression and Purification*, 33(2), pp.311-325.

Moiseeva, E. V., Rapoport, E. M., Bovin, N. V. 2005. Galectins as markers of aggressiveness of mouse mammary carcinoma: towards a lectin target therapy of human breast cancer. *Breast cancer research and treatment*. 91 (3), pp227-241.

Montanier, C., Money, V.A., Pires, V.M.R., Flint, J.E., Pinheiro, B.A., Goyal, A., Prates, J.A.M., Izumi, A., Stalbrand, H., Morland, C., Cartmell, A., Kolenova, K., Topakas, E., Dodson, E.J., Bolam, D.N., Davies, G.J., Fontes, C.M.G.A. and Gilbert, H.J. 2009. The Active Site of a Carbohydrate Esterase Displays Divergent Catalytic and Noncatalytic Binding Functions. *Plos Biology*, 7(3), pp.687-697.

Morell, A. G., Gregoria.G, Scheinbe.Ih. 1971. Role of Sialic Acid in Determining Survival of Glycoproteins in Circulation. *Journal of Biological Chemistry*. 246 (5), pp1461-&.

Moreira, I.S., Fernandes, P.A. and Ramos, M.J. 2007. Hot spots-A review of the protein-protein interface determinant amino-acid residues. *Proteins-Structure Function and Bioinformatics,* 68(4), pp.803-812.

Morrison, K.L. and Weiss, G.A. 2001. Combinatorial alanine-scanning. *Current Opinion in Chemical Biology,* 5(3), pp.302-307.

Moser, F., Irwin, D., Chen, S. and Wilson, D. B. 2008. Regulation and characterization of Thermobifida fusca carbohydrate-binding module proteins E7 and E8. *Biotechnology and bioengineering*. 100 (6), pp1066-1077.

Muchmore, E. A., Diaz, S. and Varki, A. 1998. A structural difference between the cell surfaces of humans and the great apes. *American Journal of Physical Anthropology*. 107 (2), pp187-198.

Mullis, K. B. and Faloona, F. A. 1987. Specific Synthesis of Dna Invitro Via a Polymerase-Catalyzed Chain-Reaction. *Methods in enzymology*. 155pp335-350.

Muntoni, F., Brockington, M., Blake, D. J. 2002. Defective glycosylation in muscular dystrophy. *Lancet*. 360 (9343), pp1419-1421.

Muzzarelli, R. A. A. 1997. Human enzymatic activities related to the therapeutic administration of chitin derivatives. *Cellular and Molecular Life Sciences*. 53 (2), pp131-140.

Muzzarelli, R. A. A. 2010. Chitins and Chitosans as Immunoadjuvants and Non-Allergenic Drug Carriers. *Marine Drugs*. 8 (2), pp292-312.

Muzzarelli, R. A. A. 2009. Chitins and chitosans for the repair of wounded skin, nerve, cartilage and bone. *Carbohydrate Polymers*. 76 (2), pp167-182.

Oh, K., Kim, Y., Nguyen, V. N. 2007. Demineralization of crab shell waste by Pseudomonas aeruginosa F722. *Process Biochemistry*. 42 (7), pp1069-1074.

Okawa, Y., Kobayashi, M., Suzuki, S. and Suzuki, M. 2003. Comparative study of protective effects of chitin, chitosan, and N-acetyl chitohexaose against Pseudomonas aeruginosa and Listeria monocytogenes infections in mice. *Biological & Pharmaceutical Bulletin*. 26 (6), pp902-904.

Orlean, P. 1987. 2 Chitin Synthases in Saccharomyces-Cerevisiae. *Journal of Biological Chemistry*. 262 (12), pp5732-5739.

Otto, D.M.E., Campanero-Rhodes, M.A., Karamanska, R., Powell, A.K., Bovin, N., Turnbull, J.E., Field, R.A., Blackburn, J., Feizi, T. and Crocker, P.R. An expression system for screening of proteins for glycan and protein interactions. *Analytical Biochemistry*, 441 (2), pp261-270.

Paulson, J. C. 1989. Glycoproteins - what are the Sugar Chains for. *Trends in biochemical sciences*. 14 (7), pp272-276.

Perez-Vilar, J. and Hill, R. L. 1999. The structure and assembly of secreted mucins. *Journal of Biological Chemistry*. 274 (45), pp31751-31754.

Perrakis, A., Tews, I., Dauter, Z. 1994. Crystal-Structure of a Bacterial Chitinase at 2.3-Angstrom Resolution. *Structure*. 2 (12), pp1169-1180.

Pugsley, A.P. 1993. The Complete General Secretory Pathway in Gram-Negative Bacteria. *Microbiological Reviews,* 57(1), pp.50-108.

Rademacher, T. W., Parekh, R. B. and Dwek, R. A. 1988. Glycobiology. *Annual Review of Biochemistry*. 57pp785-838.

Rapoport, E. M., Kurmyshkina, O. V. and Bovin, N. V. 2008. Mammalian galectins: Structure, carbohydrate specificity, and functions. *Biochemistry-Moscow*. 73 (4), pp393-405.

Reis, C. A., Osorio, H., Silva, L. 2010. Alterations in glycosylation as biomarkers for cancer detection. *Journal of clinical pathology*. 63 (4), pp322-329.

Rhim, A. D., Stoykova, L., Glick, M. C. and Scanlin, T. F. 2001. Terminal glycosylation in cystic fibrosis (CF): A review emphasizing the airway epithelial cell. *Glycoconjugate journal*. 18 (9), pp649-659.

Rinaudo, M. 2006. Chitin and chitosan: Properties and applications. *Progress in Polymer Science*. 31 (7), pp603-632.

Rodda, D.J. and Yamazaki, H. 1994. Poly(vinyl Alcohol) as a Blocking-Agent in Enzyme Immunoassays. *Immunological Investigations*, 23(6-7), pp.421-428.

Romaguera, A., Menge, U., Breves, R. and Diekmann, H. 1992. Chitinases of Streptomyces-Olivaceoviridis and Significance of Processing for Multiplicity. *Journal of Bacteriology*. 174 (11), pp3450-3454.

Rosas, A.L., Nosanchuk, J.D., Gomez, B.L., Edens, W.A., Henson, J.M. and Casadevall, A. 2000. Isolation and serological analyses of fungal melanins. *Journal of Immunological Methods,* 244(1-2), pp.69-80.

Rosenfeld, R., Bangio, H., Gerwig, G. J. 2007. A lectin array-based methodology for the analysis of protein glycosylation. *Journal of Biochemical and Biophysical Methods*. 70 (3), pp415-426.

Rudd, P. M., Elliott, T., Cresswell, P. 2001. Glycosylation and the immune system. *Science*. 291 (5512), pp2370-2376.

Ruiz-Herrera, J., González-Prieto, J. M. and Ruiz-Medrano, R. 2002. Evolution and phylogenetic relationships of chitin synthases from yeasts and fungi. *FEMS Yeast Research*. 1 (4), pp247-256.

Ruiz-Herrera, J. and Ortiz-Castellanos, L. 2010. Analysis of the phylogenetic relationships and evolution of the cell walls from yeasts and fungi. *FEMS Yeast Research*. 10 (3), pp225-243.

Rusch, S.L. and Kendall, D.A. 2007. Interactions that drive Sec-dependent bacterial protein transport. *Biochemistry,* 46(34), pp.9665-9673.

Ryan, B. 2006. Site Directed Mutagenesis Studies of Horseradish Peroxidase. PhD thesis. Dublin City University.

Rye, P.D. and Bovin, N.V. 1998. Carbohydrate affinity PAGE for the study of carbohydrate-binding proteins. *BioTechniques,* 25(1), pp.146-151.

Sahai, A. S. and Manocha, M. S. 1993. Chitinases of Fungi and Plants - their Involvement in Morphogenesis and Host-Parasite Interaction. *FEMS microbiology reviews*. 11 (4), pp317-338.

Saito, A., Miyashita, K., Biukovic, G. and Schrempf, H. 2001. Characteristics of a Streptomyces coelicolor A3(2) extracellular protein targeting chitin and chitosan. *Applied and Environmental Microbiology*. 67 (3), pp1268-1273.

Sanders, N. N., Eijsink, V. G. H., van den Pangaart, P. S. 2007. Mucolytic activity of bacterial and human chitinases. *Biochimica Et Biophysica Acta-General Subjects*. 1770 (5), pp839-846.

Sasaki, C., Varum, K. M., Itoh, Y. 2006. Rice chitinases: sugar recognition specificities of the individual subsites. *Glycobiology*. 16 (12), pp1242-1250.

Scanlin, T. F. and Glick, M. C. 2000. Terminal glycosylation and disease: Influence on cancer and cystic fibrosis. *Glycoconjugate journal*. 17 (7-9), pp617-626.

Schaffer, C., Graninger, M. and Messner, P. 2001. Prokaryotic glycosylation. *Proteomics*. 1 (2), pp248-261.

Schaffer, C. and Messner, P. 2001. Glycobiology of surface layer proteins. *Biochimie*. 83 (7), pp591-599.

Schirm, M., Kalmokoff, M., Aubry, A. 2004. Flagellin from Listeria monocytogenes is glycosylated with beta-O-linked N-acetylglucosamine. *Journal of Bacteriology*. 186 (20), pp6721-6727.

Schirm, M., Soo, E. C., Aubry, A. J. 2003. Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori. *Molecular microbiology*. 48 (6), pp1579-1592.

Schnellmann, J., Zeltins, A., Blaak, H. and Schrempf, H. 1994. The Novel Lectin-Like Protein Chb1 is Encoded by a Chitin-Inducible Streptomyces-Olivaceoviridis Gene and Binds Specifically to Crystalline Alpha-Chitin of Fungi and Other Organisms. *Molecular microbiology*. 13 (5), pp807-819.

Schwartz, A. L. 1984. The Hepatic Asialoglycoprotein Receptor. *CRC critical reviews in biochemistry*. 16 (3), pp207-233.

Seidl, V., Huemer, B., Seiboth, B. and Kubicek, C. P. 2005. A complete survey of Trichoderma chitinases reveals three distinct subgroups of family 18 chitinases. *Febs Journal*. 272 (22), pp5923-5939.

Serganova, I. S., Polosina, Y. Y., Kostyukova, A. S. 1995. Flagella of Halophilic Archaea - Biochemical and Genetic-Analysis. *Biochemistry-Moscow*. 60 (8), pp953-957.

Sethuraman, N. and Stadheim, T. A. 2006. Challenges in therapeutic glycoprotein production. *Current opinion in biotechnology*. 17 (4), pp341-346.

Sharon, N. and Lis, H. 2004. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*. 14 (11), pp53R-62R.

Shen, Z. and Jacobs-Lorena, M. 1998. A Type I Peritrophic Matrix Protein from the Malaria Vector Anopheles gambiae Binds to Chitin. CLONING, EXPRESSION, AND CHARACTERIZATION. *J. Biol. Chem.* 273 (28), pp17665-17670

Shepherd, V. L., Campbell, E. J., Senior, R. M. and Stahl, P. D. 1982. Characterization of the Mannose Fucose Receptor on Human Mononuclear Phagocytes. *Journal of the Reticuloendothelial Societ*y. 32 (6), pp423-431.

Sheridan, C. 2007. Commercial interest grows in glycan analysis. *Nature biotechnolog*y. 25 (2), pp145-146.

Shevchik, V.E., RobertBaudouy, J. and Condemine, G. 1997. Specific interaction between OutD, an Erwinia chrysanthemi outer membrane protein of the general secretory pathway, and secreted proteins. *Embo Journal*, 16(11), pp.3007-3016.

Shokri, A., Sanden, A.M. and Larsson, G. 2003. Cell and process design for targeting of recombinant protein into the culture medium of Escherichia coli. *Applied Microbiology and Biotechnology*, 60(6), pp.654-664.

Schrempf, H. 2001. Recognition and degradation of chitin by streptomycetes. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology*, 79(3-4), pp.285-289.

Sikorski, P., Hori, R. and Wada, M. 2009. Revisit of alpha-Chitin Crystal Structure Using High Resolution X-ray Diffraction Data. *Biomacromolecules*. 10 (5), pp1100-1105.

Singh, R. S., Tiwary, A. K. and Kennedy, J. F. 1999. Lectins: Sources, activities, and applications. *Critical Reviews in Biotechnology*. 19 (2), pp145-178.

Sleytr, U. B. and Thorne, K. J. I. 1976. Chemical Characterization of Regularly Arranged Surface-Layers of Clostridium-Thermosaccharolyticum and Clostridium-Thermohydrosulfuricum. *Journal of Bacteriology*. 126 (1), pp377-383.

Smart, J. D., Nantwi, P. K. K., Rogers, D. J. and Green, K. L. 2002. A quantitative evaluation of radiolabelled lectin retention on oral mucosa in vitro and in vivo. *European Journal of Pharmaceutics and Biopharmaceutics*. 53 (3), pp289-292.

Smith, A. C., de Wolff, J. F., Molyneux, K. 2006. O-glycosylation of serum IgD in IgA nephropathy. *Journal of the American Society of Nephrology*. 17 (4), pp1192-1199.

Sorensen, A. L., Rumjantseva, V., Nayeb-Hashemi, S. 2009. Role of sialic acid for platelet life span: exposure of beta-galactose results in the rapid clearance of platelets from the circulation by asialoglycoprotein receptor-expressing liver macrophages and hepatocytes. *Blood*. 114 (8), pp1645-1654.

Spik, G., Bayard, B., Fournet, B., Strecker, G., Bouquelet, S. and Montreuil, J. 1975. Studies on glycoconjugates. LXIV. Complete structure of two carbohydrate units of human serotransferrin. *FEBS Letters*, 50(3), pp.296-299.

Spik, G., Debruyne, V., Montreuil, J., van Halbeek, H. and Vliegenthart, J.F.G. 1985. Primary structure of two sialylated triantennary glycans from human serotransferrin. *FEBS Letters*, 183(1), pp.65-69.

Spiro, R.G. 1965. Carbohydrate Units of Thyroglobulin. *Journal of Biological Chemistry*, 240(4), pp.1603-&.

Spiro, R. G. 2002. Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology*. 12 (4), pp43R-56R.

Stahl, P. and Schlesinger, P. 1979. Mannose-N-Acetylglucosamine Receptor - Plasma Clearance and Macrophage Uptake of Glycoconjugates and Lysosomal Glycosidases. *Federation proceedings*. 38 (3), pp467-467.

Steiner, D., Forrer, P., Stumpp, M.T. and Pluckthun, A. 2006. Signal sequences directing cotranslational translocation expand the range of proteins amenable to phage display. *Nature Biotechnology*, 24(7), pp.823-831.

Stephenson, K. 2005. Sec-dependent protein translocation across biological membranes: evolutionary conservation of an essential protein transport pathway (Review). *Molecular Membrane Biology*, 22(1-2), pp.17-28.

Stoykova, L. I. and Scanlin, T. F. 2008. Cystic fibrosis (CF), Pseudomonas aeruginosa, CFTR and the CF glycosylation phenotype: A review and update. *Current Organic Chemistry*. 12 (11), pp900-910.

Studentsov, Y.Y., Schiffman, M., Strickler, H.D., Ho, G.Y.F., Pang, Y.Y.S., Schiller, J., Herrero, R. and Burk, R.D. 2002. Enhanced enzyme-linked immunosorbent assay for detection of antibodies to virus-like particles of human papillomavirus. *Journal of Clinical Microbiology*, 40(5), pp.1755-1760.

Sun, W., Dai, X., Zheng, Y., Wang, J., Hou, L., Du, J. and Hu, H. 2011. On-column refolding purification of DT389-hIL13 recombinant protein expressed in Escherichia coli. *Protein Expression and Purification,* 75(1), pp.83-88.

Suzuki, K., Sugawara, N., Suzuki, M. 2002. Chitinases A, B, and C1 of Serratia marcescens 2170 produced by recombinant Escherichia coli: Enzymatic properties and synergism on chitin degradation. *Bioscience Biotechnology and Biochemistry*. 66 (5), pp1075-1083.

Suzuki, K., Suzuki, M., Taiyoji, M. 1998. Chitin binding protein (CBP21) in the culture supernatant of Serratia marcescens 2170. *Bioscience Biotechnology and Biochemistry*. 62 (1), pp128-135.

Suzuki, K., Taiyoji, M., Sugawara, N. 1999. The third chitinase gene (chiC) of Serratia marcescens 2170 and the relationship of its product to other bacterial chitinases. *Biochemical Journal*. 343, pp587-596.

Svergun, D. I., Becirevic, A., Schrempf, H. 2000. Solution structure and conformational changes of the Streptomyces chitin-binding protein (CHB1). *Biochemistry*. 39 (35), pp10677-10683.

Takaya, N., Yamazaki, D., Horiuchi, H. 1998. Intracellular chitinase gene from Rhizopus oligosporus: molecular cloning and characterization. *Microbiology-Sgm*. 144, pp2647-2654.

Tarentino, A.L. and Plummer Jr., T.H. 1994. Enzymatic deglycosylation of asparagine-linked glycans: Purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum *IN:* William J. Lennarz, G.W.H. (ed.) *Methods in Enzymology Vol 4.* Academic Press, pp.44-57.

Taylor, P. R., Gordon, S. and Martinez-Pomares, L. 2005. The mannose receptor: linking homeostasis and immunity through sugar recognition. *Trends in immunology*. 26 (2), pp104-110.

Techkarnjanaruk, S. and Goodman, A. E. 1999. Multiple genes involved in chitin degradation from the marine bacterium Pseudoalteromonas sp. strain S91. *Microbiology-Sgm*. 145, pp925-934.

Thompson, S. E., Smith, M., Wilkinson, M. C. and Peek, K. 2001. Identification and characterization of a chitinase antigen from Pseudomonas aeruginosa strain 385. *Applied and Environmental Microbiology*. 67 (9), pp4001-4008.

Tsioptsias, C., Tsivintzelis, I., Papadopoulou, L. and Pallayiotou, C. 2009. A novel method for producing tissue engineering scaffolds from chitin, chitin-hydroxyapatite, and cellulose. *Materials Science & Engineering C-Biomimetic and Supramolecular Systems*. 29 (1), pp159-164.

Tsujibo, H., Orikoshi, H., Baba, N. 2002. Identification and characterization of the gene cluster involved in chitin degradation in a marine bacterium, Alteromonas sp strain O-7. *Applied and Environmental Microbiology*. 68 (1), pp263-270.

Ubhayasekera, W., Tang, C. M., Ho, S. W. T. 2007. Crystal structures of a family 19 chitinase from Brassica juncea show flexibility of binding cleft loops. *Febs Journal*. 274 (14), pp3695-3703.

Updyke, T. 1984. Immunoaffinity isolation of membrane antigens with biotinylated monoclonal antibodies and immobilized streptavidin matrices. *Journal of Immunological Methods*, 71(1), pp.83.

Upreti, R. K., Kumar, M. and Shankar, V. 2003. Bacterial glycoproteins: Functions, biosynthesis and applications. *Proteomics*. 3 (4), pp363-379.

Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M. F. 2005a. The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation. *Journal of Biological Chemistry*. 280 (31), pp28492-28497.

Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H. K. 2005b. Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21. *Journal of Biological Chemistry*. 280 (12), pp11313-11319.

Vaaje-Kolstad, G., Bunaes, A. C., Mathiesen, G. and Eijsink, V. G. H. 2009. The chitinolytic system of Lactococcus lactis ssp lactis comprises a nonprocessive chitinase and a chitin-binding protein that promotes the degradation of alpha- and beta-chitin. *Febs Journal*. 276 (8), pp2402-2415.

Van Damme, E. J. M., Peumans, W. J., Barre, A. and Rouge, P. 1998. Plant lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Reviews in Plant Sciences*. 17 (6), pp575-692.

Vanderschaeghe, D., Festjens, N., Delanghe, J. and Callewaert, N. 2010. Glycome profiling using modern glycomics technology: technical aspects and applications. *Biological chemistry*. 391 (2-3), pp149-161.

Varki, A. 1993. Biological Roles of Oligosaccharides - all of the Theories are Correct. *Glycobiology*. 3 (2), pp97-130.

Veronico, P., Gray, L. J., Jones, J. T. 2001. Nematode chitin synthases: gene structure, expression and function in Caenorhabditis elegans and the plant parasitic nematode Meloidogyne artiellia. *Molecular Genetics and Genomics*. 266 (1), pp28-34.

Virji, M. 1997. Post-translational modifications of meningococcal pili. Identification of common substituents: Glycans and alpha-glycerophosphate - A review. *Gene*. 192 (1), pp141-147.

Virji, M., Saunders, J. R., Sims, G. 1993. Pilus-Facilitated Adherence of Neisseria-Meningitidis to Human Epithelial and Endothelial-Cells - Modulation of Adherence Phenotype Occurs Concurrently with Changes in Primary Amino-Acid-Sequence and the Glycosylation Status of Pilin. *Molecular microbiology*. 10 (5), pp1013-1028.

Viaplana, E., Rebordosa, X., Pinol, J. and Villaverde, A. 1997. Secretion-dependent proteolysis of recombinant proteins is associated with inhibition of cell growth in Escherichia coli. *Biotechnology Letters,* 19(4), pp.373-377.

Vogt Jr., R.V., Phillips, D.L., Omar Henderson, L., Whitfield, W. and Spierto, F.W. 1987. Quantitative differences among various proteins as blocking agents for ELISA microtiter plates. *Journal of Immunological Methods,* 101(1), pp.43-50.

Wanchoo, A., Lewis, M. W. and Keyhani, N. O. 2009. Lectin mapping reveals stage-specific display of surface carbohydrates in in vitro and haemolymph-derived cells of the entomopathogenic fungus Beauveria bassiana. *Microbiology-Sgm*. 155pp3121-3133.

Wang, D. N. 2003. Carbohydrate microarrays. *Proteomics*. 3 (11), pp2167-2175.

Watanabe, T., Kanai, R., Kawase, T. 1999. Family 19 chitinases of Streptomyces species: characterization and distribution. *Microbiology-Uk*. 145pp3353-3363.
Watanabe, T., Kimura, K., Sumiya, T. 1997. Genetic analysis of the chitinase system of Serratia marcescens 2170. *Journal of Bacteriology*. 179 (22), pp7111-7117.

Ward, A., Sanderson, N.M., O'Rielly, J. 2000. The amplification, expression, identification, purification, assay, and properties of hexahistidine-tagged bacterial membrane transport proteins. In: Baldwin SA, ed. Membrane Transport–A Practical Approach. Oxford: Oxford University Press 1999: PP141-166.

Watanabe, T., Kobori, K., Miyashita, K. 1993. Identification of Glutamic Acid-204 and Aspartic Acid-200 in Chitinase-A1 of Bacillus-Circulans Wl-12 as Essential Residues for Chitinase Activity. *Journal of Biological Chemistry*. 268 (25), pp18567-18572.

Weis, W. I. and Drickamer, K. 1996. Structural basis of lectin-carbohydrate recognition. *Annual Review of Biochemistry*. 65pp441-473.

Weitz-Schmidt, G., Stokmaier, D., Scheel, G., Nifant'ev, N.E., Tuzikov, A.B. and Bovin, N.V. 1996. An E-Selectin Binding Assay Based on a Polyacrylamide-Type Glycoconjugate. *Analytical Biochemistry,* 238(2), pp.184-190.

Wells, J.A. 1990. Additivity of Mutational Effects in Proteins. *Biochemistry,* 29(37), pp.8509-8517.

Werner, S., Steiner, U., Becher, R. 2002. Chitin synthesis during in planta growth and asexual propagation of the cellulosic oomycete and obligate biotrophic grapevine pathogen Plasmopara viticola. *FEMS microbiology letters*. 208 (2), pp169-173.

Werz, D. B., Ranzinger, R., Herget, S. 2007. Exploring the structural diversity of mammalian carbohydrates ("Glycospace") by statistical databank analysis. *Acs Chemical Biology*. 2 (10), pp685-691.

Wessels, J. G. H. 1993. Wall Growth, Protein Excretion and Morphogenesis in Fungi. *New Phytologist*. 123 (3), pp397-413.

Winterburn and Phelps, C. F. 1972. Significance of Glycosylated Proteins. *Nature*. 236 (5343), pp147.

Whitaker, J. R. 1963. Determination of Molecular Weights of Proteins by Gel Filtration on Sephadex. *Analytical Chemistry*. 35 (12), pp1950.

Wongmadden, S.T. and Landry, D. 1995. Purification and Characterization of Novel Glycosidases from the Bacterial Genus Xanthomonas. *Glycobiology,* 5(1), pp.19-28.

Wright, C. 1992. Crystal-Structure of a Wheat-Germ-Agglutinin Glycophorin-Sialoglycopeptide Receptor Complex - Structural Basis For Cooperative Lectin-Cell Binding. *The Journal of Biological Chemistry*, 267(20), pp.14345.

Wu, G. H., Lu, Z. H., Xie, X. and Ledeen, R. W. 2004. Susceptibility of cerebellar granule neurons from GM2/GD2 synthase-null mice to apoptosis induced by glutamate excitotoxicity and elevated KCl: Rescue by GM1 and LIGA20. *Glycoconjugate journal*. 21 (6), pp305-313.

Yu, C., Lee, A. M., Bassler, B. L. and Roseman, S. 1991. Chitin Utilization by Marine-Bacteria - a Physiological-Function for Bacterial Adhesion to Immobilized Carbohydrates. *Journal of Biological Chemistry*. 266 (36), pp24260-24267.

Zeltins, A. and Schrempf, H. 1997. Specific interaction of the Streptomyces chitin-binding protein CHB1 with alpha-chitin - The role of individual tryptophan residues. *European Journal of Biochemistry*. 246 (2), pp557-564.

Zhang, H.Z. and Donnenberg, M.S. 1996. DsbA is required for stability of the type IV pilin of enteropathogenic Escherichia coli. *Molecular Microbiology,* 21(4), pp.787-797.

Zhang, L., Moo-Young, M. and Chou, C.P. 2011. Molecular manipulation associated with disulfide bond formation to enhance the stability of recombinant therapeutic protein. *Protein Expression and Purification,* 75(1), pp.28-39.