

A PROTEOMIC ANALYSIS OF FIBRE
DEGRADATION AND ASSIMILATION BY
BUTYRIVIBRIO PROTEOCLASTICUS

by

Jonathan Craig Dunne

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Cell and Molecular Biosciences

Victoria University of Wellington

2010

Abstract

Butyrivibrio proteoclasticus B316^T is a Gram-positive, lignocellulose degrading bacterium that is prevalent in the rumen of animals grazing pasture, and is one of only a few rumen microbes known to degrade and utilise xylan *in vitro*. Xylan is a hemicellulose that comprises up to 45% of the polysaccharide component of ruminant forages. Often as little as 30% of the total energy content of forages is utilised by the ruminant due to poor hemicellulose degradation by the fibrolytic rumen microbes. An opportunity exists to improve forage degradation in the rumen, which is predicted to improve the productivity of forage fed ruminants. A clearer understanding of the strategies employed by fibrolytic rumen microbes to degrade and utilise lignocellulose is important in realising this goal.

Almost 10% of the *B. proteoclasticus* genome encodes proteins involved in polysaccharide metabolism and transport, which includes 134 fibrolytic enzymes that are active upon plant fibre. Many of these are clustered into one of 36 polysaccharide utilisation loci that also contain transmembrane transporters, transcriptional regulators, environmental sensors and genes involved in further polysaccharide metabolism. Gel-based and gel-free proteomic analyses of the cytosolic, cell-associated, and secreted fractions of cells grown on xylan were used to identify proteins involved in the degradation, assimilation, and metabolism of hemicellulose. A set of 416 non-redundant proteins were identified, which included 12 extracellular and 24 cytosolic polysaccharidases, and 59 proteins involved in the uptake and further metabolism of polysaccharide degradation products, many of which were substrate-binding protein components of ATP-driven transporter systems. In cells grown on xylan, several of these proteins displayed significant protein abundance changes relative to cells grown on the monomeric sugar xylose, in a pattern that reflected the growth substrates used.

A model of xylan degradation by *B. proteoclasticus* based on these results hypothesises that *B. proteoclasticus* attacks the xylan backbone and main substituent groups of hemicellulose in the extracellular space, assimilates the xylooligosaccharides and performs the final stages of degradation within the cell. These results provide insight into a xylan degrading enzyme system that has evolved to efficiently degrade and utilise hemicellulose, extend our understanding of the enzymes that are likely to play important roles in hemicellulose degradation, and support the notion that *Butyrivibrio* species are important contributors to rumen fibre degradation.

Acknowledgements

First and foremost I would like to say a huge thank you to Dr. Bill Jordan, for your guidance, support, and unending patience. Your faith in me throughout the whole project, and especially during the writing up period means a great deal. I would also like to say a big thanks to Dr. Graeme Attwood, for your supervision and support, and giving me the opportunity to take on the project in the first place.

I would like to make special acknowledgement of the work of Dong Li, who grew the cultures, and harvested and fractionated the bacteria, in several instances at very short notice. Thank you very much. And thanks to Dr. Judy Bond for providing guidance and assistance in the early part of my project, for access to the membrane proteome data, and your helpful discussions.

I am extremely grateful for the assistance of Dr. Bill Kelly. Your guidance in several areas of the data analysis has been invaluable, and our discussions during the project have always given me considerable motivation. Your attention to detail during the proofreading and corrections stage, and your critical comments and suggestions has always been gratefully received. Thanks also to Dr. Sinead Leahy, for providing CAI values, genomic and protein subcellular location data, which helped a great deal with the bioinformatic analysis.

To Dr. Lifeng Peng, thank you very much for your help with the gel-free data collection and analysis. It's safe to say I couldn't have done that without you. Even more so, thanks for all the conversations over the years, project and non-project related. You are a very good travel companion.

Thank you to members of Bill's lab. Especially to Dr. Cliff Young for so many hours of enlightening conversation, and teaching me everything you knew about mass spectrometry. To Dr. Pisana Rawson for helping me to keep things in perspective, and Danyl McLauchlan for making sure the computers always did as they were supposed to. Thanks also to Hannah and Sarah for livening up the lab over the past year or so.

Thanks to Anne-Marie Knibbeler for your very thorough proofreading.

I would like to gratefully acknowledge the financial support of AgResearch, Victoria University of Wellington, and the Foundation for Research, Science, and Technology, which has allowed me complete this study.

To Mum, Dad, and Doshie. Thank you for everything! None of this would have been possible without your support and encouragement. I promise I'll have more faith.

And to Christine. Thank you. I can't wait to start our 'normal' life together.

Table of contents

Abstract	iii
Acknowledgements.....	v
Table of contents	vii
List of Tables	xiii
List of Figures	xv
Abbreviations	xx
Chapter 1 Introduction.....	1
1.1 Overview.....	2
1.2 The ruminant digestive tract.....	2
1.2.1 Ruminant microbial symbiosis	3
1.2.2 Factors affecting ruminant digestion and energy supply	4
1.3 The plant cell wall	5
1.3.1 Type-I and Type-II plant cell walls.....	5
1.3.2 Cellulose	5
1.3.3 Hemicellulose	8
1.3.3.1 Glucuronoarabinoxylan	8
1.3.3.2 Xyloglucan	9
1.3.3.3 Mannans and glucomannans	9
1.3.3.4 Mixed linkage glucans	10
1.3.4 Pectin.....	10
1.3.5 Cell wall structural proteins.....	11
1.3.6 Lignin	11
1.4 Lignocellulose degradation	12
1.4.1 Cellulose degradation.....	13
1.4.2 Hemicellulose degradation	15
1.4.3 Pectin degradation.....	15
1.5 Recalcitrance of rumen forages to microbial degradation	16
1.5.1.1 Diferulate cross-linking within grass hemicellulose	16
1.5.1.2 Diferulate cross-links between grass hemicellulose and lignin	17
1.5.1.3 The impact of feruloylation on lignocellulose degradation	18
1.6 The rumen microbial ecosystem	19
1.6.1 Rumen microbial diversity	19
1.6.2 The fibre degrading rumen microbiome	20
1.6.3 The hemicellulolytic rumen bacteria	22
1.6.3.1 <i>Butyrivibrio fibrisolvens</i>	22

1.6.3.2 <i>Prevotella bryantii</i> and <i>Prevotella ruminicola</i>	24
1.6.4 The cellulolytic rumen bacteria.....	25
1.6.4.1 <i>Ruminococcus flavefaciens</i> , <i>Ruminococcus albus</i> , and <i>Fibrobacter succinogenes</i>	25
1.6.4.2 <i>Eubacterium cellulosolvens</i>	26
1.7 <i>Butyrivibrio proteoclasticus</i> B316^T.....	27
1.7.1 Growth substrate utilisation and fermentation	28
1.7.2 The <i>B. proteoclasticus</i> genome sequence	29
1.7.2.1 The <i>B. proteoclasticus</i> glycobiome	32
1.7.2.2 Polysaccharide utilisation loci	32
1.8 Proteomics	33
1.8.1 Two-dimensional electrophoresis	34
1.8.2 Mass spectrometry	34
1.8.2.1 Peptide mass fingerprinting	35
1.8.2.2 Liquid chromatography-tandem mass spectrometry	36
1.8.3 Bacterial proteomics.....	37
1.8.3.1 Proteomics of fibrolytic bacteria	37
1.8.3.2 Rumen proteomics	38
1.9 The theoretical total proteome of <i>B. proteoclasticus</i>	38
1.9.1 Protein function classification of the predicted <i>B. proteoclasticus</i> proteome.....	39
1.9.2 Subcellular distribution of the predicted <i>B. proteoclasticus</i> proteome.....	40
1.10 Project aims.....	41
Chapter 2 Materials and Methods	43
2.1 Materials.....	44
2.2 Bacterial strains and culture conditions.....	45
2.2.1 Modified M704 medium.....	45
2.3 <i>B. proteoclasticus</i> subcellular fractionation.....	46
2.3.1 Secreted protein extraction	46
2.3.2 Optimisation of <i>B. proteoclasticus</i> cell-associated protein extraction	46
2.3.3 <i>B. proteoclasticus</i> cytosolic protein extraction	47
2.3.3.1 <i>B. proteoclasticus</i> soluble protein extraction.....	47
2.4 Sample purification.....	48
2.4.1 Optimisation of <i>B. proteoclasticus</i> secreted protein purification.....	48
2.4.1.1 Protein precipitation using TCA	48
2.4.1.2 Protein precipitation using ammonium sulphate.....	49
2.4.1.3 Centrifugal concentration	49
2.4.2 <i>B. proteoclasticus</i> cell-associated protein purification.....	49
2.4.3 Optimisation of <i>B. proteoclasticus</i> cytosolic protein purification.....	50

2.5 2-DE of <i>B. proteoclasticus</i> cytosolic subcellular fractions	50
2.5.1 Rehydration and IEF of pI 3-5.6 NL, 5.3-6.5, and 3-10 NL IPG strips.....	51
2.5.2 Rehydration and IEF of pI 6-11 IPG strips.....	52
2.5.3 IPG strip equilibration and second-dimension electrophoresis.....	52
2.6 Protein detection.....	52
2.6.1 Colloidal CBB G-250 total protein staining	53
2.6.2 Flamingo™ fluorescent total protein staining	53
2.7 Image acquisition	53
2.7.1 Colloidal CBB G-250 stained gel image acquisition	53
2.7.2 Flamingo™ stained gel image acquisition.....	54
2.8 Image analysis	54
2.8.1 Protein spot detection	54
2.8.2 Protein spot editing	54
2.9 2-DE gel image matching strategy	55
2.9.1 Technical replicate gel image matching	55
2.9.2 Biological replicate gel image matching	56
2.10 Differential protein abundance analysis	56
2.10.1 Data extraction, missing values and data imputation	56
2.10.2 Raw spot volume normalisation.....	57
2.10.3 Data transformation.....	57
2.10.4 Statistical analysis assumption testing.....	58
2.10.5 Parametric statistical testing	58
2.10.6 Non-parametric statistical testing.....	59
2.10.7 Protein abundance fold change calculation	60
2.11 Protein digestion and peptide purification	60
2.12.1 Sample preparation.....	61
2.12.2 Peptide mass fingerprinting	62
2.12.3 PMF processing and database interrogation	62
2.12.4 Positive protein match criteria	63
2.13 Reverse phase HPLC and MS/MS (1-D LC-MS/MS)	64
2.13.1 1-DE LC-MS/MS	64
2.13.2 MS/MS data analysis.....	65
2.14 Polysaccharide adsorption assay	65
2.14.1 Substrate preparation.....	65
2.14.2 <i>B. proteoclasticus</i> secreted proteome polysaccharide adsorption assay	66
2.15 Bioinformatics	66

Chapter 3 The extracellular <i>B. proteoclasticus</i> proteome.....	69
3.1 Introduction.....	70
3.1.1 Secreted polysaccharidases	70
3.1.1.1 Carbohydrate-binding modules	71
3.1.2 Uptake of lignocellulose derived soluble sugars-ABC transporters.....	71
3.2 Results	73
3.2.1 Theoretical 2-DE	73
3.2.2 Gel-based analysis of the <i>B. proteoclasticus</i> secreted proteome.....	75
3.2.2.1 Optimisation of <i>B. proteoclasticus</i> secreted protein purification.....	75
3.2.2.2 2-DE of proteins recovered from the <i>B. proteoclasticus</i> culture medium	76
3.2.3 The effect of growth substrate on the secreted proteome of <i>B. proteoclasticus</i>	87
3.2.3.1 Abundance patterns of secreted polysaccharidases	87
3.2.3.2 Abundance patterns of non-carbohydrate degrading secreted proteins	96
3.2.4 Gel-based analysis of the <i>B. proteoclasticus</i> cell-associated proteome.....	97
3.2.4.1 Optimisation of <i>B. proteoclasticus</i> cell-associated protein extraction	97
3.2.4.2 2-DE of proteins recovered from the <i>B. proteoclasticus</i> cell-associated extract	99
3.2.5 Cellulose affinity assay of <i>B. proteoclasticus</i> secreted proteins	108
3.2.6 Bioinformatics of the polysaccharidases and differentially abundant sugar substrate-binding proteins identified in the <i>B. proteoclasticus</i> extracellular proteome .	115
3.3 Summary	122
Chapter 4 The cytosolic <i>B. proteoclasticus</i> proteome	123
4.1 Introduction.....	124
4.2 Results	124
4.2.1 The theoretical cytosolic proteome of <i>B. proteoclasticus</i>	124
4.2.2 Gel-based analysis of the <i>B. proteoclasticus</i> cytosolic proteome	125
4.2.2.1 2-DE spot detection replicate gel matching	126
4.2.2.2 Polysaccharide degrading enzymes identified by MALDI-TOF in the <i>B. proteoclasticus</i> cytosolic proteome	129
4.2.3 Gel-free analysis of the <i>B. proteoclasticus</i> cytosolic proteome	136
4.2.3.1 1-D LC-MS/MS protein identification and control of false positive identification rate.....	137
4.2.3.2 Polysaccharide degrading enzymes identified by 1-D LC-MS/MS	139
4.2.4 Bioinformatics of the identified <i>B. proteoclasticus</i> cytosolic proteome.....	142
4.2.4.1 Subcellular distribution of the polysaccharidases identified in the <i>B. proteoclasticus</i> cytosol	142
4.2.4.2 Functional domains in the cytosolic polysaccharidases	143
4.2.4.3 BLASTp analysis of the polysaccharidases identified in the cytosolic <i>B. proteoclasticus</i> proteome	147

4.2.4.4 The predicted function of all identified cytosolic proteins	147
4.2.4.5 Comparative analysis of the gel-based and gel-free datasets.....	152
4.2.4.6 Codon adaptation index, protein pI, and protein identification	155
4.3 Summary	157
Chapter 5 The effect of growth substrate on the <i>B. proteoclasticus</i> cytosolic proteome.....	159
5.1 Introduction.....	160
5.2 Statistical analysis of differential protein abundance patterns	160
5.2.1 Narrow pI range analysis of the <i>B. proteoclasticus</i> cytosolic proteome	160
5.2.2 2-DE spot volume normalisation	162
5.2.3 Distribution of missing values in 2-DE	163
5.2.4 Missing value substitution rate in 2-DE	164
5.2.5 2-DE data transformation	165
5.2.6 Multiple hypothesis testing and controlling the false discovery rate	167
5.3 Relative abundance analysis of the mid-log phase cytosolic proteome by 2-DE MALDI-TOF	170
5.4 Relative abundance analysis of the stationary phase cytosolic proteome by 2-DE.....	177
5.5 Relative abundance analysis of the mid-log phase cytosolic proteome by 1-D LC-MS/MS	183
5.6 Bioinformatics of the differentially abundant sugar substrate-binding proteins identified in the <i>B. proteoclasticus</i> cytosol.....	188
5.7 Summary	195
Chapter 6 Discussion	197
6.1 Overview.....	198
6.1.1 Proposed mechanism of GAX degradation by <i>B. proteoclasticus</i>	202
6.2 A proteomic view of extracellular fibre degradation by <i>B. proteoclasticus</i>	207
6.2.1 The role of secreted xylanolytic enzymes	207
6.2.1.1 Endoxylanase Xyn10B	207
6.2.1.2 Xylosidase/arabinofuranosidase Xsa43J and glycoside hydrolase family 30 GH30A.....	208
6.2.1.3 Xylosidase/arabinofuranosidase Xsa43A	209
6.2.1.4 Endo-1,4- β -xylanase and xylosidase Mxy10-43A, and endo-1,4- β -xylanase Xyn10C.....	210
6.2.2 The possible role of secreted non-xylanolytic enzymes.....	211
6.2.2.1 Endo-1,4- β -glucanase Cel5C, pectate lyase Pel1A, and pectin methylesterase Pme8B	211
6.2.3 Extracellular proteins may be localised to the external cell surface <i>in vivo</i>	215

6.2.4 Carbohydrate-binding module containing proteins, cellulose adsorption, and substrate adherence	216
6.2.4.1 Cell adherence to plant polysaccharides	218
6.2.5 Proteolytic enzymes in the <i>B. proteoclasticus</i> proteome	219
6.3 A proteomic view of sugar assimilation and cytosolic degradation by <i>B. proteoclasticus</i>.....	220
6.3.1 Polysaccharide utilisation loci	220
6.3.2 Polysaccharide utilisation locus 3	223
6.3.2.1 -xylosidase Xyl3A.....	224
6.3.2.2 α -D-Glucuronidase Agu67A.....	226
6.3.3 Polysaccharide utilisation locus 5	228
6.3.3.1 Xylosidase/arabinofuranosidase Xsa43H	228
6.3.4 Polysaccharide utilisation locus 10	229
6.3.4.1 Endo-1,4- β -xylanase and esterase Xyn10D	230
6.3.5 Polysaccharide utilisation locus 29	231
6.3.5.1 Feruloyl esterase Est1E	231
6.3.6 ABC transporter mediated uptake of xylooligosaccharides	232
6.3.6.1 Polysaccharide utilisation locus 11	234
6.4 Non-PUL polysaccharidases and their role in hemicellulose degradation	236
6.4.1 α -L-Arabinofuranosidase Arf51A	236
6.4.2 Xylosidase/arabinofuranosidase Xsa43E.....	237
6.5 Additional enzymes identified in the <i>B. proteoclasticus</i> proteome.....	238
6.6 A proteomic view of <i>B. proteoclasticus</i> central metabolism.....	239
6.6.1 Central metabolism	239
6.6.2 Butyrate synthesis by <i>B. proteoclasticus</i>	243
6.7 Features of the complete identified protein dataset	243
6.7.1 The <i>B. proteoclasticus</i> acidic proteome signature.....	243
6.7.2 The importance of a multiplex approach for improved proteome coverage	244
6.8 Future directions	246
6.9 Conclusion	248
References	251
Appendix A	285
Appendix B.....	287
Appendix C	328

List of Tables

Table 1.1. Estimated composition of Type-I and Type-II primary and secondary plant cell walls.	6
Table 1.2. Summary of the types of hemicellulolytic and cellulolytic enzymes important for lignocellulose degradation.	13
Table 1.3. Summary of the major lignocellulose degrading bacteria of the bovine rumen.	22
Table 1.4. Summary of the four replicons comprising the <i>B. proteoclasticus</i> genome, and distribution of genes involved in plant carbohydrate metabolism and transport.	30
Table 2.1. Composition of the buffers tested for cell-associated protein extraction. ...	47
Table 2.2. Running conditions for Immobiline DryStrip IEF on IPGphor II unit.	51
Table 2.3. Summary of the robotic and manual in-gel protein digestion protocols.	61
Table 3.1. Secreted carbohydrate active proteins identified in the mid-log and stationary phase harvested <i>B. proteoclasticus</i> culture medium.	83
Table 3.2. Predicted non-carbohydrate active secreted proteins identified in the mid-log and stationary phase harvested <i>B. proteoclasticus</i> culture medium.	84
Table 3.3. Summary of the polysaccharidases and proteins uniquely identified in the <i>B. proteoclasticus</i> cell-associated fraction.	105
Table 3.4. Summary of the identified secreted proteins capable of binding to crystalline cellulose	111
Table 3.5. BLASTp homology of the predicted secreted proteins identified in the mid-log and stationary phase harvested <i>B. proteoclasticus</i> culture medium.	118
Table 4.1. Summary of the polysaccharidases identified by 2-DE/MALDI-TOF in the <i>B. proteoclasticus</i> cytosol.	133
Table 4.2. Mean peptide and protein false positive rates for the BioWorks and Scaffold2™ analysis of the 1-D LC-MS/MS raw data files.	138
Table 4.3. Summary of polysaccharide degrading enzymes identified in the mid-log phase harvested <i>B. proteoclasticus</i> cytosolic proteome by 1-D LC-MS/MS.	140
Table 4.4. BLASTp analysis of the 28 polysaccharide degrading enzymes identified in the mid-log and stationary phase <i>B. proteoclasticus</i> cytosolic proteomes by 2-DE/MALDI-TOF.	148
Table 5.1. Summary of the biological replicate samples and <i>pI</i> ranges used to examine the cytosolic proteome of mid-log phase harvested, <i>B. proteoclasticus</i> cells.	161
Table 5.2. Total numbers of matched spot groups per cytosolic <i>pI</i> 3-5.6 and <i>pI</i> 5.3-6.5 replicate gel sets submitted for statistical analysis of protein abundance profiles.	165
Table 5.3. Effect of power ^{1/5} transformation on the variance versus mean dependency of spot groups in the 2-DE data sets.	167
Table 5.4. Normality of matched spot groups after power ^{1/5} transformation using the Shapiro-Wilk goodness-of-fit test.	167

Table 5.5. Estimation of the differential abundance false discovery rate.....	168
Table 5.6. Summary of the differential abundance analysis of the mid-log and stationary phase <i>B. proteoclasticus</i> cytosolic proteomes at a 5% false discovery rate.....	170
Table 5.7. Summary of the differentially abundant proteins between xylan and xylose, or xylan and no added carbon growth conditions, identified in the mid-log phase cytosolic proteome.....	174
Table 5.8. Summary of the differentially abundant proteins between xylan and xylose, or xylan and no added carbon growth conditions, identified in the stationary phase cytosolic proteome.....	182
Table 5.9. Summary of proteins differentially or uniquely abundant between xylan and xylose growth conditions, identified in the mid-log phase cytosolic proteome by 1-D LC-MS/MS.....	186
Table 5.10. BLASTp analysis of the differentially abundant SBPs identified in the <i>B. proteoclasticus</i> cytosol.....	194
Table 6.1. Predicted catalytic activity of the identified extracellular <i>B. proteoclasticus</i> polysaccharidases.....	199
Table 6.2. Predicted catalytic activity of the identified cytosolic <i>B. proteoclasticus</i> polysaccharidases.....	200
Table 6.3. Predicted polysaccharide utilisation proteins identified in the <i>B. proteoclasticus</i> proteome.....	222
Table B.2. Summary of the non-polysaccharide degrading proteins identified in the <i>B. proteoclasticus</i> cytosol by 2-DE.....	290
Table B.3. Summary of the non-polysaccharidase degrading proteins identified in the <i>B. proteoclasticus</i> cytosol by 1-D LC-MS/MS.....	302
Table B.4. Summary of the non-polysaccharide degrading proteins identified in the <i>B. proteoclasticus</i> cell-associated fraction.....	319
Table B.5. BLASTp analysis of the differentially abundant proteins detected in the mid-log phase <i>B. proteoclasticus</i> cytosol by 2-DE.....	323
Table B.6. BLASTp analysis of the differentially abundant proteins detected in the stationary phase <i>B. proteoclasticus</i> cytosol by 2-DE.....	325
Table B.7. BLASTp analysis of the differentially abundant proteins detected in the mid-log phase <i>B. proteoclasticus</i> cytosol by 1-D LC-MS/MS.....	326

List of Figures

Figure 1.1. Simplified schematic diagram of the ruminant digestive tract.	3
Figure 1.2. Simplified chemical structure of cellulose, glucuronoarabinoxylan, xyloglucan, and mixed linkage glucans.	7
Figure 1.3. Summary of the enzyme activities important for the degradation of rumen forage material.	14
Figure 1.4. Simplified chemical structure of several of the inter-polymer cross-links found in lignocellulose.	17
Figure 1.5. Bacterial diversity in the rumen microbial ecosystem.	20
Figure 1.6. Ultrastructure of <i>B. proteoclasticus</i> B316T cells.	28
Figure 1.7. Summary of the 16S rRNA gene phylogeny of <i>B. proteoclasticus</i> and closely related strains within the <i>Clostridial</i> subcluster XIVa.	29
Figure 1.8. The genome atlas of <i>Butyrivibrio proteoclasticus</i> B316 ^T	31
Figure 1.9. The theoretical 2-DE total proteome of <i>B. proteoclasticus</i>	39
Figure 1.10. Summary of the predicted subcellular location of the carbohydrate active proteins produced by <i>B. proteoclasticus</i>	40
Figure 2.1. Summary of the three-step <i>B. proteoclasticus</i> cell fractionation procedure used to obtain secreted, cell-associated, and cytosolic subcellular protein fractions.	46
Figure 2.2. Summary of the two-step procedure for the harvest of soluble proteins. ..	48
Figure 2.3. Summary of the hierarchical procedure used to match technical and biological replicate gels.	55
Figure 2.4. Definition of matched protein spot groups.	56
Figure 2.5. Flow chart summarising decision making process for statistical testing of protein abundance profiles.	59
Figure 3.1. Schematic representation of a hypothetical fibrolytic enzyme synthesised by an anaerobic bacterium.	70
Figure 3.2. Summary of the theoretical <i>B. proteoclasticus</i> secreted proteome.	74
Figure 3.3. 1-DE analysis of the effect of methods for the purification of proteins from the culture medium of <i>B. proteoclasticus</i> cells.	76
Figure 3.4. Narrow-range 2-DE of <i>B. proteoclasticus</i> secreted proteins harvested at mid-log phase from cells grown in the presence of 0.1% xylan.	79
Figure 3.5. Narrow-range 2-DE of <i>B. proteoclasticus</i> secreted proteins harvested at mid-log phase from cells grown in the presence of 0.5% xylose.	80
Figure 3.6. Narrow-range 2-DE of <i>B. proteoclasticus</i> secreted proteins harvested at stationary phase from cells grown in the presence of 0.1% xylan.	81
Figure 3.7. Narrow-range 2-DE of <i>B. proteoclasticus</i> secreted proteins harvested at stationary phase from cells grown in the presence of 0.5% xylose.	82
Figure 3.8. MALDI-TOF peptide mass fingerprints of protein extracted from spot 3, spot 4, and spot 5.	88

Figure 3.9. MASCOT summary of peptide sequence coverage of protein extracted from spot 3, spot 4, and spot 5.	89
Figure 3.10. MASCOT summary of MS/MS Xyn10B identification and peptide sequence coverage after analysis of protein extracted from spot 4.....	90
Figure 3.11. CLUSTALW multiple sequence alignment of the GH10 domain (N-terminal and central regions) in <i>B. proteoclasticus</i> Xyn10B with the top 10 ranking BLASTp homologues.	92
Figure 3.12. CLUSTALW multiple sequence alignment of the GH10 domain (C-terminal region) containing region of <i>B. proteoclasticus</i> Xyn10B with the top 10 ranking BLASTp homologues.	93
Figure 3.13. CLUSTALW multiple sequence alignment of the CBM9 domain containing region of <i>B. proteoclasticus</i> Xyn10B with the top 10 ranking BLASTp homologues.....	94
Figure 3.14. MS identification of Xsa43J in the <i>B. proteoclasticus</i> culture medium. .	95
Figure 3.15. 1-DE analysis of the effect of buffers used for <i>B. proteoclasticus</i> cell-associated protein extraction.....	98
Figure 3.16. Polysaccharide degrading enzymes and proteins uniquely identified in the <i>B. proteoclasticus</i> cell-associated fraction by pI 3-5.6 2-DE/MALDI-TOF MS.	100
Figure 3.17. Colloidal CBB G-250 stained 2-DE separation of proteins harvested by 2% CHAPS extraction of mid-log phase, xylose grown cells.	101
Figure 3.18. Colloidal CBB G-250 stained 2-DE separation of proteins harvested by 2% CHAPS extraction of mid-log phase, no added carbon grown cells.	102
Figure 3.19. MASCOT summary of MALDI-TOF Amy13A identification of protein excised from spot CA7 and spot CA8.	106
Figure 3.20. MS identification of Xyn10B from the cell-associated fraction.....	107
Figure 3.21. 1-DE analysis of polysaccharide adsorption assay of xylan-grown <i>B. proteoclasticus</i> secreted proteins.	108
Figure 3.22. MS identification of Xyn10B after cellulose affinity assay.	112
Figure 3.23. MS identification of Xyn10C after cellulose affinity assay.	113
Figure 3.24. MS identification of Mxy10-43A after cellulose affinity assay.	114
Figure 3.25. Functional domains of the polysaccharidases and carbohydrate binding proteins identified in <i>B. proteoclasticus</i> extracellular proteome.	116
Figure 3.26. Functional domains of differentially abundant substrate-binding proteins identified in the <i>B. proteoclasticus</i> culture medium.	117
Figure 4.1. Theoretical 2-DE map of the <i>B. proteoclasticus</i> cytosolic proteins.	125
Figure 4.2. 2-DE analysis (pI 3-10) of cytosolic proteins extracted from differentially cultured, mid-log phase harvested <i>B. proteoclasticus</i> cells.	126
Figure 4.3. Mean number of spots detected in the cytosolic 2-DE analyses.	127
Figure 4.4. The percentage cumulative frequency of matched spot groups in the 2-DE analyses.....	128

Figure 4.5. Polysaccharide degrading enzymes identified in the <i>B. proteoclasticus</i> cytosol by pI 3-5.6 2-DE/MALDI-TOF MS.....	131
Figure 4.6. Polysaccharide degrading enzymes identified in the <i>B. proteoclasticus</i> cytosol by pI 6-11 2-DE/MALDI-TOF MS.....	132
Figure 4.7. MS identification of Bga2B, Bgl3C, and Man2A.....	135
Figure 4.8. MS identification of Bga35B and Xyl3A.	136
Figure 4.9. Protein identification overlap between BioWorks and Scaffold2™ analysis of the mid-log phase, cytosolic <i>B. proteoclasticus</i> proteome.	139
Figure 4.10. Functional domains of predicted secreted polysaccharidases identified in the <i>B. proteoclasticus</i> cytosol.	143
Figure 4.11. Functional domains of polysaccharidases identified in the <i>B. proteoclasticus</i> cytosol.	145
Figure 4.12. Functional domains of polysaccharidases identified in the <i>B. proteoclasticus</i> cytosol.	146
Figure 4.13. Protein function summary of all proteins identified in the <i>B. proteoclasticus</i> cytosol.	151
Figure 4.14. Protein identification overlap between the 2-DE MALDI-TOF and 1-D LC-MS/MS analyses.....	152
Figure 4.15. Two-dimensional plot of all proteins identified by 2-DE and/or 1-D LC-MS/MS analysis of the mid-log phase harvested <i>B. proteoclasticus</i> cytosol.	153
Figure 4.16. Analysis of the mean protein size for the subsets of cytosolic proteins identified by one or both proteomic techniques.	154
Figure 4.17. Summary of the number of proteins identified in the cytosolic proteome by 2-DE MALDI-TOF only, 1-D LC-MS/MS only, and common to both techniques, grouped according to predicted biological function.	154
Figure 4.18. Scatter plot of the relationship between cytosolic protein codon adaptation index and pI value.	156
Figure 4.19. The frequency distribution of <i>B. proteoclasticus</i> cytosolic proteins as a function of codon adaptation index.	156
Figure 5.1. The efficiency of normalisation in correcting for differences in mean spot volume values due to analytical variation.....	162
Figure 5.2. The relationship between the number of missing values per matched spot group, and the mean normalised spot volume.....	163
Figure 5.3. Analysis of the proportion of imputed data values as a function of missing value frequency per spot group.	164
Figure 5.4. The efficiency of different transformations in correcting for variance versus mean dependency.	166
Figure 5.5. The cumulative frequency of 2-DE significant tests and expected false discoveries as a function of percentage FDR.....	169
Figure 5.6. Differentially abundant spots between the mid-log phase harvested xylan grown cytosolic proteome and the alternate two growth conditions.	171

Figure 5.7. Differentially abundant spots between the mid-log phase harvested xylan and xylose grown cytosolic proteomes.....	172
Figure 5.8. Differentially abundant spots between the mid-log phase harvested xylan and no added carbon grown cytosolic proteomes.	173
Figure 5.9. Effect of growth substrate and growth phase on pullulanase, Pul13A (Bpr_III161) abundance.	176
Figure 5.10. Differentially abundant spots between the stationary phase harvested xylan grown cytosolic proteome and the alternate two growth conditions. ...	179
Figure 5.11. Differentially abundant spots between the stationary phase harvested xylan and xylose grown cytosolic proteomes.	180
Figure 5.12. Differentially abundant spots between the stationary phase harvested xylan and no added carbon grown cytosolic proteomes.....	181
Figure 5.13. The transformation efficiency in correcting for variance versus mean dependency in the LC-MS/MS data set by log ₁₀ transformation.	184
Figure 5.14. Cumulative frequency of 1-DE LC-MS/MS significant tests and expected number of false discoveries as a function of percentage FDR.	185
Figure 5.15. Functional domains of differentially abundant ABC transporter system proteins identified in the <i>B. proteoclasticus</i> cytosol.	190
Figure 5.16. CLUSTALW neighbour-joining phylogram of the differentially abundant SBPs identified in the <i>B. proteoclasticus</i> proteome.....	191
Figure 5.17. CLUSTALW multiple sequence alignment of the N-terminal half of the differentially abundant SBP_Bac_1 domain containing SBPs identified in the <i>B. proteoclasticus</i> proteome.	192
Figure 5.18. CLUSTALW multiple sequence alignment of the C-terminal half of the differentially abundant SBP_Bac_1 domain containing SBPs identified in the <i>B. proteoclasticus</i> proteome.	193
Figure 6.1. A simplified schematic representation of extracellular plant polysaccharide metabolism by <i>B. proteoclasticus</i>	204
Figure 6.2. A simplified schematic representation of the ATP-driven transport systems likely to mediate the assimilation of xylan derived soluble sugars.....	205
Figure 6.3. A simplified schematic representation of intracellular xylooligosaccharide degradation by <i>B. proteoclasticus</i>	206
Figure 6.4. Organisation of the <i>B. proteoclasticus</i> PUL 3.....	224
Figure 6.5. Organisation of the <i>B. proteoclasticus</i> PUL 5.....	228
Figure 6.6. Organisation of the <i>B. proteoclasticus</i> PUL 10.....	229
Figure 6.7. Organisation of the <i>B. proteoclasticus</i> PUL 29.....	231
Figure 6.8. Organisation of the <i>B. proteoclasticus</i> PUL 11.....	235
Figure 6.9. A proteomic view of <i>B. proteoclasticus</i> central metabolism.	242
Figure C.1. Cel5C sequence coverage of tryptic peptides extracted from spot 1 and spot 2.	328
Figure C.2. MASCOT summary of MALDI-TOF Pme8B identification and peptide sequence coverage of protein extracted from spot 7.....	329

Figure C.3. MASCOT summary of MALDI-TOF PellA identification and peptide sequence coverage of protein extracted from spot 6.	330
--	-----

Abbreviations

1-DE, one-dimensional electrophoresis	MOPS, 3-(N-morpholino)-
2-DE, two-dimensional electrophoresis	propanesulfonic acid
ABC, ATP Binding Cassette	MOWSE, molecular weight search
ACN, acetonitrile	MS, mass spectrometry
CAZy, carbohydrate active enzyme	NL, non-linear
CBB, Coomassie Brilliant Blue	PHX, predicted highly expressed
CBM, carbohydrate-binding module	PL, polysaccharide lyase
CDS, coding sequence	PMF, peptide mass fingerprint
CE, carbohydrate esterase	pNP, <i>p</i> -Nitrophenol
CHAPS, 3-[(3-cholamidopropyl)- dimethylammonio]-1- propanesulfonate	PUL, polysaccharide utilisation locus
CHCA, α -cyano-4-hydroxycinnamic acid	PULs, polysaccharide utilisation loci
CUT, carbohydrate uptake transporter	SBP, substrate-binding protein
DTT, dithiothreitol	SCVFA, short-chain volatile fatty acid
DUF, domain of unknown function	SDS, sodium dodecyl sulphate
EC, Enzyme Commission	TCA, trichloroacetic acid
EDTA, ethylenediaminetetraacetic acid	TEMED, N,N,N',N'-Tetramethyl- ethylenediamine
EPS, exopolysaccharide	TFA, trifluoroacetic acid
ESI, electrospray ionisation	TMD, transmembrane domain
FDR, false discovery rate	
GAX, glucuronoarabinoxylan	
GH, glycosyl hydrolase	
HPLC, high-performance liquid chromatography	
IAA, iodoacetamide	
IEF, isoelectric focusing	
IPG, immobilised pH gradient	
LDS, lithium dodecyl sulphate	
MALDI-TOF, matrix-assisted-laser- desorption/ionization time-of-flight	
MLG, mixed linkage glucan	

Chapter 1

Introduction

1.1 Overview

In 2007, 26% of the world's land area was devoted to agricultural pasture that provided forage for more than 1.3 billion cattle, and 1.8 billion sheep and goats. In the same year 46% of the New Zealand landmass was being used for year round grazing of 9.7 million cattle and 38.6 million sheep and goats, producing almost 1.5 million tonnes of meat and approximately 15 billion litres of milk. The economic value of agricultural production to the New Zealand economy is more than \$14 billion per annum, and by 2050 the worldwide demand for meat and milk products is expected to double.

New Zealand agriculture relies heavily on pasture-based forages for feeding ruminant livestock, with very little supplementary feed material such as grains or high energy concentrates. Although pasture-based forages and grains contain almost the same amount of total energy per unit of dry matter, the net energy value of forages is significantly lower and often highly variable (Barriere *et al.*, 2003, van Soest, 1994). The difference results mainly from the high lignocellulose content of forage material, and the inefficient degradation of this fibrous and insoluble component of the plant cell wall within the ruminant digestive tract. In some cases as little as 30% of the total energy content of forages is utilised by the ruminant (Barriere *et al.*, 2003). Furthermore, if the degradation rate is very slow, intact forage accumulates in the rumen and limits voluntary feed intake (Dado and Allen, 1995, Ellis, 1978).

An opportunity therefore exists to improve lignocellulose degradation in the rumen, and it is predicted that such an improvement will lead to the enhanced productivity of ruminant livestock that rely on forage as their primary source of nutrition. This scenario is supported by several studies showing that enhanced *in vivo* degradation of high fibre forage diets leads to improved ruminant performance in terms of both animal weight gain and milk production (Beauchemin *et al.*, 1995, Cruywagen and van Zyl, 2008, Titi and Lubbadah, 2004). Maximising lignocellulose degradation in the rumen and therefore improving the conversion of low quality, high fibre forages to premium quality animal products for human consumption is of paramount importance to the New Zealand agricultural sector and the New Zealand economy.

1.2 The ruminant digestive tract

The Pecora or true ruminants include cattle, sheep, goats, deer; also antelope, gazelles, and giraffes. Their digestive tract is characterised by a highly specialised

fore-stomach comprised of a rumen, reticulum, omasum, and abomasum (Figure 1.1). The rumen is the first and largest of the fore-stomach compartments, at 100 L or more in cattle and up to 10 L in sheep. The reticulum is considered a simple rumen extension, and is connected to the omasum by a short tunnel-like structure. The abomasum is the ruminant's true stomach, and histologically is similar to that of monogastric mammals. The proximal location of the rumen facilitates the regurgitation of partially digested lignocellulose for further mastication (rumination). This process triggers the flow of copious amounts of saliva from the mouth to the fore-stomach, which in pasture-grazed ruminants buffers the rumen at pH 6.5-6.8, and is essential for efficient rumen function.

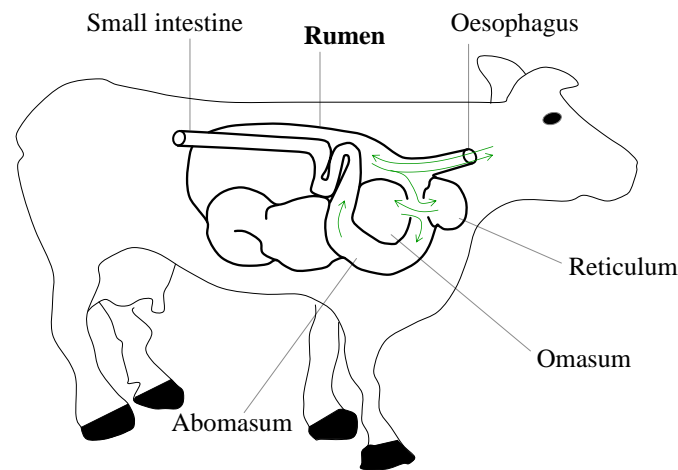


Figure 1.1. Simplified schematic diagram of the ruminant digestive tract. The passage of ingested material from the oesophagus to the small intestine, via the rumen, reticulum, omasum, and abomasum is shown by the green arrows.

1.2.1 Ruminant microbial symbiosis

The rumen environment permits a unique symbiosis between the ruminant and a wide variety of polysaccharolytic, proteolytic, and fermentative microorganisms, which allows the host to utilise energy rich and highly abundant lignocellulosic material as the main dietary component. The rumen is the primary site of microbial mediated lignocellulose degradation and fermentation, which generates large quantities of short chain volatile fatty acids (SCVFAs) and supports microbial growth and cell turnover.

Almost all the acetic, butyric, and propionic acids that are produced as microbial fermentation end products are absorbed by the ruminant across the rumen epithelium and transported to the liver and other body tissues. Acetic acid is oxidised in most

body tissues to generate ATP, and is the major source of acetyl-CoA for ruminant lipid synthesis. Most butyric acid is absorbed from the rumen wall as β -hydroxybutyric acid, and is oxidised in many tissues for energy production. Propionic acid is utilised almost exclusively in the liver, where it is a major substrate for gluconeogenesis.

Microbial biomass that is flushed distally from the rumen to the small intestine provides the ruminant with more than half of the total source of high quality protein (Ulyatt *et al.*, 1980). The digestion and uptake of microbial protein is of particular significance because ingested plant protein cannot be utilised directly by the ruminant. Plant-derived protein is hydrolysed and deaminated to ammonia and will be rapidly excreted by the ruminant unless assimilated by rumen microorganisms (Cotta and Hespell, 1986).

1.2.2 Factors affecting ruminant digestion and energy supply

Four interconnected factors are recognised as limiting lignocellulose degradation and microbial cell turnover in the rumen and each is the focus of intensive research. They are: (1) host factors that mediate the availability of nutrients through mastication, salivation and digesta kinetics; (2) the chemical composition and degradability of the insoluble component of plant material; (3) the diet-dependent composition of the rumen microbial ecosystem; (4) the polysaccharolytic capability of the rumen microbial ecosystem (Cheng *et al.*, 1991). The fibre intake rate and extent of mastication by ruminants affect the rate of passage through the digestive tract, with higher intakes resulting in lower overall fibre digestion (Russell *et al.*, 1992a). The effect of the chemical composition on plant cell wall digestibility is highlighted by studies of brown-midrib mutants of maize (Barriere *et al.*, 2004), which have demonstrated that improved fibre degradability can lead to improved animal performance. Recently, rumen inoculation of genetically and non-genetically modified bacteria harbouring improved fibrolytic or metabolic capabilities has been attempted with the aim of improving the composition and fibrolytic capability of the rumen ecosystem (Gobius *et al.*, 2002, Krause *et al.*, 1999, Krause *et al.*, 2001, Varel *et al.*, 1995). Most of these attempts have been empirical and have not translated into improved fibre degradation *in vivo* or improved ruminant performance, and it has become clear that our current understanding of the composition and polysaccharolytic capability of the rumen microbial ecosystem is highly superficial in comparison to its ecological complexity.

The purpose of this research project is to contribute to our understanding of the polysaccharolytic capability of the rumen microbial ecosystem. A better understanding of the metabolic processes performed by the fibrolytic rumen microbiota is expected to generate new opportunities to enhance forage degradation, manipulate rumen function, and improve the productivity of forage fed ruminants.

1.3 The plant cell wall

Plant cell walls contribute up to 80% of the dry weight of plant material, which is the principal source of nutrition for forage fed ruminants. In all plants, the cell wall is a complex network of cellulose fibres surrounded by a matrix of non-cellulosic structural polysaccharides, proteins, phenolic compounds, and lignin, the amounts and chemical composition of which vary greatly between plant species, cell types, and stages of maturity (Carpita and Gibeaut, 1993, Carpita, 1996). A better understanding of fibre degradation and assimilation by the fibrolytic rumen microbiota necessitates a description of the structural and chemical complexity of plant polysaccharides.

1.3.1 Type-I and Type-II plant cell walls

The primary cell wall of flowering plants can be divided into two broad categories that possess significantly different chemical and structural properties (Table 1.1) (Smith and Harris, 1999, Stinard and Nevins, 1980). Type-I cell walls, comprised of cellulose microfibrils embedded in a xyloglucan (XyG) rich hemicellulose as well as significant amounts of pectin and structural proteins, are found in dicotyledonous and noncommelinoid monocotyledonous plants and gymnosperms. Type-II cell walls, comprised of cellulose embedded in large quantities of glucuronoarabinoxylan (GAX), moderate levels of phenolic compounds, and low levels of pectin and structural proteins, are distinct in that they are found only in commelinoid monocotyledonous plants such as the grasses (Poaceae). The primary cell wall of grasses may also contain significant levels of mixed linkage glucans (MLGs) (Jung, 2003, MacAdam and Grabber, 2002). The secondary cell wall of grasses can contribute up to 50% of dry weight (Brown and Saxena, 2000) and is composed mainly of cellulose, GAX, and lignin.

1.3.2 Cellulose

Cellulose is the energy rich, structural polysaccharide that provides tensile strength to the plant cell wall, and can comprise up to 40% of the dry weight. Cellulose is

synthesised as single strands of up to 15,000 (1→4)-β-D-glucose monomers (Figure 1.2) that spontaneously form insoluble microfibrils (Pizzi and Eaton, 1985). In plants, these microfibrils are 3-5 nm wide and contain a variable number of cellulose chains that are held in a crystalline structure by numerous intra- and inter-chain hydrogen bonds and van der Waals forces (Cousins and Brown, 1995). Individual thin microfibrils may also associate with one another to form a larger single microfiber (Ha *et al.*, 1998). The outer parts of the cellulose microfibrils are often less crystalline (more amorphous) than the core regions, which may be due to interactions with surrounding polymers, or the imperfect association of the individual cellulose chains (O’Sullivan, 1997). Broad microfibrils also tend to have a lower percentage of amorphous regions compared to thinner ones, which is in agreement with the amorphous regions being found predominantly at the surface (O’Sullivan, 1997). Cellulose provides the structural framework of the plant cell wall, and is in intimate association with a number of non-cellulosic cell wall components.

Table 1.1. Estimated composition of Type-I and Type-II primary and secondary plant cell walls.^a

Cell wall component	Primary wall		Secondary wall	
	Type-I	Type-II	Type-I	Type-II
Cellulose	15-30 ^{b,c,d}	20-30 ^{b,e}	45-50 ^b	35-45 ^{b,f}
Hemicellulose				
Glucuronoarabinoxylans	5 ^b	20-40 ^c	20-30 ^{b,g}	40-50 ^{b,g}
Mixed linkage glucans	None	10-30 ^c	None	Minor
Xyloglucans	20-25	1-5 ^{b,c}	Minor	Minor
Mannans/glucomannans	5-10 ^c	Minor	3-5 ^g	Minor
Pectin	20-35 ^c	5 ^b	0.1 ^b	0.1 ^b
Structural proteins	10 ^{c,d}	1 ^c	Minor	Minor
Phenolic compounds	Minor	1-5 ^{b,c}	Minor	0.5-1.5 ^b
Lignin	Minor	Minor	7-10 ^b	20 ^b
Silica	None	None	Variable	5-15 ^b

^a Values are the percentage of plant cell-wall dry weight. The table was adapted from (Vogel, 2008). Estimates were obtained from a number of sources as detailed.

^b (Ishii, 1997).

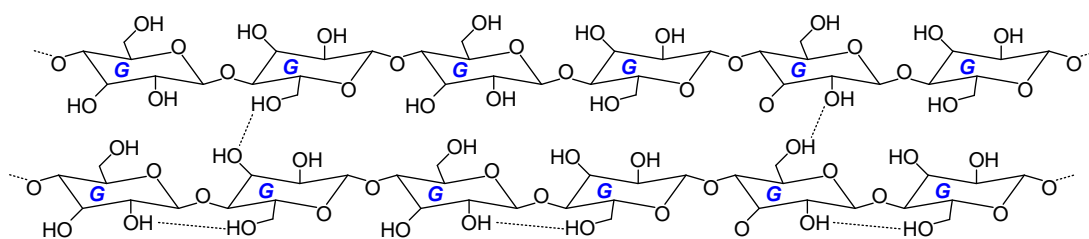
^c (O’Neil and York, 2003).

^d (Zablackis *et al.*, 1995).

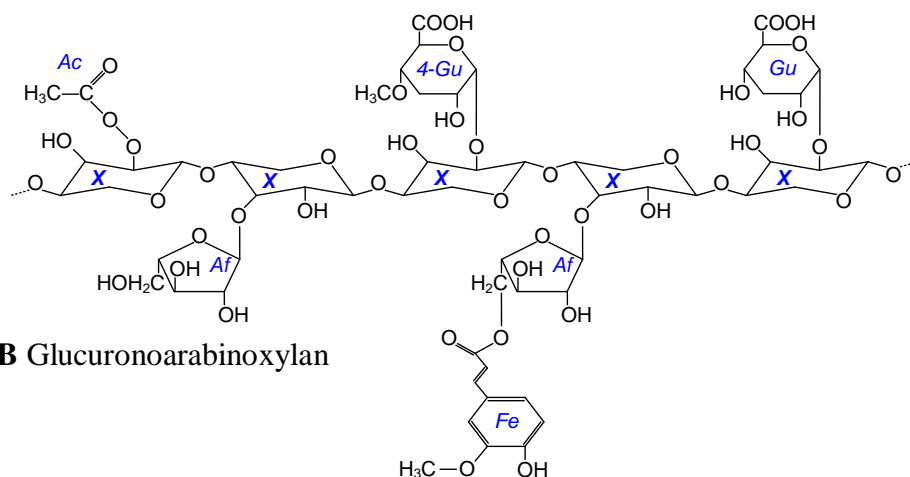
^e (Mitchell *et al.*, 2007).

^f (Hatfield *et al.*, 1999).

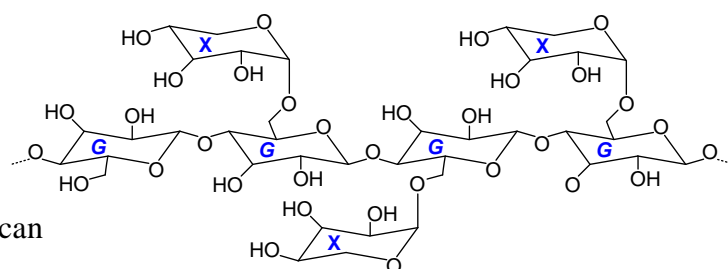
^g (Ebringerova *et al.*, 2005).



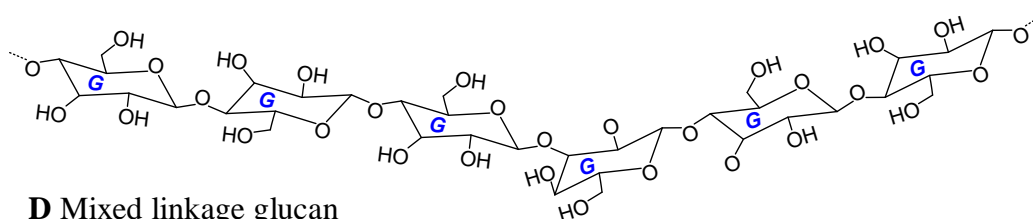
A Cellulose



B Glucuronoarabinoxylan



C Xyloglucan



D Mixed linkage glucan

Figure 1.2. Simplified chemical structure of cellulose (A), glucuronoarabinoxylan (B), xyloglucan (C), and mixed linkage glucans (D). Dotted lines denote potential hydrogen bonds. Ac., O-acetyl; Af, α -L-arabinofuranose; Fe, ferulic acid; G, glucose; 4-Gu, 4-O-methyl-D-glucuronic acid; Gu, glucuronic acid; X, xylopyranose (xylose).

1.3.3 Hemicellulose

Hemicellulose comprises 20%-35% of total plant biomass, and is poorly soluble in water or dilute acids. Unlike cellulose, it is a family of chemically heterogeneous and highly branched polymers of pentoses, hexoses, and sugar acids that interlace with one another and strengthen the plant cell wall (Kabel *et al.*, 2007, Linder *et al.*, 2003). The types and relative amounts of hemicellulose vary significantly with plant type, cell type, and growth state (Table 1.1). Hemicellulose interacts with neighbouring cellulose microfibrils in a structurally dependent manner (Nakamura *et al.*, 2002a), and may also interact with surrounding pectins (Atalla *et al.*, 1993). The interaction of hemicelluloses (and other non-cellulosic cell wall structural polymers) with cellulose is referred to as the sticky network model (Cosgrove, 2000), which describes the tethering of hemicellulose to the cellulose microfibrils by two mechanisms: being trapped in the microfibril during crystallisation, and formation of multiple hydrogen bonds between hemicellulose and cellulose after polymer deposition. Evidence also suggests that hemicellulose plays a regulatory role during the initial phase of cellulose microfibril formation (Whistler and Richards, 1970).

1.3.3.1 Glucuronoarabinoxylan

GAX in grasses consists of a homopolymeric backbone of (1→4)-β-D-xylopyranose monomers (Aspinall, 1980, Kulkarni *et al.*, 1999, Li *et al.*, 2000) substituted to varying degrees with (1→2)-α-linked and (1→3)-α-linked L-arabinofuranosyl groups (Vietor *et al.*, 1992), O-acetyl, and (1→2)-α-D-glucuronic acid or 4-O-methyl-D-glucuronic acid groups (Borneman *et al.*, 1990, Hartley *et al.*, 1990b, Mueller-Harvey *et al.*, 1986). The α-L-arabinofuranosyl side groups may also be esterified at the O-2 and/or O-5 positions with phenolic polymers such as ferulic acid and to a lesser extent, its non-methoxylated analogue *p*-coumaric acid (Kulkarni *et al.*, 1999).

The level of arabinosyl substitution decreases as cell elongation stops, and a concomitant increase in the percentage of ferulated arabinosyl units is observed (Carpita and Gibeau, 1993, Carpita, 1996). GAX of grasses and annual plants exhibit extensive α-L-arabinan and/or α-L-arabinofuranosyl substitution (Hartley and Ford, 1989, Mueller-Harvey *et al.*, 1986), and high levels of ferulic and *p*-coumaric acids (Akin *et al.*, 1990c). The cell walls of warm season grasses such as Bermuda and Switchgrass are particularly rich in arabinofuranose linked phenolic acids (Brillouet and Joseleau, 1987).

GAX in cereals are substituted mainly with (1→2)- α -glucuronic or 4-O-methyl-D-glucuronic acids and arabinose (Grabber *et al.*, 1995, Grabber *et al.*, 2000, Iiyama *et al.*, 1994, Kamisaka *et al.*, 1990, Lam *et al.*, 1992, MacAdam and Grabber, 2002), and the starchy endosperm of cereal grains contains mainly α -L-arabinans. The hemicellulose substituent groups sterically inhibit the aggregation of the xylan backbone leading to the formation of an extended, asymmetrical cell wall matrix (Dea *et al.*, 1973). In particular, the extent of arabinose and/or glucuronic acid substitution of xylans from different sources has been shown to correlate with the water solubility of the polymer (Andrewartha *et al.*, 1979, Hespell and Cotta, 1995).

1.3.3.2 Xyloglucan

Xyloglucan is the main hemicellulose in the primary cell wall of many dicotyledonous plants, and is found at low levels in Type II primary cell walls. The xyloglucan backbone consists of a repeating pattern of four (1→4)- β -D-glucose monomers that are substituted with (1→6)- α -xylopyranose residues. In Type-I cell walls the xylose residues are often further substituted with (1→2)- α -D-galactose and (1→2)- α -L-fucose monomers, but no evidence for such substitution in Type-II cell walls has been presented. Most xyloglucans are tightly hydrogen bonded to neighbouring cellulose microfibrils (Pauly *et al.*, 1999). Mannan based xylans, such as glucomannan (a heteropolymer of D-glucose and D-mannose) and galactoglucomannan (a heteropolymer of D-galactose, D-glucose, and D-mannose) constitute up to 10% of Type-I cell walls, and a minor component of Type-II walls.

1.3.3.3 Mannans and glucomannans

In spite of mannans and glucomannans comprising only minor fractions of Type-II primary and secondary cell walls, they are suggested to play important functional roles. Although mannan and glucomannan chemical composition is highly species- and growth stage specific, they are predominantly straight chain polysaccharides characterised by a backbone of (1→4)- β -linked D-mannose and D-glucose monomers in a ratio of approximately 1.6:1 (Katsuraya *et al.*, 2003). Depending on the plant species, mannose residues within the main chain may be substituted with galactose monomers, or be acetylated. Glucomannans are one of the few structural polymers that are enriched in the walls of maize coleoptile epidermal cells compared to mesophyll cells, and the interactions between glucomannans and surrounding highly substituted GAX are implicated in stem elongation (Carpita *et al.*, 2001). The wall of barley

endosperm cells also contains significant amounts of mannose containing polysaccharides (Lazaridou *et al.*, 2008). The ratio of glucomannans to total polysaccharide content is positively influenced by environmental factors such as high mean temperature and rainfall, and their presence may be a determinant of cell – cell interaction, and overall endosperm strength and rigidity.

1.3.3.4 Mixed linkage glucans

MLGs are widely spread among the Poales (the taxonomic order that contains the grasses) but are not found in other plants (Carpita, 1996), and are a rare example of a developmental stage-specific structural polysaccharide (Staudte *et al.*, 1983, Woodward and Fincher, 1983). They are unbranched glucose homo-polymers containing both (1→3)-β-D-linked and (1→4)-β-D-glucose monomers (Figure 1.2). Approximately 70% of the polymer consists of (1→4)-β-linked cellotriose units connected by single (1→3)-β-linkages (Carpita and Gibeaut, 1993). The remainder contains longer cellodextrins that are also connected by (1→3)-β-linkages. MLGs are tightly integrated with cellulose and non-cellulosic polysaccharides (Kim *et al.*, 2000), and are found in the primary walls of elongating cells where they reach maximum abundance during the period of maximum growth rate (Carpita *et al.*, 2001). As the rate of cell elongation slows, the MLGs are degraded and are no longer a major component of the plant cell wall. Consequently, MLGs found in the primary cell wall of grasses are believed to play a role in cell wall growth and remodelling (Meier and Reid, 1982). MLGs are also a major component of endosperm cells, where they act as storage carbohydrates that are hydrolysed during seed germination (Ridley *et al.*, 2001).

1.3.4 Pectin

Pectin is the most structurally complex family of polysaccharides in nature, and is implicated in a broad range of functions including cell wall structure and expansion, plant defence, cell-cell signalling, and leaf abscission (Jarvis, 1984). In grasses, pectin is found mainly in the primary cell wall where there are two types (O'Neill *et al.*, 1990). Homogalacturonan (HG) is a linear chain of (1→4)-α-D-galacturonic acid (GalA) residues that may be partially methylesterified or O-acetylated (Carpita, 1989, Shibuya and Nakane, 1984). Rhamnogalacturonan I (RG-I) is a more complex heteropolymer comprised of a backbone of repeating (1→2)-α-L-rhamnosyl-(1→4)-α-D-GalA subunits. The RG-I backbone can be heavily substituted with complex side

chains containing large amounts of arabinofuranosyl units and arabinogalactans (Saulnier and Thibault, 1999). Arabinogalactans are heterogeneous but may have a backbone of (1→3)-β-linked D-galactose residues substituted with (1→6)-β-D-galactose and/or (1→3)-α-L-arabinofuranose. Furthermore, similar to GAX the side chains of RG-I may be esterified with ferulic and *p*-coumaric acids (Coenen *et al.*, 2007, Nakamura *et al.*, 2002b). Arabinogalactan can be present in large amounts in the cell wall of grasses, and may also be classified as a hemicellulose although it originates from the side chains of pectin. It is believed that pectic polysaccharides are covalently cross-linked to one another, probably by their backbones (Duan *et al.*, 2004, Nakamura *et al.*, 2002a). It is also possible that both HG and RG-I are covalently linked to, or very tightly associated with, other plant cell wall polysaccharides including hemicellulose (Nakamura *et al.*, 2002a) and cellulose (Zykwinska *et al.*, 2007).

1.3.5 Cell wall structural proteins

Cell wall structural proteins (CWP) comprise only 1% of the primary cell wall of grasses, but may play a strengthening and stabilising role following cell elongation. They are thought to form an independent network that assists in assembly and restructuring of the cell wall. The threonine-hydroxyproline-rich glycoproteins (THPRG) are the grass homologue of the ubiquitous extensin family of proteins. Upon secretion into the cell wall they rapidly become insoluble and may form intramolecular cross-links (Cassab and Varner, 1987, Lamport, 1986), and may interact with surrounding cellulose microfibrils (Qi *et al.*, 1995) and pectin (Showalter, 1993). Glycine-rich (GRP) and proline-rich proteins (PRP) are often co-localised, and may form tyrosine dependent intra- and inter-molecular linkages. Isodityrosine cross-links may also form between PRP and THPRG (Ye and Varner, 1991). Furthermore, both GRP and PRP have been implicated as potential nucleation sites for cell wall lignification (Showalter, 1993), and the PRP may also interact ionically with surrounding pectin molecules (Chabannes *et al.*, 2001, Jones *et al.*, 2001).

1.3.6 Lignin

Lignin is an integral component of the secondary cell wall of grasses and is crucial for maintaining cell structural integrity (Barriere *et al.*, 2003), water transport and protecting against plant pathogens (Ralph *et al.*, 2001). Furthermore, it is impervious to enzymatic degradation by the rumen microbiota, and its presence in the grass cell

wall is believed to be a contributing factor to the recalcitrance of ruminant forages. Lignins are comprised mainly of guaiacyl and syringyl subunits (Grabber *et al.*, 2004), and grass lignin also contains a moderate quantity of p-hydroxyphenyl units that are found in minimal levels in dicotyledonous plants (Campbell and Sederoff, 1996). The quantity and composition of lignin varies considerably between plant species, cell types, and cell wall layers, and with plant maturity, and is controlled by a variety of developmental and environmental cues (Grabber *et al.*, 1998a). Importantly, ferulic acids are also incorporated into grass lignins in a structurally dependent manner during development (Ralph *et al.*, 1995) via ether bonds involving the ferulic acid hydroxyl groups (Kondo *et al.*, 1990, Scalbert *et al.*, 1985).

1.4 Lignocellulose degradation

The variable structure and chemical composition of lignocellulosic material necessitates the concerted action of several types of polysaccharide degrading enzymes (polysaccharidases) for its complete degradation. These enzymes are either O-glycoside hydrolases (GH) that hydrolyse glycosidic bonds between carbohydrate monomers or between carbohydrate and non-carbohydrate moieties, carbohydrate esterases (CE) which hydrolyse the ester linkages of ferulic acid or acetate side chains, or polysaccharide lyases (PL) that hydrolyse polysaccharide chains by β -elimination. GH hydrolysis takes place via general acid catalysis leading to an overall retention or inversion at the site of catalytic activity (Henrissat, 1991, Henrissat and Davies, 1997, Henrissat, 1998), while CEs are divided into two classes, those in which the sugar behaves as an acid or alcohol, such as pectin methyl esterases and acetylxyylan esterases respectively (for review see Lynd *et al.*, (2002)). Polysaccharidases are classified by the International Union of Biologists and Molecular Biologists (IUBMB) (Barrett, 1999) and designated by Enzyme Commission (EC) number (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>). A second classification system utilising hydrophobic cluster analysis catalogues GHs into 118 GH families (as of August 2010) (Carbohydrate Active enZymes Database, <http://www.cazy.org/>) (Cantarel *et al.*, 2009). CAZy family members exhibit common catalytic domain tertiary structure and hydrolytic stereospecificity, which permits confident functional prediction of newly identified enzymes. The types of fibre degrading enzymes important for rumen forage degradation, and the chemical bonds hydrolysed by each class of enzyme are summarised in Table 1.2 and Figure 1.3 respectively.

Table 1.2. Summary of the types of hemicellulolytic and cellulolytic enzymes important for lignocellulose degradation.

Enzyme	Substrate	CAZy family ^a	EC number ^b
Endo- β -1,4-xylanase	(1 \rightarrow 4)- β -D-Xylan	GH5, 8, 10, 11, 43	3.2.1.8
Exo- β -1,4-xylosidase (β -xylosidase)	(1 \rightarrow 4)- β -D-Xylooligomers, xylobiose	GH3, 30, 39, 43, 51, 52, 54	3.2.1.37
α -L-Arabinofuranosidase	α -L-Arabinofuranosyl (1 \rightarrow 2) or (1 \rightarrow 3) xylooligomers	GH3, 43, 51, 54, 62	3.2.1.55
α -Glucuronidase	Glucuronic acid, and 4-O-methyl-D-glucuronic acid	GH4, 67, 115	3.2.1.139
Acetyl xylan esterase	O-Acetyl xylan	CE 1, 2, 3, 4, 5, 6, 7	3.1.1.72
Ferulic/ <i>p</i> -coumaric acid esterase	Ferulic/ <i>p</i> -coumaric acid esterified arabinofuranosyl xylooligomers	CE 1	3.1.1.73
α -Galactosidase	(1 \rightarrow 6)- α -D-Galactose	GH4, 27, 36, 57, 110	3.2.1.22
Endo- β -1,4-glucanase	(1 \rightarrow 4)- β -D-Glucose	GH5, 6, 7, 8, 9, 12, 44, 45, 48, 51, 61, 74	3.2.1.4
Cellodextrinase	(1 \rightarrow 4)- β -D-Glucose	GH1, 3	3.2.1.74
Cellobiohydrolase		GH5, 6, 9, 48	3.2.1.91
β -Glucosidase	(1 \rightarrow 4)- β -D-Cellodextrins and cellobiose	GH1, 3, 9	3.2.1.21

^a CAZy database classification (<http://www.cazy.org/>).

^b Enzyme commission number (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>).

1.4.1 Cellulose degradation

Three types of enzymes catalyse cellulose hydrolysis to glucose. Endo- β -1,4-glucanases (EC 3.2.1.4) cleave (1 \rightarrow 4)- β -D-glycosidic bonds at internal positions along the cellulose chain generating cellodextrins of variable chain lengths. Exoglucanases such as cellodextrinases (EC 3.2.1.74) and cellobiohydrolases (3.2.1.91) act on the reducing or non-reducing ends of cellulose in a processive manner, which liberates glucose and cellobiose, respectively (Barr *et al.*, 1996). Exoglucanases are also capable of degrading microcrystalline cellulose, possibly by stripping cellulose microfibrils from the microcrystalline structure (Teeri, 1997). β -Glucosidases (3.2.1.21) complete the degradation process by hydrolysing soluble cellobiose and cellodextrins to glucose.

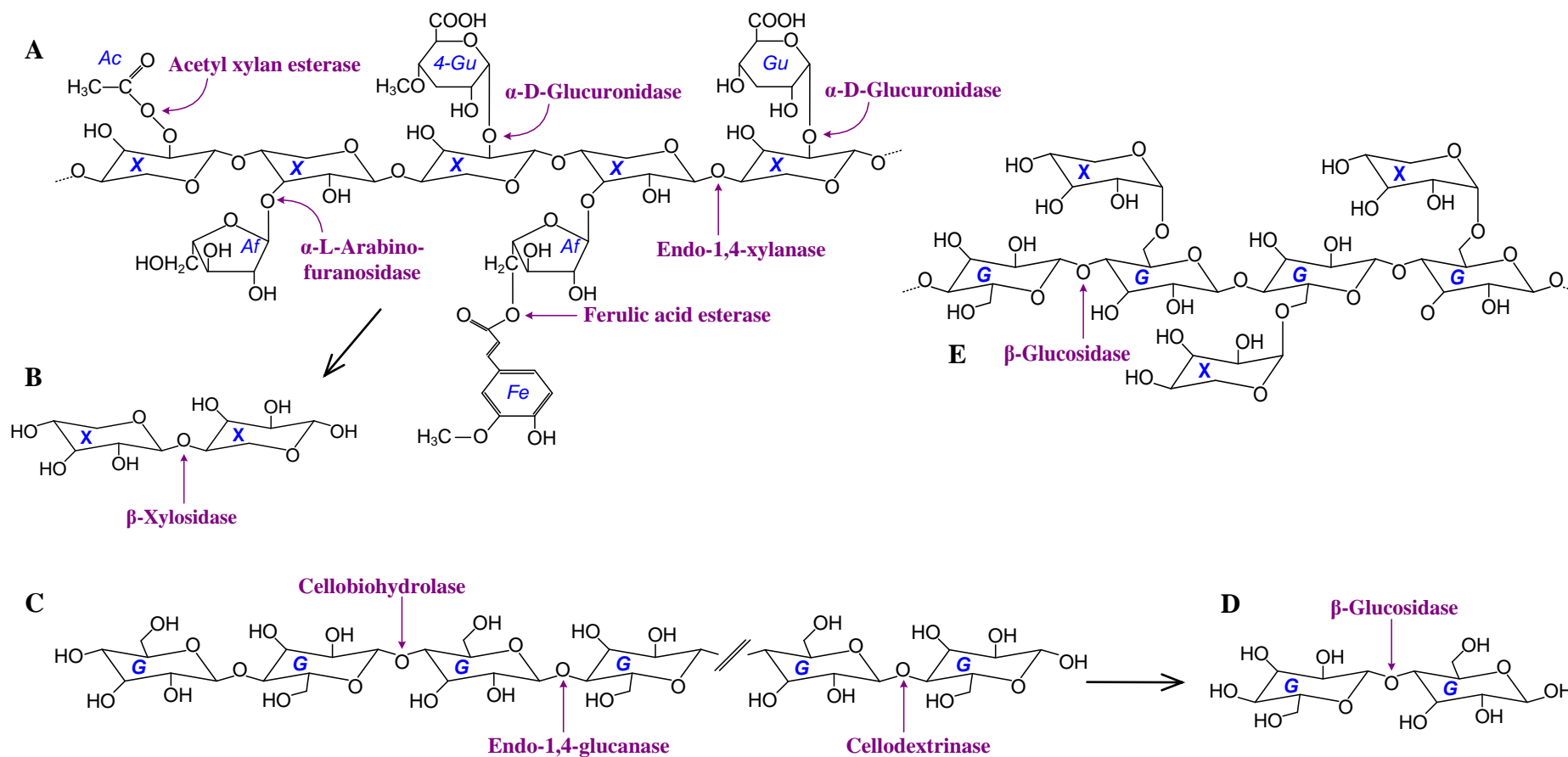


Figure 1.3. Summary of the enzyme activities important for the degradation of rumen forage material. Sites of attack are shown for the enzymes involved in the hydrolysis of GAX (A), xylobiose (B), cellulose (C), cellobiose (D), and xyloglucan (E). Ac., O-acetyl; Af, α -L-arabinofuranoside; Fe, ferulic acid; G, glucose; 4-Gu, 4-O-methyl-D-glucuronic acid; Gu, glucuronic acid; X, xylopyranose.

1.4.2 Hemicellulose degradation

Efficient hemicellulose hydrolysis requires a variety of enzymes with overlapping but different substrate specificities (Wong *et al.*, 1988), which reflects the complex and diverse chemical structure of the polysaccharide. (1→4)-β-D-xylanxylanohydrolases (endoxylanases) (EC 3.2.1.8) cleave internal linkages within the xylan backbone (Mackenzie *et al.*, 1987), generating xylobiose and xylooligosaccharides, which are further degraded to xylobiose and/or xylose by (1→4)-β-D-xylan-xylohydrolases (β-D-xylosidases) (3.2.1.37) acting on non-reducing polysaccharide chain ends. The diverse range of hemicellulose substituents are hydrolysed from the xylan backbone by the activity of α-D-glucuronidases (EC 3.2.1.139), α-L-arabinofuranosidases (EC 3.2.1.55), acetylxylan esterases (EC 3.1.1.72), ferulic acid (EC 3.2.1.73) and *p*-coumaric acid esterases (EC 3.2.1.-). Ferulic acid esterases may be particularly important in ruminant fibre degradation in that they hydrolyse the ester bond between ferulic acids and hemicellulose substituents, which is hypothesised to result in increased accessibility of hemicellulose to other fibre degrading enzymes (Black *et al.*, 1996, Janecek *et al.*, 2003). Microbial expressed endoxylanases are classified within GH families 5, 8, 10, 11 and 43 with members of families 10 and 11 being the most prevalent and well characterised. The activity of family 10 and 11 endoxylanases upon hemicellulose appears to be strongly influenced by the site of xylopyranosyl unit substitution. Depending on the type of xylan backbone substitution, GH 10 enzymes are capable of cleaving (1→4)-β-linkages immediately adjacent to the substituted xylopyranosyl residue, while family 11 enzymes are capable of cleaving the (1→4)-β-linkage two positions away from the site of substitution only (Biely *et al.*, 1997).

1.4.3 Pectin degradation

Three classes of enzymes are required for complete pectin degradation. Depolymerases act directly upon the pectin backbone, and are subdivided into lyases (acting by β-elimination) and hydrolases (Sakai *et al.*, 1993), while esterases hydrolyse acetyl, methyl, or ferulic acids substituents from the main chains of the (1→4)-α-galacturonic acids. Pectate lyases (EC 4.2.2.2) are a key enzyme in bacterial pectin hydrolysis and are involved in the initial stages of degradation (Kluskens *et al.*, 2003). Pectate lyases cleave the (1→4)-α-linkages between galacturonic acids and may preferentially attack regions of unmethylated polygalacturonate (pectate), although they can be active upon regions with low levels of methylation (Tardy *et al.*, 1997).

Pectate lyases are classified in five of the 22 polysaccharide lyase CAZy database families (1, 2, 3, 9, and 10), with the largest number of enzymes present in the PL1 family. The hydrolases involved in pectin degradation include endopolygalacturonases (EC 3.2.1.15), exopolygalacturonases (EC 3.2.1.67) and exo-poly- α -galacturonases (EC 3.2.1.82), and rhamnogalacturonases (EC 3.2.1.-). The former three enzymes cleave at random internal positions, and at the non-reducing chain ends in polygalacturonans, while rhamnogalacturonases hydrolyse bonds between D-galacturonic acid and L-rhamnose in Rhamnogalacturonan I. The majority of hydrolases are classified in the GH28 family, and members of the GH2, GH53, GH78, GH88, GH105 families also target pectin. Acetylsterases (EC 3.1.1.6) remove C-2 and C-3 attached acetyl groups, and methylsterases act on methoxylated galacturonic acids releasing methanol, which can then be used as a substrate by methanogens and acetogenic bacteria. Acetylsterase activity is enhanced when esterified pectins are pre-treated with methylsterase, suggesting there is an optimised sequence of esterase activities for pectin degradation (Shevchik and Hugouvieux-Cotte-Pattat, 2003).

1.5 Recalcitrance of rumen forages to microbial degradation

An array of covalent linkages between hemicellulose substituents, and between hemicellulose and components of the surrounding structural polymers creates a formidable structural barrier to degradation of the lignocellulose component of forages by rumen microbes.

1.5.1.1 Diferulate cross-linking within grass hemicellulose

In grasses, dimerisation of ferulic acid residues attached to arabinosyl side chains extensively cross-link neighbouring hemicellulose backbone chains (Figure 1.4A). The first diferulate cross-link identified was the (5→5)-dehydrodimer (Harris and Hartley, 1976), and since then many alternative diferulates have been identified and characterised (Ralph *et al.*, 1994). More recently, the presence of arabinoxylan-bound esterified ferulate trimers, tetramers, and larger oligoferulates in maize bran and maize culture cells have been reported (Burr and Fry, 2009, Fry *et al.*, 2000, Rouau *et al.*, 2003). The latter two studies found that the oligoferulates were up to 4-fold more prevalent than diferulate cross-links, which led the authors to propose that oligoferulates were the main contributors to hemicellulose cross-linking in maize cell walls. Arabinoxylan-bound ferulate linkages are thought to play an important role in

cell wall stiffening and growth cessation (MacAdam and Grabber, 2002), and pathogen resistance (Lyons *et al.*, 1993).

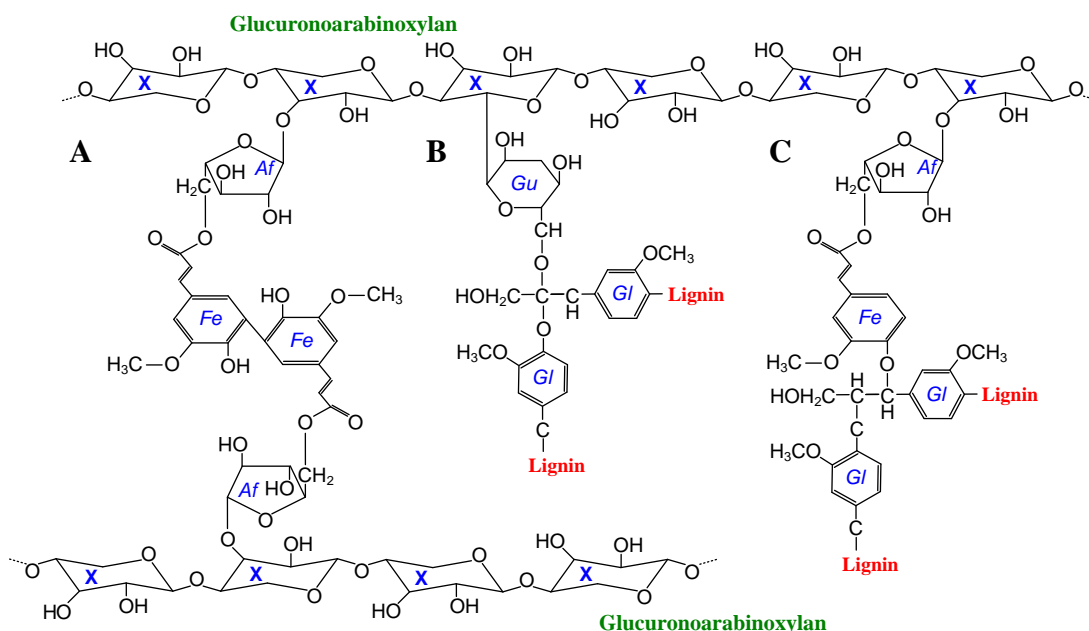


Figure 1.4. Simplified chemical structure of several of the inter-polymer cross-links found in lignocellulose. Af, α -L-arabinofuranoside; Fe, ferulic acid; Gl, Guaiacyl lignin monomer; Gu, glucuronic acid; X, xylopyranose.

1.5.1.2 Diferulate cross-links between grass hemicellulose and lignin

The bifunctional ability of arabinoxylan-bound ferulic acids to form ester and phenol coupling reactions allows the formation of polysaccharide-ferulate-lignin complexes (Figure 1.4C) (Jacquet *et al.*, 1995, Scalbert *et al.*, 1985). The presence of ether linkages between ferulated arabinoxylans and lignin in wheat was confirmed after hot alkaline treatment of the cell walls released ether-linked ferulates (Lam *et al.*, 1992). Ralph *et al* (1995) demonstrated in ryegrass that ferulated arabinosyl residues actively coupled with both syringyl and guaiacyl lignin subunits, and Lam *et al* (2001) subsequently determined that the majority of arabinosyl-linked ferulic acids in mature oat internodes are ester-linked at the benzyl position of lignin side chains. Furthermore, model studies have shown that (5 \rightarrow 5)-coupled diferulates are also capable of copolymerising with the coniferyl alcohol monomers (Grabber *et al.*, 2002). The ferulate monomers attached to α -L-arabinofuranosyl residues are believed to play an important role as initiation sites for lignin deposition, and in anchoring lignin within the cell wall (Ralph *et al.*, 1995, Trethewey *et al.*, 2005). In addition, there is evidence for the existence of direct ester linkages between the carboxyl groups of

uronic acids, such as glucuronic acid in GAX in grasses, and free hydroxyl groups in lignin (Figure 1.4B) (Iiyama *et al.*, 1994).

1.5.1.3 The impact of feruloylation on lignocellulose degradation

The di- and oligoferulate cross-links within hemicellulose, and between hemicellulose and the surrounding lignin are believed to be a primary determining factor in the degradability of high fibre forage material, by restricting the accessibility of polysaccharide degrading enzymes to their substrate. Hartley (1972) was the first to demonstrate a negative correlation between the extent of hemicellulose and/or lignin feruloylation on the one hand, and grass cell wall digestibility on the other, and subsequent studies have confirmed this relationship.

Jung *et al* (1991) used synthetically ferulated oat-spelt xylan to show that hemicellulose degradability by rumen fluid was negatively correlated with the concentration of esterified phenolic acids, and Lam *et al* (2003) extended the correlation to degradation of the complete cell wall of two common forage grass species. Furthermore, Lam *et al* (2003) concluded that the phenolic acid concentration was a stronger determinant of cell wall degradability than lignin content. Vailhe *et al.* (2000) showed in tall fescue that the youngest cells had the lowest ferulic acid content, and the highest degradability, and determined that hemicellulose feruloylation had a greater impact on the rate, rather than the extent of tall fescue cell wall degradation, although both variables were significantly impacted.

The correlation between the prevalence of ferulated arabinoxylans in Bermudagrass and cell wall degradation by rumen bacteria was demonstrated using UV absorption and histochemical staining (Akin *et al.*, 1990a, Akin *et al.*, 1990b, Ames *et al.*, 1992, Hartley *et al.*, 1990a). Cells that had a spectral or staining pattern characteristic of high levels of cell wall phenolic acids showed slow and partial degradation by rumen microorganisms, while cell walls with little or no phenolics were rapidly and extensively degraded. The fact that chemical treatments that specifically cleave diferulate linkages enhance grass cell wall digestibility *in vitro* supports these studies (Fritz *et al.*, 1991, Jung *et al.*, 1992, Morrison, 1991).

Grabber *et al.* (1998a) confirmed the influence of diferulate linkages by artificially increasing ferulate dimerisation in the walls of non-lignified maize cell suspensions from 18 to 40%, which significantly reduced the enzymatic hydrolysis of xylan, cellulose, and pectin. Subsequent inhibition of diferulate bridge formation in synthetically lignified maize cell suspensions led to a concomitant increase in cell wall

digestibility (Grabber *et al.*, 1998b). Furthermore, using vacuole-targeted expression of an *Aspergillus niger* ferulic acid esterase (FAE) in Italian ryegrass and tall fescue, the release of monomeric and dimeric ferulic acids from cell walls following cell death and a concomitant increase in cell wall digestibility was observed in both species (Buanaфина *et al.*, 2006, Buanaфина *et al.*, 2008). The common conclusion drawn by each of these research groups, that hemicellulose cross-linking is a determining factor in the rate and extent of plant fibre degradation, is in agreement with earlier kinetic studies demonstrating that cellulose hydrolysis by rumen microbes is a first order reaction in terms of cellulose concentration, implying that cellulose concentration is the rate limiting factor (Fisher *et al.*, 1989, Maglione *et al.*, 1997, van Soest, 1973, Waldo *et al.*, 1972, Weimer *et al.*, 1990). Cellulose digestion is therefore not limited by either the composition or fibrolytic capability of the cellulolytic microbial population, but rather the surface area of cellulose exposed to the rumen microbiota (Weimer, 1998). Consequently, processes that improve the access of rumen cellulolytic bacteria to cellulose, such as enhancing hemicellulose digestion and hydrolysis of diferulate cross-links between cell wall polymers, are predicted to improve the overall rate and extent of fibre degradation. For these reasons, rumen microbes that have the capacity to degrade the hemicellulosic component of grass cell walls are an important focus of research, and are potential targets for the manipulation of rumen function with the aim of enhancing ruminant productivity.

1.6 The rumen microbial ecosystem

The rumen microbiota consists mainly of obligate anaerobes that act synergistically to rapidly degrade and ferment lignocellulosic biomass. Cell numbers in the bovine rumen can reach a remarkable 10^{11} bacterial cells, 10^9 methanogenic archaea, 10^6 protozoa, and 10^3 anaerobic fungi per millilitre of rumen fluid (Flint, 2004, Hobson, 1989). Culture dependent methods have identified at least 370 bacterial and archeal species, 40 protozoal, and five fungal species (Orpin and Joblin, 1997, Stewart *et al.*, 1997, Williams and Coleman, 1997).

1.6.1 Rumen microbial diversity

In spite of the wealth of information generated by culture dependent analyses, cultivation-independent molecular surveys have recently revealed the true extent of the microbial diversity with the rumen ecosystem, and have demonstrated that as little as 11% of the total bacterial taxa have been cultured (Cho *et al.*, 2006, Edwards *et al.*,

2004, Janssen and Kirs, 2008, Shin *et al.*, 2004a, Shin *et al.*, 2004b, Shin *et al.*, 2004c, Skillman *et al.*, 2006, Tajima *et al.*, 1999, Wright *et al.*, 2007). Several hundred to thousands of uncultivated individual species are believed to be present in the rumen, with the majority falling within either the low G+C Gram-positive *Firmicutes* or the Gram-negative *Bacteroidetes* (Figure 1.5). The *Firmicutes* are represented mainly by the *Clostridial* clusters XIVa, IV, and IX, while *Prevotella* spp. are particularly numerous members of the *Bacteroidetes* (Edwards *et al.*, 2004). Through these studies, it has become apparent that the presently cultivated and characterised rumen microorganisms may not accurately represent the numerically dominant nor functionally significant members of the microbial community (Attwood *et al.*, 2008). Consequently, our knowledge of the bacterial species that are important contributors to rumen fibre degradation and by extension our understanding of the enzyme systems used by these bacteria to degrade lignocellulose is at best rudimentary. Examination of the fibrolytic system of novel rumen microbes that are capable of degrading lignocellulosic material, in particular the hemicellulosic component, will provide a clearer understanding of the metabolic processes involved in plant cell wall degradation by the fibrolytic rumen microbiota.

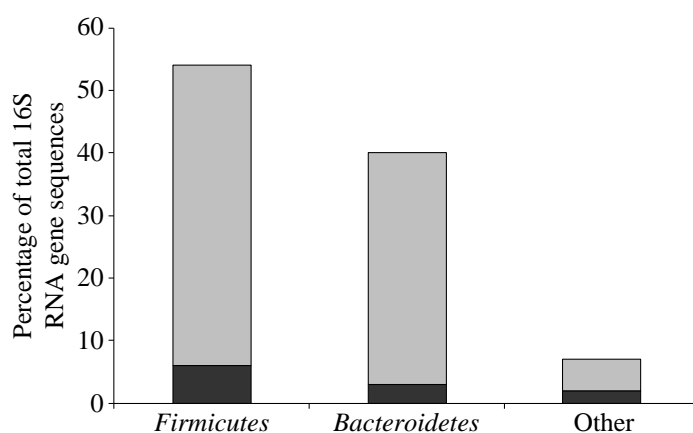


Figure 1.5. Bacterial diversity in the rumen microbial ecosystem. The percentage of 16S rRNA gene sequences represented by cultivated (dark grey) and uncultivated (light grey) bacteria is shown for the two predominant phyla, and minor or unknown phyla. Figure adapted from Edwards *et al.*, (2008).

1.6.2 The fibre degrading rumen microbiome

The rumen microbiome is the set of rumen microbes, their genomes, and environmental interactions, and is composed of: 1) microbes associated with the

rumen liquor (planktonic phase); 2) microbes loosely or tightly associated with ingested plant material (adherent phase); and 3) microbes associated with the rumen epithelium. Microbial attachment is a key process in lignocellulose degradation (McAllister *et al.*, 1994) and bacteria and protozoa have been shown to attach to plant material within five minutes of ingestion (Bonhomme, 1990, Koike *et al.*, 2003a). Adherence to plant particles has been shown to be a prerequisite for efficient degradation of plant cell wall polysaccharides (Jun *et al.*, 2007, Miron *et al.*, 1998, Mosoni and Gaillard-Martinie, 2001), with adhesion defective mutants typically displaying reduced or abolished cellulolytic capability (Devillard *et al.*, 2004, Koike *et al.*, 2003a). Fibre adherent bacteria account for 70-80% of the total microbial biomass, and are estimated to be responsible for approximately 90% of the cellulase and xylanase activity, 70% of the amylase activity, and 75% of the protease activity in the rumen (Akin, 1980, Craig *et al.*, 1987, Minato *et al.*, 1993). Consequently, fibre adherent bacteria are considered the predominant degraders of lignocellulosic plant material in the rumen (Flint, 2004, Foroozandeh *et al.*, 2009). The rumen protozoa and fungi undoubtedly play important fibrolytic roles, especially during the initial stages of degradation of large substrate particles (van Soest, 1994).

Molecular surveys of the bacterial composition of the three rumen microbial fractions have demonstrated that they are markedly different from each other, with representatives of the *Firmicutes* accounting for up to 75% of the adherent population while the planktonic phase is dominated by the *Bacteroidetes* (Cho *et al.*, 2006, Tajima *et al.*, 1999). Culture-independent studies focused only on fibre adherent bacteria confirmed the prevalence of *Firmicute* species within the fibre adherent microbiome, especially when ruminants were fed an all-grass diet, and surprisingly also found that the commonly cultivated fibrolytic species were often not detected (Koike *et al.*, 2003b, Larue *et al.*, 2005).

Current understanding of the enzymatic processes that mediate rumen forage degradation is derived mainly from molecular and biochemical *in vitro* analyses of single genes and/or gene products produced by culturable microbes. These analyses have mostly been conducted on single bacterial species grown in pure culture, and have demonstrated that relatively few bacteria are capable of efficiently degrading the cellulosic component of plant cell walls. Still fewer are capable of degrading and utilising the non-cellulosic components such as hemicellulose and pectin (Stewart *et al.*, 1997). The bacteria currently considered the predominant fibrolytic bacteria in the

rumen of cattle fed a high forage diet are summarised in Table 1.3. It should be noted that Table 1.3 does not constitute an exhaustive list, as less well characterised bacteria such as other *Butyrivibrio* and *Pseudobutyrvibrio* sp., *Clostridium* sp., *Eubacterium ruminantium* and *Lachnospira multipara* are also present and undoubtedly play an important role in rumen fibre degradation.

Table 1.3. Summary of the major lignocellulose degrading bacteria of the bovine rumen.

Phylum	Species	Predominant Substrate ^a	Predominant niche	Main fermentation products ^b
<i>Firmicutes</i> (Gram-positive)	<i>Butyrivibrio fibrisolvens</i>	P, S, X	Solid	Formate/butyrate/acetate/lactate ^c
	<i>Ruminococcus flavefaciens</i>	C, P, X	Solid	Formate/succinate
	<i>Ruminococcus albus</i>	C, P, X	Solid	Formate/succinate
	<i>Eubacterium cellulosolvens</i>	C, X	Solid	Lactate/formate
<i>Bacteroidetes</i> (Gram-negative)	<i>Prevotella ruminicola</i>	P, S, X	Liquid	Propionate/succinate
	<i>Prevotella bryantii</i>	P, S, X	Liquid	Propionate/succinate
<i>Fibrobacter</i> (Gram-negative)	<i>Fibrobacter succinogenes</i>	C, P	Solid	Formate/succinate
	<i>Fibrobacter intestinalis</i>	C	Solid	Formate/succinate

^a C, cellulose; P, pectin; S, starch; X, xylan.

^b Formate is converted to acetate through the metabolic activity of resident acetogens. Succinate is converted to propionate through the action of bacteria such as *Selenomonas ruminantium* (Weimer, 1998).

^c Different *B. fibrisolvens* species give different predominant fermentation products.

1.6.3 The hemicellulolytic rumen bacteria

1.6.3.1 *Butyrivibrio fibrisolvens*

Butyrivibrio fibrisolvens belongs to the genetically diverse *Butyrivibrio/Pseudobutyrvibrio* genus (Clostridial cluster XIVa) (family Lachnospiraceae) and is considered the most efficient degrader of hemicellulose amongst the presently culturable rumen bacteria (Kopecny *et al.*, 2001, Orpin *et al.*, 1985). Furthermore, members of the *Butyrivibrio/Pseudobutyrvibrio* are the predominant butyrate producing bacteria in the rumen (Asanuma *et al.*, 2005), and utilise exclusively one of two butyrate synthesis pathways (Diez-Gonzalez *et al.*, 1999,

Paillard *et al.*, 2007a). Isolates with *Butyrivibrio*-like phenotypic and metabolic qualities were originally all classified as *Butyrivibrio fibrisolvens*, although individual cultures were known to have diverse characteristics. As taxonomic methods have been refined this group of cultures have been reclassified into the *Butyrivibrio* and *Pseudobutyrvibrio* genera, both of which now contain several different species. (Kopečný *et al.*, 2001, Moon *et al.*, 2008).

B. fibrisolvens cells are prevalent in the rumen ecosystem (Reilly *et al.*, 2002), especially so when the animal is grazed on high fibre forage (Mrazek *et al.*, 2006). Many strains produce an array of fibrolytic enzymes that collectively degrade the xylan backbone of hemicellulose (Lin and Thomson, 1991a, Lin and Thomson, 1991b, Mannarelli *et al.*, 1990, Utt *et al.*, 1991) and hydrolyse the majority of substituent groups that mediate the interaction between hemicellulose and the surrounding cell wall polymers (Dalrymple and Swadling, 1997, Hespell and Obryan, 1992, Utt *et al.*, 1991). In particular, a number of arabinofuranosidases and esterases have been characterised that are likely to hydrolyse the di- and oligoferulate linkages that negatively influence forage degradation by the rumen microbiome (Dalrymple and Swadling, 1997, Hespell and Obryan, 1992, Hespell and Cotta, 1995, McSweeney *et al.*, 1998, Utt *et al.*, 1991).

B. fibrisolvens strains are versatile in their ability to degrade xylans from a variety of intact forages with differing chemical and physical properties (Coen and Dehority, 1970, Hespell and Cotta, 1995, Miron and Benghedalia, 1993, Miron *et al.*, 1994). Hespell and Cotta (1995) showed that both the rate and extent of xylan degradation by *B. fibrisolvens* H17c are not negatively influenced by either the type of xylan backbone substitution or the water solubility of the polymer. They also demonstrated that xylans containing high proportions of arabinose and/or glucuronic acids, as well as xylans of variable solubility were all rapidly and extensively degraded. *B. fibrisolvens* strains are also capable of growing on a range of simple sugars and plant polysaccharides including pectins, mannans, and starch, in addition to hemicellulose (Hespell, 1991, Kasperowicz, 1994, Marounek and Duskova, 1999), although in spite of producing a number of cellulose degrading enzymes (Berger *et al.*, 1989, Berger *et al.*, 1990) all strains are incapable of utilising cellulose *in vitro* (Attwood and Reilly, 1995). Recently, culture independent analyses have shown that the majority of the rumen *B. fibrisolvens* population adheres to the solid-fraction of the

rumen contents (Koike *et al.*, 2003b, Tajima *et al.*, 1999, Whitford *et al.*, 1998), although the mechanisms of substrate attachment are currently unknown.

A feature of many *B. fibrisolvens* strains is a high proteolytic activity primarily of the serine type (Attwood and Reilly, 1995, Cotta and Hespell, 1986, Sales *et al.*, 2000). Cotta and Hespell (1986) examined the relative proteolytic activity in several *Butyrivibrio* and *Pseudobutyrvibrio* species, and found the highest levels in strains including *B. fibrisolvens* H17c and 12, and *Pseudobutyrvibrio* sp. 49. In most strains the majority of the proteolytic activity was detected in the culture medium rather than being cell associated or cytosolic.

1.6.3.2 *Prevotella bryantii* and *Prevotella ruminicola*

Many prevalent rumen microbes that are unable to degrade intact plant cell walls thrive by metabolic cross-feeding, deriving energy from the initial degradation products of carbohydrates released by the primary fibre degraders, or from the conversion of fermentation products (Osborne and Dehority, 1989). The non-cellulolytic, Gram-negative *Prevotella bryantii* (Dehority, 1966) is such a species. Although it is considered an important contributor to hemicellulose degradation, the ability to utilise xylan is limited strictly to soluble substrates (Miyazaki *et al.*, 1997). The related *P. ruminicola* also degrades insoluble xylan *in vitro* (Dodd *et al.*, 2009) although evidence of insoluble xylan degradation in liquid culture is lacking. Furthermore, *P. ruminicola* preferentially degrades and utilises soluble xylooligosaccharides over intact xylan, and at a rate exceeding that of other xylanolytic rumen bacteria (Cotta, 1993). *Prevotella* species are capable of utilising other non-cellulosic cell wall polymers such as pectin (Kasperowicz, 1994, Marounek and Duskova, 1999) and mixed linkage glucans (Fields *et al.*, 1998).

Both *P. bryantii* and *ruminicola* produce multiple xylanolytic enzymes, including endoxylanases (Flint *et al.*, 1997, Gasparic *et al.*, 1995b, Whitehead, 1993), an α -L-arabinofuranosidase (Gasparic *et al.*, 1995a), an oxygen sensitive β -xylosidase (Gasparic *et al.*, 1995b) and a bi-functional xylanase-ferulic acid esterase (Dodd *et al.*, 2009). Recently, transcriptomic analysis of xylan degradation by *P. bryantii* exposed a cluster of genes that are up-regulated by growth on wheat arabinoxylan relative to a mixture of its monosaccharide components, and led to the functional characterisation of a novel GH5 endoxylanase that was previously annotated as a hypothetical protein (Dodd *et al.*, 2010). Fractionation studies of *P. bryantii* demonstrate that the majority of xylanolytic activity is located within the cell membrane or periplasm, and not at the

cell surface, which is in keeping with the hypothesis that this organism utilises smaller soluble oligosaccharides rather than actively degrading intact plant cell walls. Support for this hypothesis comes from the fact that no polysaccharide attachment mechanism has been identified in *Prevotella* spp., and that cells are found most abundantly in the rumen liquor.

1.6.4 The cellulolytic rumen bacteria

1.6.4.1 *Ruminococcus flavefaciens*, *Ruminococcus albus*, and *Fibrobacter succinogenes*

The Gram-positive *Ruminococcus flavefaciens* and *R. albus* (Antonopoulos *et al.*, 2003, Hungate, 1957) of Clostridial cluster IV (family *Ruminococcaceae*) (Aurilia *et al.*, 2000, Doerner and White, 1990, Flint *et al.*, 1993, Pettipher and Latham, 1979) and the Gram-negative *Fibrobacter succinogenes* (Bryant and Doetsch, 1955) are together considered the predominant rumen cellulose degraders. Strains of all three species produce many well characterised cellulolytic enzymes that are active upon plant-derived cellulose and cellooligosaccharides. Growing *F. succinogenes* cultures are known to degrade cellulose at a rate superior to that of most other organisms (Halliwell and Bryant, 1963, Maglione *et al.*, 1997).

The draft genome sequences of *R. flavefaciens* FD-1 (Berg Miller *et al.*, 2009) and *F. succinogenes* (Morrison *et al.*, 2003) have revealed the presence of 95 and more than 100 coding sequences (CDSs) encoding polysaccharide degrading enzymes, respectively, the majority of which are responsible for the degradation of cellulosic substrates. Strains of all three cellulolytic species also produce enzymes that may degrade non-cellulosic components of lignocellulose, therefore enhancing the cellulolytic activity of the bacteria (Aurilia *et al.*, 2000, Flint *et al.*, 1993, Matte and Forsberg, 1992, McSweeney *et al.*, 1998, Nakamura *et al.*, 2002c, Paradis *et al.*, 1993, Romaniec *et al.*, 1989, Zhang and Flint, 1992).

Real-time PCR assays have recently confirmed that all three cellulolytic species cells are enriched in the fibre adherent fraction of the rumen microbiome (Yang *et al.*, 2009). A feature of *R. flavefaciens* is the presence of a large molecular weight multi-enzyme complex termed the cellulosome (Lamed *et al.*, 1983), which retains high concentrations of fibrolytic enzymes on the bacterial cell surface, and promotes synergy between cellulolytic and/or non-cellulolytic enzymes recognising different polysaccharide substrates (Murashima *et al.*, 2002a). A carbohydrate-binding module (CBM) within at least one of the enzymes comprising the *R. flavefaciens* cellulosome

facilitates attachment of the bacterial cell to cellulose (Rincon *et al.*, 2001). *R. flavefaciens* also utilises a distinct cell surface associated carbohydrate-binding protein that also mediates attachment of the bacterial cell to crystalline cellulose (Rincon *et al.*, 2007).

R. albus appears to use a variety of modes of attachment to cellulosic substrates. A cellulosome-like complex that contained at least 15 proteins was identified in strain F-40 (Ohara *et al.*, 2000a), and characterised F-40 fibrolytic enzymes contain the requisite dockerin and CBMs suggestive of bacterial cell attachment to cellulose (Ohara *et al.*, 2000b, Taguchi *et al.*, 2004a). Cellulose adhesion defective *R. albus* 8 mutants have been generated that lack two cell-bound cellulolytic enzymes, neither of which contain dockerin domains, which suggests that non-cellulosome associated secreted fibrolytic enzymes may play a direct role in cell attachment to cellulose (Devillard *et al.*, 2004). Furthermore, in certain growth conditions *R. albus* 8 is capable of adhesion to crystalline cellulose, and also synthesises a Type IV fimbrial (Pil) protein that is localised to the outer surface of the cell membrane (Pegden *et al.*, 1998). In the presence of a soluble cellulose analogue both the cell- and Pil-protein attachment to crystalline cellulose is diminished, suggesting that the cell-associated Pil-protein constitutes a third method of cell/substrate interaction.

F. succinogenes binds tightly to the plant cell wall, although there is no evidence that cells produce a cellulosome complex. Differential abundance analysis of outer membrane proteins extracted from wild-type and adhesion defective mutants (Jun *et al.*, 2007) demonstrated that *F. succinogenes* S85 expresses a Type IV fimbrial similar to that expressed by *R. albus* (Pegden *et al.*, 1998) which may play a role in cell adhesion to crystalline cellulose. Furthermore, at least one cell associated cellulose binding protein (Gong *et al.*, 1996) and one endoglucanase (Jun *et al.*, 2007) appear to mediate interaction with crystalline and amorphous cellulose respectively.

1.6.4.2 *Eubacterium cellulosolvens*

Eubacterium cellulosolvens (family *Lachnospiraceae*) is an abundant cellulolytic bacterium in the rumen of forage fed animals (van Gylswyk, 1990) that produces large amounts of lactic acid, as well as formic and butyric acids, and utilises acetic and propionic acids when grown in the presence of cellulose (van Gylswyk and van der Toorn, 1986). Some strains are capable of growth on xylan, xylose, and arabinose, but these substrates do not support high fermentation rates, and no lactic acid is produced (van Gylswyk and van der Toorn, 1986).

When *E. cellulosolvens* cells are cultured in cellulose-only containing medium, they produce a large extracellular protuberance that binds to cellulose and contains cellulase activity (Blair and Anderson, 1999). Separate studies have characterised four cellulose degrading proteins including a secreted endoglucanase (Cel5A) versatile in its ability to degrade a range of cellulosic substrates, as well as oat-spelt xylan, and two cell associated enzymes displaying enzymatic activity characteristic of typical endoglucanases (Taguchi *et al.*, 2008, Toyoda and Minato, 2002, Toyoda *et al.*, 2005, Yoda *et al.*, 2005). Cel5A possesses cellulose binding ability that is important for the catalytic function of the enzyme (Yoda *et al.*, 2005), and the remaining two enzymes are capable of binding a range of cellulosic substrates as well as oat-spelt xylan (Yoshimatsu *et al.*, 2007). Cell wall-binding domains in the later two proteins are likely to mediate the attachment of the enzymes to the bacterial cell surface and consequently adherence of the cells to insoluble substrate (Toyoda *et al.*, 2003, Toyoda *et al.*, 2005).

1.7 *Butyrivibrio proteoclasticus* B316^T

Butyrivibrio proteoclasticus B316^T (formerly *Clostridium proteoclasticum*) (Attwood *et al.*, 1996, Moon *et al.*, 2008) is a low G+C, butyrate-producing anaerobic bacterium that is prevalent in the rumen of animals grazing fresh pasture and silage-based diets (Paillard *et al.*, 2007b, Reilly and Attwood, 1998), and commonly detected in rumen bacterial 16S rDNA libraries (Edwards *et al.*, 2004). Cells are rod-shaped with a single sub-polar flagellum, and have a Gram-positive cell wall architecture although established cultures appear Gram-negative (Figure 1.6) (Attwood *et al.*, 1996).

B. proteoclasticus cells were first isolated from the rumen contents of a New Zealand cow grazing on ryegrass-clover pasture (Attwood and Reilly, 1995) and had high cell-associated and/or cytosolic proteolytic activity, mainly of the serine type, although lower amounts of cysteine and metalloprotease activity were also detected (Attwood *et al.*, 1996). Rapid plant protein degradation in the rumen can lead to up to 50% loss of available protein and inefficient utilisation of pasture nitrogen (Macrae and Ulyatt, 1974). As a consequence *B. proteoclasticus* was examined initially with the view of being a target for improving protein utilisation in New Zealand ruminants.

Based on near full-length 16S rRNA gene sequence analysis, *B. proteoclasticus* is taxonomically most closely related to the fibrolytic *Butyrivibrio/Pseudobutyrvibrio*

group of bacteria (family *Lachnospiraceae*), which belongs to the *Clostridial* rRNA subcluster XIVa (Figure 1.7). DNA G+C content was determined to be 40.04 mol%, which is very similar to that of closely related *Butyrivibrio* strains DSM 10304, and H17c(SA) that was subsequently reclassified as a *B. proteoclasticus* strain (Moon *et al.*, 2008).

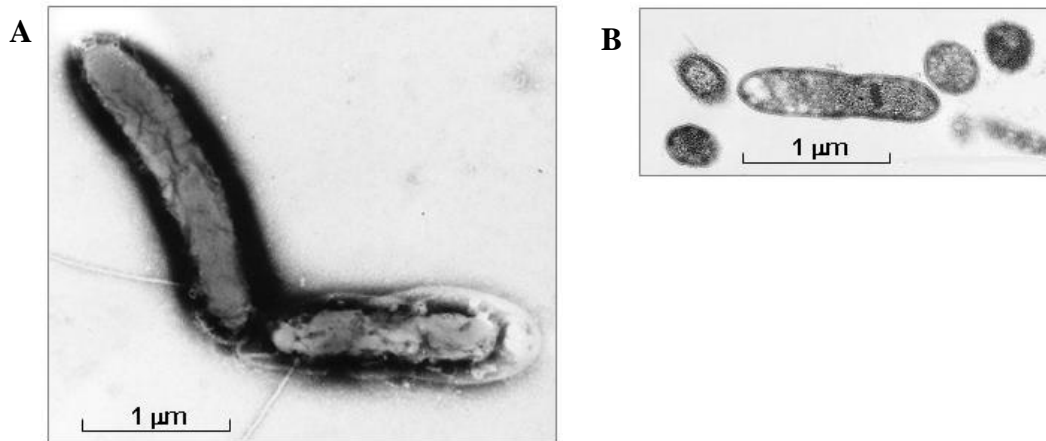


Figure 1.6. Ultrastructure of *B. proteoclasticus* B316T cells. Electron micrographs of negatively stained *B. proteoclasticus* cells showing the sub-polar flagellum (A), and thin-sectioned cells showing the Gram-positive cell wall structure (B). Figure courtesy of Attwood *et al.* (1996).

1.7.1 Growth substrate utilisation and fermentation

The initial characterisation of *B. proteoclasticus* together with more recent studies have demonstrated that cells growing *in vitro* display substrate utilisation and fermentative profiles similar to closely related *Butyrivibrio* species (Attwood *et al.*, 1996, Moon *et al.*, 2008). *B. proteoclasticus* cells can utilise a variety of soluble mono- and disaccharides including arabinose, cellobiose, galactose, glucose, rhamnose, sucrose, and xylose. Furthermore, the bacterium is one of a small number of rumen fibrolytic microbes capable of efficiently degrading and utilising the insoluble, complex polysaccharide xylan, and is capable of utilising pectin and starch also (Attwood *et al.*, 1996, Moon *et al.*, 2008). Notably, *B. proteoclasticus* is unable to degrade or utilise crystalline cellulose *in vitro*.

A metabolic footprinting approach was used to examine the growth substrate utilisation pattern of *B. proteoclasticus* cells cultured in medium containing glucose, xylan, and pectin as the major substrates (Villas-Boas *et al.*, 2006). After a 24 hour culture period *B. proteoclasticus* generated increased levels of xylose, rhamnose, α -

mannoside, and β -arabinoside which indicated that xylan and pectin were rapidly degraded. When compared to the metabolic footprint of the fibrolytic *B. fibrisolvens* AR10, *B. proteoclasticus* degraded more fibre, liberated more than 10-fold the amount of xylose, and utilised almost 10-fold more glucose than its close relative.

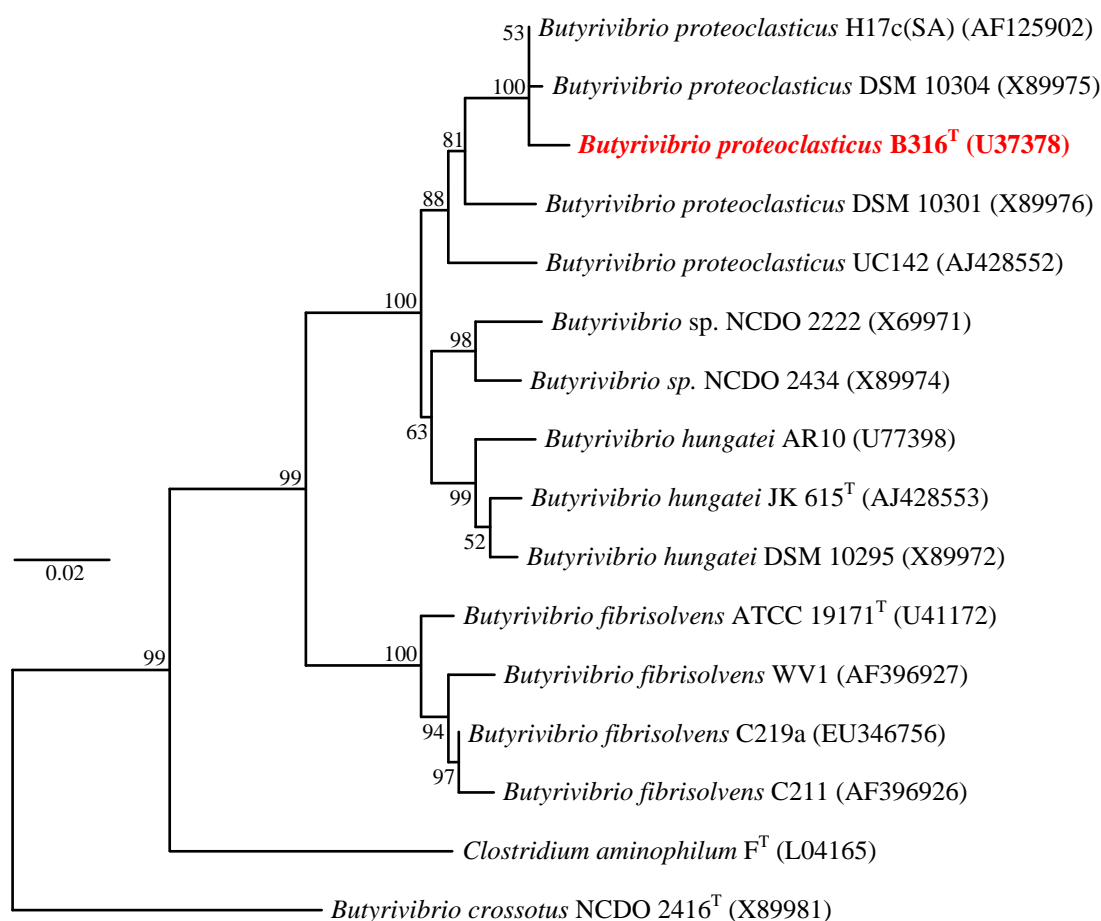


Figure 1.7. Summary of the 16S rRNA gene phylogeny of *B. proteoclasticus* and closely related strains within the *Clostridial* subcluster XIVa. Percentage bootstrap values are shown at nodes if >50%. GenBank accession numbers for each species are shown in brackets. Bar = 0.02 nucleotide substitutions per site. Figure adapted from Moon *et al.* (2008).

1.7.2 The *B. proteoclasticus* genome sequence

The 4.4 Mb *B. proteoclasticus* genome has recently been sequenced by the Rumen Microbial Genomics laboratory, Grasslands Research Centre, Palmerston North, New Zealand (Kelly *et al.*, 2010). It encodes 3813 CDSs that are distributed among the main chromosome (BPc1), a chromid (BPc2) (Harrison *et al.*, 2010) and two megaplasms (pCY360 and pCY186) (Table 1.4 and Figure 1.8).

Table 1.4. Summary of the four replicons comprising the *B. proteoclasticus* genome, and distribution of genes involved in plant carbohydrate metabolism and transport.^a

	BPc1	BPc2	pCY360	pCY186	Total
Size (bp)	3 554 804	302 358	361 399	186 325	4 404 886
G+C%	40.21	40.04	38.95	38.10	39.33
CDSs	2939	251	425	198	3813
rRNAs	4	2	-	-	6
tRNAs	47	2	-	1	50
GHs	89	15	-	-	104
CEs	19	-	-	-	19
GH/CE ^b	4	-	-	-	4
PLs	1	1	-	-	2
GTs	2	-	-	-	2
Other carbohydrate active proteins	83	6	-	-	89
Transporter proteins	85	11	-	-	96

^a Table adapted from Kelly *et al.* (2010).

^b Enzymes contain multiple catalytic domains annotated as GH or CE.

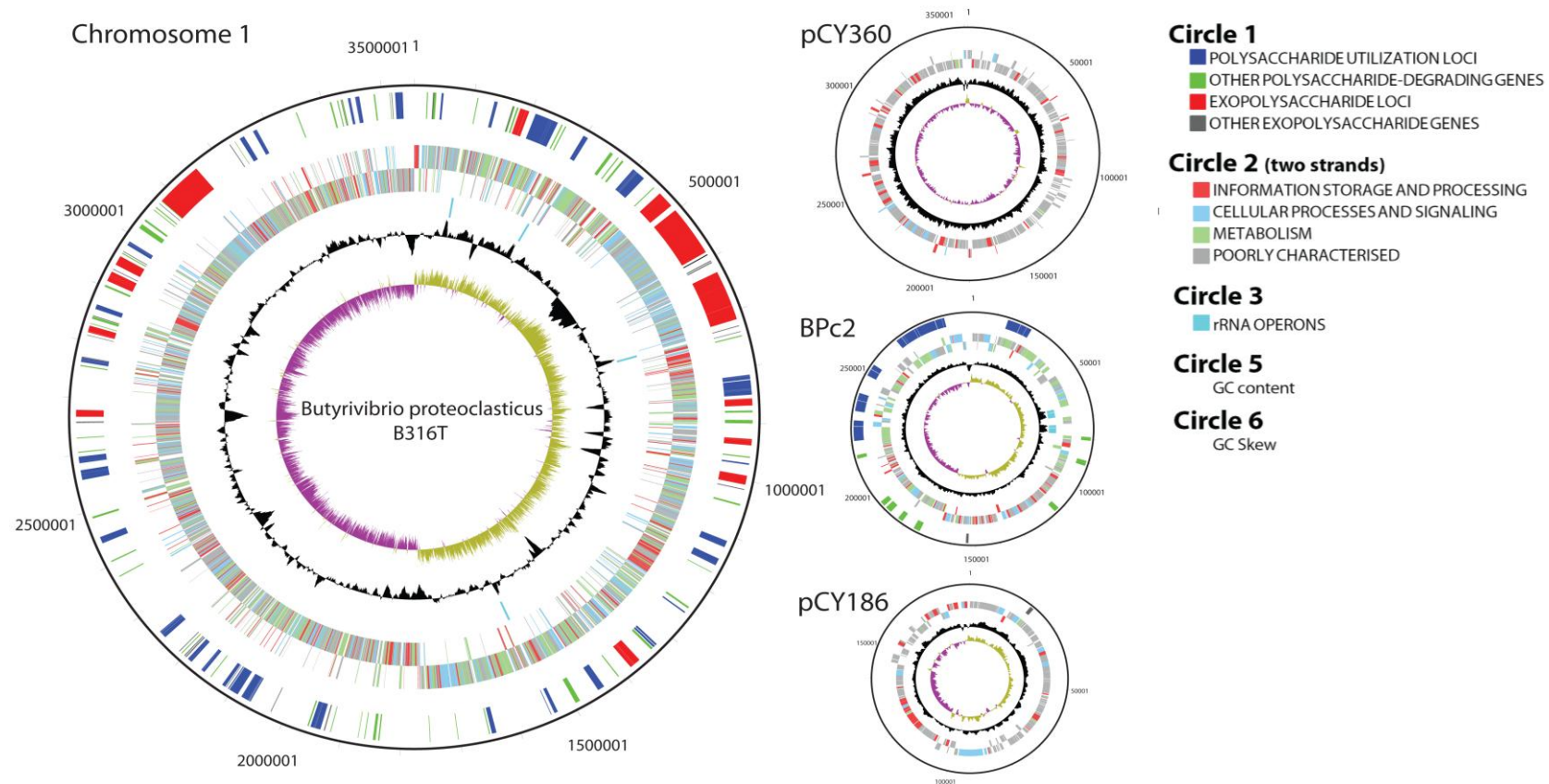


Figure 1.8. The genome atlas of *Butyrivibrio proteoclasticus* B316^T. The four replicons that make up the *B. proteoclasticus* genome and the location of the genes predicted to encode proteins involved in polysaccharide degradation are shown. The four rRNA operons on the main chromosome and the two on BPc2 are also shown. The colour coding of the genomic features in circle 2 represents the different Clusters of Orthologous Groups (COG) categories. Figure from Kelly *et al.* (2010).

1.7.2.1 The *B. proteoclasticus* glycobiome

A feature of the genome sequence that is consistent with the growth characteristics and substrate utilisation patterns of the bacterium is the set of genes devoted to the breakdown and reassembly of complex polysaccharides (glycobiome). As of August 2009, the glycoside hydrolase content of the *B. proteoclasticus* genome ranked tenth among the 841 bacterial genome sequences recorded in the CAZy database, while the glycosyl transferase (GT) content was greater than all other recorded bacterial species (Kelly *et al.*, 2010). The *B. proteoclasticus* genome encodes 316 proteins involved in carbohydrate metabolism and transport, 134 of which contain one or more GH, CE, or PL or GT catalytic domains that target plant structural polysaccharides. More than 89% of these are located on chromosome 1, with the remainder found on BPc2 (Table 1.4). The megaplasמידs are hypothesised to encode properties important for survival in the rumen environment, but make no detectable contribution to the *B. proteoclasticus* glycobiome.

Based on their catalytic domains each of these polysaccharidases are classified into 44 different CAZy families, which exceeds the number of CAZy families represented in the cellulolytic *R. flavefaciens* (Berg Miller *et al.*, 2009) as well as the fibre-adherent rumen metagenome (Brulc *et al.*, 2009). The GH2, GH3, GH13, GH43 families are the most highly represented, with 43 enzymes classified within the four families alone. In addition, five polysaccharidases are currently not matched to any CAZy family, and three genes are present that encode carbohydrate-binding proteins that may also play a role in carbohydrate degradation. *B. proteoclasticus* encodes five GH25 lysozymes that are likely to be involved in microbial cell wall hydrolysis, but not lignocellulose degradation.

1.7.2.2 Polysaccharide utilisation loci

Almost half the genes in the *B. proteoclasticus* genome that encode lignocellulose degrading enzymes are clustered within one of 36 regions (Figure 1.8, Circle 1) termed polysaccharide utilisation loci (PULs) (Bjursell *et al.*, 2006). The starch utilisation system (Sus) locus of the Gram-negative gut bacterium *Bacteroides thetaiotaomicron* was the first PUL to be identified (Reeves *et al.*, 1997), and was later shown to encode a set of proteins that formed a cell-envelope associated multiprotein system that enabled the bacterium to bind and degrade starch (Bjursell *et al.*, 2006). A feature of the Sus locus was the coordinated action of several gene products involved in substrate-binding and degradation (Bjursell *et al.*, 2006, Martens *et al.*, 2008). It is

now recognised that PULs encode cell envelope systems that usually include one or more polysaccharide degrading enzymes, together with energy dependent, cell-associated transmembrane transporters, transcriptional regulators, environmental sensors and genes involved in further polysaccharide metabolism. Despite the abundance of PULs in *Bacteroides* species, the substrate specificities of all but the Sus PUL and a homologous starch specific PUL in *Bacteroides fragilis* (Spence *et al.*, 2006) remain poorly defined. The clustering of genes involved in oligosaccharide breakdown presumably allows coordinated control of enzyme production, substrate transport and intracellular metabolism, and highlights the fact that the synergistic activity of several proteins can be more sophisticated and effective than their individual functions in isolation. Although it is becoming apparent that the presence and activity of PULs are a common strategy in fibrolytic Gram-positive (Berg Miller *et al.*, 2009) and Gram-negative (Martens *et al.*, 2009a) bacteria, *B. proteoclasticus* is the first member of the *Butyrivibrio/Pseudobutyrvibrio* assemblage in which PULs have been identified. Collectively the *B. proteoclasticus* PUL clustered proteins comprise 7.2% of the total proteome. Similar to *B. thetaiotaomicron* most of the *B. proteoclasticus* PULs contain components of ATP-binding cassette (ABC) transport systems, transcriptional regulators such as AraC and LacI family proteins, environmental sensors including two-component histidine kinases and response regulator proteins, and genes involved in further oligosaccharide metabolism.

1.8 Proteomics

The proteome is ‘the entire protein complement encoded by a genome, cell, tissue, or organism. It is the set of proteins produced by an organism at a given time, in response to a defined set of environmental conditions, and includes all protein isoforms and post-translationally modified proteins. The development of high throughput DNA sequencing technologies has led to the rapid accumulation of large volumes of genome sequence data for many bacterial species, and has dramatically advanced our understanding of many cellular processes. In spite of this, functional genomics does not provide information of the physiological processes occurring at a specific point in time, which are performed by the highly dynamic set of synthesised proteins. Expression proteomics aims to identify and quantify the abundance of all proteins encoded by a genome, assess their cellular location, and examine potential post-translational modifications. As a complement to functional genomics, expression

proteomics is vital to enhancing our understanding of cellular function. The foundations of contemporary proteomics lie with the development of two-dimensional electrophoresis (2-DE) by Klose (1975) and O'Farrell (1975), and methods to efficiently convert large, intact biomolecules into volatile gas phase ions such that they may be analysed by mass spectrometry (MS).

1.8.1 Two-dimensional electrophoresis

The power of 2-DE is the ability to separate complex protein mixtures into single protein species using each protein's unrelated properties of isoelectric point (*pI*) and molecular weight (MW). 2-DE can resolve hundreds to thousands of proteins on a single polyacrylamide gel, and can rapidly expose differences in protein abundance, identify protein isoforms, and elucidate the presence of post-translational modifications (PTM's) such as phosphorylation and glycosylation. The development of immobilised pH gradient (IPG) strips has had a major impact on the utility of the 2-DE technique (Bjellqvist *et al.*, 1982). Wide and medium range IPG strips such as IPGs 3–10, 4–9, or 4–7 are excellent for the analysis of simple proteomes or when an overview of more complex proteomes is required, while narrow range overlapping IPG strips enhance the resolving power of the first dimension focusing. Using this approach, Cho *et al* (2003) resolved 1237 individual protein species between *pI* 3.5 and 5.5 when analysing a cytosolic protein sample of the halophilic bacterium *Halobacterium salinarum*. When a *Saccharomyces cerevisiae* total protein extract was analysed using a series of IPG strips covering as little as one pH unit per IPG strip, 2286 individual protein spots were visualised, compared with 755 when using a standard single pH 3-10 gradient IPG strip (Wildgruber *et al.*, 2000).

1.8.2 Mass spectrometry

In recent years MS technologies have evolved as the dominant method of protein identification within the field of proteomics. Two "soft ionisation" methods, matrix-assisted-laser-desorption-ionisation (MALDI) (Karas and Hillenkamp, 1988) and electrospray ionisation (ESI) (Fenn *et al.*, 1989) are used routinely to introduce analytes into mass spectrometers. MALDI relies on the co-crystallisation of a peptide sample with acidified matrix upon a sample plate. Sample preparation for MALDI based mass spectrometry is straightforward and sample ionisation is robust in the presence of contaminating substances such as salts and detergents. Furthermore, the crystallised sample is stable and can be re-analysed if required. ESI requires the

solubilised sample to be maintained in the liquid phase and is considerably more sensitive to contaminating substances. Furthermore, the liquid phase nature of the preparation means the sample cannot be re-analysed. The major benefits of ESI include the high ionisation efficiency relative to MALDI, and the ability to interface the ESI-MS to a high-resolution liquid-chromatographic separation apparatus.

1.8.2.1 Peptide mass fingerprinting

MALDI ionisation is most often coupled to a Time-of-Flight (TOF) mass analyser (MALDI-TOF), and used to rapidly measure peptide masses and identify the parent protein by a process termed peptide mass fingerprinting (PMF). The underlying principle of PMF is the comparison of experimentally derived peptide masses with theoretically calculated peptide masses generated through *in silico* digestion of translated genomic sequences, using the expected cleavage specificities of site-specific proteolytic enzymes (Yamazaki and Tove, 1977, Yates *et al.*, 1993). PMF search algorithms perform protein database searches and return a list of possible protein matches ranked according to variables including the number of matched peptide masses, the size of the individually matched peptides, the peptide mass error, the database size, and the number of non-matched masses. Molecular Weight Search (MOWSE) (Pappin *et al.*, 1993) is a sophisticated scoring algorithm implemented in the PMF search algorithms Mascot (<http://www.matrixscience.com>) (Perkins *et al.*, 1999), and Protein Prospector (<http://prospector.ucsf.edu/prospector/mshome.htm>) (Clauser *et al.*, 1999). ProFound is a commonly used PMF search algorithm (<http://prowl.rockefeller.edu>) (Zhang and Chait, 2000), which uses the Bayesian theory to rank search results according to their probability of occurrence. An evaluation of the commonly used PMF algorithms concluded that Mascot and ProFound were superior to Protein Prospector (Chamrad *et al.*, 2004).

A limitation of PMF for unambiguous protein identification is the probabilistic nature of the search process. Measured peptide masses can randomly match to peptides from a sequence database due to the necessity to set a maximum allowable mass tolerance, and as a result there is always a risk of obtaining false-positive protein identification. This probabilistic approach also necessitates the analysis of peptides derived from single protein species wherever possible, which is why 2-DE is often coupled with MALDI-TOF. When several proteins are identified as being present in a single gel-plug, manual analysis of the individual identifications is necessary. Furthermore, MALDI ionisation is biased towards arginine carboxy-terminating

peptides, and typically results in only 30-45% coverage of the complete protein sequence (Resing and Ahn, 2004). 2-DE is also biased against hydrophobic proteins, basic proteins, and proteins below 15 kDa. An alternative to the 2-DE MALDI-TOF approach is to maintain protein or peptide mixtures in the liquid phase, separate the components using high-pressure liquid chromatography (HPLC), and analyse the eluate using tandem mass spectrometry (MS/MS).

1.8.2.2 Liquid chromatography-tandem mass spectrometry

The power of liquid chromatography tandem mass spectrometry (LC-MS/MS) is the ability to obtain peptide sequence data that is difficult and sometimes impossible to generate by MALDI-TOF. Comparison of MS/MS spectral data with predicted peptide sequences by peptide fragmentation fingerprinting (PFF) can be used to identify proteins with high confidence, sometimes when as few as two peptides are matched to the parent protein. Furthermore, it is often possible to identify proteins by LC-MS/MS that are difficult to detect using a 2-DE MALDI-TOF approach, such as hydrophobic (Wolff *et al.*, 2008), basic, and low molecular weight proteins (Kuntumalla *et al.*, 2009).

Quantitative LC-MS/MS based proteomic analyses typically use either isotopic protein labelling prior to chromatographic separation, such as isobaric tagging for relative and absolute quantitation (iTRAQ) (Ross *et al.*, 2004) and stable isotope labelling with amino acids in cell culture (SILAC) (Ong *et al.*, 2002), or a label-free approach that uses data acquired during separation and analysis to derive an estimation of relative protein abundances (Old *et al.*, 2005). A benefit of the latter approach is the elimination of potential bias induced by variable protein labelling or detection of labelled sample. Label-free quantitation utilises measurement of variables such as mass spectral peak intensities (Chelius and Bondarenko, 2002) and spectral counting (Zhang *et al.*, 2006). Peak intensities of peptide ions can correlate with protein abundances, while spectral counting methods estimate protein abundance by comparing the number of MS/MS spectra assigned to each protein, based on the assumption that the number of observed peptides correlates with protein abundance (Rappsilber *et al.*, 2002). LC-MS/MS is routinely used to identify single protein species within complex mixtures without prior subcellular fractionation (Francis *et al.*, 2005, Wang *et al.*, 2004).

1.8.3 Bacterial proteomics

Proteomic technologies have been used to examine the proteomes of many industrially and medically important bacteria cultured under a variety of defined conditions. The model Gram-negative and Gram-positive organisms, *Escherichia coli* and *Bacillus subtilis* have naturally been the focus of extensive proteomic examination, as have other important *Bacillus* species such as *B. anthracis* and *B. licheniformis* (Chitlaru and Shafferman, 2009, Deutscher and Saier, 2005, Gilois *et al.*, 2007, Hecker *et al.*, 2008, Ohlmeier S, 2000, Tam *et al.*, 2006, Voigt *et al.*, 2007). Several groups have demonstrated the utility of a combined gel-based and gel-free proteomic approach in achieving improved proteome coverage of *Bacillus* species and other Gram-positive bacteria (Hahne *et al.*, 2008, Jungblut *et al.*, 2010, Schmidt *et al.*, 2004, Wolff *et al.*, 2006).

Proteomic technologies have been utilised in the analysis of several important human gastrointestinal bacteria, including *Fusobacterium varium* and *Lactobacillus plantarum*, as well as the polysaccharolytic *Bifidobacterium longum* (Cohen *et al.*, 2006, Potrykus *et al.*, 2008, Yuan *et al.*, 2006). In particular, analyses of the cytosolic, exported, and cell envelope proteins of *B. longum*, and the adaptation of cells to low pH environments has revealed important insights into the glycosyl hydrolases produced by this bacterium (Ruiz *et al.*, 2009, Sanchez *et al.*, 2007, Sanchez *et al.*, 2008, Yuan *et al.*, 2006). An LC-MS/MS label-free quantitative proteomics approach was used recently to examine the cytosolic proteomes of *B. longum* strains differing in their heat shock resistance (Guillaume *et al.*, 2009).

1.8.3.1 Proteomics of fibrolytic bacteria

Proteomics approaches have been used with success to investigate the fibrolytic enzyme systems of a number of polysaccharide degrading bacteria. Murashima and co-workers identified several cellulosomal enzymes expressed by the cellulolytic soil bacterium *Clostridium cellulosolvens* (Murashima *et al.*, 2002b), and subsequently demonstrated using 2-DE/MALDI-TOF that the cellulosome composition and relative enzyme abundance are both regulated by the growth substrate (Han *et al.*, 2004, Han *et al.*, 2005). Similarly, the cellulosome composition of *C. cellulolyticum*, and the effect of polymeric substrates on the organisation of extracellular multienzyme complexes in *Paenibacillus curdlanolyticus* have recently been determined (Blouzard *et al.*, 2007, Waeonukul *et al.*, 2008). The effect of growth substrates on the relative abundance of fibrolytic enzymes has also been examined for a diverse range of

bacteria including *Bacillus* sp. K1, *Cellulomonas flavigena*, and the marine bacterium *Saccharophagus degradans* (Chu *et al.*, 2000, Sanchez-Herrera *et al.*, 2007, Sato *et al.*, 2007, Taylor *et al.*, 2006).

1.8.3.2 Rumen proteomics

In spite of the variety of gut bacteria having been the subject of intensive proteomic efforts, and the importance of the rumen microflora to the energy supply of the host, the application of proteomic technologies to the study of the prevalent fibrolytic bacteria of the bovine rumen has been limited. Devillard *et al* (2004) used 2-DE and MALDI-TOF MS to investigate the mechanisms underlying attachment of *Ruminococcus albus* 8 to cellulosic substrates. Later, the group used a similar approach to demonstrate the increased abundance of the molecular chaperone GroEL within the cell-membrane associated proteome when cells were exposed to linoleic and conjugated linoleic acids (Devillard *et al.*, 2006). Following the publication of the completed genome sequence of the Gram-negative, non-fibrolytic, succinic acid producing rumen bacterium *Mannheimia succiniciproducens* (Hong *et al.*, 2004), the proteome of this bacterium has been the subject of recent attention. 2-DE reference maps of the whole cellular, membrane, and secreted proteins allowed the identification of more than 200 proteins, and revealed important insights into growth phase dependent physiological alterations (Lee *et al.*, 2006). This information was subsequently used to engineer a mutant strain that harbours improved succinic acid producing capability (Lee and Lee, 2010). There are currently no reports of proteomics being applied to the analysis of the polysaccharide degrading enzyme systems of fibrolytic rumen bacteria.

1.9 The theoretical total proteome of *B. proteoclasticus*

The *pI* value and size of the 3813 *B. proteoclasticus* protein products ranges from 3.17 to 13.41, and 3.4 kDa to 656 kDa, respectively (Kelly *et al.*, 2010). When each protein is plotted on a theoretical 2-DE map (Figure 1.9) a bimodal *pI* distribution and a moderate acidic skew is observed, with 73% of the proteins predicted to have a *pI* value below 7. Twenty-two percent of the theoretical proteome has *pI* values between 8 and 11, and a region of fewer proteins (5.2% of the total) is found between *pI* 7 and 8. Only 28 proteins have a *pI* value greater than 11. The size distribution of the total proteome is consistent with other Gram-positive bacteria (Voigt *et al.*, 2007, Wolff *et al.*, 2007), with 92% of all proteins predicted to fall within 10 to 150 kDa.

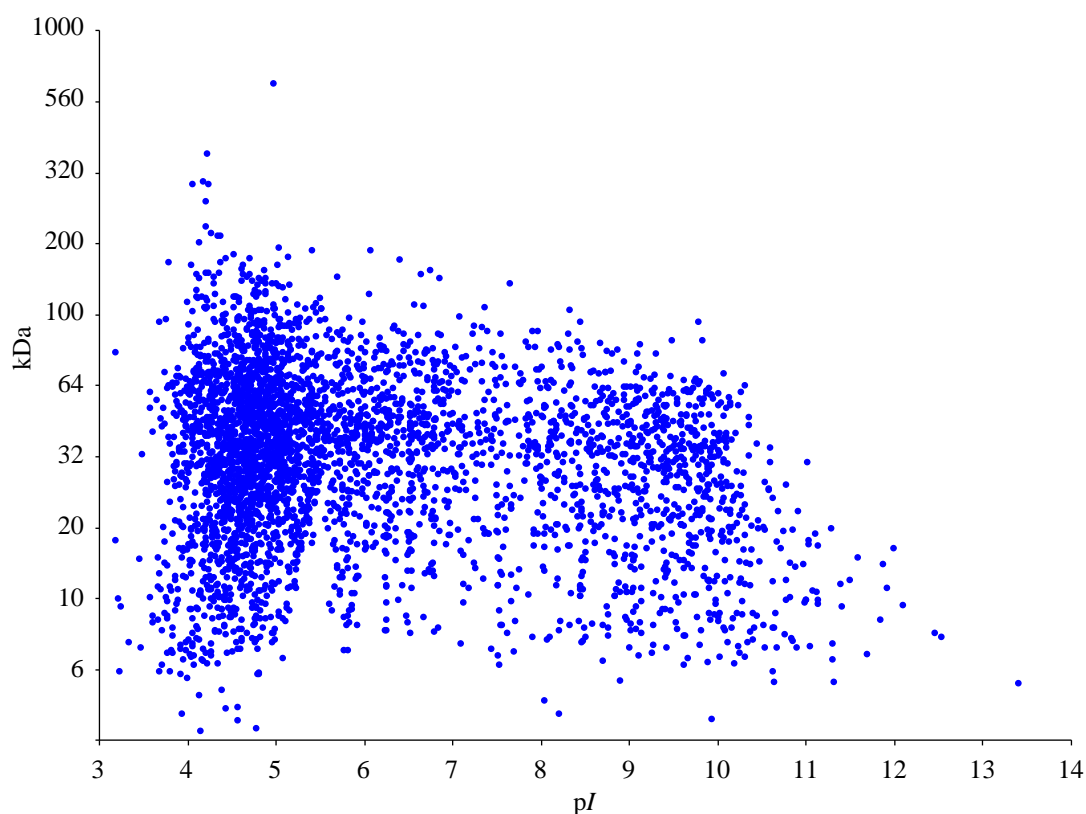


Figure 1.9. The theoretical 2-DE total proteome of *B. proteoclasticus*. The pI and size values for each protein were calculated using the Emboss analysis package (<http://emboss.sourceforge.net/index.html>). Each spot represents a single predicted protein. The y-axis is presented in logarithmic scale to represent separation of proteins by 2-DE.

1.9.1 Protein function classification of the predicted *B. proteoclasticus* proteome

Using a combination of Pfam (<http://pfam.sanger.ac.uk/>), Basic Local Alignment Search Tool (BLASTp) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), and Tigrfam (<http://blast.jcvi.org/web-hmm/>), and analysis of the full length protein sequences, each protein has been categorised into one of 21 functional groups (Kelly *et al.*, 2010) (summarised in Appendix A, Table A.1), which demonstrates that 5.7% (220 proteins) of the *B. proteoclasticus* proteome is devoted to carbohydrate metabolism. This set of proteins consists of 131 polysaccharide degrading enzymes, 44 enzymes involved in glycolysis, gluconeogenesis, or the pentose phosphate pathway, and others that are active upon monosaccharides and aminosugars, or involved in glycogen biosynthesis. Two-hundred and eighty-two proteins were predicted to be involved in cellular transport, 96 of which were elements of systems involved in the transmembrane transport of sugars or oligosaccharides. The genes encoding 80 of these transporter

proteins were also components of a PUL. *B. proteoclasticus* may produce 39 proteins mediating protein degradation, including one subtilisin family and two Htra family serine proteases, three cysteine proteases, and at least six metallopeptidases. Notably, 45% of the *B. proteoclasticus* proteome is currently annotated as hypothetical or encodes proteins with no known function (Figure 1.10).

1.9.2 Subcellular distribution of the predicted *B. proteoclasticus* proteome

Within the set of 220 proteins involved in carbohydrate metabolism, 82% are predicted to be localised to the cytosol (Figure 1.10), of which 91 are lignocellulose-degrading enzymes that collectively represent 34 GH, CE, PL or GT families. Within the cytosol, members of the GH2, GH3, GH13, GH31, and GH43 families are the most prevalent. Notably, all of the GH2 and GH31 family enzymes and most of the GH3, GH13, and GH43 enzymes are localised to the cytosol. In addition, all five polysaccharidases unmatched to a CAZy family are also predicted cytosolic enzymes.

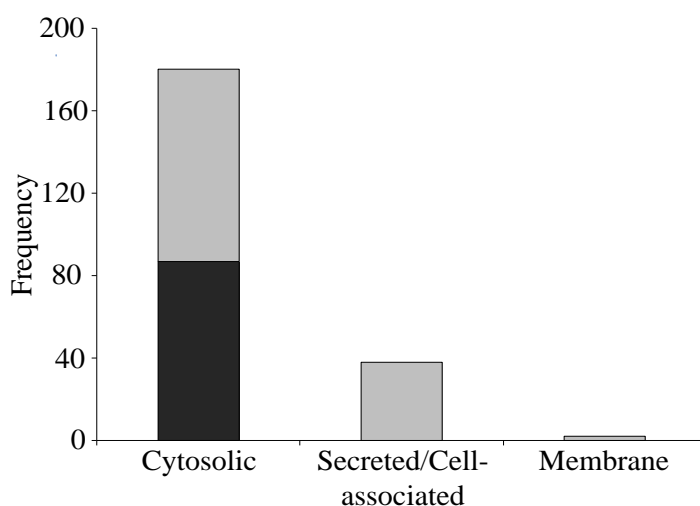


Figure 1.10. Summary of the predicted subcellular location of the carbohydrate active proteins produced by *B. proteoclasticus*. Total bar height shows the frequency of carbohydrate active proteins localised to each subcellular compartment. Light grey portion of each bar denotes the frequency of polysaccharide degrading enzymes. This data was compiled with the assistance of Dr. Bill Kelly and Dr. Sinead Leahy.

All 38 carbohydrate metabolising proteins that are predicted to be secreted are plant polysaccharidases, and are dominated by members of the GH3, GH5, GH10, and GH13 families. All members of the GH5 and GH10 families are secreted, as well as both carbohydrate binding proteins. Furthermore, 35% of the proteins predicted to be involved in the transmembrane-transport of oligosaccharides are predicted to be

secreted into the periplasmic or extracellular space. The two membrane bound polysaccharidases are the GH3 β -glucosidase Bgl3D, and the GH115 α -glucuronidase Gh115A, which have three, and one, transmembrane domain (TMD) respectively, and no identifiable signal peptide.

1.10 Project aims

It is predicted that a proteomic examination of the fibre degrading enzyme system of *B. proteoclasticus* will be a step toward mapping the complex conversion of plant biomass into milk and meat for human consumption. It is hypothesised that *B. proteoclasticus* cells grown on GAX synthesise a consortium of polysaccharide processing proteins that enable them to rapidly degrade and assimilate the substrate, and that the abundance of many of these proteins is sensitive to growth substrate. We anticipate that some of the proteins involved in polysaccharide processing may be similar to those produced by other prevalent hemicellulose degrading microbes, and that the production of novel fibrolytic enzymes will be found.

The goals of this research were 1) to identify *B. proteoclasticus* proteins that are involved in the degradation, assimilation, and metabolism of hemicellulose; and 2) to examine the relative abundance of those proteins in cells grown *in vitro* in the presence of GAX and the soluble monomeric sugar xylose. A combination of gel-based and gel-free proteomic analyses were used to examine the cytosolic and extracellular fractions of *B. proteoclasticus* cells. Central to achieving these goals was the sequencing and *in silico* translation of the *B. proteoclasticus* genome, which was completed by the AgResearch Rumen Microbial Genomics team at the Grasslands Research Centre, Palmerston North, New Zealand. The functionally annotated protein sequence database used throughout this research was “B316_GOLD_June2008”.

The predicted subcellular distribution of the carbohydrate active proteins produced by *B. proteoclasticus* dictated that the scope of this PhD project was limited to the analysis of the cytosolic and extracellular compartments. In a parallel study, a proteomic analysis of the membrane compartment of *B. proteoclasticus* was conducted by Dr. Judy Bond.

Chapter 2

Materials and Methods

2.1 Materials

Congo red was purchased from Ajax Finechem (NSW, Australia). MALDI-TOF MS Peptide Mass Fingerprint (PMF) Calibration Mixture 2 was obtained from Applied Biosciences (Foster City, CA, USA). Ultra-filtered water ($18.3 \text{ M}\Omega\cdot\text{cm}^{-1}$) was obtained from a Barnstead EasyPure UV filtration system (Barnstead International, Dubuque, IA, USA). Acetone, ammonium hydrogen carbonate, ammonium sulphate, bromophenol blue, CBB G-250, orthophosphoric acid, and TCA were purchased from BDH Laboratory Supplies (Poole, England, UK). Acrylamide, ammonium persulphate, DTT, Flamingo™ Fluorescent Stain Reagent, glycine, iodoacetamide (IAA), SDS, TEMED, Tris, and Tween® 20 were purchased from Bio-Rad (Hercules, CA, USA). Microtitre plates (96 well) were obtained from Corning Life Sciences (Corning, NY, USA). Perfect-Pure® C18 tips were purchased from Eppendorf AG (Hamburg, Germany). Ammonium citrate, apple pectin (76282), Avicel® PH-101, cellulase (from *Aspergillus niger*), phenol, and TFA were obtained from Fluka (Buchs, Switzerland). DryStrip cover fluid, Immobiline DryStrips (pI 3-10, non-linear (NL), pI 3-5.6 NL, pI 5.3- 6.5 NL and pI 6- 11 NL), IPG buffer, and 2D Quant kit were purchased from GE Healthcare (Uppsala, Sweden). LDS sample buffer (4x), sample reducing agent (10x), electrophoresis running buffer antioxidant, Benchmark™ and Peppermint Stick™ molecular weight standards, MOPS electrophoresis running buffer (20x), NuPAGE® Novex 4-12% Bis-Tris ZOOM® 2-DE gels, and NuPAGE® Novex 4-12% Bis-Tris 10 well and 12 well 1-DE (one-dimensional electrophoresis) gels were from Invitrogen (Carlsbad, CA, USA). HPLC-grade ACN, formic acid, glacial acetic acid, glycerol, HPLC-grade methanol, ProteoExtract™ protein precipitation kit, and HPLC-grade water were purchased from Merck (Darmstadt, Germany). Centricon® Ultracel® YM-10 Centrifugal Concentrator units were from Millipore (Billerica, MA, USA). Complete® Protease Inhibitor and Modified Sequencing-Grade Trypsin (Cat #11 521 187 001) were purchased from Roche Applied Science (Mannheim, Germany). CHAPS, CHCA, EDTA, D-(–)-fructose (F0127), pectinase (from *Aspergillus aculeatus*, P2611), Sephadex G-250®, thiourea, urea, xylanase (from *Thermomyces lanuginosus*, X2753), xylan (from oat-spelts, 95590), and xylose (X0627) were from Sigma-Aldrich (St. Louis, MO, USA).

2.2 Bacterial strains and culture conditions

Butyrivibrio proteoclasticus cells were used in all experiments and all cultures were prepared by Mr. Dong Li within the Rumen Functional Genomics Laboratory, AgResearch Limited, Grasslands Research Centre (Palmerston North, New Zealand). Cells were grown under strictly anaerobic conditions at 37°C, in modified M704 culture medium (500 mL) supplemented with either 0.1% xylan, 0.5% xylose, or in unsupplemented modified M704 medium. Cell cultures were harvested at mid-log phase (optical density 600 nm = 0.5) or stationary phase (optical density 600 nm = 0.7). Three biological replicate cultures were prepared for each growth substrate, harvested at each time point.

2.2.1 Modified M704 medium

Modified M704 medium was prepared by Mr. Dong Li and contained the following components: 0.075% (w/v) K_2HPO_4 , 0.2% (w/v) trypticase peptone (BBL), 0.2% (w/v) yeast extract, 0.0001% (w/v) haemin, 0.4% (w/v) Na_2CO_3 , 0.0001% (w/v) resazurin, 0.025% (w/v) cysteine-HCl x H_2O , 0.025% (w/v) Na_2S x 9 H_2O , 0.09% (w/v) NaCl, 0.045% (w/v) $(NH_4)_2SO_4$, 0.012% $CaCl_2$ x 2 H_2O , 0.02 (w/v) $MgSO_4$ x 7 H_2O , 0.17% (v/v) acetic acid, 0.06% (v/v) propionic acid, 0.04% (v/v) butyric acid, 0.01% (v/v) *n*-valeric acid, 0.01% (v/v) *iso*-butyric acid, 0.01% (v/v) DL-2-methyl butyric acid, and 0.01% (v/v) *iso*-valeric acid.

All ingredients except cysteine-HCl x H_2O , Na_2S x 9 H_2O , and Na_2CO_3 were dissolved in water and the medium was adjusted to pH 6.7 with 1 M HCl. The medium was boiled for 5 min and then cooled to room temperature under CO_2 gas flow. The Na_2CO_3 and cysteine-HCl x H_2O was then added and the medium was equilibrated with CO_2 to pH 6.8. The medium was then dispensed under CO_2 into 250 mL serum bottles, which were sealed and autoclaved at 121 °C for 15 min. Immediately before inoculation a sterile solution of Na_2S x 9 H_2O (1% w/v) was prepared separately and added to the medium (2.5% v/v).

2.3 *B. proteoclasticus* subcellular fractionation

B. proteoclasticus secreted, cell-associated, and cytosolic subcellular fractions were prepared as summarised in Figure 2.1.

2.3.1 Secreted protein extraction

The secreted protein of *B. proteoclasticus* cultures was harvested by centrifugation at 3,000 x g for 30 min at 4°C. The culture medium was collected and Complete[®] Protease Inhibitor was added immediately (one tablet per 50 mL culture volume) to minimise proteolysis. The medium was separated into 50 mL aliquots and frozen immediately at -80°C. All subsequent sample handling was performed at 4°C.

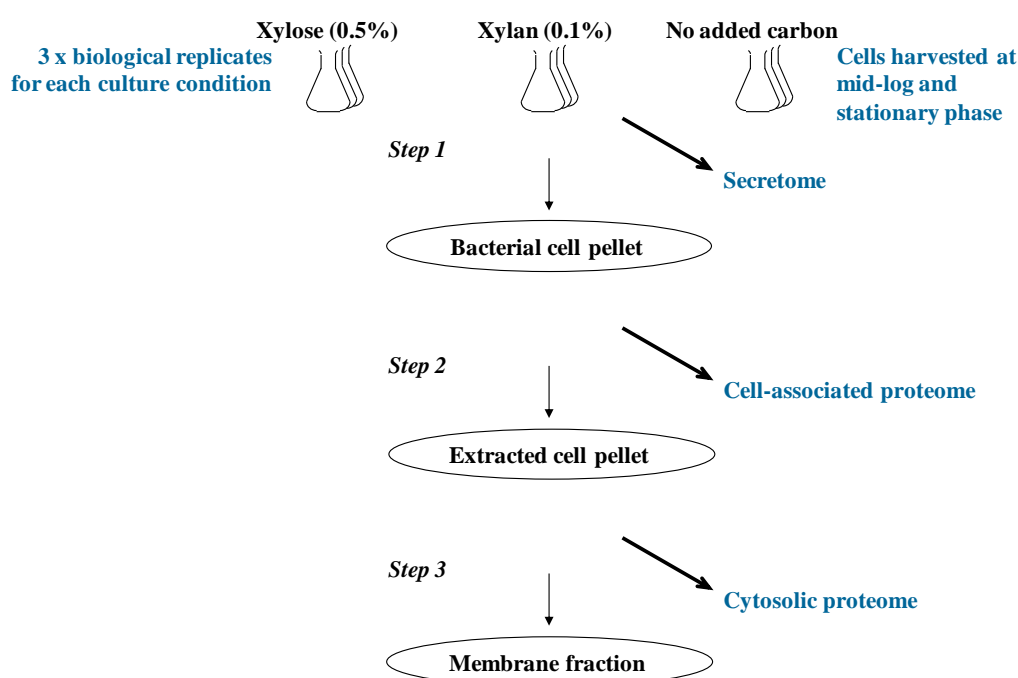


Figure 2.1. Summary of the three-step *B. proteoclasticus* cell fractionation procedure used to obtain secreted, cell-associated, and cytosolic subcellular protein fractions. For simplicity, one set of flasks depicts mid-log and stationary phase growth conditions.

2.3.2 Optimisation of *B. proteoclasticus* cell-associated protein extraction

The performance of five buffers for extraction of cell-associated proteins from the external surface of the *B. proteoclasticus* cell was examined. The bacterial cell pellet obtained after harvesting the culture supernatant (section 2.3.1) was resuspended in 5 mL of ice-cold 50 mM Tris-HCl buffer (pH 7.0), and divided into five 1 mL aliquots. Each aliquot was centrifuged at 5,000 x g for 10 min at 4°C, and the supernatant

removed. Each cell pellet was then resuspended in 2.6 mL of one of five extraction buffers (Table 2.1). Each cell suspension was incubated at 4°C for 30 min with end-over-end mixing. Intact cells were pelleted by centrifugation at 13,600 x g for 30 min at 4°C. The supernatant removed, separated into 400 µL aliquots and stored at -80°C. Cell-associated extracts were purified as described in Section 2.4.2 and were analysed by 1-D and 2-DE to identify the most efficient extraction method. Routine extractions of *B. proteoclasticus* cell-associated proteins following optimisation were performed by Mr Dong Li.

Table 2.1. Composition of the buffers tested for cell-associated protein extraction.

Buffer	Supplement ^a	Reference
Ice-cold 50 mM Tris-HCl buffer (pH 7.0)	2% CHAPS and Complete [®] Protease Inhibitor	Hansmier <i>et al.</i> , (2004) using CHAPS in place of SDS
200 mM Glycine (pH 2.2)	Complete [®] Protease Inhibitor	McCoy <i>et al.</i> , (1975)
5 M Lithium chloride	Complete [®] Protease Inhibitor	Lortal <i>et al.</i> , (1992)
8 M Urea	None	Wright <i>et al.</i> , (2005)
Ice-cold 50 mM Tris-HCl buffer (pH 7.0)	50 mM EDTA and Complete [®] Protease Inhibitor	Wright <i>et al.</i> , (2005)

^a Complete[®] Protease Inhibitor was used at a concentration of 1 tablet/10ml buffer volume.

2.3.3 *B. proteoclasticus* cytosolic protein extraction

The cell pellet remaining after extraction and harvest of the cell-associated proteins (section 2.3.2) was resuspended in 10 mL of ice-cold 50 mM Tris-HCl (pH 7.0) lysis buffer containing 2% (w/v) CHAPS and Complete[®] Protease Inhibitor. Silicone beads (100 mg) were added to the cell suspension, and the resuspended cells were lysed by bead-beating for 5 min, using a FastPrep FP120-230 bead beater (MP Biomedicals, Irvine, CA, USA). The disrupted cell suspension was centrifuged at 13,600 x g for 30 min at 4°C to pellet the silicone beads and insoluble material. The supernatant was removed, separated into 400 µL aliquots, and stored immediately at -80°C. Routine extraction of *B. proteoclasticus* cytosolic proteins was performed by Mr Dong Li.

2.3.3.1 *B. proteoclasticus* soluble protein extraction

The soluble fraction of the *B. proteoclasticus* proteome was harvested using a two phase extraction summarised in Figure 2.2. The bacterial cell pellet obtained after harvest of the culture supernatant (Section 2.3.1) was resuspended in 10 mL of ice-cold 50 mM Tris-HCl (pH 7.0) lysis buffer containing 2% (w/v) CHAPS, Complete[®]

Protease Inhibitor, and 100 mg of silicone beads, and the cells lysed by bead-beating for 5 min. The disrupted cell suspension was centrifuged at 13,600 x g for 30 min at 4°C to pellet the silicone beads and membrane fraction. The supernatant was removed and stored immediately at -80°C. The membrane fraction was also stored at -80°C for future analysis, which was performed by Dr. Judy Bond.

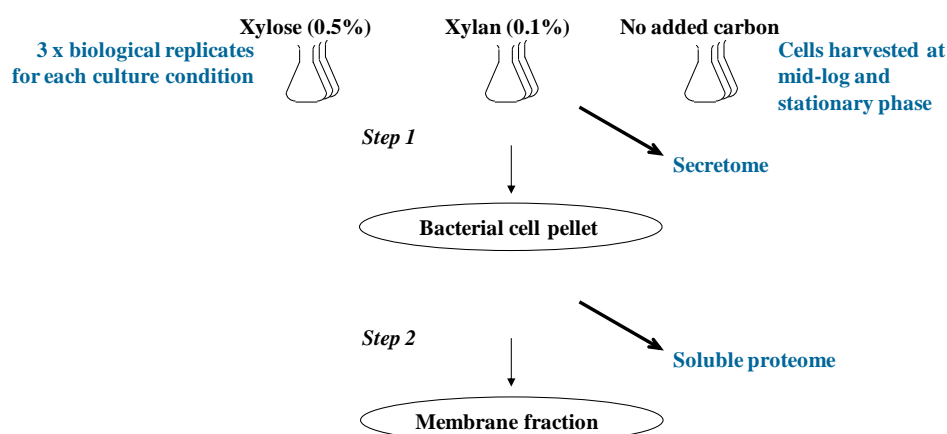


Figure 2.2. Summary of the two-step procedure for the harvest of soluble proteins. For simplicity, one set of flasks depicts mid-log and stationary phase growth conditions.

2.4 Sample purification

2.4.1 Optimisation of *B. proteoclasticus* secreted protein purification

Three methods suitable for the purification of low concentration protein from large solution volumes were trialled for purification of *B. proteoclasticus* secreted protein.

2.4.1.1 Protein precipitation using TCA

This method was essentially as described by Voigt *et al.* (2006). Aliquots of secreted proteins were thawed on ice and centrifuged at 6,000 x g for 60 min at 4°C to remove any cell debris. The cleared culture medium (45 mL) was transferred to a fresh 50 mL Falcon tube, and protein precipitation was affected by the step-wise addition of ice cold 100% TCA (w/v) to a TCA final concentration of 10% (w/v), followed by overnight incubation at 4°C. Precipitated protein was pelleted by centrifugation at 6,000 x g for 60 min, washed three times with ice-cold 90% ethanol, and allowed to air dry overnight at room temperature. For 2-DE analysis, each dry protein pellet was resuspended in 7 M urea, 2 M thiourea, 2% (w/v) CHAPS and 50 mM DTT, and incubated at 4°C overnight with vigorous shaking to ensure maximal protein re-suspension.

2.4.1.2 Protein precipitation using ammonium sulphate

Ammonium sulphate precipitation was as described by Jiang *et al* (2004), using 60% (w/v) ammonium sulphate final concentration. Aliquots of secreted proteins were thawed on ice and centrifuged at 6,000 x g for 60 min, and 45 mL of the cleared culture medium was transferred to a fresh 50 mL Falcon tube. Protein precipitation was affected by the slow addition of 16.6 g of ammonium sulphate, over a 60 min period, with vigorous shaking. The mixture was incubated overnight at 4°C. Precipitated protein was pelleted by centrifugation at 6,000 x g for 60 min, washed three times with ice-cold 90% ethanol, and air dried overnight at room temperature.

2.4.1.3 Centrifugal concentration

B. proteoclasticus secreted protein purification was performed using a Centricon® Centrifugal Concentrator unit with a 10,000 molecular weight cut-off ultra-filter membrane (Millipore, Billerica, MA). To equilibrate the concentrator prior to use, 2 mL of HPLC-grade water was loaded and centrifuged at 5,000 x g until the liquid volume had reduced to 100 to 200 µL. Cleared culture medium (2 mL) was then loaded and centrifuged until the solution volume had reduced to 100 to 200 µL. The concentrator was recharged with another 2 mL of cleared culture medium, and centrifuged again. This process was repeated until up to 20 mL of cleared culture medium had passed through the concentrator. Dialysis of the culture medium solution against a 2-DE compatible urea solution was achieved by loading the device with 2 mL of 7 M urea, and centrifuging at 5,000 x g until the solution volume was no greater than 100 µL. The device was then inverted and centrifuged at 5,000 x g for 10 min to collect the purified protein solution. The protein preparation was stored immediately at -80°C.

2.4.2 *B. proteoclasticus* cell-associated protein purification

Cell-associated protein was precipitated using the TCA/acetone procedure, as described by Schwarz *et al* (2007). The cell-associated protein extract was mixed with 4 x volumes of -20°C acetone containing 10% TCA (w/v), vortexed vigorously for 5 min, and incubated overnight at -20°C. The precipitated protein was pelleted by centrifugation at 13,600 x g for 60 min at 4°C. The protein pellet was washed 3 x 60 min in ice-cold 90% acetone at -20°C with regular vortexing. Care was taken to ensure the pellet was removed from the wall of the tube, and fully disrupted during this

process. The washed pellet was allowed to dry in a stream of clean air at room temperature for 10 min.

2.4.3 Optimisation of *B. proteoclasticus* cytosolic protein purification

The performance of two methods for purification of the *B. proteoclasticus* cytosolic and soluble proteins was examined. These were the TCA/acetone procedure described in Section 2.4.2, and the phenol/methanol/ammonium acetate procedure as described by Carpentier *et al.* (2005). The cytosolic or soluble protein extract was mixed with an equal volume of ice-cold Tris-HCl buffered phenol (pH 8.0) and vortexed vigorously for 30 min at 4°C. The mixture was centrifuged at 8,000 x g for 5 min to separate the phases, and the protein containing phenolic phase was removed and placed in a fresh 1.5 mL microcentrifuge tube. An equal volume of ice-cold 50 mM Tris-HCl Buffer (pH 7.0) was added to the phenol and the mixture vigorously vortexed for 30 min at 4°C. The mixture was centrifuged at 8,000 x g for 5 min and the phenol phase removed and placed in a fresh 1.5 mL microcentrifuge tube. Proteins were precipitated by the addition of 4 x volume of -20°C methanol containing 100 mM ammonium acetate. The mixture was vigorously vortexed for 60 sec and incubated overnight at -20°C. The precipitated protein was pelleted by centrifugation at 13,600 x g for 60 min at 4°C, and the protein pellet was air-dried in a fume cupboard overnight. The protein pellet was resuspended in 7 M urea, 2 M thiourea, 2% (w/v) CHAPS and 50 mM DTT and incubated overnight at 4°C with shaking.

2.5 2-DE of *B. proteoclasticus* cytosolic subcellular fractions

An overview of each *B. proteoclasticus* cytosolic subcellular fraction (Section 2.3) was obtained by 2-DE using pI 3-10 NL 7 cm IPG strips. Analytical technical replicate¹ gels were run in triplicate for each biological replicate² sample using narrow range pI 3-5.6 NL and 5.3-6.5 7 cm IPG strips. The presence of low abundance basic *B. proteoclasticus* proteins was examined by 2-DE using pI 6-11 7 cm IPG strips. All IEF was performed using an IPGphor II IEF unit with a 12-lane IEF manifold (GE Healthcare, Uppsala, Sweden).

¹ A technical replicate gel is the 2-DE separation of the same protein sample isolated from a single bacterial culture.

² A biological replicate gel is the 2-DE separation of independently isolated samples from replicate bacterial cultures.

2.5.1 Rehydration and IEF of pI 3-5.6 NL, 5.3-6.5, and 3-10 NL IPG strips

When pI 3-5.6 NL, 5.3-6.5, and 3-10 NL 7 cm IPG strips were used, the total protein concentration of each sample was assayed immediately prior to IPG strip rehydration using the 2D Quant kit (GE Healthcare, Uppsala, Sweden), as per the manufacturer's instructions. Fresh IEF Rehydration Buffer containing 7 M urea, 2 M thiourea, 2% CHAPS (w/v), 20 mM DTT, and a trace amount of bromophenol blue was added to each sample to achieve a final concentration of 0.8 $\mu\text{g. } \mu\text{L}^{-1}$. IPG Buffer (the pI range of which was matched to the pI range of the IPG strips) was added to each sample to a final concentration of 0.5% (v/v). Each sample was vortexed vigorously for 60 sec, and centrifuged at 13,000 x g for 5 min to pellet any remaining insoluble material, and 125 μL applied to a dehydrated IPG strip using an IPG strip-reswelling tray (GE Healthcare, Uppsala, Sweden). Care was taken to avoid trapping air bubbles against the gel surface of the strip. Each strip was then covered with 3 mL of mineral oil to prevent rehydration buffer crystallisation, and passively rehydrated at room temperature for 16 h. Following rehydration, each IPG strip was checked for uniform reswelling and if acceptable was subjected to IEF. The IPGphor II unit was set up according to the manufacturer's instructions for passively rehydrated IPG strips, using anode and cathode paper wicks blotted with 150 μL of HPLC-grade water. IEF was conducted for 9-11 kVh as detailed in Table 2.2.

Table 2.2. Running conditions for Immobiline DryStrip IEF on IPGphor II unit.^a

IPG strip range (pI)	Phase	Voltage ^b	kVh ^c
3 – 10 NL	1- Step	200	200
	2- Gradient	1000	300
	3- Gradient	5000	4500
	4 Step	5000	4000
5.3 – 6.5	1- Step	200	200
	2- Gradient	1000	300
	3- Gradient	5000	4500
	4 Step	5000	6000
6-11	1- Step	200	400
	2- Gradient	1000	600
	3- Gradient	5000	4500
	4 Step	5000	6000

^a All IPG strips were run at 20 °C, and a maximum current 50 mA per strip.

^b This value is the maximum potential voltage attained during each phase. The actual voltage at end of phase may have been lower.

^c Phase concluded when kVh target achieved.

2.5.2 Rehydration and IEF of pI 6-11 IPG strips

When pI 6-11 IPG strips were used each was rehydrated with 125 μL of 7 M urea, 2 M thiourea, 2% (w/v) CHAPS, 10% (v/v) isopropanol, 5% (v/v) glycerol, 3% (w/v) DTT, 0.5% (v/v) IPG buffer, and a trace of bromophenol blue. Strips were covered with 3 mL of mineral oil and allowed to rehydrate at room temperature for 16 h. Immediately prior to IEF, the concentration of each protein sample was assayed using a 2D Quant kit, and 7 M urea, 2 M thiourea, 2% (w/v) CHAPS, 10% (v/v) isopropanol, 5% (v/v) glycerol, 3% (w/v) DTT, and a trace of bromophenol blue was added to achieve a final protein concentration of 2.4 $\mu\text{g. } \mu\text{L}^{-1}$. IPG Buffer (pI 6-11) was then added to the protein sample to a final concentration of 0.5% (v/v). Each sample was vortexed vigorously for 60 sec, and centrifuged at 13,200 x g for 5 min. The IPGphor II unit was set up according to the manufacturer's instructions for cup-loaded strips, using anode and cathode paper wicks blotted with 150 μL HPLC-grade water, and fresh pI 6-11 IEF Rehydration Buffer containing 3% (w/v) DTT respectively. Each sample (125 μL) was cup-loaded onto each IPG strip, and IEF conducted for 11.5 kVh as summarised in Table 2.2.

2.5.3 IPG strip equilibration and second-dimension electrophoresis

Immediately following IEF, each IPG strip was placed in a 15 mL Falcon tube, and focused proteins were reduced for 15 min using 2 mL of Invitrogen 1x LDS Buffer containing 1x Reducing Agent, then alkylated for 15 min using 2 mL of 1x LDS Sample Buffer containing 100 mM iodoacetamide. Each strip was rinsed briefly in 1x NuPAGE MOPS electrophoresis running buffer, then positioned on top of a 1 mm thick, NuPAGE[®] Novex 4-12% Bis-Tris ZOOM[®] gel. Standard IPG strip orientation was with the anode end to the left. Second dimension electrophoresis was performed using an Invitrogen NuPAGE[®] MOPS Buffer system at constant 200 V until the dye front reached 1 mm from the base of the gel.

2.6 Protein detection

All 2-DE gels used for total protein abundance analysis were stained with colloidal CBB G-250 (Anderson *et al.*, 1991). Several analytical 1-DE and 2-DE gels were stained using the Flamingo[™] Fluorescent Gel Staining Reagent, which permits total protein visualisation below the detection limit of colloidal CBB G-250 and other fluorescent based stain reagents (Berkelman, 2006).

2.6.1 Colloidal CBB G-250 total protein staining

Following 2-DE, technical replicate gel sets were placed in a clean, air-tight container and fixed overnight in 50% (v/v) EtOH: 2% (v/v) orthophosphoric acid with gentle shaking. The fixative was then discarded and the gels washed 3 x 30 min with ultra-filtered water. Proteins were visualised by staining with 0.1% (w/v) CBB G-250 in 34% (v/v) methanol, 2% (v/v) orthophosphoric acid containing 17% (w/v) ammonium sulphate for five days with gentle shaking. Immediately prior to protein visualisation, colloidal CBB G-250 stained gels were washed 3 x 60 min in ultra-filtered water.

2.6.2 Flamingo™ fluorescent total protein staining

Following electrophoresis, gels to be stained using the Flamingo™ Fluorescent Gel Stain were placed in a clean, airtight container and fixed overnight in 40% (v/v) ethanol: 10% (v/v) acetic acid with gentle shaking. The fixative was then discarded and the gels washed 3 x 10 min in ultra-filtered water. This washing step was not included in the protocol supplied by the manufacturer, but significantly enhanced the staining uniformity. A working stain solution was prepared by diluting one volume of 10x Flamingo™ Fluorescent Gel Stain Stock Solution with nine volumes of ultra-filtered water, and 50 mL of this solution was used per gel. Gels were stained at room temperature overnight in the dark, with gentle shaking. Immediately prior to protein visualisation, each gel was washed for 10 min in the dark with 100 mL of 0.1% (v/v) Tween 20, then 2 x 10 min in ultra-filtered water. This was an optional step suggested by the manufacturer, and enhanced background stain removal.

2.7 Image acquisition

The apparatus used for image acquisition was dependent on the staining method used to visualise 1-DE and 2-DE separated protein.

2.7.1 Colloidal CBB G-250 stained gel image acquisition

Colloidal CBB G-250 stained gels were scanned using a Molecular Dynamics Personal Densitometer SI (Sunnyvale, CA, USA) at 12 bits per pixel digital resolution and 50 µm pixel size. Technical replicate gels were always imaged in a single scan. Immediately after scanning, gels were placed in re-sealable plastic bags containing 5 mL of HPLC-grade water and stored at 4 °C.

2.7.2 Flamingo™ stained gel image acquisition

A Fujifilm FLA-5100 Fluorescent Imaging system was used to scan Flamingo™ Fluorescent total protein stained gels at an excitation wavelength of 532 nm (SHG green laser). The imaging system was controlled by Image Reader FLA 5,000 Series (Version 1.0) software. Laser intensity of 600 V was optimal for all applications.

2.8 Image analysis

2.8.1 Protein spot detection

Image Master™ 2D Platinum software (Version 5.0) (GE Healthcare, Uppsala, Sweden) was used for image analysis of all 2-DE gels in this project. Automatic spot detection parameters for colloidal CBB G-250 stained gels were set at “smooth”¹ equal to two, “saliency”² equal to 0.5, and “minimum pixel area”³ equal to 100. Automatic spot detection parameters for Flamingo™ total protein stained gels were set at smooth equal to three, saliency equal to 0.8, and minimum pixel area equal to 100. These parameters were decided upon after extensive analysis of the effect of small changes in each of the three parameters on spot border detection accuracy and spot resolving power.

2.8.2 Protein spot editing

Automatic spot detection was performed, followed by manual deletion of dust spots and the molecular weight marker lane. Gel regions that contained horizontally smeared proteins were retained to maintain the total protein detected on each gel. When necessary, manual splitting of highly abundant overlapping spots was performed. Manual editing was performed with caution to minimise loss of data.

¹ The “smooth” parameter fixes the number of times ImageMaster™ will smooth the image before detecting spots, using a smooth-by-diffusion algorithm. The “smooth” parameter was optimised to detect all real spots and split as many as possible overlapping spots, without being concerned about artifact spots.

² The “saliency” parameter is a measure of spot curvature, and indicates how far a spot stands out with respect to its local environment. In general, real spots have high saliency values and background and artefact spots have low saliency values. Saliency is highly dependent on the gel image, and must be optimised accordingly.

³ The “minimum pixel area” is the number of pixels used to depict each spot. Artifact spots, for example dust particles, that have an area smaller than a specified threshold (expressed as the number of pixels) are not detected.

2.9 2-DE gel image matching strategy

A hierarchical spot matching process was used to generate matched protein spot groups between substrate or growth phase-dependent 2-DE biological replicate gel sets (Figure 2.3).

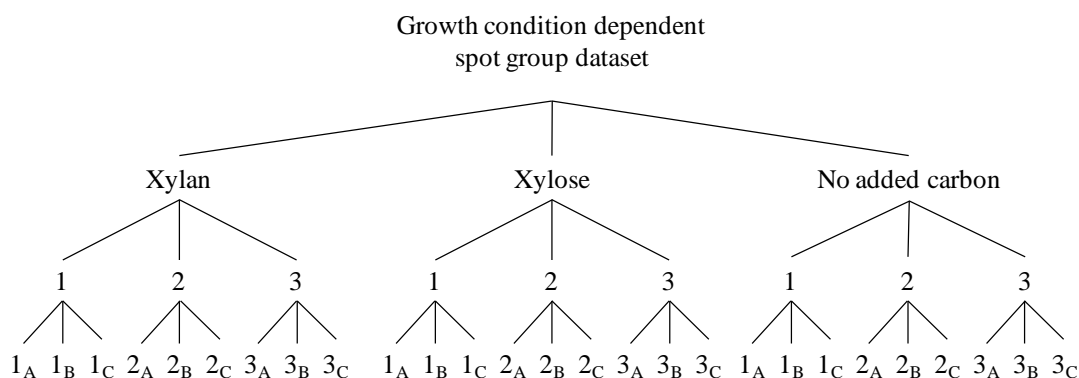


Figure 2.3. Summary of the hierarchical procedure used to match technical and biological replicate gels. Numbers (1-3) represent biological replicate gel sets, and numbers with subscript letters (A-C) represent technical replicate gels. Note that the technical replicates are nested beneath the biological replicates.

2.9.1 Technical replicate gel image matching

Protein spots were first matched between technical replicate gels ($n = 3$) (Figure 2.4). Three to five carefully selected spots were used as landmark groups to initiate gel matching. Four criteria were used to select landmark spots- a) spots were present in all technical and biological replicate gels for a given pI range; b) spots were widely separated on the gel image; c) spots clearly corresponded between all gel images; d) spots were small, sharp, and well defined. The most clearly resolved 2-DE image of each technical replicate gel set was chosen as the set reference gel and the landmark spots matched between gels. All spots present in the reference gel were then automatically matched to the two remaining technical replicate gels. In cases where localised gel deformation caused small regions of spot mismatching, incorrectly matched groups were deleted. A landmark spot group was created in the deformed region and a second automatic matching cycle performed. All technical replicate spot groups were examined manually, and where necessary additional manual spot group matching performed.

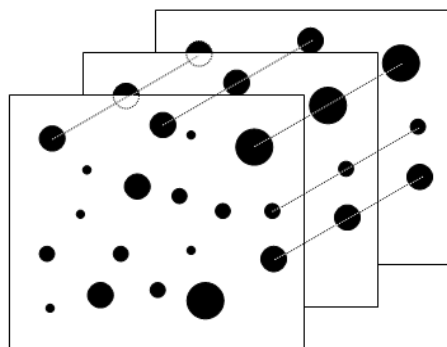


Figure 2.4. Definition of matched protein spot groups. Broken lines indicate grouping of equivalent protein spots between two or more replicate 2-DE gels.

2.9.2 Biological replicate gel image matching

Using the reference gel for each technical replicate gel set, spot groups were matched between biological replicate gel sets ($n = 3$). As above, a reference gel was selected, and all spots present in this gel were automatically matched to the two remaining gels. Finally, using the reference gel selected for each biological replicate gel set, biological replicate spot groups were automatically matched between growth conditions, thus generating a growth condition dependent spot group dataset for each detected protein species (Figure 2.3). Each growth condition was matched to all others to ensure proteins unique to a particular growth condition were characterised.

2.10 Differential protein abundance analysis

Statistical analysis of matched biological replicate spot groups was performed using SPSS 14.0 (SPSS, Chicago, IL, USA). Parametric and non-parametric tests were used to characterise statistically significant substrate and growth phase dependent changes in *B. proteoclasticus* protein abundance profiles.

2.10.1 Data extraction, missing values and data imputation

Raw protein spot volumes were exported from Image Master Platinum V5.0 to Microsoft Excel. A biological replicate spot group was deemed reliable for statistical analysis if no more than 2 spots within a potential spot group were undetected across all biological replicate gels, and no greater than one missing value was derived from the same technical replicate gel set. If these conditions were met, it was assumed the undetected spots were most likely due to them being below the limit of detection. Consequently, the missing values for each matched protein spot group were imputed by inserting the lowest raw spot volume (threshold) value detected on the gel that generated the missing value, which is similar to the method employed by Meleth *et al.*

(2005). Matched protein spot groups containing more than two missing values were removed from the automated statistical analysis.

2.10.2 Raw spot volume normalisation

To correct for technical variability between 2-DE replicate gels, each raw spot volume was normalised to the sum of all detectable spot volumes derived from the respective gel, according to Equation 2.1.

$$\text{Normalised spot volume (V}_i\text{)} = \text{raw spot volume} / \text{total detectable spot volume}$$

Equation 2.1 Normalisation of raw spot volume values against total detectable protein.

2.10.3 Data transformation

Data normalisation using a constant sum constraint creates a compositional dataset, which consequently presents problems for parametric testing due to the introduction of artificial correlations between previously unrelated variables. To account for this phenomenon, Aitchison (1982) proposed a log-ratio transformation for compositional data after demonstrating transformation of this kind produced a dataset exhibiting a multivariate normal distribution. Log₁₀ transformation of the normalised raw spot volumes was performed as described in Equation 2.2. To ensure the normalised spot volume data was suitable for parametric statistical testing, other transformation methods were also tested. These were power^{1/3} and power^{1/5} transformations according to Equation 2.3 and Equation 2.4.

$$\log_{10} V_i = \log_{10} [V_i / (100 - V_i)]$$

where V_i = individual spot volume.

Equation 2.2. Log₁₀ transformation of normalised spot volumes, to account for the creation of a compositional dataset.

$$\log^{1/3} V_i = [V_i / (100 - V_i)]^{1/3}$$

where V_i = individual spot volume.

Equation 2.3. Power^{1/3} transformation of normalised spot volumes, to account for the creation of a compositional dataset.

$$\log^{1/5} V_i = [V_i/(100 - V_i)]^{1/5}$$

where V_i = individual spot volume.

Equation 2.4. Power^{1/5} transformation of normalised spot volumes, to account for the creation of a compositional dataset.

2.10.4 Statistical analysis assumption testing

The parametric tests used to investigate significant substrate and growth phase dependent protein abundance changes carry two assumptions; that the sample groups being compared are derived from populations that exhibit normal distributions, and display equality of variance (homoscedasity). The normal distribution of the \log_{10} transformed normalised spot volumes for each spot group within a growth condition dependent spot group dataset was assessed using the Shapiro-Wilk Goodness of Fit test (Shapiro and Wilk, 1965a), which is considered suitable for small sample sizes. Homoscedasity of matched protein spot groups was assessed by constructing scatter plots of standard deviations as a function of the means of the untransformed, and \log_{10} transformed normalised spot group volumes. A Pearson Correlation co-efficient was calculated for each plot and used to report the standard deviation versus means dependency.

2.10.5 Parametric statistical testing

The decision making process for appropriate statistical testing is summarised in Figure 2.5. If each biological replicate spot group within a growth condition dependent spot group dataset generated a Shapiro-Wilk test p value > 0.05, the dataset for that spot was deemed suitable for parametric testing. When biological replicate spot groups were detected in three or more growth conditions, a two-factor nested ANOVA was used to detect statistically significant differences between spot group means. A nested ANOVA was used to account for the relationship between technical and biological replicates for each growth condition. A probability $p < 0.01$ was chosen as the cut-off for statistical significance. For each nested ANOVA, a Levene's Homogeneity-of-Variance test (Levene, 1960) was also conducted to ensure samples were derived from homoscedasic populations. Spot group datasets with a ANOVA $p < 0.01$, and a Levene-test p value > 0.05, were submitted to Bonferroni-adjusted Multiple Comparison Post-hoc testing (Hochberg, 1988) to characterise the growth condition dependent protein abundance changes. Bonferroni Correction is a stringent method of

addressing the multiple testing problem, that is, the unacceptably high accumulation of type I errors (rejecting H_0 when H_0 is true) when performing multiple statistical testing, thus controlling the family-wise error rate. The correction adjusts the significance level for each test according to the total number of statistical tests being performed. Spot groups datasets failing the homogeneity-of-variance assumption (Levene-test p value < 0.05), were submitted to Tamhane's T2 Post-hoc testing (Tamhane, 1979). Tamhane's T2 conducts conservative t -test-based pairwise comparisons, but does not assume equality of variance between sample groups. In the event that biological replicate spot groups were detected in only two growth conditions, an independent samples Students t -test was conducted. If the Levene-test p value > 0.05 , the equality of variance test statistic was used. If the p value < 0.05 , the unequal variance test statistic was used.

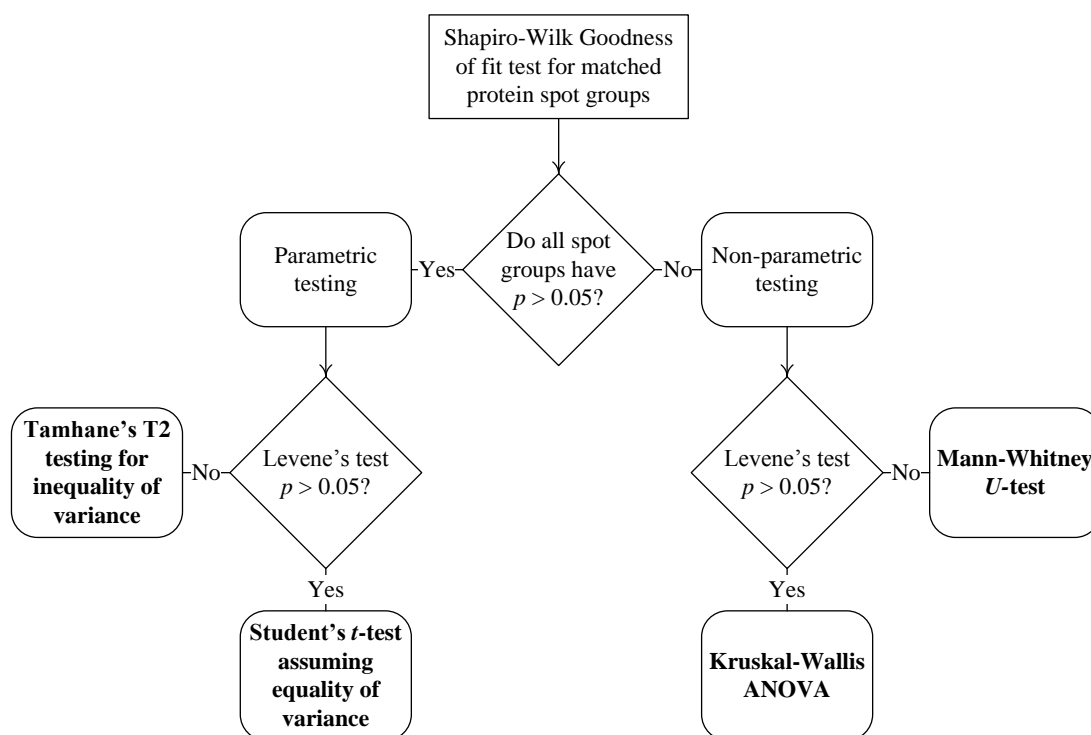


Figure 2.5. Flow chart summarising decision making process for statistical testing of protein abundance profiles.

2.10.6 Non-parametric statistical testing

When at least one member of a growth condition dependent spot group exhibited a Shapiro–Wilk test p value < 0.05 , the growth condition dependent dataset was deemed unsuitable for parametric statistical testing. As such, the non-parametric Kruskal-Wallis one-way ANOVA and Wilcoxon-Mann-Whitney U tests were used to assess

the statistical significance of changes detected between *B. proteoclasticus* protein abundance profiles, depending on whether three or more, or two biological replicate spot groups were detected across all growth conditions, respectively.

2.10.7 Protein abundance fold change calculation

The fold change ratio for significant protein abundance changes was reported relative to the growth substrate eliciting the greater abundance level.

Fold change ratio = (larger mean normalised spot volume – 1x standard deviation)/(lesser mean normalised spot volume + 1x standard deviation)

Equation 2.5 Calculation of protein abundance fold change ratio for statistically significant protein abundance changes.

2.11 Protein digestion and peptide purification

All solutions used for in-gel protein digestion were HPLC-grade. CBB G-250 stained protein spots were excised using a One Touch Plus Manual Spot Picking pipette (The Gel Company, San Francisco, CA, USA), with a tip internal diameter of 1.5 mm. The majority of excised gel plugs were processed for mass spectrometry using an Ettan™ Digester Robotic Digestion Platform (GE Healthcare, Uppsala, Sweden). A number of gel plugs were also processed manually; in which case an in-gel protein digestion protocol based on that of Shevchenko *et al* (1996) was used with modifications. Excised gel plugs were placed in either a 96 well microtitre plate (for robotic digestion), or a 0.65 mL microcentrifuge tube (for manual digestion) and the protein digested and peptides extracted as indicated in Table 2.3.

Table 2.3. Summary of the robotic and manual in-gel protein digestion protocols.^a

Phase	Step	Robotic Digestion	Manual digestion
Pre-digest	Destain	45 min x 3 50 mM NH ₄ HCO ₃ : 50% MeOH (v/v)	60 min x 3 50 mM NH ₄ HCO ₃ : 50% ACN (v/v)
	Dehydration	Air-dry	100% ACN
Digestion	Trypsin	10 µL trypsin (6 nm/µL) in 20 mM NH ₄ HCO ₃	2 µL trypsin (25 nm/µL) in 50 mM NH ₄ HCO ₃
	Digest buffer	None	50 mM NH ₄ HCO ₃ (10 µL)
	Time	5 h	Overnight
Peptide extraction	Extraction 1	40 µL 0.1% TFA : 50% ACN (v/v)	40 µL 0.1% TFA : 50% ACN (v/v)
	Extraction 2	40 µL 0.1% (v/v) TFA : 50% ACN (v/v)	40 µL 0.1% TFA (v/v)
	Extraction 3	40 µL 0.1% TFA : 50% ACN (v/v)	40 µL 100% ACN
	Drying	Air-dry, room temperature	Vacuum centrifuge, 37°C

^a Modified Sequencing-Grade Trypsin (Cat #11 521 187 001), Roche Applied Science (Mannheim, Germany).

2.12 MALDI-TOF mass spectrometry

2.12.1 Sample preparation

CHCA was used as the matrix for analysis of extracted peptides. Saturated CHCA matrix solution was prepared fresh by the addition of 10 mg of CHCA to 1 mL of 0.1% (v/v) TFA: 50% (v/v) ACN. The matrix solution was spiked with 0.75 µL of PMF Calibration Mixture 2 to provide peptide masses for internal calibration. The mixture was vortexed vigorously for 60 sec and centrifuged at 13,600 x g for 5 min to pellet un-dissolved CHCA crystals. The saturated matrix solution (200 µL) was carefully removed and placed in a fresh 0.65 mL microcentrifuge tube. Robotically digested samples were resuspended in 1.25 µL of saturated CHCA matrix solution, and the peptide/matrix mixture spotted onto a stainless steel, 96 well MALDI-TOF target plate (Foster City, CA, USA). The mixture was allowed to crystallise in a stream of clean air at room temperature. PerfectPure[®] C18 purified peptide mixtures were mixed 1:1 with saturated CHCA matrix solution and carefully vortexed to ensure maximal sample/matrix mixing. The peptide/matrix mixture (2 µL) was spotted onto the MALDI-TOF target plate and allowed to crystallise in a stream of clean air at room temperature.

2.12.2 Peptide mass fingerprinting

Positive ion tryptic PMF were acquired using a Voyager DE Pro MALDI-TOF mass spectrometer (Applied Biosystems, Foster City, CA, USA) equipped with a 337 nm nitrogen laser operating at 3 Hz. The following instrument settings were optimised for the tryptic peptide mass range: 20 kV acceleration voltage, 75% grid voltage, 0.02% guide wire voltage, 180 ns delay time, reflectron detector, 650 m/z low mass gate. PMF were obtained by averaging 200 individual mass spectra measuring across a 700 to 3500 m/z mass range. A minimum of three PMF were acquired from every tryptic peptide mixture analysed.

2.12.3 PMF processing and database interrogation

All PMF were processed manually using Data Explorer software, Version 4.0 (Applied Biosystems, Forster City, CA, USA). The two highest quality spectra from the minimum three obtained from each tryptic peptide mixture were selected for protein database interrogation. Internal calibration against the three Calibration Mixture 2 peaks, observed at m/z 1296.6853, 2093.0867, and 2465.1989 was performed. When internal calibration was not possible due to an inability to detect all three Calibration Mixture 2 peaks, external calibration against a closely positioned Calibration Mixture 2 sample, followed by internal calibration against one or two observable Calibration Mixture 2 peaks was performed. The protonated monoisotopic peak mass list was then derived from each PMF, using the Data Explorer peak deisotoping function. Finally, monoisotopic masses matching at two decimal places to Bovine Trypsin Autolytic Peptide or CHCA matrix masses (Harris *et al.*, 2002) were discarded. Each peak mass list was analysed in-house against the *B. proteoclasticus* “B316_GOLD_June2008” protein sequence database using the MASCOT search engine (<http://www.matrixscience.com/>) (Perkins *et al.*, 1999), which incorporates the MOWSE scoring algorithm (Pappin *et al.*, 1993). The “B316_GOLD_June2008” database was constructed from the closed and functionally annotated *B. proteoclasticus* genome sequence, and contained the translated amino acid sequences of the 3813 *B. proteoclasticus* CDSs, plus the 3813 randomised protein sequences of the same size and amino acid composition.

Search parameters used were: maximum one missed tryptic cleavage site, monoisotopic peptides only, maximum +/- 50 ppm mass tolerance, and carbamidomethylation of cysteine residues as a fixed modification. Database searches were initially performed with no variable modifications permitted. If a positive protein

match was obtained, the search was subsequently performed allowing for oxidised methionine residues as a variable modification. Each PMF was simultaneously searched against a randomised sequence decoy database of the same length as that for *B. proteoclasticus*. While this method is not a true indicator of the false positive match rate, it is informative especially for protein identifications with a MOWSE score close to the significance threshold (Elias *et al.*, 2005). When a PMF obtained from a spot excised from a secreted protein 2-DE separation did not return a positive identification, the PMF was then searched against the eukaryotic division of the NCBI non-redundant database. This was done in an attempt to clarify whether spots that were visible in 2-DE separations of secreted proteins did not originate from proteins that may have been present in the culture medium prior to use.

2.12.4 Positive protein match criteria

A positive protein match was considered reliable only when both PMF derived from a single sample generated a MOWSE score above the minimum threshold, which was set at 62. This value was chosen as Mascot predicts the following false positive identification rates: 5% at a score of 52, 0.5% at 62, 0.05% at 72, and 0.0005% at 92. The Mascot prediction is based on ideal mass spectra and does not consider trypsin-derived or other frequently occurring mass peaks. Proteins with maximum MOWSE scores between 52 and 62 were manually examined, considering the following additional criteria: (i) the expectation score, (ii) the orders of magnitude between the expectation of the top ranked match, and subsequent possible matches (iii) the number of matching peptides and the resultant full length protein sequence coverage (iv) the proportion of identified peptides with missed cleavage sites (v) the root mean square error of the identified peptides (vi) the correlation between the theoretical and observed *pI* and molecular weight of the protein identification (vii) identifications obtained from the decoy database and (viii) the number of modified peptides contributing to the protein identification. All instances of multiple proteins identified in a single gel plug were also manually examined, and the criteria described above was also used to assess the significance of the one protein identification independent of the other. In particular, the correlation between the theoretical *pI* and molecular weight of both proteins was examined, and identifications were accepted only when these protein parameters were in agreement. Furthermore, the set of unique peptides contributing to the protein identification was determined. If two or more peptides were not unique the protein identification with the lower MOWSE score was discarded.

2.13 Reverse phase HPLC and MS/MS (1-D LC-MS/MS)

1-DE and 2-DE separated proteins were analysed by 1-D reverse phase HPLC-ESI MS/MS. All lyophilised tryptic peptide mixtures, irrespective of digestion method, were prepared for 1-D HPLC-MS/MS analysis by Perfect Pure C18 tip purification as described in Section 2.9.1. Peptides were eluted from the C18 tip into 5 μL 70% (v/v) ACN : 0.1% (v/v) formic acid, to which was added 65 μL 0.1% (v/v) formic acid. Each sample was vortexed for 30 sec and centrifuged at 13,600 $\times g$ for 10 min to pellet any remaining particulate material. An aliquot (65 μL) was carefully removed and placed in a 200 μL HPLC sample vial and care was taken to avoid trapping air bubbles at the base of the vial.

1-D reverse phase HPLC was conducted using a Dionex UltiMate[®] 3,000 Nano HPLC system (LC Packings, The Netherlands) fitted with an Acclaim[®] PepMap[™] silica based C18 stationary phase analytical nano-column (75 μm i.d. \times 15 cm, 100 Å pore size). 65 μL of each sample was automatically injected using a Well Plate Autosampler equipped with a dual-needle 100 μL syringe set for full loop injection. Peptides were loaded onto the C18 nano-column at a flow rate of 200 $\text{nL}\cdot\text{min}^{-1}$. Separation of 1-DE derived peptide mixtures was achieved using a mobile phase gradient constructed from 0.1% formic acid (Buffer A) and 0.1% formic acid in 80% ACN (Buffer B): linear gradient 0 to 15% Buffer B over 25 min; 15 to 35% Buffer B 25 to 67 min; 35 to 70% Buffer B 67 to 111 min; 70 to 100% Buffer B 111 to 126 min. Separation of 2-DE derived peptide mixtures was achieved using a mobile phase gradient constructed from 0.1% formic acid (Buffer A) and 0.1% formic acid in 80% ACN (Buffer B): 0 to 15% linear gradient Buffer B over 25 min; 15 to 35% linear gradient Buffer B 25 to 67 min; 35 to 100% linear gradient Buffer B 67 to 77 min. On-line peptide identification was achieved using a Finnigan LTQ[™] Linear Ion-trap equipped with a nano-spray ion source (Thermo Scientific, Waltham, MA, USA). Automated Data Dependent[™] MS/MS was performed where the five most abundant precursor ions detected within a single full MS scan from m/z 400 to m/z 2,000 were selected for Pulsed Q-dissociation analysis, using 35% collision energy. Dynamic exclusion was utilised to prevent repeated MS/MS analysis of high abundance peptides.

2.13.1 1-DE LC-MS/MS

1-DE analysis was performed using a NuPAGE MOPS Buffer system (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instructions. Each sample was

mixed 2.6:1 with 4 x LDS Sample Buffer. 10 x Sample Reducing Agent (Invitrogen, Carlsbad, CA, USA) was then added to achieve a 1x final concentration. Each sample was vortexed vigorously for 60 sec then heated for 10 min at 70°C. All samples were centrifuged at 13,600 x g for 2 min to pellet any insoluble material, and 40 µL of each was loaded into a well of a 1.5 mm NuPAGE® Novex 4-12% Bis-Tris 10-well ZOOM® Gel. Electrophoresis was performed at constant 200 V until the dye front reached 1 mm from the base of the gel. Each gel was then removed from the plastic casing and processed immediately for colloidal CBB G-250 staining as described in Section 2.6. After gels were stained and imaged, the 1-DE lanes of interest was dissected into 40-45 slices of 1.5 to 2 mm thickness. Each gel slice was placed in a well of a 96 well microtitre plate, and protein digestion and extraction performed using an Ettan™ Digester Robotic Digestion platform as described in Section 2.11. Each tryptic peptide digest was then prepared for LC-MS/MS analysis as described in Section 2.14.

2.13.2 MS/MS data analysis

Raw MS/MS data files were processed using BioWorks 3.3.1. MS/MS ion searches were performed against the *B. proteoclasticus* sequence database using MASCOT. Search parameters were maximum 1 missed tryptic cleavage site, carbamidomethylation of cysteine residues as a fixed modification, methionine oxidation as a variable modification, no specific quantitation, maximum 2 Da peptide mass tolerance, maximum 0.5 Da MS/MS mass tolerance, 2+ and 3+ monoisotopic peptide charge. No precursor ion m/z was defined, and a decoy database was used to obtain an estimation of the false discovery rate.

2.14 Polysaccharide adsorption assay

2.14.1 Substrate preparation

Washed microcrystalline cellulose was prepared by incubating 100 mg of Avicel® PH-101 with 5 mL of ice-cold 50 mM phosphate buffer (pH 7.0) for 15 min with vortexing. The suspension was centrifuged at 5,000 x g for 5 min, and the wash process repeated for a total of three washes. Sephadex G-250® was prepared by incubating 100 mg of Sephadex G-250® with 5 mL of room temperature 50 mM phosphate buffer (pH 7.0) for 24 h with gentle mixing. The suspension was centrifuged at 1,000 x g for 20 min, and washed twice using the same buffer for 20 min.

2.14.2 *B. proteoclasticus* secreted proteome polysaccharide adsorption assay

Secreted protein harvested at stationary phase from xylan *B. proteoclasticus* was assayed for plant structural polysaccharide binding potential. Stationary phase growth medium (50 mL) was thawed on ice and supplemented with 50 mL of 250 mM phosphate buffer (pH 7.0). The buffered stationary phase growth medium was then mixed directly with 100 mg of washed microcrystalline cellulose. A second aliquot of stationary phase growth medium (50 mL) was incubated with 100 mg of washed Sephadex G-250[®] which acted as a negative control. Protein/substrate adsorption was achieved by incubating the mixture for 120 min at 4°C with gentle end over end mixing. Following incubation, the cellulose was recovered by centrifugation at 5,000 x g for 20 min at 4°C, and washed 5 x 30 min with 2 mL of ice-cold 50 mM sodium phosphate buffer (pH 7.0). The Sephadex G-250[®] was recovered by centrifugation at 1,000 x g for 20 min at 4°C, and washed 5 x 30 min with 2 mL of ice-cold 50 mM sodium phosphate buffer (pH 7.0). Each of the wash fractions was recovered and stored at -20°C. Substrate bound proteins were eluted in 2 mL of 1 x LDS (Invitrogen, Carlsbad, CA) sample buffer containing 20 mM DTT for 60 min at 4°C with vigorous shaking. The supernatant was recovered from the cellulose by centrifugation at 5,000 x g for 20 min at 4°C and stored at -20°C. The supernatant was recovered from the Sephadex G-250[®] by centrifugation at 1,000 x g for 20 min at 4°C and stored at -20°C.

Proteins present in each wash and elution fraction were precipitated using the ProteoExtract[™] protein precipitation kit, according to the manufacturer's instructions. Precipitated protein was pelleted by centrifugation at 13,600 x g for 60 min at 4°C, washed twice in 70% ethanol, and air-dried for 30 min before resuspension in 20 µL 1x LDS buffer containing 20 mM DTT in preparation for 1-DE. Immediately prior to 1-DE the samples were denatured by heating at 70°C for 10 min. Gels were stained using colloidal CBB G-250 as described in Section 2.6.1.

2.15 Bioinformatics

Signal peptides were predicted by SignalP 3.0 (www.cbs.dtu.dk/services/SignalP/) (cut-off score $p > 0.5$), LipoP 1.0 (<http://www.cbs.dtu.dk/services/LipoP/>), and pattern searching as described by Sutcliffe and Harrington for Gram-positive bacteria (Bendtsen *et al.*, 2004, Juncker *et al.*, 2003, Sutcliffe and Harrington, 2002). Membrane-spanning domains were predicted using the TMHMM 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/) and SOSUI/G (<http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiG/sosuiGsubmit.html>) utilities (Hirokawa *et al.*, 1998, Krogh *et al.*,

2001). Proteins containing at least one membrane spanning domain distinct from the N-terminal signal peptide were defined as integral membrane proteins. Protein isoelectric point and molecular weight values were calculated using Emboss iep (<http://emboss.sourceforge.net/index.html>) and Protein Molecular Weight (http://www.bioinformatics.org/sms2/protein_mw.html) respectively. Functional domains were identified using Pfam, Tigrfam, and BLASTp analysis (Altschul *et al.*, 1990, Finn *et al.*, 2008, Haft *et al.*, 2003). Protein sequence alignments were performed, and phylograms constructed using ClustalW (Larkin *et al.*, 2007) Settings for sequence alignment were gap penalty = 10, end gaps excluded, gap extension penalty = 0.2, and gap separation penalty = 4. The Gonnet250 database was used for each multiple alignment.

The codon adaptation index (CAI) was calculated as follows: A codon usage table for 40 highly expressed *B. proteoclasticus* reference genes (translation elongation factors *tufA*, *tsf*, *fusA*, and 37 ribosomal protein encoding genes *rplA-rplF*, *rplI-rplT*, *rpsB-rpsT*, as used by Sharp *et al.* (2005)) was created by Emboss-CUSP (<http://bioweb.pasteur.fr/docs/EMBOSS/cusp.html>), which was then used to calculate a CAI value for each predicted protein by Emboss-CAI (<http://emboss.sourceforge.net/apps/release/5.0/emboss/apps/cai.html>). The genome was then ranked by descending CAI value, and the top 382 genes (10% of the genome) were designated predicted highly expressed (PHX) genes.

Chapter 3

The extracellular *B. proteoclasticus* proteome

3.1 Introduction

The size and insolubility of lignocellulosic material dictates that the degradation to soluble oligo- and monosaccharides takes place in the extracellular environment, which requires the secretion of a variety of polysaccharide degrading enzymes across the bacterial cell wall. Utilisation of the released oligosaccharides and soluble sugars necessitates uptake across the bacterial cell wall, which in Gram-positive bacteria is mediated by a variety of extracellular substrate-binding proteins linked to dedicated sugar transport systems. Secreted proteins may also play important roles in processes such as substrate attachment, cell-cell interactions, and defence. An examination of the secreted and cell-associated components of the *B. proteoclasticus* proteome, herein referred to as the extracellular proteome, is therefore an important step to characterising the fibre degrading enzyme system of the bacterium.

3.1.1 Secreted polysaccharidases

The majority of fibre degrading enzymes secreted by anaerobic rumen bacteria have a distinctive modular architecture that contains one or more catalytic domains in association with a variety of non-catalytic modules (Figure 3.1).

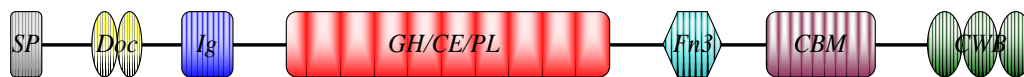


Figure 3.1. Schematic representation of a hypothetical fibrolytic enzyme synthesised by an anaerobic bacterium. The modular structure of a hypothetical anaerobic bacteria secreted glycosyl hydrolase is shown, which may contain many or all of the following domains; Doc, dockerin module; Ig, bacterial Ig-like domain; GH/CE/PL, glycosyl hydrolase/carbohydrate esterase/polysaccharide lyase catalytic domains respectively; Fn3, fibronectin Type-III module; CBM, carbohydrate binding domain; SLH, CWB, cell wall binding repeats. Domain size, location, and arrangement are representative only. Figure adapted from Schwarz *et al.* (2004).

In addition to signal peptides, the non-catalytic modules can be any of the following; dockerin domains (Doc), immunoglobulin (Ig)-like domains, fibronectin type III (Fn3) domains, CBMs, cell wall binding (CWB) repeat regions, and domains of unknown function (DUF). They can be positioned anywhere within the enzyme, although CBM and CWB domains are in most cases found towards the C-terminus.

3.1.1.1 Carbohydrate-binding modules

CBMs are well characterised domains of 35 to 180 residues that target the enzyme to different regions of the insoluble polysaccharide, dramatically increasing enzyme concentration at the site of catalysis, and improving enzyme substrate interaction and activity (Black *et al.*, 1996, Bolam *et al.*, 1998). As with catalytic GH domains, CBMs are classified into functional families based on amino acid similarity and three dimensional structure (Sun *et al.*, 1998), and there are currently 61 CBM families catalogued in the CAZy database. CBMs recognise a variety of lignocellulose degrading enzymes including crystalline cellulose, non-crystalline cellulose, chitin, (1→3)-β-D-glucans and mixed linkage glucans, xylans, mannans, galactan and starch, while some CBMs display ‘lectin-like’ specificity and bind to a variety of cell-surface glycans. The ligand specificity of the CBM may or may not be identical to the substrate specificity of the associated catalytic module (Sun *et al.*, 1998). Such an arrangement is likely to target the enzyme to multiple regions of complex insoluble polysaccharides, therefore enhancing overall catalytic potential.

CBMs have also been grouped into three types based on structural and functional similarities. Type-I CBMs are “surface-binding” modules that bind to the surface of highly crystalline cellulose and/or chitin in a thermodynamically dependent manner (Creagh *et al.*, 1996); Type-II “glycan-chain-binding” modules interact with insoluble and soluble glycan chains rather than crystalline surfaces, and bind to extended grooves of clefts. The binding ability is influenced by the degree of polymerisation, with increased substrate affinities up to hexasaccharides and weak interaction with oligosaccharides of three monomers or less (Notenboom *et al.*, 2001a). Type-III CBM modules lack the extended binding site ability of Type-II modules, and instead have lectin-like properties that bind optimally to mono-, di- or trisaccharides. Notably, all members of the CBM9 family are all Type-III molecules, and have so far been discovered exclusively in xylanases.

3.1.2 Uptake of lignocellulose derived soluble sugars-ABC transporters

The rapid assimilation of oligosaccharides and soluble sugars liberated from lignocellulose by the activity of secreted polysaccharidases are important to rumen bacterial metabolism. ATP Binding Cassette (ABC) transporter systems couple solute transport against a concentration gradient with the energy derived from ATP hydrolysis. The frequency of ABC transporter encoding genes in the *B. proteoclasticus* genome indicates the bacterium relies heavily on these systems for

sugar uptake in the rumen. ABC transport systems are ubiquitous among prokaryotes and comprise one of the largest known paralogous protein superfamilies (Dassa and Bouige, 2001). ABC uptake systems mediate the translocation of a wide variety of essential nutrients and osmoprotectants, including mono- and oligosaccharides, amino acids and oligopeptides, organic and inorganic ions, metal ions, and vitamins. They comprise four “core” domains; two hydrophobic TMDs that form the substrate translocation channel, and two highly conserved, hydrophilic nucleotide-binding domains (NBDs) that are located on the cytoplasmic face of the cell membrane (Schneider and Hunke, 1998). Bacterial ABC uptake systems also utilise a high-affinity substrate-binding protein (SBP), which in Gram-positive bacteria SBPs are often soluble lipoproteins that are tethered to the external surface of the cell wall by N-terminal cysteine acylation, or fused directly to the membrane transporter (van der Heide and Poolman, 2002).

SBPs are essential for efficient transmembrane transport irrespective of the external solute concentration (Dabard *et al.*, 2001, Davidson *et al.*, 1992). They restrict substrate diffusion away from the cell and therefore enhance substrate interaction with the membrane bound translocation machineries (Monedero *et al.*, 2008). Moreover, SBP/substrate interaction improves SBP/transporter affinity (Chen *et al.*, 2001), and conformational changes in SBPs induced by substrate-binding also stimulates NBD ATP hydrolysis, both of which enhance the activity of the associated transporter system (Davidson *et al.*, 1992).

SBP dependent ABC transporter uptake systems comprise two sub-families (<http://www.tcdb.org/>) (Saier *et al.*, 2009), which are differentiated by the nature of their substrates, core subunit composition, and conserved sequence motifs. Members of the Carbohydrate Uptake Transporter 1 (CUT1) family transport a variety of di- and oligosaccharides, and are composed of the four core domains plus the SBP, while systems belonging to the CUT2 family transport almost exclusively monosaccharides and possess single copies of the TMD and NBD plus the SBP. The wide variety of substrates assimilated by ABC transporter uptake systems is mediated primarily by SBP specificity, although TMDs may also possess low affinity substrate-binding domains (Shuman, 1982). Most SBPs bind one or a family of related substrates, for example the CUT1 family maltose and xylooligosaccharide transporters of *Lactobacillus casei* (Shulami *et al.*, 2007) and *Geobacillus stearothermophilus* (Doeven *et al.*, 2004) respectively. Some SBP dependent import systems also show

versatility in their ability to handle a variety of structurally unrelated substrates. Such versatility may be achieved by a single SBP interacting with a variety of substrates (Higgins and Ames, 1981, van der Heide and Poolman, 2002), or promiscuous membrane bound transporter complexes interacting with several SBPs of variable specificity (Stephenson, 2005).

3.2 Results

3.2.1 Theoretical 2-DE

A Type-I or Type-II N-terminal secretory signal peptide was detected in 537 *B. proteoclasticus* proteins, 200 of which were predicted to be transmembrane proteins due to the presence of at least one membrane-spanning domain distinct from the signal peptide. The theoretical *pI* distribution of the remaining 337 secreted proteins showed a strong skew towards the acidic *pI* range (Figure 3.2A), where 294 proteins (87%) had a *pI* value between 3 and 5.6. The theoretical size distribution of the secretome ranged between 6 and 371 kDa and 94% of the proteins were between 10 and 150 kDa. Examination of the predicted protein function demonstrated that almost a quarter of the *B. proteoclasticus* secretome is devoted to lignocellulose degradation and the uptake of released soluble sugars (Figure 3.2B). The *B. proteoclasticus* secretome included 35 polysaccharide-degrading enzymes that contained catalytic domains collectively representing 14 GH, four CE, and two PL families. The GH3, 5, 10, and 13 classifications are the most well represented, and the complete repertoire of secreted enzymes contains catalytic domains classified within 21 different CAZy families. There are representatives of all three types of enzymes (GHs, CEs, and PLs) necessary for the complete lignocellulose degradation. All three carbohydrate binding proteins encoded by the *B. proteoclasticus* genome are also predicted to be secreted proteins.

All but three of secreted polysaccharidases had a theoretical *pI* value below 5.5, and all were between 40 and 294 kDa (Figure 3.2A). More than 16% of the secretome were components of one or more ABC transport systems and included 34 sugar-specific SBPs, some of which are likely to mediate the uptake of lignocellulose derived soluble sugars. Strikingly, all 34 SBPs had a theoretical *pI* value below 4.7. Almost 5% of the secretome was involved in protein processing, but surprisingly for a highly proteolytic bacterium, only one extracellular protease and one peptidase were predicted. More than half the *B. proteoclasticus* secretome consisted of hypothetical proteins or proteins of unknown function.

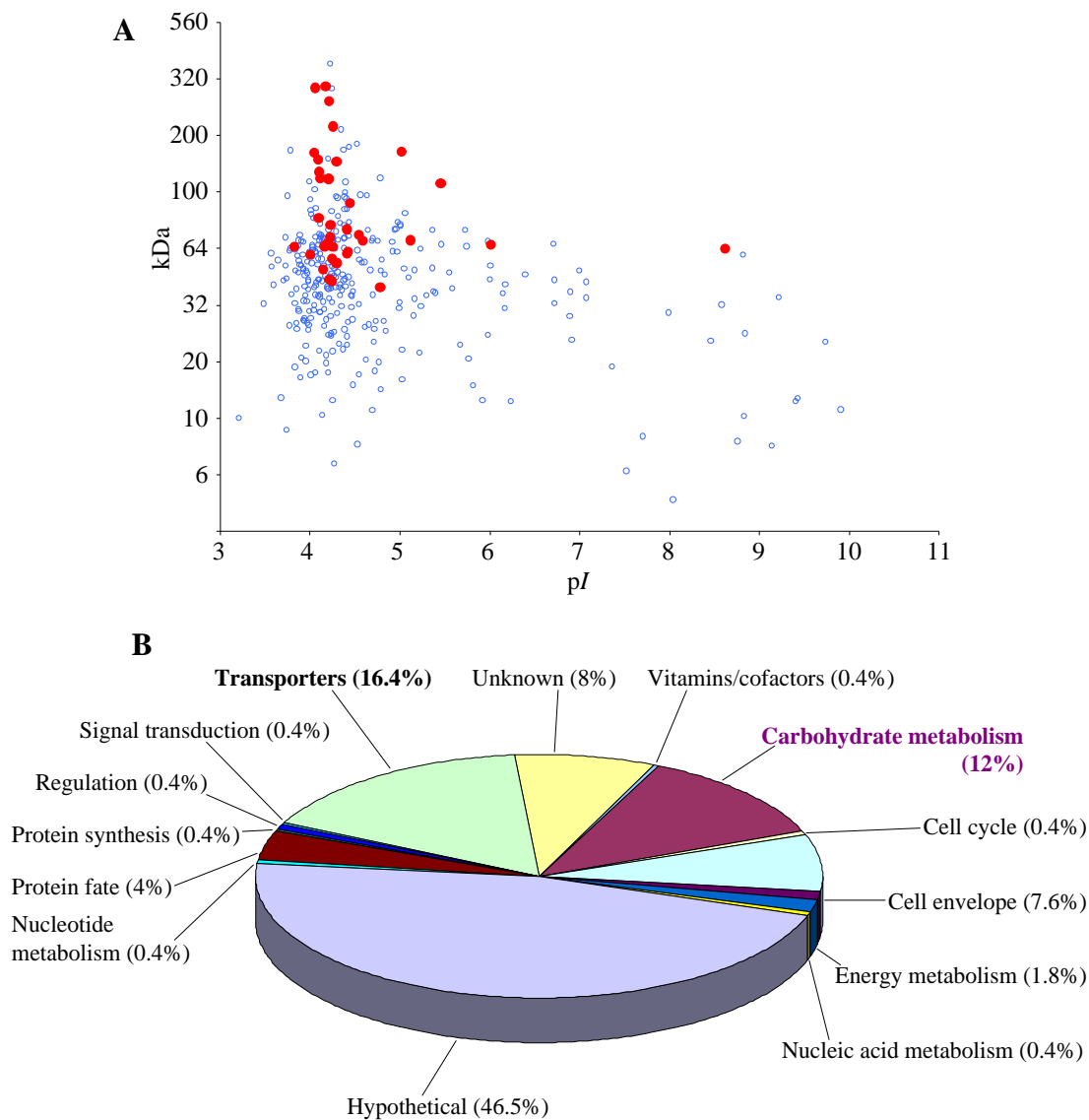


Figure 3.2. Summary of the theoretical *B. proteoclasticus* secreted proteome. Theoretical 2-DE (A) and protein function summary (B) of the 337 predicted *B. proteoclasticus* secreted proteins. Red symbols in (A) represent predicted secreted polysaccharide-degrading enzymes.

The codon adaptation index (CAI) (Sharp and Li, 1987) is a measure of the extent of synonymous codon usage bias, and in bacterial genomes is positively correlated with mRNA expression and relative protein abundance (Futcher *et al.*, 1999, Goetz and Fuglsang, 2005, Ishihama *et al.*, 2008, Jansen *et al.*, 2003, Kanaya *et al.*, 1999). A CAI value for each predicted secreted protein was calculated using 40 highly expressed *B. proteoclasticus* reference genes, which were used as an estimator of relative protein abundance.

Relative protein abundance estimation using the CAI demonstrated that 20% of the secretome were the products of predicted highly expressed genes, and almost half of these were either polysaccharide-degrading enzymes or ABC transporter solute binding proteins. The predicted highly expressed polysaccharidases included enzymes targeting the xylan backbone, xylooligomers, and backbone substituents, as well as cellulose, starch, and pectin. Secreted proteins that were the products of predicted non-highly expressed genes were dominated by hypothetical proteins, proteins of unknown function, and those involved in cell envelope biogenesis.

3.2.2 Gel-based analysis of the *B. proteoclasticus* secreted proteome

2-DE protein separation in combination with MALDI-TOF mass spectrometry and ESI-MS/MS was used to identify *B. proteoclasticus* secreted proteins, and proteins that may be attached to the external cell surface. To examine the effect of simple and complex plant polysaccharides on secreted protein abundance patterns, cells were grown *in vitro* in modified M704 culture medium supplemented with either 0.1% xylan or 0.5% xylose. The effect of growth in each of these conditions was analysed in cells harvested at mid-log and stationary phase.

3.2.2.1 Optimisation of *B. proteoclasticus* secreted protein purification

Three methods suitable for the purification of low concentration protein from large solution volumes were trialled for purification of *B. proteoclasticus* secreted proteins. The proteins present in 45 mL of freshly thawed *B. proteoclasticus* culture medium harvested from xylan grown cells were precipitated using either TCA at a final concentration of 10% (w/v) as described by Voigt *et al.* (2006), or ammonium sulphate at a final concentration of 60% (w/v) as described by Jiang *et al.* (2004). In both cases the precipitation was performed overnight at 4°C. After harvesting and washing the precipitated proteins, each protein pellet was resuspended in 20 µL of 1 x SDS sample buffer (containing 20 mM DTT) and analysed by 1-D SDS-PAGE. Secreted protein purification was also attempted using a Centricon® Centrifugal Concentrator unit with a 10,000 Dalton cut-off ultra-filter membrane. After repeated attempts this method was found to be unsuitable because the filter membrane of the device rapidly clogged and prevented the passage of the fluid through the device. Furthermore, the retained sample became progressively contaminated with particulate matter that was likely to be undigested oat-spelt xylan suspended in the culture medium. As a consequence this method was discarded from the optimisation analysis.

Loading the total protein mass precipitated by the TCA and ammonium sulphate methods allowed direct comparison of the efficiency of each procedure, and a clear quantitative difference between the two methods was found (Figure 3.3). Replicate experiments confirmed that the TCA precipitation method was significantly more effective for the precipitation of *B. proteoclasticus* secreted proteins and was therefore selected as the most appropriate method for the preparation of secreted proteins for all future analyses.

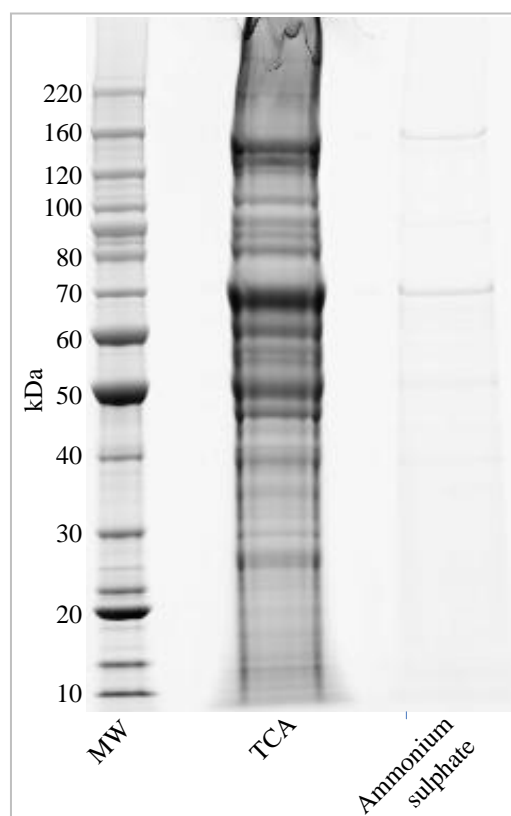


Figure 3.3. 1-DE analysis of the effect of methods for the purification of proteins from the culture medium of *B. proteoclasticus* cells. Total protein recovered from 45 mL of culture medium harvested from xylan grown, mid-log phase harvested cells by TCA and ammonium sulphate was analysed by 1-D SDS-PAGE, and the gel was stained with CBB-G250.

3.2.2.2 2-DE of proteins recovered from the *B. proteoclasticus* culture medium

The pI 3-10 2-DE spot pattern of proteins recovered from the culture medium of xylan-grown, mid-log phase harvested cells was similar to the theoretical secretome 2-DE map, with approximately 70% of the detectable proteins focused between pI 3.0 and 5.6. Consequently, *B. proteoclasticus* secreted proteins recovered from the culture medium of mid-log phase (Figure 3.4 and Figure 3.5) and stationary phase (Figure 3.6

and Figure 3.7) harvested cells grown in the presence of xylan (Figure 3.4 and Figure 3.6) and xylose (Figure 3.5 and Figure 3.7), were separated by 2-DE using pI 3.0-5.6 IPG strips.

In each gel set an average of 195 individual protein spots were detected, 111 of which were identified as the products of 74 *B. proteoclasticus* CDSs. Thirty of these non-redundant gene products contained a Type-I or Type-II N-terminal signal peptide, which included five polysaccharidases and two carbohydrate-binding proteins (CBPs) (Figure 3.4-Figure 3.7, and Table 3.1). All but two of these were the product of a predicted highly expressed gene. The five secreted polysaccharide-degrading enzymes were identified by matching at least 13 peptide masses to the full-length protein sequences with a maximum mass error of 50 ppm, and achieving a minimum 20% sequence coverage. The *B. proteoclasticus* genome encodes two secreted CBPs at loci Bpr_I0736 and Bpr_I1599 (chromosome 1), and both were detected with 15% and 16% coverage respectively. All seven carbohydrate active proteins contained a Type-I signal peptide (probability score = 1) and none contained an identifiable TMD. In addition, 16 ABC transporter system substrate-binding proteins were identified, four of which were detected in one or more protein spots that were between 3.7 and 7.6-fold more abundant in the 2-DE separations of protein harvested from xylan grown cells, or were unique to the xylan condition. Although the SBP product of Bpr_I1720 was identified in spot 18 that was detected uniquely in the xylan growth condition, the protein was also identified in several other spots that were not differentially abundant between growth conditions, and was therefore deemed to not be a differentially abundant protein in the *B. proteoclasticus* extracellular proteome. The only secreted subtilisin family serine protease, one cell-surface protein, and three hypothetical proteins were also identified (Figure 3.4-Figure 3.7, and Table 3.2).

The majority of the 44 predicted cytosolic proteins identified in the *B. proteoclasticus* culture medium (Appendix B, Table B.1) were involved in carbohydrate utilisation, energy metabolism, nucleotide and nucleic acid metabolism, and protein synthesis. Relative abundance analysis of the complete set of identified proteins demonstrated that the 20 most abundant 2-DE protein spots were all secreted proteins. Furthermore, theoretical protein abundance estimation using the CAI showed that 75% of the identified cytosolic proteins were the products of predicted highly expressed genes.

In an attempt to clarify whether unidentified protein spots that were visible in the 2-DE separations of secreted proteins originated from proteins that may have been present in the culture medium prior to use, each unidentified PMF was searched against the eukaryotic division of the NCBI non-redundant database. In no case was a positive identification obtained.

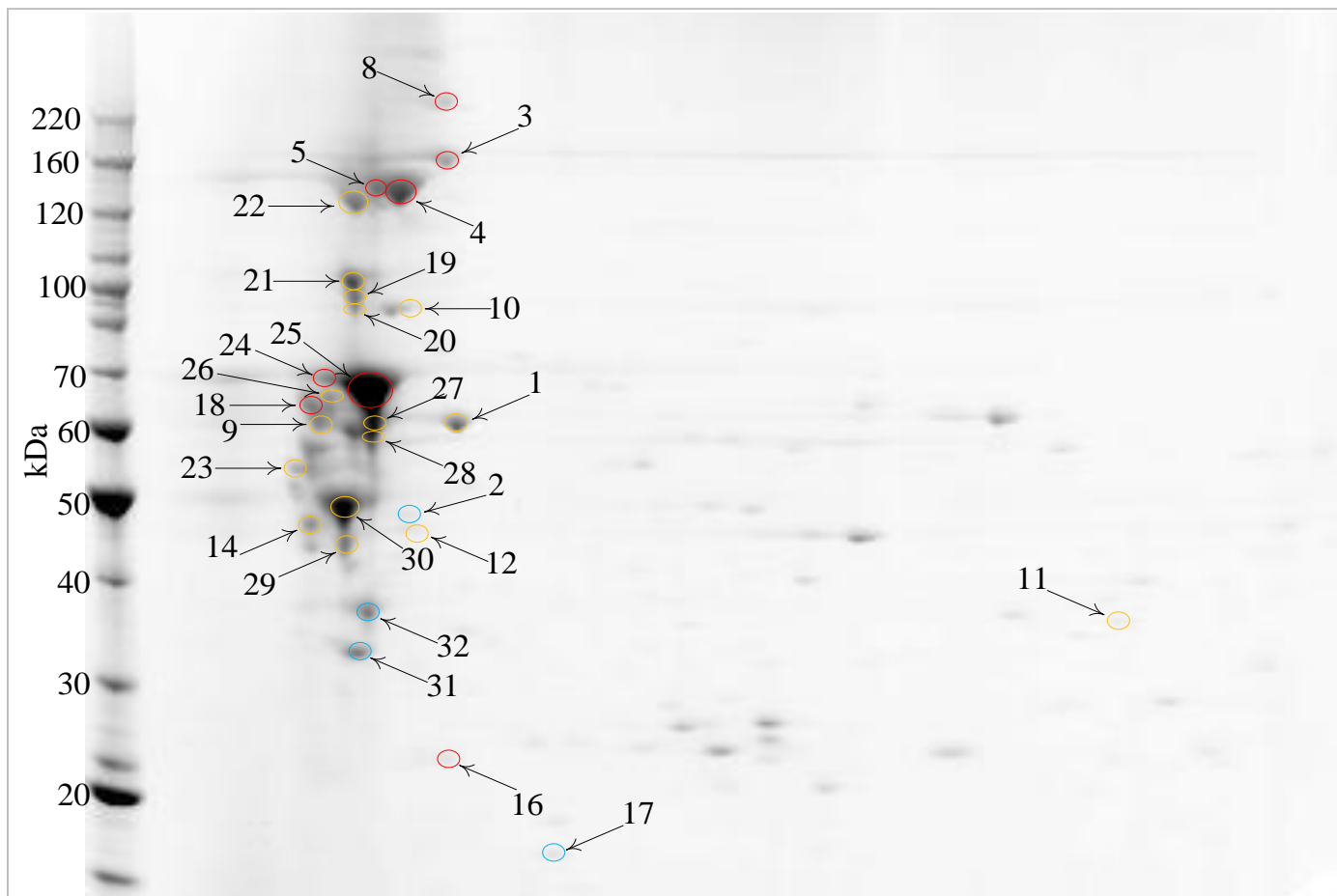


Figure 3.4. Narrow-range 2-DE of *B. proteoclasticus* secreted proteins harvested at mid-log phase from cells grown in the presence of 0.1% xylan. Red, blue, and gold circles denote spots with increased, decreased, and unchanged abundance respectively in the xylan growth condition relative to xylose. Circled spots are summarised in Table 3.1 and Table 3.2.

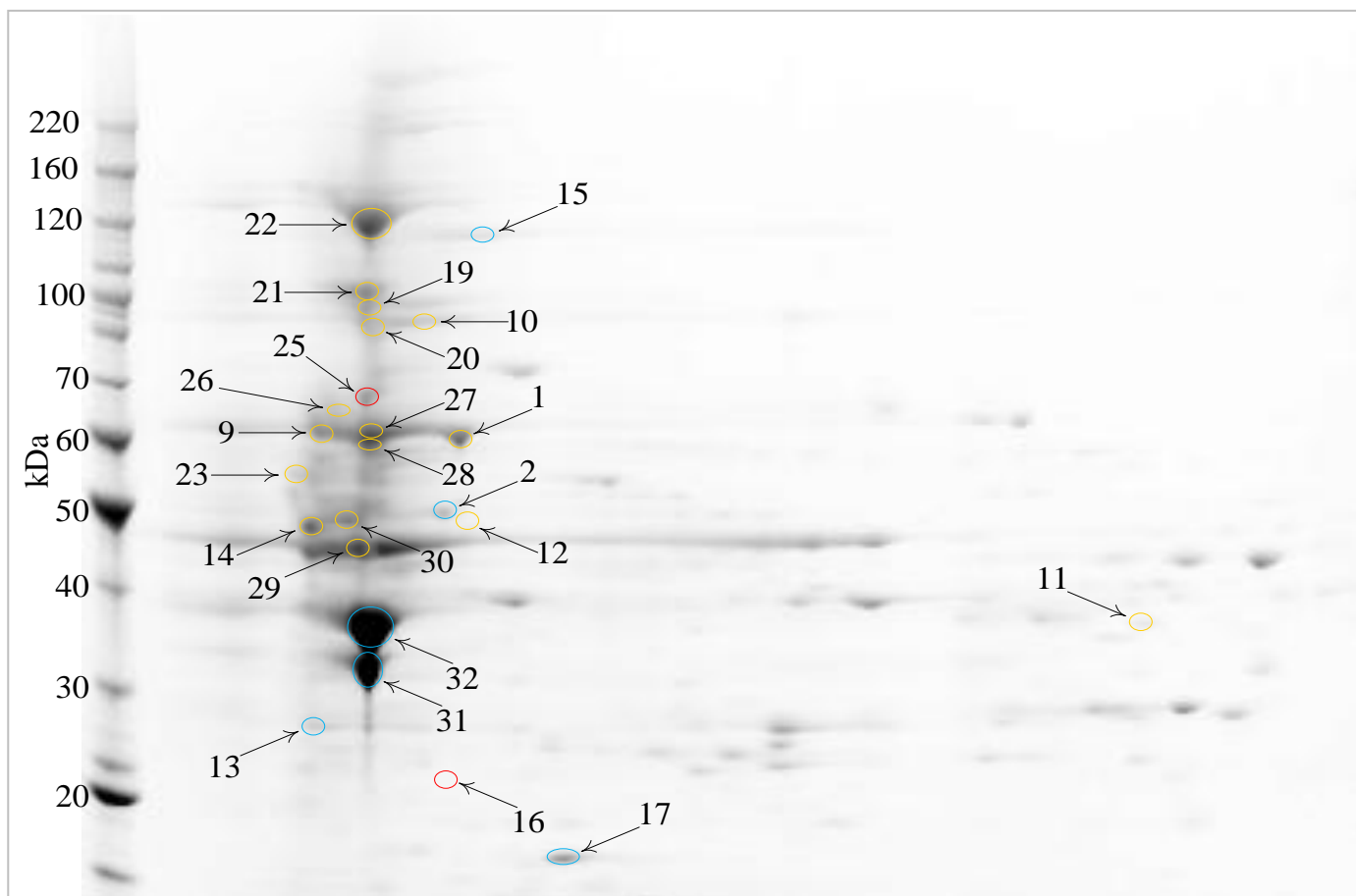
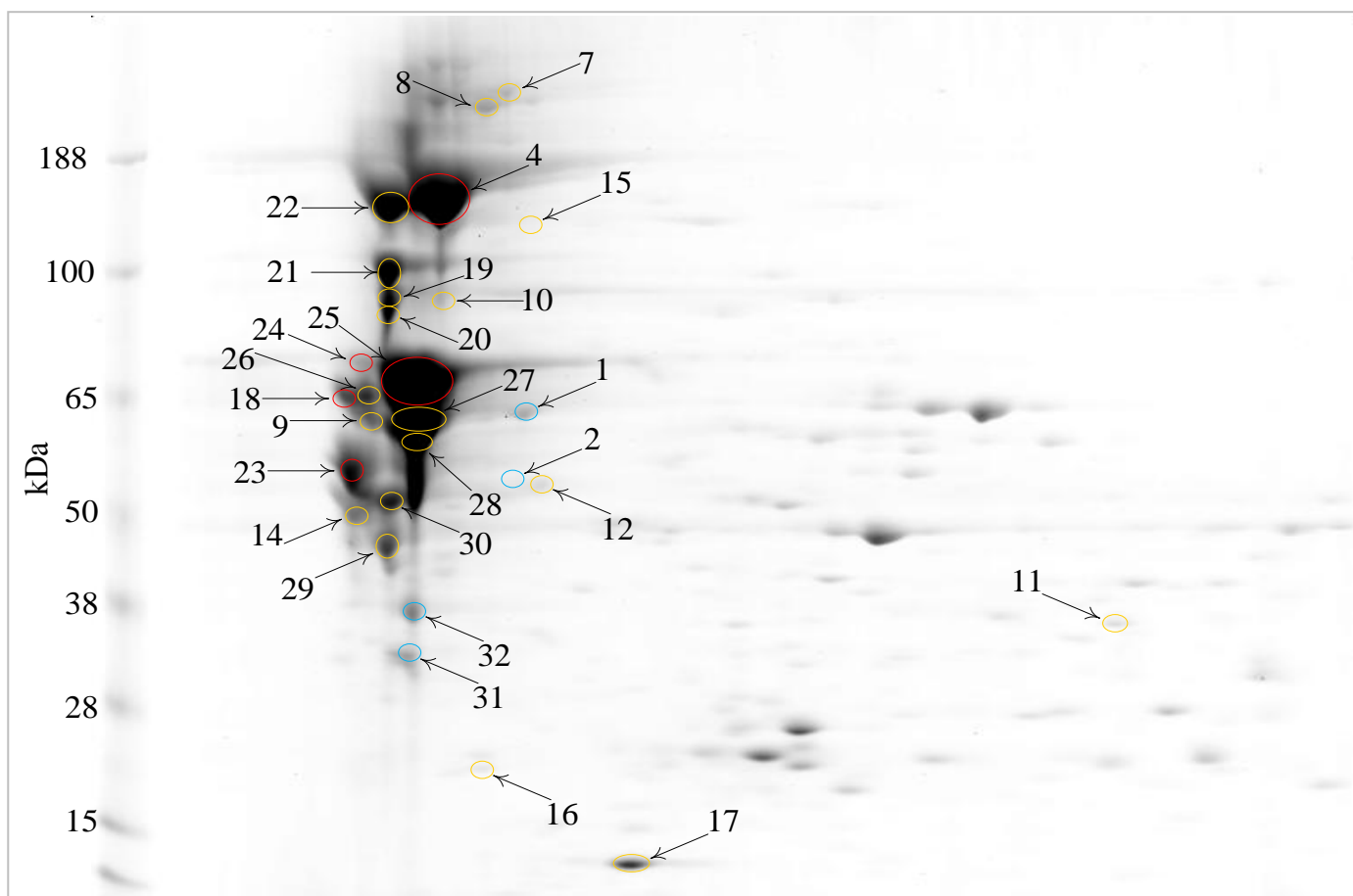


Figure 3.5. Narrow-range 2-DE of *B. proteoclasticus* secreted proteins harvested at mid-log phase from cells grown in the presence of 0.5% xylose. Red, blue, and gold circles denote spots with increased, decreased, and unchanged abundance respectively in the xylan growth condition relative to xylose. Circled spots are summarised in Table 3.1 and Table 3.2.



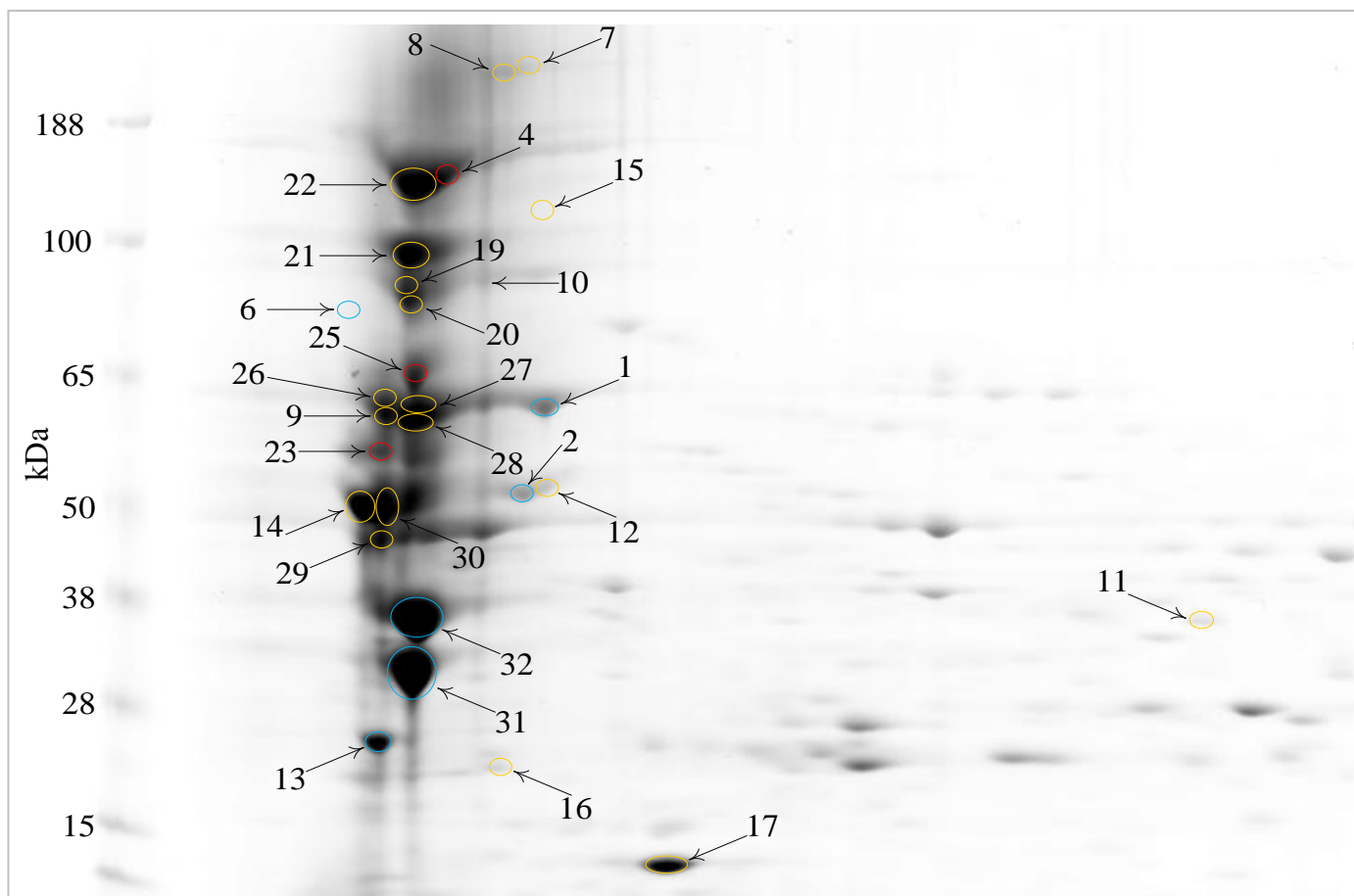


Figure 3.7. Narrow-range 2-DE of *B. proteoclasticus* secreted proteins harvested at stationary phase from cells grown in the presence of 0.5% xylose. Red, blue, and gold circles denote spots with increased, decreased, and unchanged abundance respectively in the xylan growth condition relative to xylose. Circled spots are summarised in Table 3.1 and Table 3.2.

Table 3.1. Secreted carbohydrate active proteins identified in the mid-log and stationary phase harvested *B. proteoclasticus* culture medium.^a

Spot	Protein	Locus	EC	PHX ^b	Score ^c	pI	kDa	Pep. ^d	Cov. ^d	ML ^e	Stat ^e
1	Endo-1,4-β-glucanase, Cel5C	Bpr_I1710	3.2.1.4	Y	1.2e-09	4.6	61.1	16/54	36%	n/c	-5.7±1.7 (0.002)
2	Endo-1,4-β-glucanase, Cel5C	Bpr_I1710	3.2.1.4	Y	7.4e-07	4.6	61.1	13/46	28%	-17.2±4.8 (0.008)	-12.1±5.3 (0.008)
3	Endo-1,4-β-xylanase, Xyn10B	Bpr_I0026	3.2.1.8	Y	1.2e-14	4.3	136.9	24/45	20%	xylan	n/d
4	Endo-1,4-β-xylanase, Xyn10B	Bpr_I0026	3.2.1.8	Y	9.6e-14	4.3	136.9	24/55	25%	xylan	17.5±2.5 (0.000)
5	Endo-1,4-β-xylanase, Xyn10B	Bpr_I0026	3.2.1.8	Y	1.5e-09	4.3	136.9	20/53	20%	xylan	n/d
6	Pectate lyase, Pel1A	Bpr_I2372	4.2.2.2	N	3.0e-13	4.4	115.5	19/56	28%	n/d	xylose
7	Pectin methyl-esterase, Pme8B	Bpr_I2473	3.1.1.11	Y	2.4e-10	4.2	294.1	24/58	14%	n/d	n/c
8	Xylosidase/arabinofuranosidase, Xsa43J	Bpr_I2935	-	Y	3.0e-18	4.2	251.9	31/61	17%	xylan	n/c
9	Carbohydrate-binding protein	Bpr_I0736	-	N	1.0e-05	4.1	57.5	6/20	15%	n/c	n/c
10	Carbohydrate-binding protein	Bpr_I1599	-	Y	4.8e-08	4.1	76.9	11/21	16%	n/c	n/c

^a Secretory signal-peptides were predicted using SignalP (Ver. 3.0) (www.cbs.dtu.dk/services/SignalP/) and LipoP (Ver. 1.0) (<http://www.cbs.dtu.dk/services/LipoP/>).

^b PHX, predicted highly expressed gene.

^c Score value is the statistical expectation that the top ranked protein match is a false positive identification.

^d Pep, the number of matched peptides masses/searched peptides masses; Cov, Protein sequence coverage of the matched peptides.

^e ML, culture medium harvested at mid-log phase (OD₆₀₀ = 0.5); Stat, culture medium harvested at stationary phase (OD₆₀₀ = 0.7). Mean fold-change ± SEM calculated from three biological replicate experiments is shown. *p*-values are shown in brackets. Xylan/xylose denotes uniquely detected in culture medium harvested from xylan / xylose grown cells respectively; n/d, not detected in either growth condition; n/c, no protein abundance change between growth conditions.

Table 3.2. Predicted non-carbohydrate active secreted proteins identified in the mid-log and stationary phase harvested *B. proteoclasticus* culture medium.

Spot	Protein	Locus	Func. ^a	PUL	PHX ^b	Score ^c	SigP ^d	pI	kDa	Pep. ^b	Cov. ^b	ML ^e	Stat ^e
11	3-hydroxybutyryl-CoA dehydrogenase, Hbd	Bpr_I2486	E	-	Y	1.2e-08	SpI	5.4	31.5	11/47	41%	n/c	n/c
12	ABC transporter SBP	Bpr_I1600	T	-	N	4.8e-05	SpII	4.2	93.5	13/46	16%	n/c	n/c
13	Amino acid ABC transporter SBP	Bpr_I2466	T	-	Y	4.9e-06	SpII	4.1	32.4	6/15	29%	xylose	xylose
14	Bmp family protein	Bpr_I1560	U	-	Y	1782	SpI	3.9	40.7	4	10%	n/c	n/c
	Sugar ABC transporter SBP	Bpr_I1667	T	-	Y	1094	SpII	4.1	47.8	5	14%		
15	Cell surface protein	Bpr_I2508	I	-	N	1.1e-06	SpI	4.6	141.4	12/31	13%	xylose	n/c
16	Hypothetical protein	Bpr_I0139	H	-	N	3.0e-08	SpI	4.7	26.7	8/34	33%	n/c	n/c
	Hypothetical protein	Bpr_I0188	H	3	N	3.9e-05	SpI	5.1	27.0	934	39%		
17	Hypothetical protein	Bpr_I2628	H	-	Y	1.5e-14	SpI	7.8	31.8	12/34	39%	-8.2±2.8 (0.004)	n/c
18	Oligopeptide ABC transporter SBP, OppA1	Bpr_I1276	T	-	Y	3195	SpII	4.3	83.3	10	15%	Xylan	Xylan
	Peptide/nickel ABC transporter periplasmic protein	Bpr_I2750	T	-	Y	1303	SpII	4.1	58.2	7	12%		
	Sugar ABC transporter SBP	Bpr_I0313	T	5	Y	5047	SpII	4.0	55.3	9	11%		
	Sugar ABC transporter SBP	Bpr_I1720	T	16	Y	3195	SpII	4.2	50.0	10	15%		
19	Oligopeptide ABC transporter SBP, OppA1	Bpr_I1276	T	-	Y	7.6e-20	SpII	4.3	83.3	24/40	36%	n/c	n/c

Table continues

Table 3.2 continued.

Spot	Protein	Locus	Func. ^a	PUL	PHX ^b	Score ^c	SigP ^d	pI	kDa	Pep. ^b	Cov. ^b	ML ^e	Stat ^e
20	Oligopeptide ABC transporter SBP, OppA1	Bpr_I1276	T	-	Y	8794	SpII	4.3	83.3	12	17%	n/c	n/c
	Oligopeptide ABC transporter SBP, OppA2	Bpr_III023	T	-	Y	1252	SpII	4.3	77.5	7	10%		
21	Oligopeptide ABC transporter SBP, OppA1	Bpr_I1276	T	-	Y	1.5e-27	SpII	4.3	83.3	28/38	42%	n/c	n/c
22	Serine protease subtilisin family	Bpr_I2629	P	-	Y	6.1e-08	SpI	4.0	153.2	26/89	24%	n/c	n/c
23	Sugar ABC transporter SBP	Bpr_I0313	T	5	Y	4180	SpII	4.0	55.3	9	10%	n/c	3.7±1.7 (0.034)
	Sugar ABC transporter SBP	Bpr_I2010	T	20	Y	1.9e-8	SpII	4.0	47.6	6	29%		
	Sugar ABC transporter SBP	Bpr_I2344	T	24	Y	1152	SpII	4.1	52.4	5	14%		
24	Sugar ABC transporter SBP	Bpr_I0182	T	3	Y	3.8e-10	SpII	4.4	63.5	15/54	29%	Xylan	Xylan
25	Sugar ABC transporter SBP	Bpr_I0182	T	3	Y	1.2e-9	SpII	4.4	63.5	14/47	35%	7.6±3.0 (0.010)	5.3±0.2 (0.000)
26	Sugar ABC transporter SBP	Bpr_I0237	T	4	Y	1087	SpII	4.1	65.1	5	10%	n/c	n/c
	Sugar ABC transporter SBP	Bpr_I0313	T	5	Y	1682	SpII	4.0	55.3	5	5%		
	Sugar ABC transporter SBP	Bpr_I1589	T	13	Y	2260	SpII	4.2	61.2	10	18%		
27	Sugar ABC transporter SBP	Bpr_I0937	T	9	Y	1.0e-5	SpII	4.4	61.5	12/41	24%	n/c	n/c
28	Sugar ABC transporter SBP	Bpr_I0937	T	9	Y	3.8e-9	SpII	4.4	61.5	16/53	32%	n/c	n/c
29	Sugar ABC transporter SBP	Bpr_I1720	T	16	Y	4737	SpII	4.2	50.0	9	20%	n/c	n/c
30	Sugar ABC transporter SBP	Bpr_I1720	T	16	Y	3706	SpII	4.2	50.0	7	18%	n/c	n/c

Table continues

Table 3.2 continued.

Spot	Protein	Locus	Func. ^a	PUL	PHX ^b	Score ^c	SigP ^d	pI	kDa	Pep. ^b	Cov. ^b	ML ^e	Stat ^e
31	Sugar ABC transporter SBP	Bpr_III244	T	34	Y	2.4e-10	SpII	4.3	35.3	11/34	48%	-4.5±0.6 (0.005)	-3.6±1.5 (0.010)
32	Xylose ABC transporter SBP	Bpr_I1173	T	10	Y	3.0e-8	SpII	4.4	38.6	10/34	34%	-4.8±1.1 (0.009)	-3.3±1.7 (0.010)

^a E, Energy metabolism; H, Hypothetical; I, Cell envelope biogenesis; P, Protein fate; T, Transporters; U, Unknown function.

^b PHX, predicted highly expressed gene; SigP, contains a signal peptide; Pep, the number of matched peptides masses/searched peptides masses; Cov, protein sequence coverage of the matched peptides.

^c Score value is the statistical expectation that the top ranked match is a false positive identification. MOWSE scores are given for LC-MS/MS identifications.

^d Secretory signal-peptides were predicted using SignalP (Ver. 3.0) (www.cbs.dtu.dk/services/SignalP/) and LipoP (Ver. 1.0) (<http://www.cbs.dtu.dk/services/LipoP/>).

^e ML, culture medium harvested at mid-log phase (OD₆₀₀ = 0.5); Stat, culture medium harvested at stationary phase (OD₆₀₀ = 0.7). Mean fold-change is shown ± SEM. *p*-values are shown in brackets. Xylan / xylose denotes uniquely detected in culture medium harvested from xylan / xylose grown cells respectively; n/d, not detected in either growth condition; n/c, no protein abundance change between growth conditions.

3.2.3 The effect of growth substrate on the secreted proteome of *B. proteoclasticus*

To examine the effect of simple and complex plant carbohydrates on protein abundance patterns *B. proteoclasticus* secreted proteins, cells were grown *in vitro* in modified M704 culture medium supplemented with either 0.1% xylan or 0.5% xylose. Preliminary experiments demonstrated that very little protein could be recovered from cells grown in unsupplemented modified M704 medium; therefore the unsupplemented growth condition was not used in this analysis. The effect of growth in xylan and xylose containing medium was analysed in cells harvested at mid-log and stationary phase using gel-based proteomics.

3.2.3.1 Abundance patterns of secreted polysaccharidases

The endoxylanase Xyn10B is the product of a predicted highly expressed gene and was the most abundant polysaccharide degrading enzyme detected in the *B. proteoclasticus* secretome at mid-log and stationary phase time points (Table 3.1, Figure 3.4 and Figure 3.6). In the 2-DE separations of proteins extracted from xylan-grown, mid-log phase harvested cells Xyn10B was identified in three distinct spots that together comprised 5.6% of the total detectable protein (Figure 3.4). The most abundant of the three spots containing Xyn10B was spot 4, which contributed more than 80% of the total Xyn10B abundance at mid-log phase. During stationary phase, the enzyme was identified in a single spot at the same 2-DE position as the mid-log phase detected spot 4, and represented 8.7% of the total protein (Figure 3.6).

In contrast to the abundance of Xyn10B in the secreted protein recovered from xylan grown cells, the enzyme was undetectable at mid-log phase when cells were grown in the presence of xylose as a sole carbon source (Figure 3.5), and was only weakly detectable during stationary phase growth (Figure 3.7). Differential abundance analysis showed Xyn10B was 17.5-fold more abundant in the secreted protein of xylan-grown cells during stationary phase growth compared to growth on xylose.

The *xyn10B* gene is predicted to encode a 136.9 kDa mature protein, which is similar to the observed 2-DE positions of spot 4 and spot 5, and smaller than the position of spot 3 (Figure 3.4 and Figure 3.6). Examination of the peptide mass fingerprints (Figure 3.8) and MASCOT sequence coverage results (Figure 3.9) of spots 3, 4, and 5 showed that six well resolved peptides detected in spot 3 that matched to the 110-residue C-terminal region of the Xyn10B were not detectable in spots 4 or 5 (Figure 3.9 A, black boxes). Furthermore, MS/MS analysis of protein extracted from spot 4 identified 21 peptides matching to Xyn10B, but also failed to detect peptides

matching to the C-terminal region (Figure 3.10). These results suggest that spot 3 is the full-length form of Xyn10B and spots 4 and 5 represent C-terminal truncated forms.

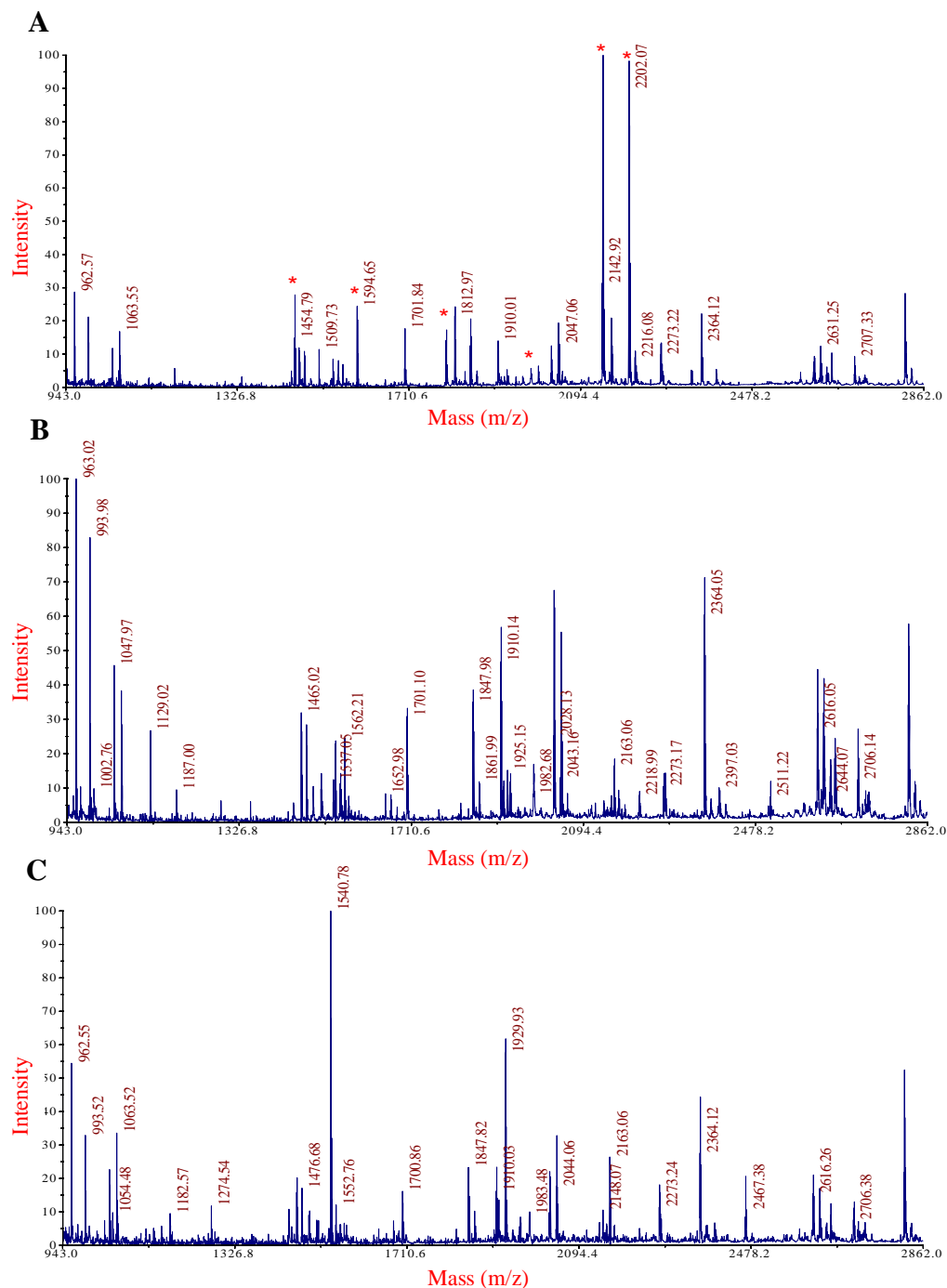


Figure 3.8. MALDI-TOF peptide mass fingerprints of protein extracted from spot 3 (A), spot 4 (B), and spot 5 (C). Peptide masses unique to spot 3 are marked with asterisks.

A Match to: Bpr_I0026 Score: 175 Expect: 1.2e-014
Bpr_I0026 endo-1,4-beta-xylanase Xyn10B 33228:36929 forward MW:136888

1	MH	KRWLG	GISS	FVMAGLLT	TVT	NVGTGVTQ	VF	AAEAATG	SSSL	VAEVEEAA	VAE
51	ES	SAETS	SVED	AADEASSE	AS	VAEEATEE	SS	DESAKEAT	DE	ASAETSAE	PA
101	EE	ENSLV	FFN	RWNQSAK	TSG	KLVSFKQ	YEE	YCYDLG	ETVP	VANVKALT	TVK
151	VS	KQDKN	VAI	KLYDADM	TEK	YANYGCDKNC	EYVFNP	TYDYG	NVRYIG	VMSM	
201	AS	GEAEY	PYG	ITIDEVV	V	QTEAPKNEE		TIVLEG	DALK	FTEAWEG	TEV
251	DG	ATLEFD	KE	WREYRV	SLGK	TLP	GP	VDKSV		KVTF	FAEANKQ
301	TEL	KADY	GKN	GSKSYTT	YPN	VTADVDA	IAI	MAMNEQ	TYPF	SVTVEK	IEVV
351	VD	TPAED	RP	QAGVEY	DIVD	LRKPVAE	LMG	DDFIIG	TAAS	YDEFGD	PLDM
401	EL	VYK	HFNGV	TLGNELK	PKDS	MLR	KDAE	IVD	YELNGE	MVPF	PVLD
451	RL	DRFV	KWNE	EHPEKKI	QVR	GHVLVW	HSQT	PSFFFH	EDYD	TSKPYV	TPDV
501	MN	K RL	EIYIR	EVAKHFT	AED	SKYKDL	FY	GW	DV	VNEAVSDG	TGTYRNASER
551	SE	WGW	VYGSQ	EFITNA	FVYA	NRYPAD	IAL	FYNDY	NETVS	SKMGGIC	QLL
601	RD	VKATP	GAR	IDGMGMQ	AHY	QIASNNPSME		QFKTA	AAKAYA	AI	VDQVQVTE
651	LDFK	GATS	SAK	DDRLAER	YKA	VYDTIRRLRE		DGVNFT	GMTI	WGVTDK	HSWL
701	QSS	NNNGG	GA	DGSAR	QYPLL	FDDYYKAKDC		FWAIA	NAGEL	EPEVKS	ITLV
751	QNV	NNDFS	SAG	NVYEFV	AGKF	IPMWS	ENGID	VKVVIK	DTT	QDD	FFTVYA
801	DD	GTGIK	SVV	VKREEA	TENE	NGYEA	VNV	VD	FEALT	TANK	VKFDIV
851	DF	LAV	FNDTT	FKHAQSS	KFF	AETVIKPLAV		VK	KGTPVIDG	EFDDEAWKTA	
901	DEL	PVAI	KVG	AKASASA	KIL	WDDEY	LYVLA	DVKD	SVLNKA	SSDAWE	QDSV
951	EI	FVD	NNNK	TSTYEADDKQ		YRINF	ENTHS	FNGTK	CEVEN	IKSEVV	VTTED
1001	GK	IEA	AALKW	TDIAP	VEGAK	VGLDIQV	NDA	DNSGKR	VGTL	NWADKTGN	GW
1051	SST	EVFG	TIL	LADGSD	VPTP	EVPTPE	TPEE	PEGTL	VTKYG	KTYVVT	TANGE
1101	KL	TGLQ	KVDG	NMYYCK	ADGK	VSKNAL	FTFE	GAVY	CAGADG	VLARNVKLEK	
1151	YHNEYFFGDD			CKAKTGFVTY		DGQDYFCKDN		GRVAKDYMIT		VGDKKYYAKK	
1201	DG	TLAKNE	QI	VKLFR	RYTFD	ENGVLIKTEK		GIF			

B Match to: Bpr_I0026 Score: 166 Expect: 9.6e-014
Bpr_I0026 endo-1,4-beta-xylanase Xyn10B 33228:36929 forward MW:136888

1	MH	KRWLG	GISS	FVMAGLLT	TVT	NVGTGVTQ	VF	AAEAATG	SSSL	VAEVEEAA	VAE
51	ES	SAETS	SVED	AADEASSE	AS	VAEEATEE	SS	DESAKEAT	DE	ASAETSAE	PA
101	EE	ENSLV	FFN	RWNQSAK	TSG	KLVSFKQ	YEE	YCYDLG	ETVP	VANVKALT	TVK
151	VS	KQDKN	VAI	KLYDADM	TEK	YANYGCDKNC	EYVFNP	TYDYG	NVRYIG	VMSM	
201	AS	GEAEY	PYG	ITIDEVV	V	QTEAPKNEE		TIVLEG	DALK	FTEAWEG	TEV
251	DG	ATLEFD	KE	WREYRV	SLGK	TLP	GP	VDKSV		KVTF	FAEANKQ
301	TEL	KADY	GKN	GSKSYTT	YPN	VTADVDA	IAI	MAMNEQ	TYPF	SVTVEK	IEVV
351	VD	TPAED	RP	QAGVEY	DIVD	LRKPVAE	LMG	DDFIIG	TAAS	YDEFGD	PLDM
401	EL	VYK	HFNGV	TLGNELK	PKDS	MLR	KDAE	IVD	YELNGE	MVPF	PVLD
451	RL	DRFV	KWNE	EHPEKKI	QVR	GHVLVW	HSQT	PSFFFH	EDYD	TSKPYV	TPDV
501	MN	K RL	EIYIR	EVAKHFT	AED	SKYKDL	FY	GW	DV	VNEAVSDG	TGTYRNASER
551	SE	WGW	VYGSQ	EFITNA	FVYA	NRYPAD	IAL	FYNDY	NETVS	SKMGGIC	QLL
601	RD	VKATP	GAR	IDGMGMQ	AHY	QIASNNPSME		QFKTA	AAKAYA	AI	VDQVQVTE
651	LDFK	GATS	SAK	DDRLAER	YKA	VYDTIRRLRE		DGVNFT	GMTI	WGVTDK	HSWL
701	QSS	NNNGG	GA	DGSAR	QYPLL	FDDYYKAKDC		FWAIA	NAGEL	EPEVKS	ITLV
751	QNV	NNDFS	SAG	NVYEFV	AGKF	IPMWS	ENGID	VKVVIK	DTT	QDD	FFTVYA
801	DD	GTGIK	SVV	VKREEA	TENE	NGYEA	VNV	VD	FEALT	TANK	VKFDIV
851	DF	LAV	FNDTT	FKHAQSS	KFF	AETVIKPLAV		VK	KGTPVIDG	EFDDEAWKTA	
901	DEL	PVAI	KVG	AKASASA	KIL	WDDEYLYVLA		DVKD	SVLNKA	SSDAWEQDSV	
951	EI	FVD	NNNK	TSTYEADDKQ		YRINF	ENTHS	FNGTK	CEVEN	IKSEVV	VTTED
1001	GK	IEA	AALKW	TDIAP	VEGAK	VGLDIQV	NDA	DNSGKR	VGTL	NWADKTGN	GW
1051	SST	EVFG	TIL	LADGSD	VPTP	EVPTPE	TPEE	PEGTL	VTKYG	KTYVVT	TANGE
1101	KL	TGLQ	KVDG	NMYYCK	ADGK	VSKNAL	FTFE	GAVY	CAGADG	VLARNVKLEK	
1151	YHNEYFFGDD			CKAKTGFVTY		DGQDYFCKDN		GRVAKDYMIT		VGDKKYYAKK	
1201	DG	TLAKNE	QI	VKLFR	RYTFD	ENGVLIKTEK		GIF			

C Match to: Bpr_I0026 Score: 124 Expect: 1.5e-009
Bpr_I0026 endo-1,4-beta-xylanase Xyn10B 33228:36929 forward MW:136888

1	MH	KRWLG	GISS	FVMAGLLT	TVT	NVGTGVTQ	VF	AAEAATG	SSSL	VAEVEEAA	VAE
51	ES	SAETS	SVED	AADEASSE	AS	VAEEATEE	SS	DESAKEAT	DE	ASAETSAE	PA
101	EE	ENSLV	FFN	RWNQSAK	TSG	KLVSFKQ	YEE	YCYDLG	ETVP	VANVKALT	TVK
151	VS	KQDKN	VAI	KLYDADM	TEK	YANYGCDKNC	EYVFNP	TYDYG	NVRYIG	VMSM	
201	AS	GEAEY	PYG	ITIDEVV	V	QTEAPKNEE		TIVLEG	DALK	FTEAWEG	TEV
251	DG	ATLEFD	KE	WREYRV	SLGK	TLP	GP	VDKSV		KVTF	FAEANKQ
301	TEL	KADY	GKN	GSKSYTT	YPN	VTADVDA	IAI	MAMNEQ	TYPF	SVTVEK	IEVV
351	VD	TPAED	RP	QAGVEY	DIVD	LRKPVAE	LMG	DDFIIG	TAAS	YDEFGD	PLDM
401	EL	VYK	HFNGV	TLGNELK	PKDS	MLR	KDAE	IVD	YELNGE	MVPF	PVLD
451	RL	DRFV	KWNE	EHPEKKI	QVR	GHVLVW	HSQT	PSFFFH	EDYD	TSKPYV	TPDV
501	MN	K RL	EIYIR	EVAKHFT	AED	SKYKDL	FY	GW	DV	VNEAVSDG	TGTYRNASER
551	SE	WGW	VYGSQ	EFITNA	FVYA	NRYPAD	IAL	FYNDY	NETVS	SKMGGIC	QLL
601	RD	VKATP	GAR	IDGMGMQ	AHY	QIASNNPSME		QFKTA	AAKAYA	AI	VDQVQVTE
651	LDFK	GATS	SAK	DDRLAER	YKA	VYDTIRRLRE		DGVNFT	GMTI	WGVTDK	HSWL
701	QSS	NNNGG	GA	DGSAR	QYPLL	FDDYYKAKDC		FWAIA	NAGEL	EPEVKS	ITLV
751	QNV	NNDFS	SAG	NVYEFV	AGKF	IPMWS	ENGID	VKVVIK	DTT	QDD	FFTVYA
801	DD	GTGIK	SVV	VKREEA	TENE	NGYEA	VNV	VD	FEALT	TANK	VKFDIV
851	DF	LAV	FNDTT	FKHAQSS	KFF	AETVIKPLAV		VK	KGTPVIDG	EFDDEAWKTA	
901	DEL	PVAI	KVG	AKASASA	KIL	WDDEYLYVLA		DVKD	SVLNKA	SSDAWEQDSV	
951	EI	FVD	NNNK	TSTYEADDKQ		YRINF	ENTHS	FNGTK	CEVEN	IKSEVV	VTTED
1001	GK	IEA	AALKW	TDIAP	VEGAK	VGLDIQV	NDA	DNSGKR	VGTL	NWADKTGN	GW
1051	SST	EVFG	TIL	LADGSD	VPTP	EVPTPE	TPEE	PEGTL	VTKYG	KTYVVT	TANGE
1101	KL	TGLQ	KVDG	NMYYCK	ADGK	VSKNAL	FTFE	GAVY	CAGADG	VLARNVKLEK	
1151	YHNEYFFGDD			CKAKTGFVTY		DGQDYFCKDN		GRVAKDYMIT		VGDKKYYAKK	
1201	DG	TLAKNE	QI	VKLFR	RYTFD	ENGVLIKTEK		GIF			

Figure 3.9. MASCOT summary of peptide sequence coverage of protein extracted from spot 3 (A), spot 4 (B), and spot 5 (C). Each was identified as the endoxylanase Xyn10B.



Figure 3.10. MASCOT summary of MS/MS Xyn10B identification (A) and peptide sequence coverage (B) after analysis of protein extracted from spot 4.

Xyn10B is between 40% and 44% identical to several uncharacterised gene products produced by *Roseburia* and *Butyrivibrio* isolates. The GH10 and CBM9 domains in Xyn10B also display good homology to corresponding domains found in XynA of *Eubacterium ruminantium* (Q47871_9FIRM) (42% and 57% respectively), and in the XynA precursor of *Thermoanaerobacterium thermosulfurigenes* (Q60046_THETU) (32% and 32% respectively), both of which have been experimentally examined. The GH10 consensus sequence (Henrissat and Bairoch, 1993) containing the conserved catalytic residues are present at positions E535 and

E650 in Xyn10B (Figure 3.11 and Figure 3.12, denoted by “*”), as are the highly conserved aromatic residues and residues involved in hydrogen bonding (Figure 3.13, denoted by “#” and “+” respectively) that mediate CBM9 attachment to cellulose (Notenboom *et al.*, 2001b).

During mid-log phase, the 252 kDa xylosidase/arabinofuranosidase Xsa43J was also detected in the secreted protein recovered from xylan-grown cells only (Figure 3.4, spot 8). In contrast, at stationary phase Xsa43J was detected at equal intensity in the culture medium harvested from cells grown on both substrates (comparison of Figure 3.6 and Figure 3.7). The 2-DE position of Xsa43J corresponded well with the theoretical *pI* and size of the enzyme and the masses of several clearly resolved peptides were matched to the N- and C-terminal regions of the mature protein sequence (Figure 3.14), which together indicate the excised spot represented the full-length protein species.

Three polysaccharidases with predicted hydrolytic activities towards cellulose or pectin were identified in the culture supernatant of xylan- or xylose-grown cells. Cel5C is predicted to be a 62 kDa endocellulase that contains a GH5 endoglucanase catalytic domain located between residues 55 and 363, and a C-terminal CBM2a domain beginning at residue 456 (Figure 3.25). At both time points Cel5C was detected in two spots located at approximately 60 kDa and 52 kDa (Figure 3.4-Figure 3.7, spot 1 and spot 2 respectively). Spot 1 was detected at equal abundance during mid-log phase, but during stationary phase was significantly less abundant in the secreted protein of xylan-grown cells. Furthermore, at both time points spot 2 was more than 12-fold less abundant in the secreted protein recovered from xylan-grown cells. Peptide sequence coverage of the two Cel5C protein forms confirmed in both cases the presence of several contiguous peptides spanning the catalytic domain (Appendix C, Figure C.1). One peptide derived from the full length protein which was not present in the lower mass form matched to the N-terminal end of the CBM2a domain. As no tryptic peptides were matched to either the N- or C-terminus, it difficult to predict where a protein truncation might occur in order to account for the presence of the lower molecular weight spot.

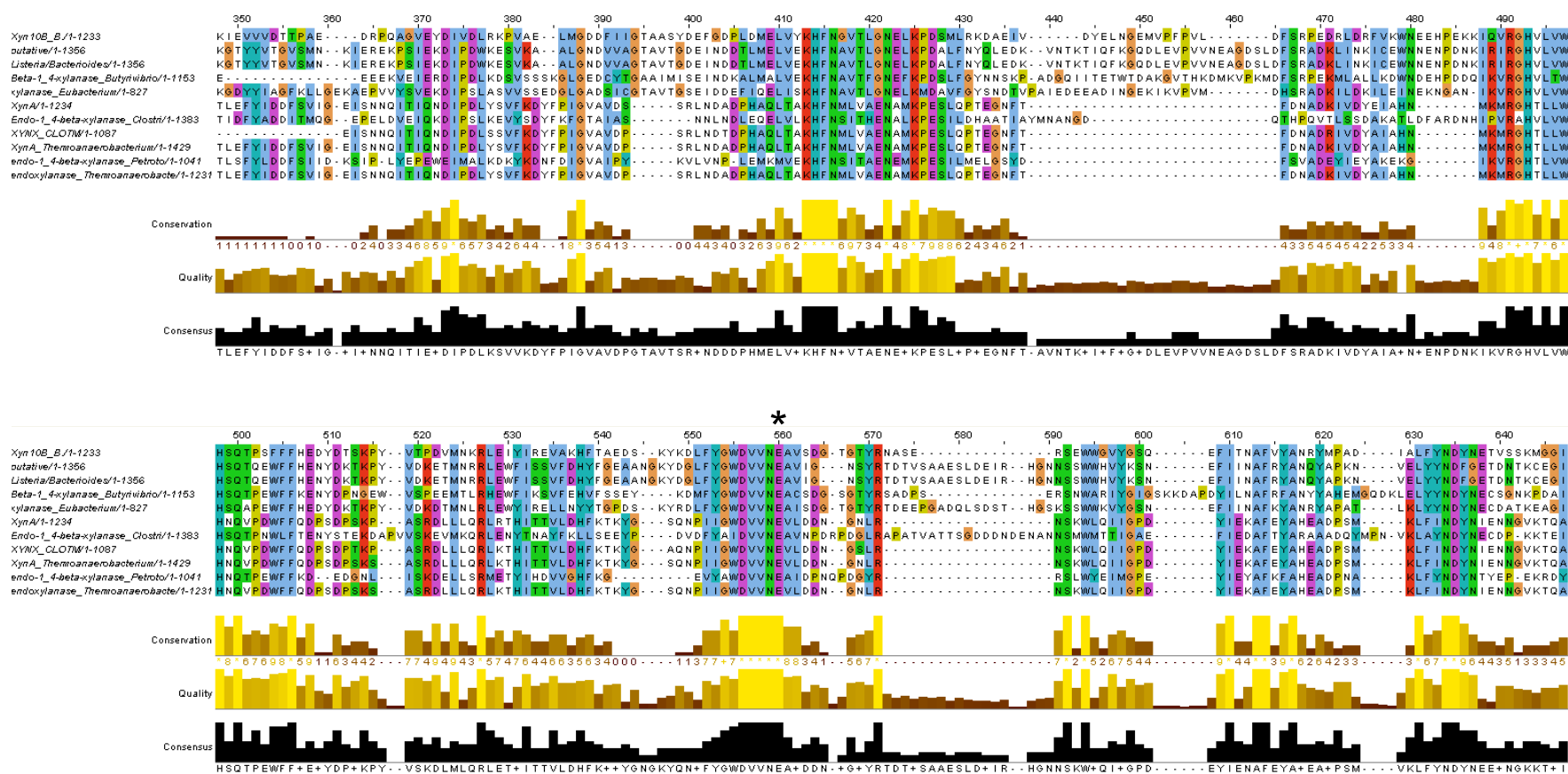


Figure 3.11. CLUSTALW multiple sequence alignment of the GH10 domain (N- terminal and central regions) in *B. proteoclasticus* Xyn10B with the top 10 ranking BLASTp homologues. Asterisk denotes the catalytic glutamate residues within the GH10 consensus sequence. Proteins are arranged by BLASTp expectation score.

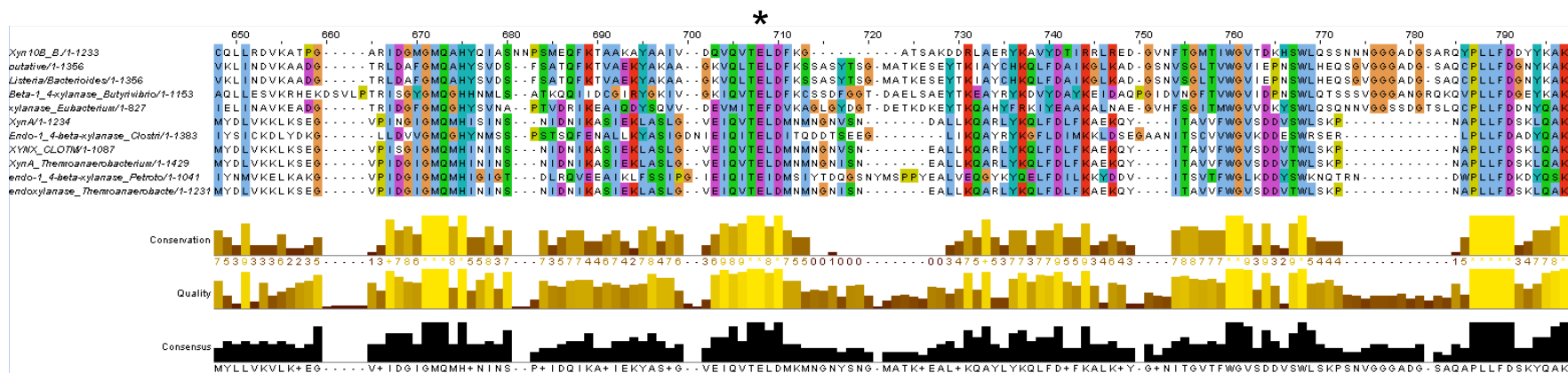


Figure 3.12. CLUSTALW multiple sequence alignment of the GH10 domain (C-terminal region) containing region of *B. proteoclasticus* Xyn10B with the top 10 ranking BLASTp homologues. Asterisk denotes the catalytic glutamate residues within the GH10 consensus sequence. Proteins are arranged by BLASTp expectation score.

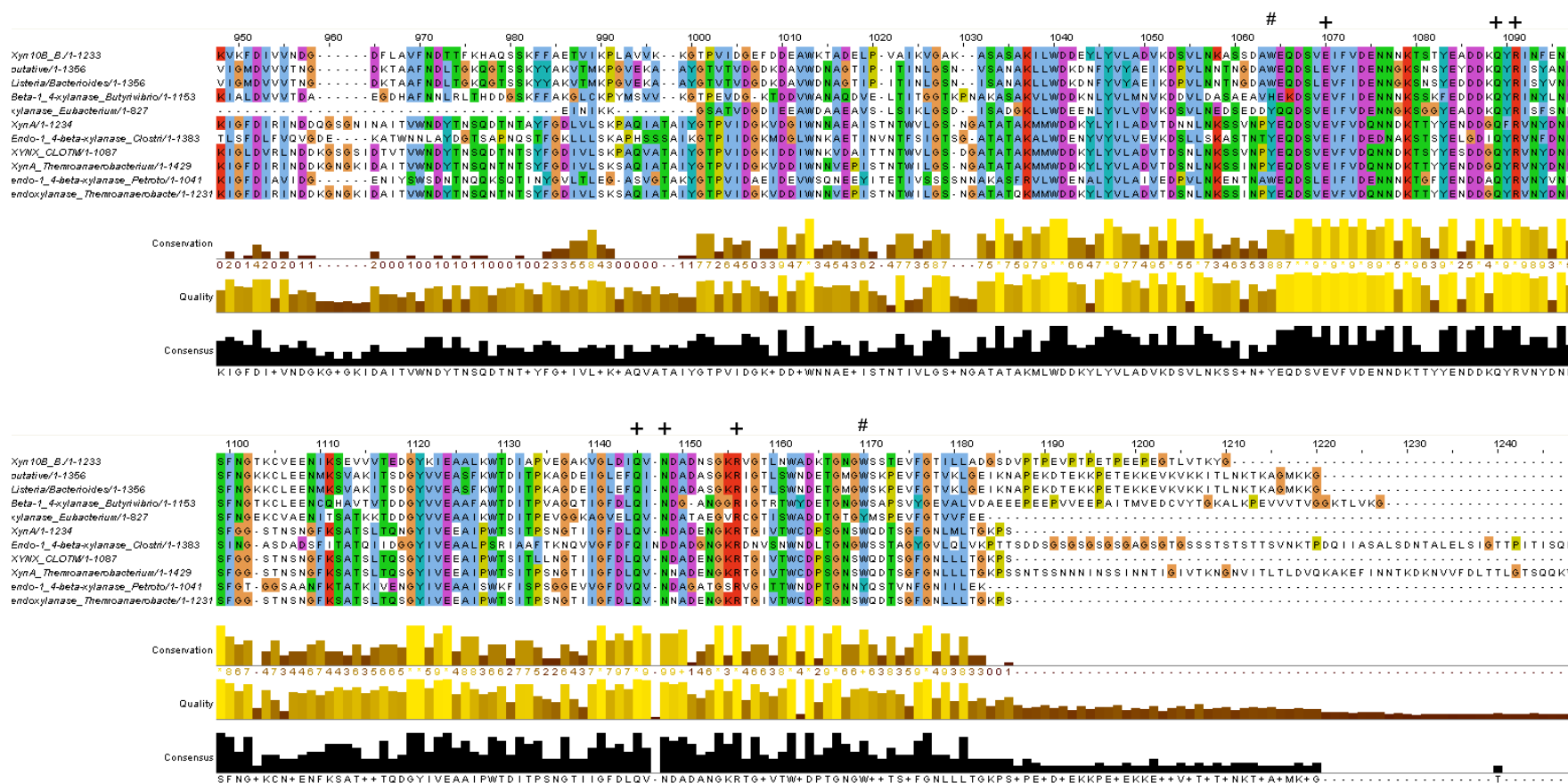
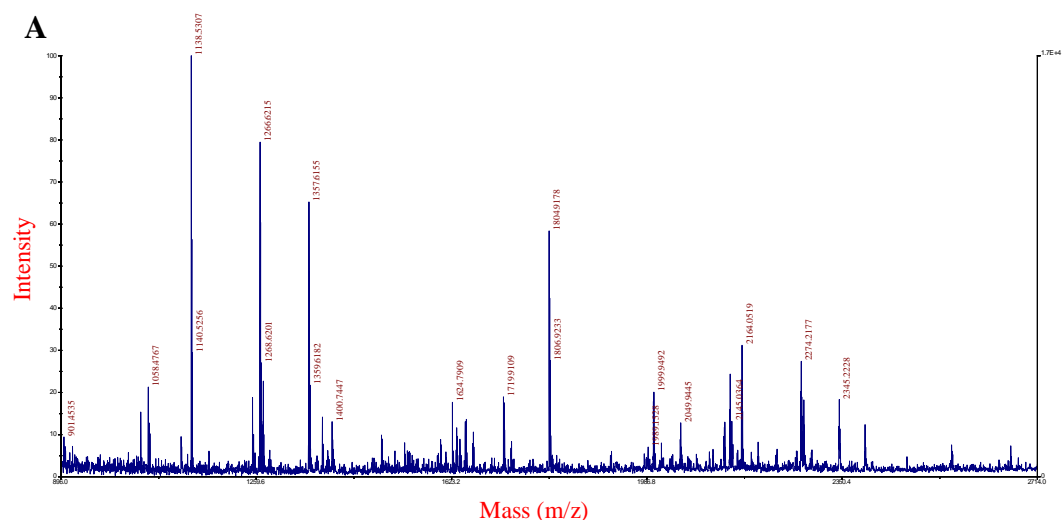


Figure 3.13. CLUSTALW multiple sequence alignment of the CBM9 domain containing region of *B. proteoclasticus* Xyn10B with the top 10 ranking BLASTp homologues. “#” and “+” symbols denote the conserved aromatic residues and residues involved in hydrogen bonding respectively. Proteins are arranged by BLASTp expectation score.



B Bpr_I2935 xylosidase/arabinofuranosidase Xsa43J 3522507:3529361 reverse MW:251937

Nominal mass (M_r): 252464; Calculated pI value: 4.46

NCBI BLAST search of [Bpr_I2935](#) against nr

Unformatted [sequence string](#) for pasting into other applications

Fixed modifications: Carbamidomethyl (C)

Variable modifications: Oxidation (M)

Cleavage by Trypsin: cuts C-term side of KR unless next residue is P

Number of mass values searched: 51

Number of mass values matched: 29

Sequence Coverage: 15%

Matched peptides shown in **Bold Red**

```

1 MQRNWRRVLA SVLGVTITFT TAFISGSAAT VTTQAAEGKV DRTVKLSPAI
51 ASVFNDITDGD GFGEFQSGFT SLCWWANRVG YSDELTKQSA TAFDKEEGL
101 GMTIGRYNVG GGDNVGESPE VPVNEKAYFY DTSSENLSYS GSKMSVSTNT
151 SLADASYSKS DADFGLTSGQ KVSFNAIGW INNLDGEIGS GDNLHYLINS
201 EEDGTYTIKM LFTLSGTNKR GVSIKVTSED MTTYSHENKE AGTEGEVVVS
251 GDDLEGDSKK DERAASEATS EEKADEAVT EASKEKDEVS EDSSDRASTA
301 SSEDISEEEET KEETTDEAAS EDSVNPVTDK ESATDKLLAV TNGESKVYTV
351 PSSEVNKNLV ASGNNNMLYV VTFSDVELKA GVNKIDIGGA DGDWCLDFVK
401 MAVIKSGEEG VLPSSDFFLH GAHAKRSDSI VPGYATDVTI INLEKHDAEY
451 YETNFVRADF DCGYANNYDW DADRNMQNVL IAAQAAGDE FGEAFSNSP
501 PYFMTVSGCS SGGVDPNVDN LREDSYTAFA CYMADVIEHW ADEGVINFQS
551 TTFMNEPYTN YWSAFNNKQE GCHFDIGESQ SRILVELNKE LEKKGIDILI
601 SASDEISIDI AITSFNALTD DAKEVVDRIID AHTYSGSNRA GLKALSESAG
651 KNLNMSEVDG TFTAAGNAGE MRAGLGFAQR IITDVNGMMP SAWIMWDAVD
701 VHIDSNNNEFD TATRESVERT LSNQDGFWGI AIADHDTQEL LLSKKYAFG
751 QFSRYIRPGY AIIGSSDSTV AAYPDKEGKV VIVAVNTAEN DQTKWKNLQD
801 FSEIGSKVIA IRTSGSRADG ENWADVTASD DIVVDKATKN VTATLKANSI
851 TTVIVEGVVY DSSAESVAV DDKNIYAAKG TLPVLPETVS VKTSKNNNTIN
901 AKVTWDLNV DLATASEVTG TLDGYGYEVT YAIKYVDENM AWYIDCNDNE
951 VGHTSYRTMD SYADLFNEVS DQKYDGTWGR LADYGAVNNS DDDPNAYGWY
1001 AYSNQSIDYT VPLEAGNYTL TFGFKENWNQ SRPMSIYAVT DGGETLIGNT
1051 NTKNGNNWN TPVYSYSQEE AGNVTFSVRK NGGPDVPLSF IRIQKNLND
1101 ELKSVLAKAQ VMDTSDYSES KSARLHSLFE RAQRLMLRAA TTQEDVDLLA
1151 KEISEFVETD GNVFSEEEIA ANDVLYLVLD VASKETTQVE EGYKLGLYQS
1201 VTDQEDAVDA GTGLKWGYGT DPSYGVVRNG GASDGTLTGT YAYMSDKGWT
1251 YERGVSGLYY SFELPDRSNN EYLVITGVKS PWSARDISVD LEGVNVESGL
1301 NLQGSILVER TYNVEVTDGT LNLYAYATNR TSSYVDPBLS YVIVKAVPEY
1351 TYETLISAIN KYRAEMEGKV YSEKSLAVFE EAVVKAQALI DVKSTDVAAI
1401 KKAYKDLEEA FKALAIVVY SSITGVEGAP LYDNNGILVQ AHGGQIQQFT
1451 IDGVTKYYWY GEDKTYDYQP VVGVHLYTST DLYNWTDEGV FFRAPIDYNE
1501 IDGKFKKAGY KADLSIFEED KYFASLYSDY KDQAADDSQY DNKLEEVYWN
1551 IAADRTVMER PKVLYNDKIG KYVMWHCDG NTKSNPTGSN YGKAKAGVAI
1601 SDSPFGPFKF LGAYKLNYSY TADHQWDSDE SAWGSVRDMN VFKDDDGTAI
1651 VMYSSDGNIN MYIAKLNDY TYLAKDQKHA VLGEDFTLNF AGASREAPAM
1701 FKYNGIYYMI TSGCTGWDPN PASYAYASSP LGFWITVNNF CTDDGANTTY
1751 RTQSTCVFFV DAAAGKFIYM GDRWNSGDL ESRYVWLPVE FLAGNKIALR
1801 SYSNWILDDEL ENKGVFEINT KLFEYALSKE DLEEKLPSEV ELTLDGSTV
1851 TKAVTWNTHS ARLVGDVEIT GTLADFNRSF AITVSMIPEK MIYFYDSASR
1901 NVLGDEEASY LTSARSVLKK QLRNASADE FATTKVSGYR GVRSSENAER
1951 YDVGFKNKSGS DINGHGFMAA GNKPIDYVFT LEEGQYIVAT GYQEWNTSR
2001 PTRITVTDAS GKELARQNT LGSSDSARLE TVSFELTEDI QVTVRVAKTG
2051 NPDPVLSFIA VIKDNLEKEP EDEGTLIKKY GKYYLVTEDE EKLTFGEVD
2101 GILRYFDENT GVMANKWVT VGDNKYRALD EGRIASNEII REYGSDDYLG
2151 EDGVLVTGLF DYNGDKYYAK ANGKIVKSGL LEIDGYYIIP DSGRLMHDI
2201 KTEVYFSEYI LGSDCKAVKG FTTYEGQRY GKANGRVAKD YNFTVDSGTY
2251 YAKNDGTAV SETITRYFKK YTFDENGKLI SVGK

```

Figure 3.14. MS identification of Xsa43J in the *B. proteoclasticus* culture medium. MALDI-TOF peptide mass fingerprints of protein extracted from spot 8 (A) and MASCOT peptide sequence coverage summary of Xsa43J identification (B).

During stationary phase only, two pectinolytic enzymes secreted by cells grown in the presence of xylan or xylose were detected at low-abundance. The 294 kDa pectin methylesterase Pme8B was identified in spot 7 (Figure 3.6 and Figure 3.7) and appeared to be the full length form of the constitutively expressed enzyme (Appendix C, Figure C.2). In contrast, spot 6 was detected only in xylose-grown, stationary phase cells, and was matched to the pectate lyase, Pel1A, which contains a PL family 1 catalytic domain (Figure 3.25). The 2-DE spot position of spot 6 was approximately 45 kDa smaller than the predicted size of the full-length size of the enzyme. The PMF sequence coverage showed that all 20 peptide masses contributing to the MASCOT identification were matched to the N-terminal half of the protein sequence (Appendix C, Figure C.3), which suggests the 2-DE detected protein was an N-terminal fragment.

The *B. proteoclasticus* genome encodes two secreted CBPs that lack any identifiable catalytic domains. Both proteins were identified in the culture medium of xylan and xylose grown cells at equivalent levels. Protein spot 9 and spot 10 (Figure 3.4-Figure 3.7) were matched to the predicted products of loci Bpr_I0736 and Bpr_I1599, respectively, and the abundance of both proteins was unaffected by growth substrate. The product of locus Bpr_I1599 in particular was detected at low relative abundance at both time points.

3.2.3.2 Abundance patterns of non-carbohydrate degrading secreted proteins

Several proteins with predicted functions other than polysaccharide degradation exhibited differential abundance patterns between the culture supernatants of xylan and xylose-grown cells. An ABC transporter SBP (Bpr_I0182) was the most abundant protein present in both the mid-log and stationary phase harvested xylan-grown culture medium where it constituted 8.1% and 17% of the total detectable protein respectively (Figure 3.4 and Figure 3.6, spot 25). At mid-log and stationary phase growth, the SBP was 7.2-fold and 5.4-fold more abundant in xylan-grown cells respectively, compared to the secreted protein recovered from cells grown on xylose. Bpr_I0182 is most similar to a family 1, extracellular solute-binding protein (SBPbac1, PF01547) secreted by *Geobacillus* sp. Y412MC10 (Table 3.5), and also shows similarity to several other SBPbac1 proteins produced by Gram-positive bacteria. Protein spot 24 also contained the product of Bpr_I0182 and was detected only in xylan-grown cells harvested at both mid-log and stationary phases of growth. During stationary phase, spot 23 was more abundant in the culture medium of xylan grown cells, and contained three ABC transporter SBPs, two of which also contained SBPbac1 domains

(Bpr_I0313 and Bpr_I2344) (Table 3.2). Several Peripla_BP_1 (PF00532) sugar binding proteins were significantly more abundant in xylose-grown cells at both mid-log and stationary phases of growth. Consistent with the culture conditions used, protein extracted from the spot 31 and 32 (Figure 3.5 and Figure 3.7) was identified as ABC transporter sugar binding proteins specific for simple sugars (RbsB) (Bpr_III244) and xylose (Bpr_I1173) respectively (Table 3.2). During stationary phase, *B. proteoclasticus* cells grown in the presence of xylose also secreted an abundant 32 kDa amino acid ABC transporter substrate-binding protein (Bpr_I2466) that was undetectable in the secreted protein recovered from cells grown on xylan (Figure 3.7, spot 13). This protein was similar to polar amino acid substrate-binding proteins produced by *Clostridial* species, and may be involved in glutamine transport.

The single subtilisin-family serine protease (EC3.4.1.62) encoded by the *B. proteoclasticus* genome was identified from protein spot 22, and was 33% identical to a peptidase synthesised by *Thermoanaerobacter* sp. X514. No detectable peptide masses were matched to the first 250 N-terminal residues, which contains 12 potential tryptic peptides larger than 800 kDa, indicating a portion of the N-terminal region may be missing from the excised protein. The protease was detected at a slightly greater abundance in xylose-grown cells at both time points and was considerably more abundant during stationary phase growth.

3.2.4 Gel-based analysis of the *B. proteoclasticus* cell-associated proteome

3.2.4.1 Optimisation of *B. proteoclasticus* cell-associated protein extraction

Five buffers were tested for *B. proteoclasticus* cell-associated protein extraction. Freshly harvested mid-log phase, xylan grown cells (cell mass pelleted from 50 mL of culture medium) were resuspended in 2.6 mL of one of the five buffers and incubated at 4°C for 30 min. After TCA/acetone precipitation of the extracted protein, 50 mg of each purified sample was analysed by 1-DE, except for the 0.2M glycine extract, where the total sample volume was analysed (Figure 3.15).

A 20 mM Tris-HCl buffer containing 1% SDS has been demonstrated to be effective for the extraction of S-layer proteins from the cell surface of the Gram-positive anaerobe *Peptostreptococcus anaerobius* (Kotiranta *et al.*, 1995). CHAPS detergent causes less cell membrane disruption than SDS, and is more compatible with 2-DE. As a consequence, the 2% CHAPS in ice-cold 50 mM Tris-HCl buffer (pH 7.0) with Complete[®] Protease Inhibitor was used as the benchmark extraction buffer for this analysis. The 200 mM Glycine (pH 2.2) buffer extracted no detectable protein.

The remaining four buffers extracted approximately equal amounts of total protein, although quantitative differences were evident between the protein profiles of each buffer extract.

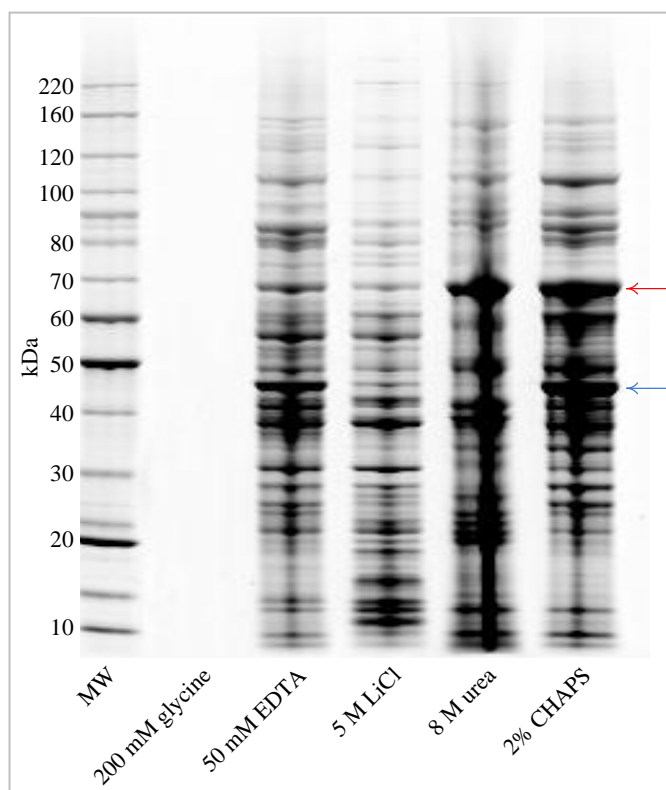


Figure 3.15. 1-DE analysis of the effect of buffers used for *B. proteoclasticus* cell-associated protein extraction. 50 μ g of extracted protein was loaded per lane, except for the 200 mM glycine (pH 2.2) buffer, where the total extraction was loaded. The gel was stained with CBB-G250. The red arrow indicates the most abundant protein in the 2% CHAPS extract. Blue arrow indicates the most abundant protein in the 50 mM EDTA buffer extract.

The most abundant protein in the 2% CHAPS and 8M urea extracts (red arrow) was weakly detected in the 50 mM EDTA and 5 M LiCl extracts. Conversely, the most abundant protein in the 50mM EDTA extract (blue arrow) was also highly abundant in the 2% CHAPS extract, but only weakly detectable in the 5 M LiCl and 8 M urea extracts. Several low molecular weight proteins were efficiently extracted by 5 M LiCl that were only weakly detectable in all other extracts. In spite of these differences, a marked improvement over the 2% CHAPS buffer extract could not be achieved with any of the other buffers tested. Furthermore, 2% CHAPS extracted several highly abundant proteins that were not as abundant in the other buffers tested, which

suggested that these may be cell-associated proteins. Consequently, the 2% CHAPS buffer was selected as the most appropriate buffer for routine extraction of *B. proteoclasticus* cell-associated proteins.

3.2.4.2 2-DE of proteins recovered from the *B. proteoclasticus* cell-associated extract

Narrow range 2-DE separations of the proteins harvested by 2% CHAPS extraction of cells grown in the xylan- (Figure 3.16), xylose- (Figure 3.17), and no added carbon (Figure 3.18) growth conditions were used in an attempt to identify cell-associated *B. proteoclasticus* proteins. One hundred and sixty two protein spots were excised from the mid-log and stationary phase separations covering the pI 3-5.6 and pI 5.3-6.5 ranges, in-gel trypsin digested, and analysed by MALDI-TOF MS or LC-MS/MS. Each tryptic-PMF or LC-MS/MS data file was searched against the *B. proteoclasticus* protein sequence database using the MASCOT search algorithm (described in Section 2.12.3). Proteins were identified in 118 spots, which were the products of 61 genes (Appendix B, Table B.4). This set of 61 non-redundant gene products contained 17 proteins that were associated with carbohydrate metabolism, and 10 with the transport of polysaccharide degradation products. Three polysaccharide-degrading enzymes were identified in five spots in the 2% CHAPS extraction of *B. proteoclasticus* cells grown in the presence of xylan (Figure 3.16 and Table 3.3, above the solid dividing line), and no additional polysaccharidases were identified in the alternate growth conditions.

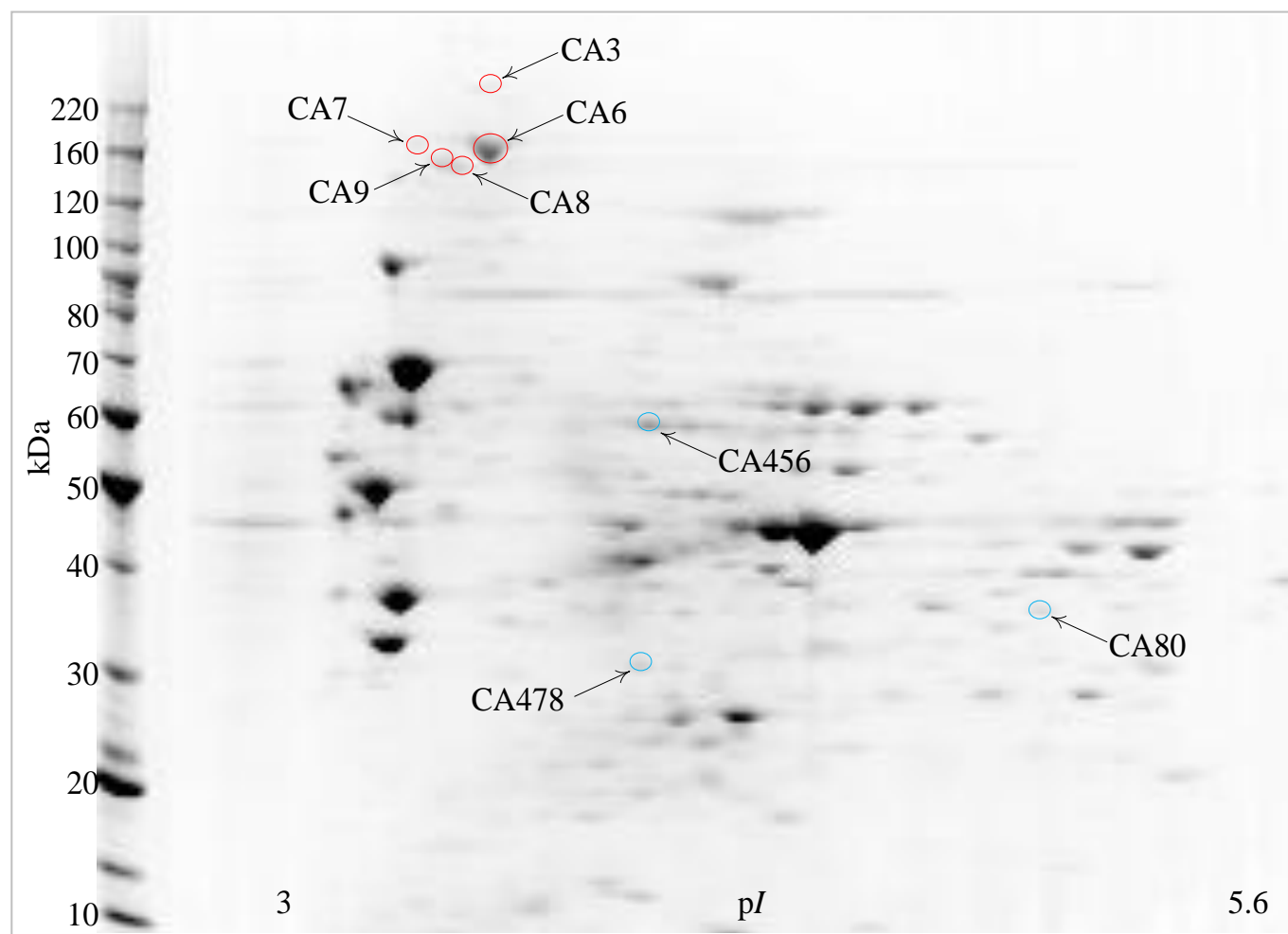


Figure 3.16. Polysaccharide degrading enzymes and proteins uniquely identified in the *B. proteoclasticus* cell-associated fraction by pI 3-5.6 2-DE/MALDI-TOF MS. Polysaccharidases, and proteins uniquely identified in the cell-associated fraction are circled in red and blue respectively.

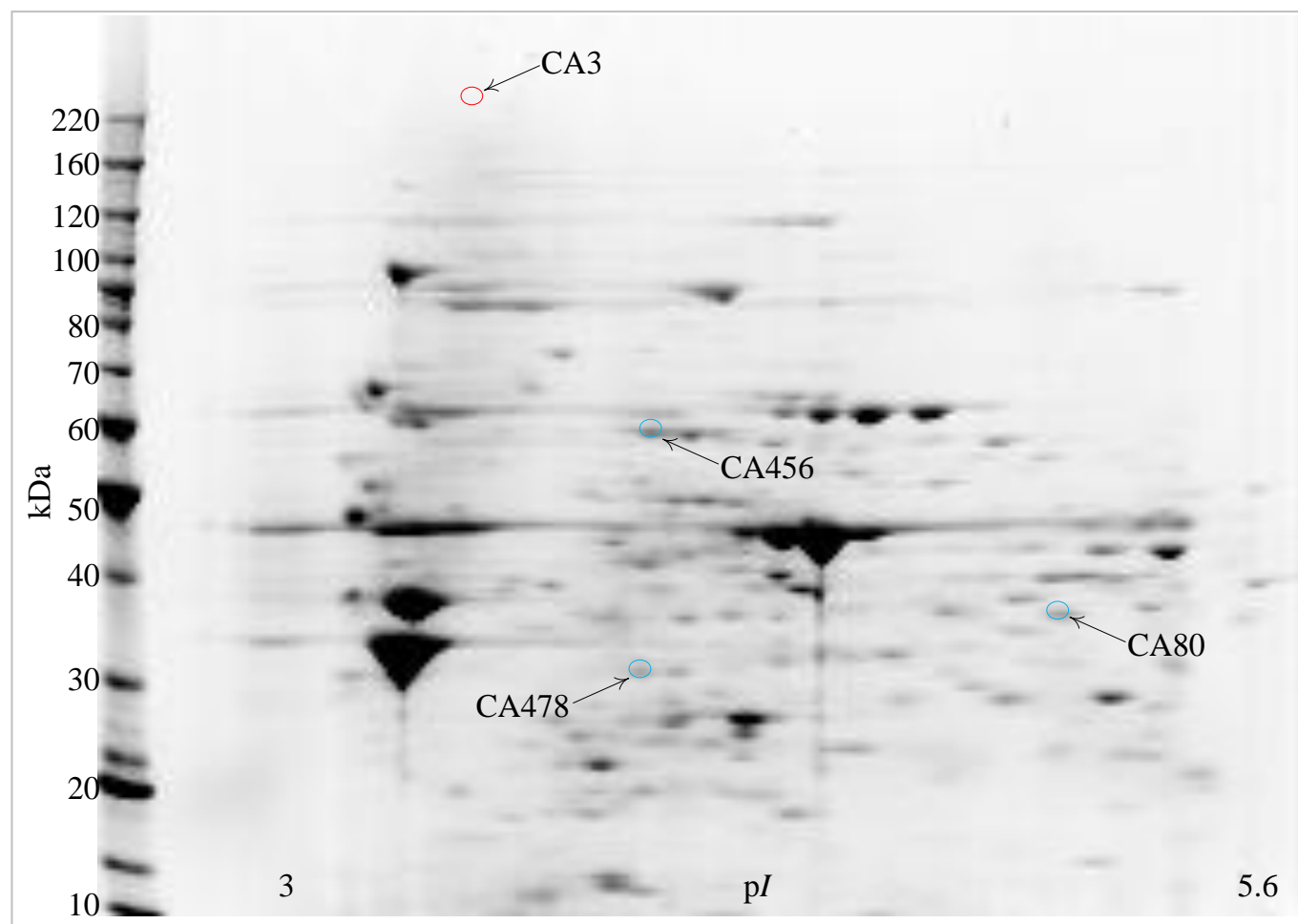


Figure 3.17. Colloidal CBB G-250 stained 2-DE separation of proteins harvested by 2% CHAPS extraction of mid-log phase, xylose grown cells. Polysaccharidases, and proteins uniquely identified in the cell-associated fraction are circled in red and blue respectively.

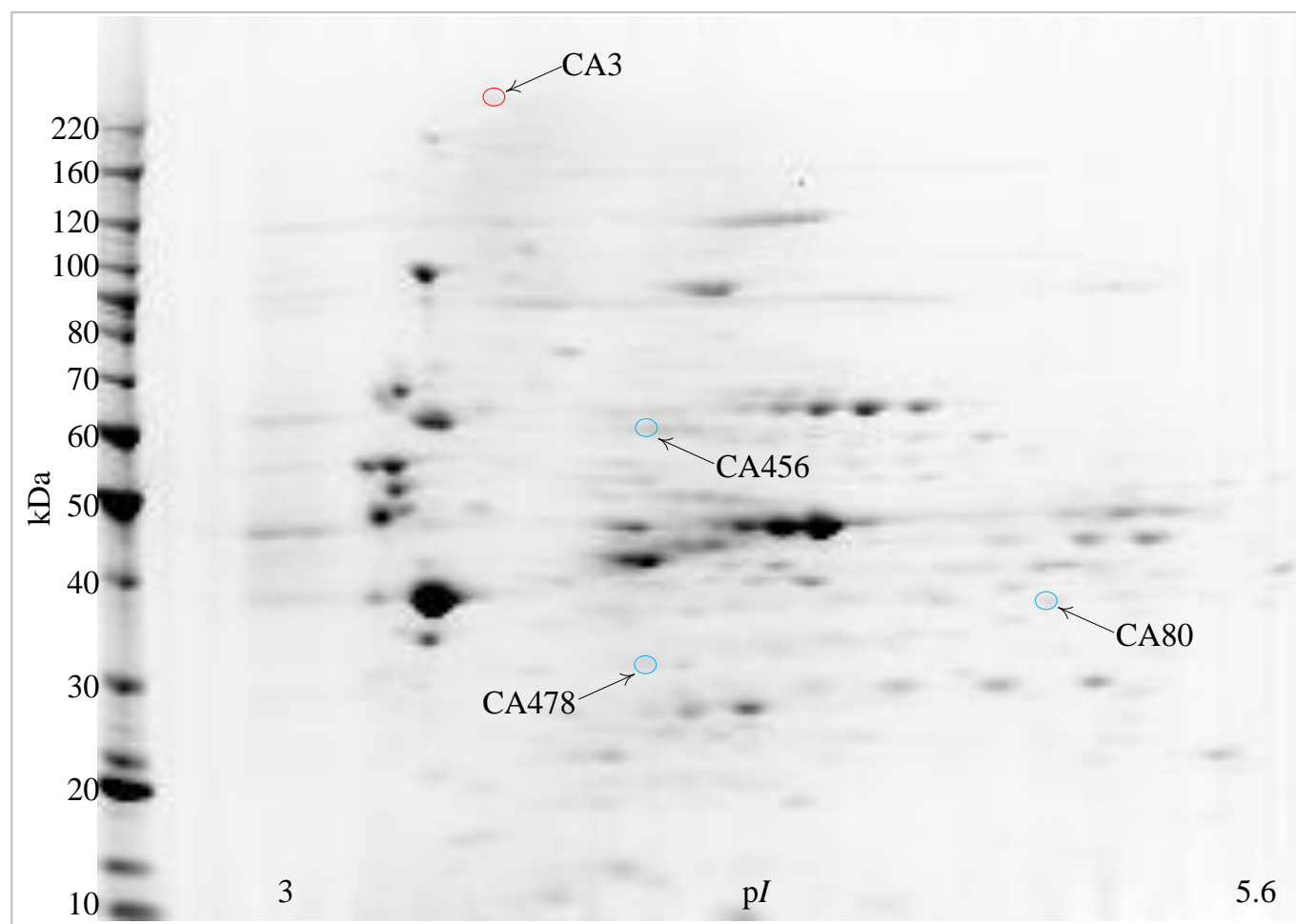


Figure 3.18. Colloidal CBB G-250 stained 2-DE separation of proteins harvested by 2% CHAPS extraction of mid-log phase, no added carbon grown cells. Polysaccharidases, and proteins uniquely identified in the cell-associated fraction are circled in red and blue respectively.

The α -amylase Amy13A (Bpr_I1087) was identified in two protein spots in the *pI* 3-5.6 2-DE separation of proteins harvested by 2% CHAPS extraction of mid-log phase, xylan grown cells (Figure 3.16, spot CA7 and spot CA8). Neither spot was detected in the xylose or no added carbon growth conditions (Figure 3.17 and Figure 3.18 respectively), indicating that the abundance of Amy13A may be influenced by growth substrate. Protein spot CA7 (Figure 3.16) had apparent *pI* and size values of 3.8 and 160 kDa respectively, which are similar to the predicted *pI* and size values of Amy13A (Table 3.3). Spot CA8 had a more basic *pI* value and was approximately 10 kDa smaller. Manual analysis of the protein sequence coverage data from spot CA7 and spot CA8 (Figure 3.19 A and B) showed that a similar set of peptides extracted from the two Amy13A protein spots were matched to the full-length protein sequence, and none were matched to the N- or C-termini. This sequence coverage makes it difficult to predict the reason for the presence of spot CA8. Nonetheless, Amy13A was not identified in any growth condition or time point in the 2-DE analysis of protein present in the culture medium, and therefore adds to the set of polysaccharidases confirmed as being present in the *B. proteoclasticus* extracellular proteome.

Xyn10B and Xsa43J (Figure 3.16) were also identified in the *pI* 3-5.6 2-DE separation of proteins harvested by 2% CHAPS extraction of mid-log phase *B. proteoclasticus* cells. Xyn10B was detected in two spots in the separation of proteins harvested from xylan grown cells (Figure 3.16, spot CA6 and spot CA9), but could not be detected in either the xylose or no-added carbon growth conditions. Protein spot CA6 had an apparent size of 160 kDa, and spot CA9 was a lower abundant spot with a size approximately 10 kDa smaller and slightly more acidic. This Xyn10B spot distribution and relative abundance pattern mimics the pattern detected for Xyn10B in the 2-DE separations of proteins present in the culture medium. Examination of the PMF and protein sequence coverage of protein extracted from spot CA6 (Figure 3.20 A and B respectively) showed that 33 well resolved tryptic peptide masses were matched to a large region of the protein, and was similar to the sequence coverage pattern obtained for Xyn10B extracted from spot 3 in the secreted protein analysis (Figure 3.9 A). Furthermore, the sequence coverage pattern of Xyn10B extracted from spot CA9 (Figure 3.20 C) was also very similar to the coverage determined for the secreted spot 4 (Figure 3.9 B), including the absence of peptides matching to the C-terminus. This supports the hypothesis that the lower molecular

weight spot identified as Xyn10B is a C-terminally truncated fragment of the full-length enzyme.

Examination of the full set of proteins identified in the 2-DE separations of 2% CHAPS proteins extracts confirmed extensive contamination with cytosolic proteins (Appendix B, Table B.4). Despite numerous replicate experiments to ensure the cytosolic protein contamination was not an experimental error, the contamination of extracted cell-associated proteins could not be reduced. Quantitative analysis of relative protein abundance would be affected by sample contamination. For this reason, together with the fact that no non-carbohydrate active secreted proteins were uniquely detected in the *B. proteoclasticus* cell-associated proteome during the initial analysis, an examination of the differentially abundant secreted or cell-associated proteins based on the 2-DE separations of 2% CHAPS proteins extracts was not conducted. Furthermore, only three other proteins identified in the cell-associated extracts were not identified in the cytosolic proteome (Table 3.3, below the solid dividing line), and all three were cytosolic proteins. A rigorous examination of all cytosolic protein abundance changes was therefore performed using 2-DE separations of cytosolic protein extracts.

Table 3.3. Summary of the polysaccharidases and proteins uniquely identified in the *B. proteoclasticus* cell-associated fraction^a.

Spot	Protein	Locus	EC	PHX ^b	Score	SigP ^c	pI	kDa	Pep. ^a	Cov. ^a
CA7	α -Amylase, Amy13A	Bpr_I1087	3.2.1.1	Y	1.3e-05	SpI	4.1	139.1	12	14%
CA8	α -Amylase, Amy13A	Bpr_I1087	3.2.1.1	Y	3.0e-07	SpI	4.1	139.1	15	17%
CA6	Endo-1,4- β -xylanase, Xyn10B	Bpr_I0026	3.2.1.8	Y	4.8e-13	SpI	4.3	136.9	33	31%
CA9	Endo-1,4- β -xylanase, Xyn10B	Bpr_I0026	3.2.1.8	Y	1.5e-08	SpI	4.3	136.9	21	22%
CA3	Xylosidase/arabinofuranosidase, Xsa43J	Bpr_I2935	-	Y	1.9e-07	SpI	4.2	251.9	15	9%
CA478	DegV family protein	Bpr_I1462		N	1217	-	4.5	31.9	4	14%
CA456	L-ribulokinase, AraB	Bpr_I2815	2.7.1.16	Y	1.5e-11	-	4.6	58.2	15	29%
CA80	Oxidoreductase GFO/IDH/MOCA family	Bpr_I0422		N	6.5e-07	-	5.0	36.9	11	29%

^a Polysaccharidases, and proteins uniquely identified are shown above and below the solid line respectively.

^b PHX, predicted highly expressed gene; Pep, the number of matched peptides masses/searched peptides masses; Cov, Protein sequence coverage of the matched peptides.

^c Secretory signal-peptides were predicted using SignalP (Ver. 3.0) (www.cbs.dtu.dk/services/SignalP/) and LipoP (Ver. 1.0) (<http://www.cbs.dtu.dk/services/LipoP/>).

A Match to: Bpr_I1087 Score: 85 Expect: 1.3e-005
Bpr_I1087 alpha-amylase Amy13A 1335030:1338818 forward MW:139111

Nominal mass (M_r): 139424; Calculated pI value: 4.34
NCBI BLAST search of [Bpr_I1087](#) against nr
Unformatted [sequence string](#) for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 29
Number of mass values matched: 12
Sequence Coverage: 14%

Matched peptides shown in **Bold Red**

```

1 MKRGKLGWRL VSAAGLSLSI FLSSIGNVST AYAMESNDAL VLDETKENTE
51 SATDASSNEA SDAEADNDTD EAITDASSKE LSAENDGASE SDSSFDEYDH
101 TALPETDEIT VTAAGELSTA KAELYTLAPR EAREADNSLV TRDSIHDGAI
151 LHAFCSWFNT IADNMADIAD AGYTAVQTSP INECLSTNPG MNLHGPDGMW
201 YYHYQPTDWV IGNYQLGSRD EFKHMCVDAD EYGVAVIVDI LPNHTTPSTG
251 NIAKALKEAA GGSEALYHTT GKIGGGYTDR LELTYYSMGG LPDVTENTG
301 FQQYFYEFLLK DCVYLGADGF RIDTAKHISL PDDPVPSDYS DAGRNTFYPN
351 MREALNDYSE EVGTRSYDEL FVYGEVLQGT NDRLAAYQQY IGGTTASNYG
401 SSLRSALSSG NLSVNRLLDY QIYDDTAYGS TYTADTEKLV TWVESHDNYM
451 NDESCEWKSI DDDMVIMGWS IIAARDAGTP LFFSRPNSS AENPYGDNLI
501 GAAGSPIYKA PEVKAVNLFR EKMGEAD EYL SNPGGNIQTL MIERYNDTVQ
551 GAIVVNAAQT RTTISTETHL SDGIYPDQVE GSNVFLVKD GVLSGSVEGE
601 GVVVLSEKID GTGKVISFYN NKNWNGVVAR VDNAEETLDT IDENDGWFPQV
651 TVLDDEFTIR FESADGKEVS PEFQITAESG TFATPDSSSEL YYSKAETEEG
701 LGHTYFVYF FNTENWGSVY TYGWLDGGAQ LFGGWPGTVA VNEGSGWYRA
751 DVKTTGEITA FNLIFNNGNG IQTVNVEGIT PDSKDIYLAV DAEKSNGLI
801 VNRYDEKESA EKALGVSGSY TTAYFYNTG WDKVCAYTWG ATALGDWPGK
851 ELTQDEEDGWY SVVLPAGPSE DLNIIFNNGN NGKQTNMCKI SDMKYRFILN
901 NGISYQKYGS KKDAMEAIAG AGDVTYETVY FYNEKADDAN WKNVYLYVFG
951 GTDGEYNNLV GTWPGKLMEK EEDSNWLR AE VPSKALES GT LTYIFNNGNG
1001 TQLDDNKNT STKNYFTFSS RDSFASKEEV YSFLGISTDD PSAPPEEPQEP
1051 EATLTKKYSK YYLVTEDGEK LTGFHEVDGI LRYFAENTGV MAINKVVTV
1101 DKRYRALDEG RIASNEIIRE YGSDYYLGED GVLVTGLFDY NGDKYYAKSN
1151 GKIVKSGLLE IDGEYIIPDS DGRMLHDVKT AIYFSEYILG SDCKAVKGFT
1201 TYEGQRYYGK ANGRVAKDYM FTVDGDYYA KNDGTLAVSE TITRFFKKYT
1251 FDENGKLISV GK

```

B Match to: Bpr_I1087 Score: 101 Expect: 3e-007
Bpr_I1087 alpha-amylase Amy13A 1335030:1338818 forward MW:139111

Nominal mass (M_r): 139424; Calculated pI value: 4.34
NCBI BLAST search of [Bpr_I1087](#) against nr
Unformatted [sequence string](#) for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: 38
Number of mass values matched: 15
Sequence Coverage: 17%

Matched peptides shown in **Bold Red**

```

1 MKRGKLGWRL VSAAGLSLSI FLSSIGNVST AYAMESNDAL VLDETKENTE
51 SATDASSNEA SDAEADNDTD EAITDASSKE LSAENDGASE SDSSFDEYDH
101 TALPETDEIT VTAAGELSTA KAELYTLAPR EAREADNSLV TRDSIHDGAI
151 LHAFCSWFNT IADNMADIAD AGYTAVQTSP INECLSTNPG MNLHGPDGMW
201 YYHYQPTDWV IGNYQLGSRD EFKHMCVDAD EYGVAVIVDI LPNHTTPSTG
251 NIAKALKEAA GGSEALYHTT GKIGGGYTDR LELTYYSMGG LPDVTENTG
301 FQQYFYEFLLK DCVYLGADGF RIDTAKHISL PDDPVPSDYS DAGRNTFYPN
351 MREALNDYSE EVGTRSYDEL FVYGEVLQGT NDRLAAYQQY IGGTTASNYG
401 SSLRSALSSG NLSVNRLLDY QIYDDTAYGS TYTADTEKLV TWVESHDNYM
451 NDESCEWKSI DDDMVIMGWS IIAARDAGTP LFFSRPNSS AENPYGDNLI
501 GAAGSPIYKA PEVKAVNLFR EKMGEAD EYL SNPGGNIQTL MIERYNDTVQ
551 GAIVVNAAQT RTTISTETHL SDGIYPDQVE GSNAVFLVKD GVLSGSVEGE
601 GVVVLSEKID GTGKVISFYN NKNWNGVVAR VDNAEETLDT IDENDGWFPQV
651 TVLDDEFTIR FESADGKEVS PEFQITAESG TFATPDSSSEL YYSKAETEEG
701 LGHTYFVYF FNTENWGSVY TYGWLDGGAQ LFGGWPGTVA VNEGSGWYRA
751 DVKTTEGITA FNLIFNNGNG IQTVNVEGIT PDSKDIYLAV DAEKSNGLI
801 VNRYDEKESA EKALGVSGSY TTAYFYNTG WDKVCAYTWG ATALGDWPGK
851 ELTQDEEDGWY SVVLPAGPSE DLNIIFNNGN NGKQTNMCKI SDMKYRFILN
901 NGISYQKYGS KKDAMEAIAG AGDVTYETVY FYNEKADDAN WKNVYLYVFG
951 GTDGEYNNLV GTWPGKLMEK EEDSNWLR AE VPSKALES GT LTYIFNNGNG
1001 TQLDDNKNT STKNYFTFSS RDSFASKEEV YSFLGISTDD PSAPPEEPQEP
1051 EATLTKKYSK YYLVTEDGEK LTGFHEVDGI LRYFAENTGV MAINKVVTV
1101 DKRYRALDEG RIASNEIIRE YGSDYYLGED GVLVTGLFDY NGDKYYAKSN
1151 GKIVKSGLLE IDGEYIIPDS DGRMLHDVKT AIYFSEYILG SDCKAVKGFT
1201 TYEGQRYYGK ANGRVAKDYM FTVDGDYYA KNDGTLAVSE TITRFFKKYT
1251 FDENGKLISV GK

```

Figure 3.19. MASCOT summary of MALDI-TOF Amy13A identification of protein excised from spot CA7 (A) and spot CA8 (B).

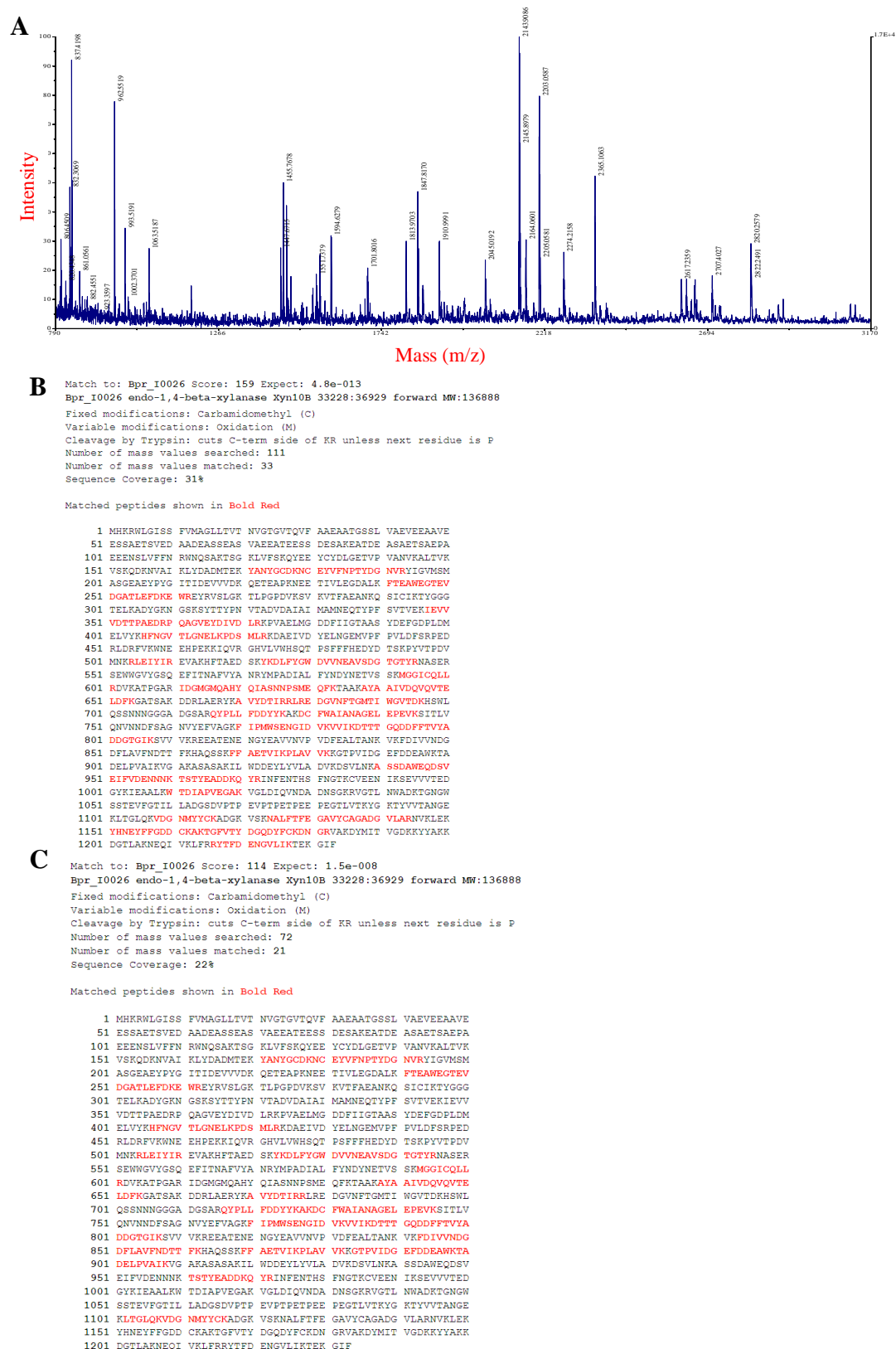


Figure 3.20. MS identification of Xyn10B from the cell-associated fraction. MALDI-TOF peptide mass fingerprint (A) and peptide sequence coverage (B) of protein extracted from spot CA6. (C) Peptide sequence coverage of protein extracted from spot CA9.

3.2.5 Cellulose affinity assay of *B. proteoclasticus* secreted proteins

The presence of CBMs in the polysaccharide-degrading enzymes and non-catalytic secreted proteins identified in the *B. proteoclasticus* extracellular proteome stimulated an examination of the cellulose-binding ability of proteins secreted by xylan-grown, stationary phase harvested *B. proteoclasticus* cells. Phosphate buffered, stationary phase harvested growth medium was mixed with Avicel® PH-101, and proteins present in the culture medium were allowed to adsorb to the substrate over a two hour period. After removal of the culture medium and washing the cellulose pellet, proteins were eluted in 1 x LDS sample buffer and analysed by 1-DE (Figure 3.21).

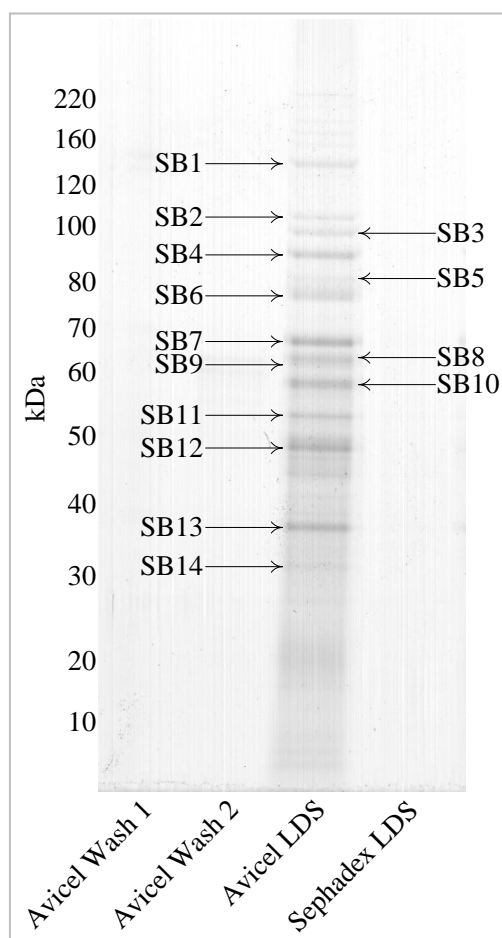


Figure 3.21. 1-DE analysis of polysaccharide adsorption assay of xylan-grown *B. proteoclasticus* secreted proteins. Avicel, Wash 1, and Wash 2 lanes show proteins removed by the final two phosphate buffer washes of Avicel™; Avicel LDS, LDS elution of proteins bound to Avicel™; Sephadex LDS, LDS elution of proteins bound to Sephadex (negative control). Annotated bands are summarised in Table 3.4.

No proteins were detected in the phosphate buffer wash fractions, or the Sephadex LDS elution used as the negative control. In contrast, 14 well resolved protein bands

were detected exclusively in the LDS eluted fraction. Each band was excised, trypsin digested, and analysed by MALDI-TOF MS. Each of the 10 proteins that were identified in 10 of the fourteen excised bands were secreted proteins (Table 3.4), and included the previously identified Xyn10B (Band SB1, and Figure 3.22) and Cel5C (identified as part of a mixture in Band SB10). In addition, two secreted polysaccharidases were identified, Xyn10C (Figure 3.23) and Mxy10-43A (Figure 3.24) that had eluded detection in the 2-DE analysis of the *B. proteoclasticus* extracellular proteome. Proteins present in bands SB5, SB9, SB13, and SB14 were not identified.

All proteins other than Mxy10-43A were detected in the 1-DE separation at the expected size, which indicated that each cellulose bound protein was the full-length form. Xyn10C is a 109.5 kDa polysaccharidase that contains a GH10 and EstD domain located at the N- and C-terminal respectively, separated by central CBM13 and CMB2a domains (Figure 3.25). Xyn10C was identified in a band of 105 kDa (Figure 3.21, band SB2) which is similar to the theoretical protein size, and in a second band that was approximately 5 kDa smaller (Figure 3.21, band SB3). Examination of the sequence coverage results confirmed that one peptide of m/z 1844.90 was detected in band SB2 and matched to the N-terminal region of Xyn10C that was not detected in protein extracted from band SB3. This data supports the possibility that band SB3 contained Xyn10C that was missing the N-terminus.

Mxy10-43A has a theoretical size of 151 kDa, but was detected in a band at approximately 75 kDa. The enzyme contains an N-terminal GH43 domain and a GH10 domain beginning at residue 710, as well as two centrally located CBM6 domains, and single CBM13 and CBM2a domains at the C-terminus (Figure 3.25). Examination of the MASCOT sequence coverage showed that the 13 peptides contributing to the Mxy10-43A identification were all matched to the C-terminal region (Figure 3.24). This data suggested that the polypeptide present in band SB6 was the C-terminal half of Mxy10-43A beginning no closer to the C-terminus than residue 732, which contained the CBM13 and CBM2a domains. Based on the sequence coverage data it is not possible to ascertain if the protein fragment contains the intact GH10 domain. The calculated pI and size of this C-terminal fragment were 4.87 and 72.8 kDa respectively, which is in agreement with the apparent size of the 1-DE band identified as Mxy10-43A.

To examine the possibility that proteins secreted by xylan-grown, stationary phase harvested *B. proteoclasticus* cells possessed xylan-binding ability the polysaccharide adsorption assay was conducted in the same manner as described above, but using oat-spelt xylan in place of Avicel[®] PH-101 as the insoluble substrate. After 1-DE analysis of the 1 x LDS sample buffer eluted fraction no proteins present in the *B. proteoclasticus* extracellular proteome were detected as being capable of binding to oat-spelt xylan.

Table 3.4. Summary of the identified secreted proteins capable of binding to crystalline cellulose

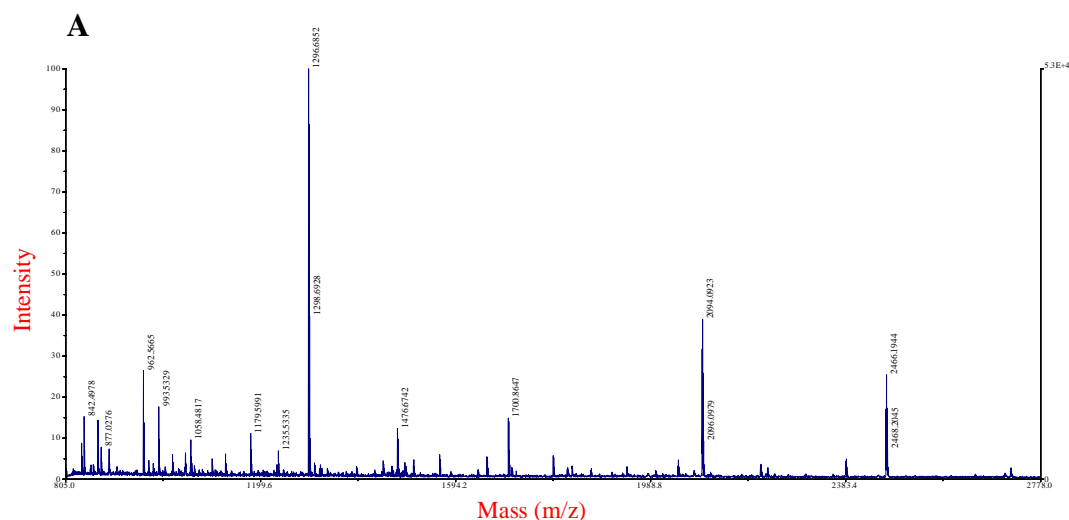
Band	Protein	Locus	EC	PHX ^a	Score ^b	SigP ^c	pI	kDa	Pep. ^a	Cov. ^a
SB1	Endo-1,4- β -xylanase, Xyn10B	Bpr_I0026	3.2.1.8	Y	5.3e-04	Y	4.3	136.9	13	11%
SB2	Endo-1,4- β -xylanase, Xyn10C	Bpr_I1008	3.2.1.8	N	6.1e-08	Y	5.5	109.5	17	22%
SB3	Endo-1,4- β -xylanase, Xyn10C	Bpr_I1008	3.2.1.8	N	3.3e-05	Y	5.5	109.5	16	23%
SB4	Oligopeptide ABC transporter substrate-binding protein, OppA1	Bpr_I1276	-	Y	9.6e-10	Y	4.0	83.1	15	21%
SB6	Endo-1,4- β -xylanase and xylosidase, Mxy10-43A	Bpr_I0737 ^d	3.2.1.8	Y	5.8e-07	Y	5.0	150.7	13	12%
SB7	Sugar ABC transporter substrate-binding protein	Bpr_I0182	-	Y	3.8e-13	Y	4.1	63.3	18	33%
SB8	Sugar ABC transporter substrate-binding protein	Bpr_I1589	-	Y	3.2e-06	Y	3.9	61.1	8	20%
SB10	Sugar ABC transporter substrate-binding protein	Bpr_I2443	-	Y	1.5e-10	Y	4.0	57.0	10	22%
	Endo-1,4- β -glucanase, Cel5C	Bpr_I1710	3.2.1.4	Y		Y	4.6	61.1	10	19%
SB11	Carbohydrate binding protein	Bpr_I0736	-	N	3.0e-07	Y	3.8	57.6	6	16%
SB12	Sugar ABC transporter substrate-binding protein	Bpr_I1720	-	Y	2.0e-04	Y	4.0	49.3	6	18%

^a PHX, predicted highly expressed gene; Pep, the number of matched peptides masses/searched peptides masses; Cov, Protein sequence coverage of the matched peptides.

^b Score value is the statistical expectation that the top ranked protein match is a false positive identification. MOWSE score is given for Mxy10-43A that was identified by LC-MS/MS.

^c Secretory signal-peptides were predicted using SignalP (Ver. 3.0) (www.cbs.dtu.dk/services/SignalP/) and LipoP (Ver. 1.0) (<http://www.cbs.dtu.dk/services/LipoP/>).

^d Predicted C-terminal fragment.



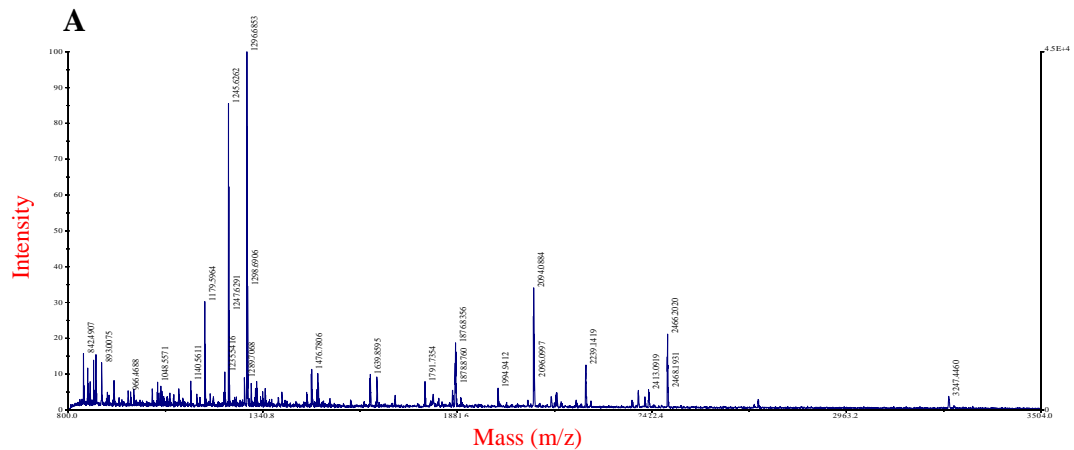
B Match to: Bpr_I0026 Score: 69 Expect: 0.00053
 Bpr_I0026 endo-1,4-beta-xylanase Xyn10B 33228:36929 forward MW:136888

```

1  MHKRWLGISS FVMAGLLTVT NVGTGVTQVF AAEAATGSSL VAEVEEAAVE
51  ESSAETSVED AAEASSEAS VAEATEESS DESAKEATDE ASAETSAEPA
101 EEENSLVFFN RWNQSAKTSG KLVFSKQYEE YCYDLGETVP VANVKALTVK
151 VSKQDKNVAI KLYDADMTEK YANYGCDKNC EYVFNPTYDG NVRVIGVMSM
201 ASGEAEYPYG ITIDEVVVDK QETEAPKNEE TIVLEGDALK FTEAWEGTEV
251 DGATLEFDKE WREYRVSLGK TLPGPDKSV KVTFAEANKQ SICIKTYGGG
301 TELKADYGKN GSKSYTTYPN VTADVDAIAI MAMNEQTYPF SVTVEKIEVV
351 VDTTPAEDRP QAGVEYDIVD LRKPVAELMG DDFIIGTAAS YDEFGDPLDM
401 ELVYKHFNGV TLGNELKPDS MLRKDAEIVD YELNGEMVPF PVLDFSRPED
451 RLDRFVKWNE EHPEKKIQVR GHVLVWHSQT PSFFFHEDYD TSKPYVTPDV
501 MNKRLEIYIR EVAKHFTAED SKYKDLFYGW DVVNEAVSDG TGTYRNASER
551 SEWGVVYGSQ EFITNAFVYA NRYMPADIAL FYNDYNETVS SKMGGICQLL
601 RDVKATPGAR IDGMGMQAHY QIASNNPSME QFKTAARAYA AIVDQVQVTE
651 LDFKGATSAK DDRLAERYKA VYDTIRRLRE DGVNFTGMTI WGVTDKHSWL
701 QSSNNNGGGA DGSARQYPLL FDDYYKAKDC FWAIANAGEL EPEVKISITLV
751 QNVNNDFSAG NVYEFVAGKF IPMWSENGID VKVVIKDTT GQDDFFTVYA
801 DDGTGIKSVV VKREATENE NGYEAVNVNP VDFEALTANK VKFDIVVNDG
851 DFLAVFNDTT FKHAQSSKFF AETVIKPLAV VKKGTPVIDG EFDDEAWKTA
901 DELPVAIKVG AKASASAKIL WDDEYLYVLA DVKDSVLNKA SSDAWEQDSV
951 EIFVDENNNK TSTYEADDKQ YRINFENTHS FNGTKCVEEN IKSEVVVTE
1001 GYKIEAALKW TDIAPVEGAK VGLDIQVNDA DNSGKRVGTL NWADKTGNGW
1051 SSTEVEFGTIL LADGSDVPTP EVPTPETPEE PEGTLVTKYG KTYVVTANGE
1101 KLTGLQKVDG NMYCKADGK VSKNALFTFE GAVYCAGADG VLARNVKLEK
1151 YHNEYFFGDD CKAKTGFVTY DGQDYFCKDN GRVAKDYMIT VGDKKYAKK
1201 DGTLAKNEQI VKLFRRYTFD ENGVLIKTEK GIF

```

Figure 3.22. MS identification of Xyn10B after cellulose affinity assay. MALDI-TOF peptide mass fingerprints (A) and MASCOT sequence coverage summary (B) of protein extracted from band SB1 and identified as Xyn10B.



B Match to: Bpr_I1008 Score: 108 Expect: 6.1e-008
Bpr_I1008 endo-1,4-beta-xylanase Xyn10C 1241289:1244279 forward MW:109456

```

1  MSMKQRNVKC FISKLLVMLL LVSTMGVCFK TNVQAAAYNL METYGAKYGY
51  SGNCVHTFML RNQNTVNSIK KDSNIVTLGN EFKPDYLLGS RSATLISVDE
101 AKRLGYIIPS NYRESTVPKI NFSTVDEAMK LCYQNGLKMR GHTLVVHWSQT
151 PTWFFRNQFS GNGGFVNQST MDARLEFYVK TVLNHVYSSQ YGSVVVYWDV
201 ANEIMHAQNS GWEAVYGNNK TNAAYVKKAF NYAYECLEYF KLTDKVKLFY
251 NDYNTYMVVN DEIKLVNYIN QGKKVCAGIG MQSHLSTSY SVDYYTQALN
301 AFIRAGFEVQ ITELDIKNNG DNDLNNYCYQ LFKNINTAKK NGGNISCITW
351 WGPSDAETWL SGQKPLIWSN IGVSKPAYDY VVKAYTDVFG NPGQSSGGSG
401 SGSSETKPSG NTNATARIED GWYIKNPAS QKYLTVEGNK ADGWNSVVIS
451 SGTGVDGQKW YVTNRSDGYV TLTSKLG NIM MDVANGENTD GANIGTYQGY
501 GGNAQQFIK TTGTNGVYTI GTKCSNAARF LDIYEKKTND GANVCQWTYN
551 GNANQQWQFE KVEENAQQNP EPTPDPEPTP EPTPTPTPDP TPVTGSGLEL
601 EYSINSWGSG YQVSYKISNN SGNTVDGWT L KVNKNQLNID SSWNVNIKDS
651 GDYYVITPVD WNKTIANGTS IEFSGIGVGQ AGSSFYVLE SAAGSSTGNE
701 SGSSSGGETG SEDPKPAPEP EPIPA PNPY NPTARIENSI PSKYSSVRYG
751 EGSGTIVNIS YTAHDTEANG RTYTKKANVY LPAGYSTDKK YNVLYLLHGI
801 GGENEWGMT GNSSTVKAIM DNLAYYGDID SFIVVTPNGK ASASGSVNSF
851 YNFGAELRND LIPYIDSHYS TNADRDHRAI AGLSMGGMQT INIGIGECMD
901 LFSYFGAFSA APTSNAASKT ASLLNGNSYP IHYFYNVCG L QDNVAYSSHS
951 QAAKNLPSVC DQFVDGQNYM WQELNGAHDF NIWYLGIFYNF AQIAFK

```

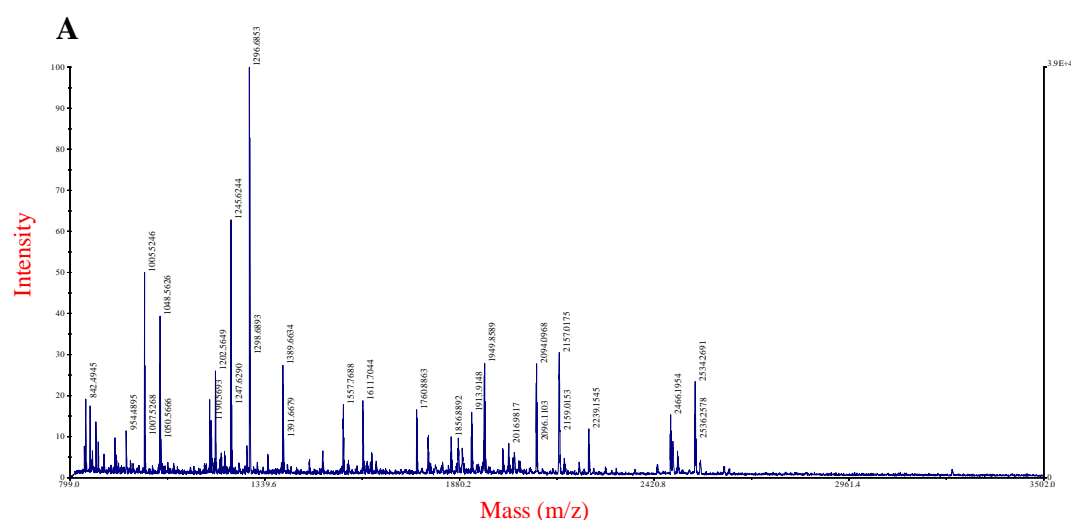
C Match to: Bpr_I1008 Score: 81 Expect: 3.3e-005
Bpr_I1008 endo-1,4-beta-xylanase Xyn10C 1241289:1244279 forward MW:109456

```

1  MSMKQRNVKC FISKLLVMLL LVSTMGVCFK TNVQAAAYNL METYGAKYGY
51  SGNCVHTFML RNQNTVNSIK KDSNIVTLGN EFKPDYLLGS RSATLISVDE
101 AKRLGYIIPS NYRESTVPKI NFSTVDEAMK LCYQNGLKMR GHTLVVHWSQT
151 PTWFFRNQFS GNGGFVNQST MDARLEFYVK TVLNHVYSSQ YGSVVVYWDV
201 ANEIMHAQNS GWEAVYGNNK TNAAYVKKAF NYAYECLEYF KLTDKVKLFY
251 NDYNTYMVVN DEIKLVNYIN QGKKVCAGIG MQSHLSTSY SVDYYTQALN
301 AFIRAGFEVQ ITELDIKNNG DNDLNNYCYQ LFKNINTAKK NGGNISCITW
351 WGPSDAETWL SGQKPLIWSN IGVSKPAYDY VVKAYTDVFG NPGQSSGGSG
401 SGSSETKPSG NTNATARIED GWYIKNPAS QKYLTVEGNK ADGWNSVVIS
451 SGTGVDGQKW YVTNRSDGYV TLTSKLG NIM MDVANGENTD GANIGTYQGY
501 GGNAQQFIK TTGTNGVYTI GTKCSNAARF LDIYEKKTND GANVCQWTYN
551 GNANQQWQFE KVEENAQQNP EPTPDPEPTP EPTPTPTPDP TPVTGSGLEL
601 EYSINSWGSG YQVSYKISNN SGNTVDGWT L KVNKNQLNID SSWNVNIKDS
651 GDYYVITPVD WNKTIANGTS IEFSGIGVGQ AGSSFYVLE SAAGSSTGNE
701 SGSSSGGETG SEDPKPAPEP EPIPA PNPY NPTARIENSI PSKYSSVRYG
751 EGSGTIVNIS YTAHDTEANG RTYTKKANVY LPAGYSTDKK YNVLYLLHGI
801 GGENEWGMT GNSSTVKAIM DNLAYYGDID SFIVVTPNGK ASASGSVNSF
851 YNFGAELRND LIPYIDSHYS TNADRDHRAI AGLSMGGMQT INIGIGECMD
901 LFSYFGAFSA APTSNAASKT ASLLNGNSYP IHYFYNVCG L QDNVAYSSHS
951 QAAKNLPSVC DQFVDGQNYM WQELNGAHDF NIWYLGIFYNF AQIAFK

```

Figure 3.23. MS identification of Xyn10C after cellulose affinity assay. PMF (A) and MASCOT sequence coverage summary (B) of protein extracted from band SB2. (C) MASCOT sequence coverage summary (B) of protein extracted from band SB3. Protein in both bands was identified as Xyn10C.



B Match to: Bpr_I0737 Score: 98 Expect: 5.8e-007
Bpr_I0737 endo-1,4-beta-xylanase and xylosidase Mxy10-43A 895531:899715

```

1  MKQRAIYGKC  NRNIAMMML  LERKNEYGKG  SSMKKRLAVA  LCLMLFGGAV
51  IGNSVDVQAA  SVVNGSYVKA  DNENNPLVTQ  NYGADPGVLV  YNDTVVYVTT
101  NDTQEFAGNN  ENTYGKINTL  NCYSSKDMVN  WTDHGSFKIA  GSSGAAKWAS
151  NSWAPCIATK  KINGKDKFFL  YFANNNGGIG  VLTADSPTGP  WSDPIGRALV
201  NGSTPGASGV  VWMFDPVLV  DSDGTGYLYF  GGGVPNGQAA  HPKTRVIKLV
251  GSNMISTSGS  AVTIDAPYLF  EDSGINKIGN  TYYYSYCSNW  NCANGFNNA
301  IEYMTSSSPM  GPFTYQGEIM  KNPGVFFSGS  TGNNHMIFE  FKGSYYMAYH
351  TRSVESKVIK  KSLGYRTTQI  DKINVSGGRI  SSLTPMSGV  SQLSYVNPYE
401  AVQAEFTMTQ  SGIGVQNGD  GVAWVNNISN  GDYTKVKGVN  FSNGLGSITV
451  NVNSRSGSTV  QVRENSPSGN  LLGTINLSNT  NGKNADFTGT  MKNVSGVKNI
501  CFVFENGFEF  DYWKAQAPGE  AAGGNQGGNE  NQNTGNQGG  NQNTGNQSS
551  DNKVECNMT  RSGQYAGVIS  SPFNGVALYA  NNDTVSFDQY  FAYDTHNVT
601  RGCSNNSNMA  KVLKIGGEE  KGTFFYGDEY  PAEYTIKSVK  HGTGVQTVEL
651  TITSDDGTDW  AAYVDLWTGN  GSGSSQGNTG  NQGGNQGNTG  NQGGNQGNTG
701  NQGQVVEANL  RNTYGAQYGY  SGTCLNLYQL  RDAGQLNFLK  THYNSITLEN
751  EMKPDALLGG  SARLISVND  KSRGYIIPDG  YSEQYVPQIN  FNTVDEVMKI
801  CYQNGLMKRA  HTLVWHSQTP  SWFFRNGYSG  NGGFVNQSTM  DKRLEMYVKT
851  VMNHVYNSQY  GSVVYAWDVA  NEVIHANNNG  WEAVYGNRT  NATYVKKAFN
901  YAYDCLEYFK  LTDSVKLFYN  DYNTRYMEVNN  VIKLVNYINQ  GKQVCAGVGM
951  QSHLGTGFPS  VDYTTALRS  FLNAGFEVQI  TEMDITNKGD  SDMATYAYNL
1001  FKNINSLKKN  GKITSTIWW  GLSDQTTWIS  NSAPLLFSRP  GTPKQAYNKV
1051  IQAYTETFGQ  PTAGGGSNTP  EINPDPGSGN  GGNQGGSQEQ  TPVVNENSTA
1101  RIEDGWYIYK  GVQSQKYLTV  EGNKADGWNS  VIISGSGVD  GQKQWYVSNVS
1151  DGYITLTSKL  GNIMMDVANG  EDTDGANVGT  YQGYGGDAQQ  FIVKNTANS
1201  YYTIGTKASG  ASKYLDCEYK  KTDDGTNVVQ  WTYNGNPNQ  WQFEKVKDET
1251  AQPGEAGNTG  SGEQTDPSQG  SGEQGGQDPG  QGGQGGEDPN  GQTDPTFVGT
1301  GLELTYSINS  WSGYQVSYK  ITNSTGNAVN  TWTLLAKDQ  LNIDSSWNVN
1351  VKTDGNYIVI  TPVDWNSYIA  NGQSIEFGSL  GVGQVGNSE  YTLN

```

Figure 3.24. MS identification of Mxy10-43A after cellulose affinity assay. PMF (A) and MASCOT sequence coverage summary (B) of protein extracted from band SB6, which was identified as the endo-1,4- β -xylanase and xylosidase Mxy10-43A.

3.2.6 Bioinformatics of the polysaccharidases and differentially abundant sugar substrate-binding proteins identified in the *B. proteoclasticus* extracellular proteome

The SignalP 3.0 (www.cbs.dtu.dk/services/SignalP/), LipoP 1.0 (<http://www.cbs.dtu.dk/services/LipoP/>), pattern searching, TMHMM 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/) and SOSUI/G (<http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiG/sosuigsubmit.html>) utilities were used to analyse each enzyme for the presence of an N-terminal secretory signal peptide and transmembrane helices. Pfam analysis was used to examine the identity and location of carbohydrate active domains within each polysaccharidase. Where significant Pfam matches were not found, NCBI protein BLASTp alignments to TIGR or COG classifications were used. Functional domains characteristic of secreted carbohydrate active proteins are present in all 10 gene products (Figure 3.25). Xsa43J and Mxy10-43A were distinct in that they each contained two dissimilar GH domains, while the remainder each possessed a single catalytic module. Five of the eight identified secreted polysaccharidases possessed Type-I cell wall binding repeat regions (CWBD1, PF01473) at their C-termini. Single copies of family 2 and family 9 carbohydrate binding domains (CBMs) were detected in Cel5C and Xyn10B respectively, and pairs of CBMs were identified in both CBPs. The functional domains identified in the 6 differentially abundant SBPs identified in the *B. proteoclasticus* culture medium formed two distinct groups (Figure 3.26). The four SBPs that were more abundant in the xylan growth condition each contained a CUT1 domain, represented by either a SBP_Bac_1 domain (PF 01547) or a MalE domain (COG2182), while the two remaining SBPs both detected in lower abundance in xylan grown cells contained a Peripla_BP_1 domain (PF00532).

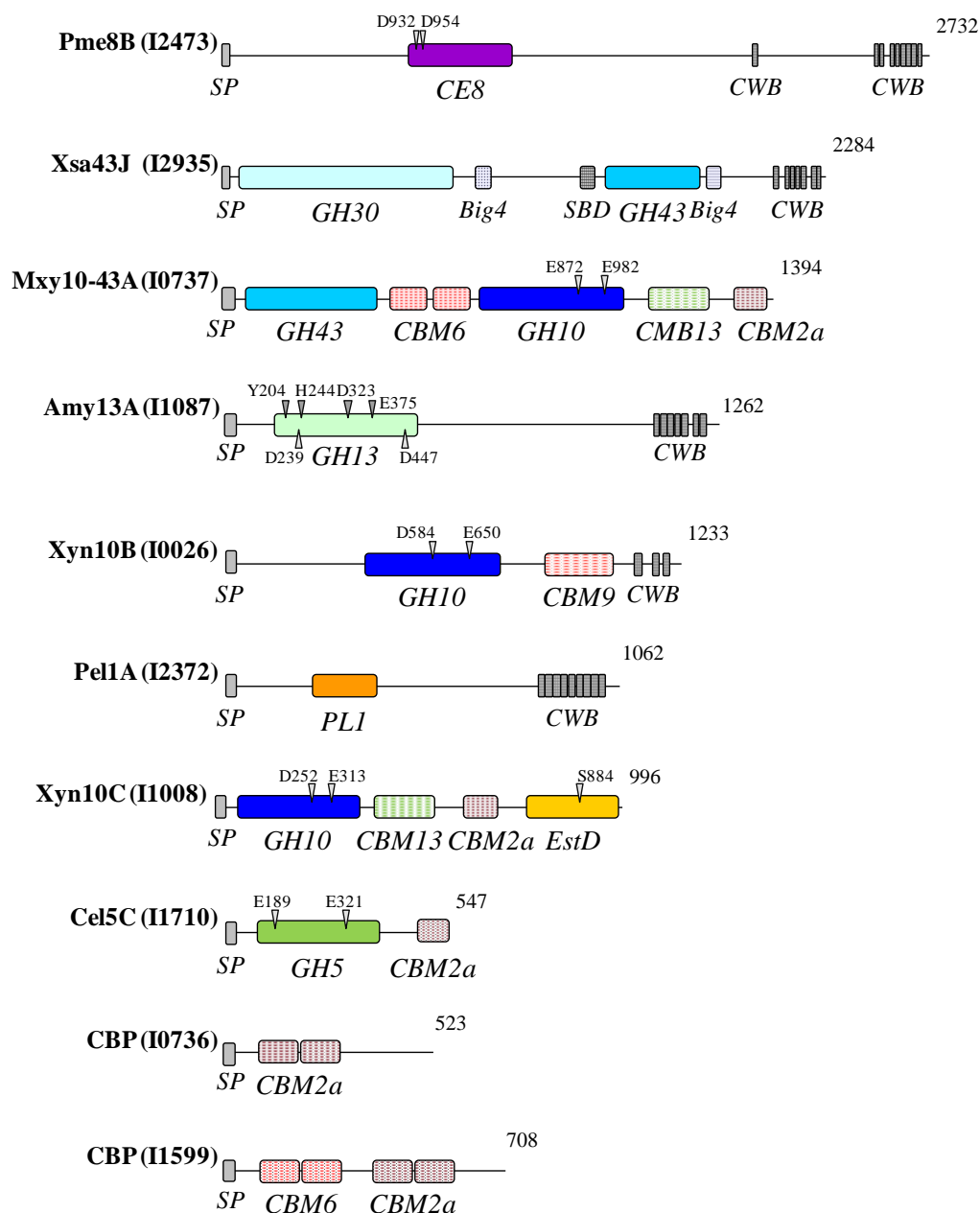


Figure 3.25. Functional domains of the polysaccharidases and carbohydrate binding proteins identified in *B. proteoclasticus* extracellular proteome. SP, signal peptide; GH5, GH family 5 (PF00150); GH10, GH family 10 (PF00331); GH13, GH family 13 (PF00128); GH43, GH family 43 (PF04616); CE8, CE family 8 (PF01095); EstD, Esterase D domain (PF00756); PL1, PL family 1 (PF00544); CWB, Type-I cell wall binding domain (PF01473); Big4, bacterial Ig-like domain-group 4 (PF07532); SBD, Uncharacterised sugar-binding domain (PF07554); CBM2a, CBM6, CBM9, and CBM13, family 2a (PF00553), family 6 (PF03422), family 9 (IPR010502), and family 13 (PF00652) carbohydrate binding modules, respectively. Residue length is indicated to the right of each enzyme. Note that Pme8B and Xsa43J are shown at two-thirds relative scale.

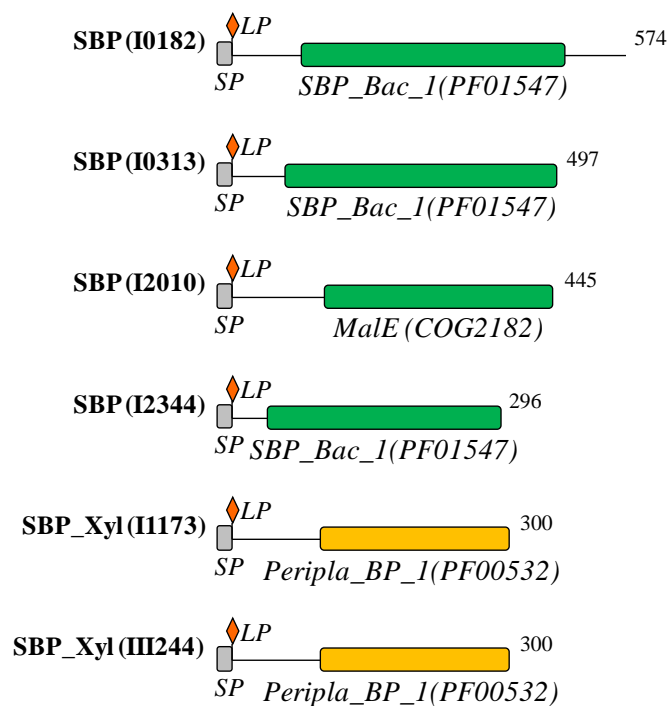


Figure 3.26. Functional domains of differentially abundant substrate-binding proteins identified in the *B. proteoclasticus* culture medium. Values to the right of each depicted protein denote number of amino acid residues. LP, lipobox motif; SP, signal peptide. Domains shown in green and gold are associated with proteins more abundant in the xylan and xylose growth conditions respectively.

Table 3.5. BLASTp homology of the predicted secreted proteins identified in the mid-log and stationary phase harvested *B. proteoclasticus* culture medium.

<i>B. proteoclasticus</i> protein	Locus	Homologue	Organism	gi	Expect	Identity	Region
<i>Carbohydrate active proteins</i>							
α -Amylase, Amy13A	Bpr_I1087	α -amylase	<i>Butyrivibrio fibrisolvens</i>	144153	1e-180	99%	1-975
Endo-1,4- β -glucanase, Cel5C	Bpr_I1710	Endoglucanase	<i>Butyrivibrio fibrisolvens</i>	39473	1e-180	99%	1-547
Endo-1,4- β -xylanase and xylosidase, Mxy10-43A	Bpr_I0737	β -1,4-xylanase	<i>Butyrivibrio fibrisolvens</i> 16/4	291519800	1e-180	69%	707-1386
Pectate lyase, Pel1A	Bpr_I2372	Pectate lyase	<i>Ruminococcus albus</i> 8	294639839	1e-180	55%	48-745
Pectin methyl-esterase, Pme8B	Bpr_I2473	Fibronectin Type-III domain-containing protein	<i>Clostridium phytofermentans</i> ISDg	160881987	1e-180	49%	1468-2452
Xylosidase/arabinofuranosidase, Xsa43J	Bpr_I2935	Cna B domain protein	<i>Clostridium saccharolyticum</i> WM1	302386974	1e-180	42%	879-2064
Endo-1,4- β -xylanase, Xyn10B	Bpr_I0026	Putative carbohydrate binding domain protein	<i>Roseburia intestinalis</i> L1-82	257413505	1e-180	44%	338-1068
Endo-1,4- β -xylanase, Xyn10C	Bpr_I1008	Endo-1,4- β -xylanase and xylosidase Mxy10-43A	<i>Butyrivibrio proteoclasticus</i> B316	302670102	1e-180	66%	34-689
Carbohydrate-binding protein	Bpr_I0736	Carbohydrate-Binding Module family 2 with two N-terminal CBM2 domains	<i>Eubacterium rectale</i>	238922693	8e-53	43%	1-253
Carbohydrate-binding protein	Bpr_I1599	1,4- β -cellobiosidase A	<i>Butyrivibrio fibrisolvens</i> 16/4	291519773	2e-44	34%	320-708

Table continues

Table 3.5 continued.

<i>B. proteoclasticus</i> protein	Locus	Homologue	Organism	gi	Expect	Identity	Region
<u><i>Non-carbohydrate active proteins</i></u>							
3-hydroxybutyryl-CoA dehydrogenase Hbd	Bpr_I2486	β -hydroxybutyryl-CoA dehydrogenase	<i>Butyrivibrio fibrisolvens</i>	52421176	3e-127	78%	1-290
ABC transporter SBP	Bpr_I1600	Extracellular solute-binding protein family 1	<i>Ruminococcus flavefaciens</i> FD-1	268611735	1e-47	32%	332-821
Amino acid ABC transporter SBP	Bpr_I2466	Amino acid ABC transporter substrate-binding protein, PAAT family	<i>Butyrivibrio fibrisolvens</i> 16/4	291518760	1e-133	79%	1-300
Bmp family protein	Bpr_I1560	Nucleoside-binding protein	<i>Butyrivibrio fibrisolvens</i> 16/4	291519926	3e-154	79%	50-393
Cell surface protein	Bpr_I2508	Cell surface protein	<i>Butyrivibrio proteoclasticus</i> B316	302671437	4e-173	38%	167-1179
Hypothetical protein	Bpr_I0139	Uncharacterized protein in bglA 3'region precursor	<i>Butyrivibrio fibrisolvens</i>	140270	1e-46	97%	24-119
Hypothetical protein	Bpr_I0188	No homologue	-	-	-	-	-
Hypothetical protein	Bpr_I2628	Hypothetical protein	<i>Butyrivibrio fibrisolvens</i> 16/4	291519088	5e-52	44%	23-280
Oligopeptide ABC transporter SBP, OppA1	Bpr_I1276	Hypothetical protein	<i>Roseburia inulinivorans</i> DSM 16841	225375060	1e-180	54%	1-757
Oligopeptide ABC transporter SBP, OppA2	Bpr_III023	Extracellular solute-binding protein family 5	<i>Paenibacillus</i> sp. JDR-2	251797709	2e-88	33%	1-687

Table continues

Table 3.5 continued.

<i>B. proteoclasticus</i> protein	Locus	Homologue	Organism	gi	Expect	Identity	Region
Peptide/nickel ABC transporter periplasmic protein	Bpr_I2750	Extracellular solute-binding protein	<i>Alkaliphilus metalliredigens</i> QYMF	150389816	3e-162	56%	47-533
Serine protease subtilisin family	Bpr_I2629	Bacterial Ig-like domain (group 2) / Subtilase family.	<i>Butyrivibrio fibrisolvens</i> 16/4	291519087	1e-180	62%	1-1347
Sugar ABC transporter substrate-binding protein	Bpr_I2344	ABC-type sugar transport system, periplasmic component	<i>Butyrivibrio fibrisolvens</i> 16/4	291518407	1e-180	81%	1-474
Sugar ABC transporter SBP	Bpr_I0182	ABC-type sugar transport system, periplasmic component	<i>Butyrivibrio fibrisolvens</i> 16/4	291518643	1e-180	63%	1-568
Sugar ABC transporter SBP	Bpr_I0237	ABC-type sugar transport system periplasmic component-like protein	<i>Geobacillus</i> sp. Y412MC10	261408632	4e-63	30%	7-575
Sugar ABC transporter SBP	Bpr_I0313	Extracellular solute-binding protein, family 1	<i>Epulopiscium</i> sp. 'N.t. morphotype B'	168334645	8e-58	36%	74-497
Sugar ABC transporter SBP	Bpr_I0937	ABC-type sugar transport system, periplasmic component	<i>Roseburia intestinalis</i> M50/1	291535091	1e-180	73%	3-547
Sugar ABC transporter SBP	Bpr_I1589	Putative bacterial extracellular solute-binding protein	<i>Roseburia intestinalis</i> L1-82	257413330	7e-88	39%	1-555
Sugar ABC transporter SBP	Bpr_I1667	Bacterial extracellular solute-binding protein	<i>Turicibacter</i> sp. PC909	293376212	6e-58	34%	3-419
Sugar ABC transporter SBP	Bpr_I1720	Extracellular solute-binding protein family 1	<i>Acetivibrio cellulolyticus</i> CD2	302592083	5e-69	34%	1-455
Sugar ABC transporter SBP	Bpr_I2010	Carbohydrate ABC transporter substrate-binding protein, CUT1 family	<i>Coprococcus</i> sp. ART55/1	295094446	1e-55	38%	1-413

Table continues

Table 3.5 continued.

<i>B. proteoclasticus</i> protein	Locus	Homologue	Organism	gi	Expect	Identity	Region
Sugar ABC transporter SBP	Bpr_III244	Monosaccharide ABC transporter substrate-binding protein, CUT2 family	<i>Eubacterium siraeum</i> V10Sc8a	291558050	9e-141	75%	2-330
Xylose ABC transporter SBP	Bpr_I1173	Putative solute-binding component of ABC transporter	<i>Clostridium phytofermentans</i> ISDg	160879730	4e-161	80%	19-359

3.3 Summary

Gel-based analysis of *B. proteoclasticus* proteins present in the culture medium after growth to mid-log or stationary phase was successful in identifying 30 Type-I or Type-II N-terminal signal peptide containing proteins, which included the endoxylanase Xyn10B and the xylosidase/arabinofuranosidase Xsa43J, the pectate lyase Pel1A and pectin methylesterase Pme8B, and the endoglucanase Cel5C. Two carbohydrate-binding proteins were also identified, which were the products of Bpr_I0736 and Bpr_I1599. Relative protein abundance analysis demonstrated that during mid-log phase growth Xyn10B was detected in the culture medium of xylan grown cells only, where it comprised 5.6% of the total detectable protein. Similarly, during stationary phase Xyn10B comprised 8.7% of the total detectable protein in the culture medium of xylan grown cells, and was 17.5-fold more abundant relative to xylose grown cells. In contrast, Cel5C and Pel1A were both less abundant in the culture medium of xylan grown cells, relative to those grown in the presence of xylose.

Gel-based analysis of *B. proteoclasticus* cell-associated proteome added to the inventory of identified polysaccharidases. The α -amylase Amy13A was identified in mid-log phase, xylan grown cells, but not in cells grown in any other growth condition, which indicates that the abundance of Amy13A may be influenced by growth substrate. Xyn10B and Xsa43J were also identified in the cell-associated proteome of mid-log phase *B. proteoclasticus* cells, and the Xyn10B relative abundance pattern supported that detected in the culture medium.

Using a cellulose affinity assay it was demonstrated that *in vitro*, at least 10 proteins secreted by xylan-grown, stationary phase harvested *B. proteoclasticus* cells are able to adsorb to crystalline cellulose. These proteins included Xyn10B and Cel5C, as well as two additional secreted polysaccharidases, which were the endoxylanase Xyn10C, and the endoxylanase and xylosidase Mxy10-43A. Both these polysaccharidases are distinctive in that they are two of only three identified secreted enzymes that contain multiple GH domains.

The identified secreted enzymes that target hemicellulose contain GH domains that hydrolyse (1→4)- β -D-linkages within the xylan backbone, or between the backbone and arabinose side chain groups. Polysaccharidases that target other hemicellulose substituents such as glucuronic acid, ferulic acid, and acetate groups were not detected in the culture medium of *B. proteoclasticus* cells.

Chapter 4

The cytosolic *B. proteoclasticus* proteome

4.1 Introduction

The activity of secreted and cell-associated polysaccharide degrading enzymes such as the endo-1,4- β -xylanase Xyn10B and xylosidase/arabinofuranosidase Xsa43J, and of a variety of ABC transporter proteins identified in the *B. proteoclasticus* culture medium is likely to result in the assimilation of substituted xylooligosaccharides and monosaccharides into the cytosol. Prior to entering the central metabolic pathways, these polymers must be further degraded to their constituent monomeric sugars by the action of cytosolic polysaccharidases. The *B. proteoclasticus* genome encodes 2854 cytosolic proteins, including 94 polysaccharidases representing 26 GH, 8 CE, and two GT families. Analysis of the cytosolic proteome will assist in understanding the processes that *B. proteoclasticus* uses to degrade and utilise polysaccharides in the rumen, and of the subsequent conversion of the fibrolytic end products to energy sources that can be utilised by the host.

4.2 Results

A theoretical two-dimensional map of the *B. proteoclasticus* cytosolic proteome was constructed and used to assist with targeting subsequent 2-DE analyses to the *pI* regions of specific interest. Proteomic analysis of the *B. proteoclasticus* cytosol was performed using 2-DE MALDI-TOF and 1-D HPLC MS/MS.

4.2.1 The theoretical cytosolic proteome of *B. proteoclasticus*

The theoretical 2-DE map of the *B. proteoclasticus* cytosolic proteome (Figure 4.1) showed an acidic *pI* distribution that was similar to that of the theoretical total proteome. Of the 2854 proteins that are localised to the cytosol, 79% (2242 proteins) possessed a predicted *pI* less than 7, and 90 of the 94 cytosolic polysaccharide-degrading enzymes encoded by the *B. proteoclasticus* genome had a predicted *pI* value below 6.8 (Figure 4.1, red spots). A particularly dense cluster of polysaccharidases was found between *pI* 4 and 5.6 (81 enzymes). The protein size distribution of the cytosolic proteome was similar to that of the total proteome. Ninety percent of the cytosolic proteome (2575 proteins) possessed a theoretical size between 10 and 150 kDa. All cytosolic polysaccharidases had a predicted size between 26 and 134 kDa. Furthermore, 148 cytosolic proteins (5.2%) were identified as components of a PUL. One-hundred and thirty-five PUL proteins had a *pI* value less than 6.5, and 145 were between 10 and 150 kDa in size. The set of 802 proteins with a predicted *pI* value greater than 6.5 were dominated by hypothetical proteins and mobile elements

(transposases), and proteins involved in nucleic acid metabolism, protein synthesis, transcriptional regulation, and cell envelope biogenesis.

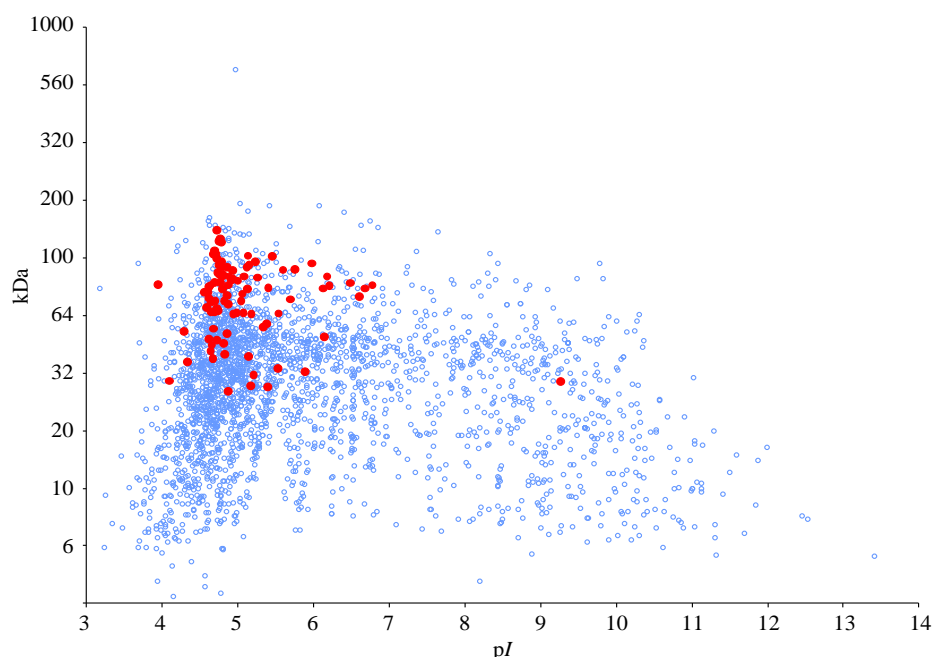


Figure 4.1. Theoretical 2-DE map of the *B. proteoclasticus* cytosolic proteins. Red spots and blue open circles represent predicted polysaccharide-degrading enzymes, and all other cytosolic proteins respectively. The y-axis is presented in logarithmic scale to represent separation of proteins by 2-DE.

Protein function analysis of the predicted cytosolic proteome demonstrated a similar distribution to the total proteome. The major differences were an increase in the relative proportion of proteins involved in amino acid biosynthesis, carbohydrate degradation, nucleic acid and nucleotide metabolism, and protein synthesis, with a concomitant decrease of proteins involved in signal transduction, and transporter proteins. Proteins of unknown function were increased in the cytosolic proteome, and hypothetical proteins were decreased.

4.2.2 Gel-based analysis of the *B. proteoclasticus* cytosolic proteome

To obtain an overview of the *B. proteoclasticus* cytosolic proteome, soluble proteins (100 µg) extracted and purified from xylan and xylose grown *B. proteoclasticus* cells harvested at mid-log phase were separated on a pI 3-10 IPG strip followed by a NuPAGE® Novex 4-12% Bis-Tris second-dimension gel. After colloidal CBB G-250 staining, an acid pI distribution was observed in the cytosolic proteomes of both the xylan (Figure 4.2 A) and xylose (Figure 4.2 B) grown cells with

almost all detectable protein possessing a pI value below approximately seven. A small number of weakly detectable spots were detected above pI 7.

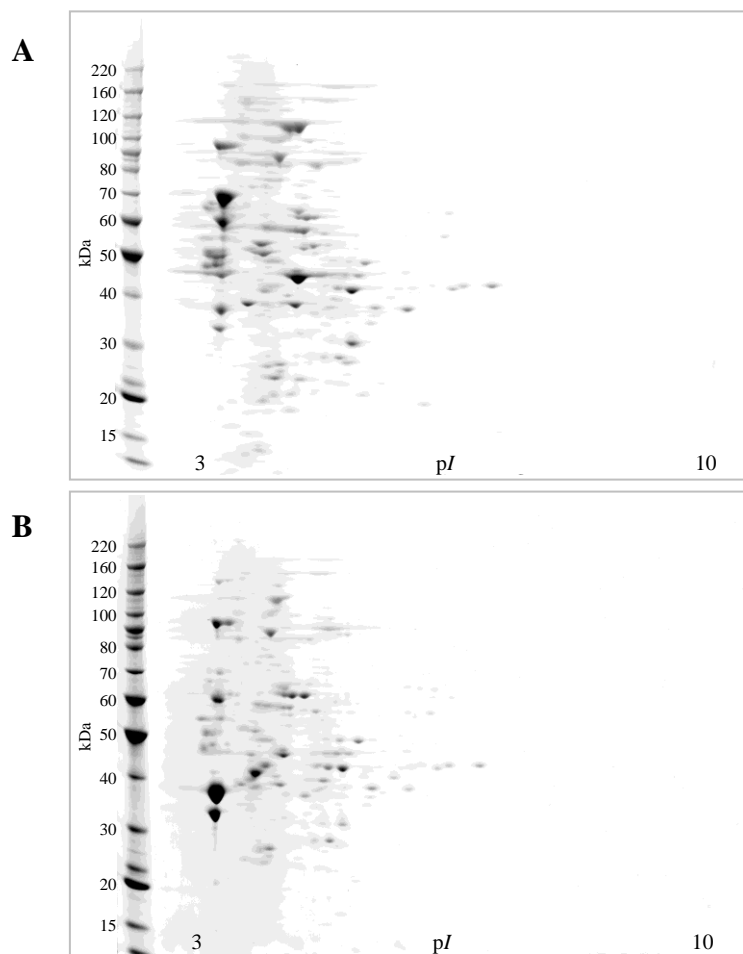


Figure 4.2. 2-DE analysis (pI 3-10) of cytosolic proteins extracted from differentially cultured, mid-log phase harvested *B. proteoclasticus* cells. Cytosolic proteins were extracted from xylan (A) and xylose (B) grown, mid-log phase harvested *B. proteoclasticus* cells, and purified by phenol/methanol/ammonium acetate precipitation. Each gel was stained with colloidal CBB G-250.

Consequently, IPG strips covering the pI ranges 3-5.6 and 5.3-6.5 were used to examine further the cytosolic proteome of *B. proteoclasticus*, and to investigate the effect of growth substrate on the protein abundance profiles of identified proteins. IPG strips covering the pI 6-11 range were also used to identify proteins that may have a basic isoelectric point.

4.2.2.1 2-DE spot detection replicate gel matching

Following 2-DE and image acquisition, protein spots were detected using Image Master™ 2D Platinum. Automated spot detection parameters were optimised for

maximum spot resolution and minimal detection of artefacts such as colloidal CBB G-250 precipitate and dust particles. Any spots detected around the borders of each gel, or within the molecular weight marker region were not considered in the analysis. In each growth condition, the mean total spot number per gel was consistent within each *pI* range, and in all growth conditions, the majority of proteins were detected in the *pI* 3-5.6 region (Figure 4.3). A slightly larger number of spots were detected in the mid-log growth phase compared to the stationary phase *pI* 3-5.6 gel sets (1264 and 1180 spots, respectively).

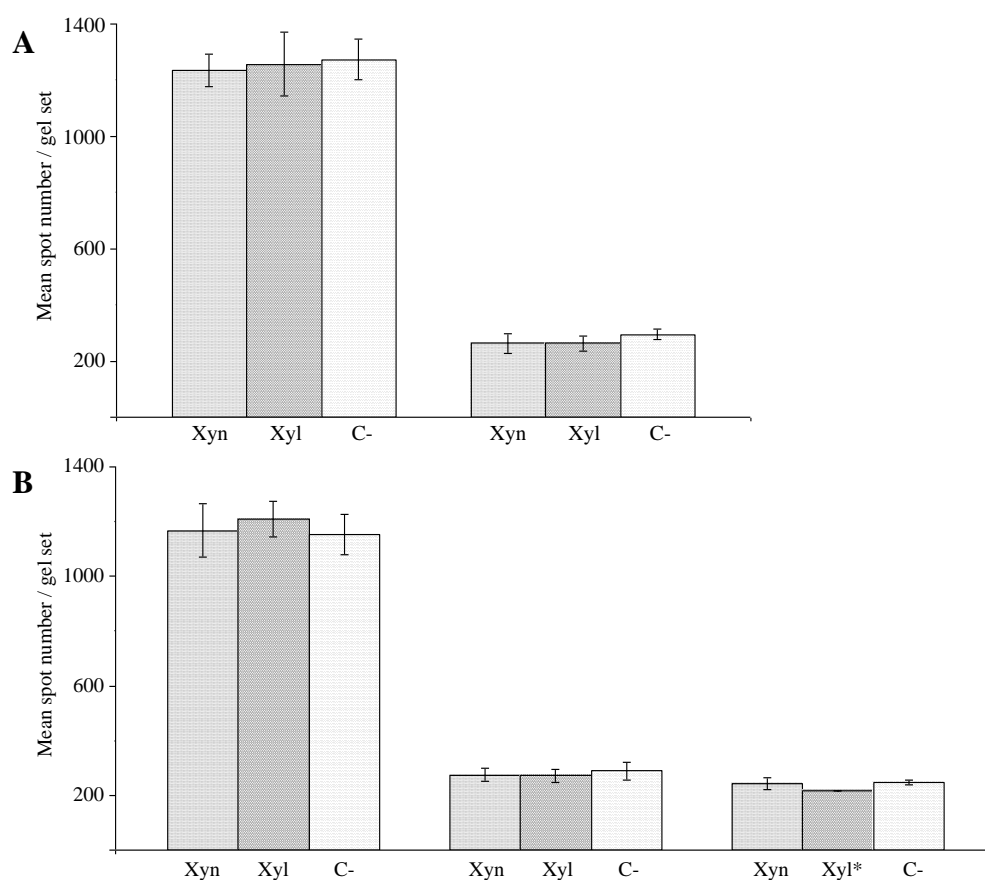


Figure 4.3. Mean number of spots detected in the cytosolic 2-DE analyses. Mean spot number detected in the mid-log (A) and stationary (B) phase cytosolic 2-DE analyses are shown. Error bars represent one standard deviation. Xn, xylan; Xl, xylose, C-, no added carbon. *The SD for the *pI* 6-11 xylose gel set was 0.7.

A matched spot group was defined as a group of spots where the spot was present in the reference gel and was also present in at least six of the eight other replicate gels. The percentage cumulative frequency of spot groups that possessed between two and nine spots per group was plotted for the mid-log phase (Figure 4.4A) and stationary

phase (Figure 4.4B) *pI* 3-5.6 gels sets. Within each *pI* range, the proportion of complete spot groups (proteins matched in all nine replicate gels per group) is similar across each growth condition. An increase in the cumulative frequency of spot groups occurred as the number of spots per group decreased. In the *pI* 3-5.6 mid-log gel sets for instance, an average of 71.6% and 79.9% of matched spot groups possessed at least eight and seven contributing spots, respectively.

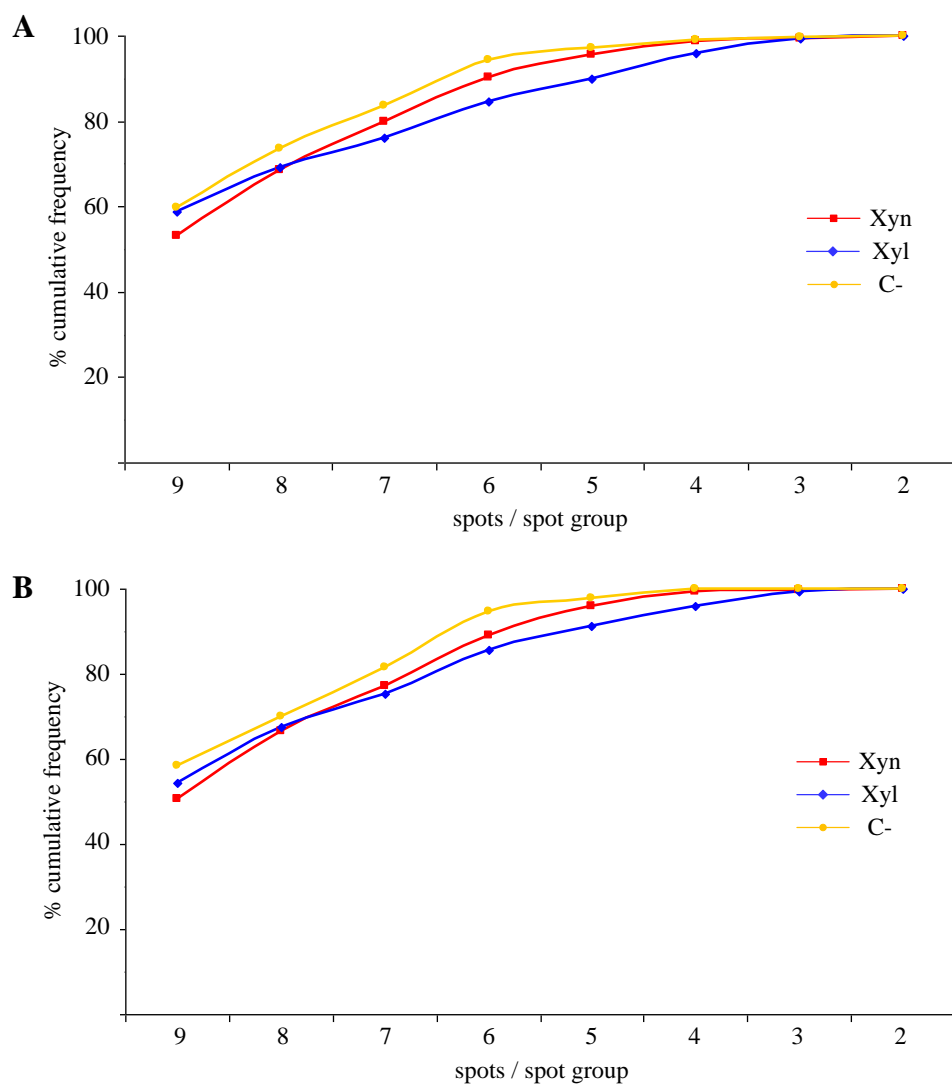


Figure 4.4. The percentage cumulative frequency of matched spot groups in the 2-DE analyses. Mid-log (A) and stationary phase (B) *pI* 3-5.6 gel sets, having the indicated number of matched spots per group (including the reference gel) are shown C-, no added carbon.

4.2.2.2 Polysaccharide degrading enzymes identified by MALDI-TOF in the *B. proteoclasticus* cytosolic proteome

2-DE separations of the cytosolic subcellular fraction harvested from xylan-grown cells were used to identify cytosolic proteins produced by *B. proteoclasticus*. In addition, spots present in the xylose, or no added carbon growth conditions that were absent in the xylan grown condition were also analysed. In total, 625 protein spots were excised from the xylan, xylose, and no added carbon, mid-log and stationary phase separations covering the *pI* 3-5.6, *pI* 5.3-6.5, and *pI* 6-11 ranges, and in-gel trypsin digested. Using the MASCOT search algorithm, each tryptic-PMF was searched against the *B. proteoclasticus* protein sequence database, which was constructed from the functionally annotated *B. proteoclasticus* genome sequence, and contained the translated amino acid sequences of the 3813 CDSs, plus the 3813 randomised protein sequences of the same size and amino acid composition (described in Section 2.12.3). Proteins were identified in 485 of the 625 spots analysed, which were the products of 223 genes. Within this set were 48 proteins devoted to carbohydrate metabolism, and 11 to the transport of polysaccharide degradation products. The presence of 20 polysaccharide-degrading enzymes within the cytosol of *B. proteoclasticus* cells grown in the presence of xylan was confirmed (Figure 4.5, Figure 4.6, and Table 4.1), and no additional enzymes were identified in either the xylose or no added carbon growth conditions. Nineteen of the 20 polysaccharidases were identified from the *pI* 3-5.6 2-DE separation (Figure 4.5). No polysaccharidases were identified in the *pI* 5.3-6.5 2-DE separation, but the glycogen debranching enzyme Glgx2 (Bpr_I1494) was identified in the *pI* 6-11 separation (Figure 4.6). The 2-DE spot positions for all but the glycoside hydrolase family 30 enzyme Gh30A (Bpr_I2937) (spot C644) agreed with the predicted *pI* and size value. Four polysaccharidases were identified in more than one spot, while spot C1120 and spot C1147 both contained more than one enzyme. Surprisingly, spot C1147 contained β -galactosidase Bga2B, β -glucosidase Bgl3C, and β -mannosidase Man2A. Analysis of the sequence coverage results confirmed that other than the *m/z* 876.46 peptide shared between Bga2B and Man2A, the sets of well resolved peptide masses matched to each protein were unique (Figure 4.7). The theoretical *pI* and size of the three enzymes identified in spot C1147 were similar, supporting the likelihood that each was present in the 2-DE spot from which it was identified. Likewise, sequence coverage of the PMF obtained from spot C1120 and the resulting sequence coverage of each identified enzyme showed that apart from the peptide mass at *m/z* 1022.52, a unique set of

peptides contributed to the identification of the β -galactosidase, Bga35B (Bpr_I2006) and the β -xylosidase Xyl3A (Bpr_I0184) (Figure 4.8). Furthermore, the probability that the identification of both enzymes in spot C1120 was a false positive result was $1.9\text{e-}20$.

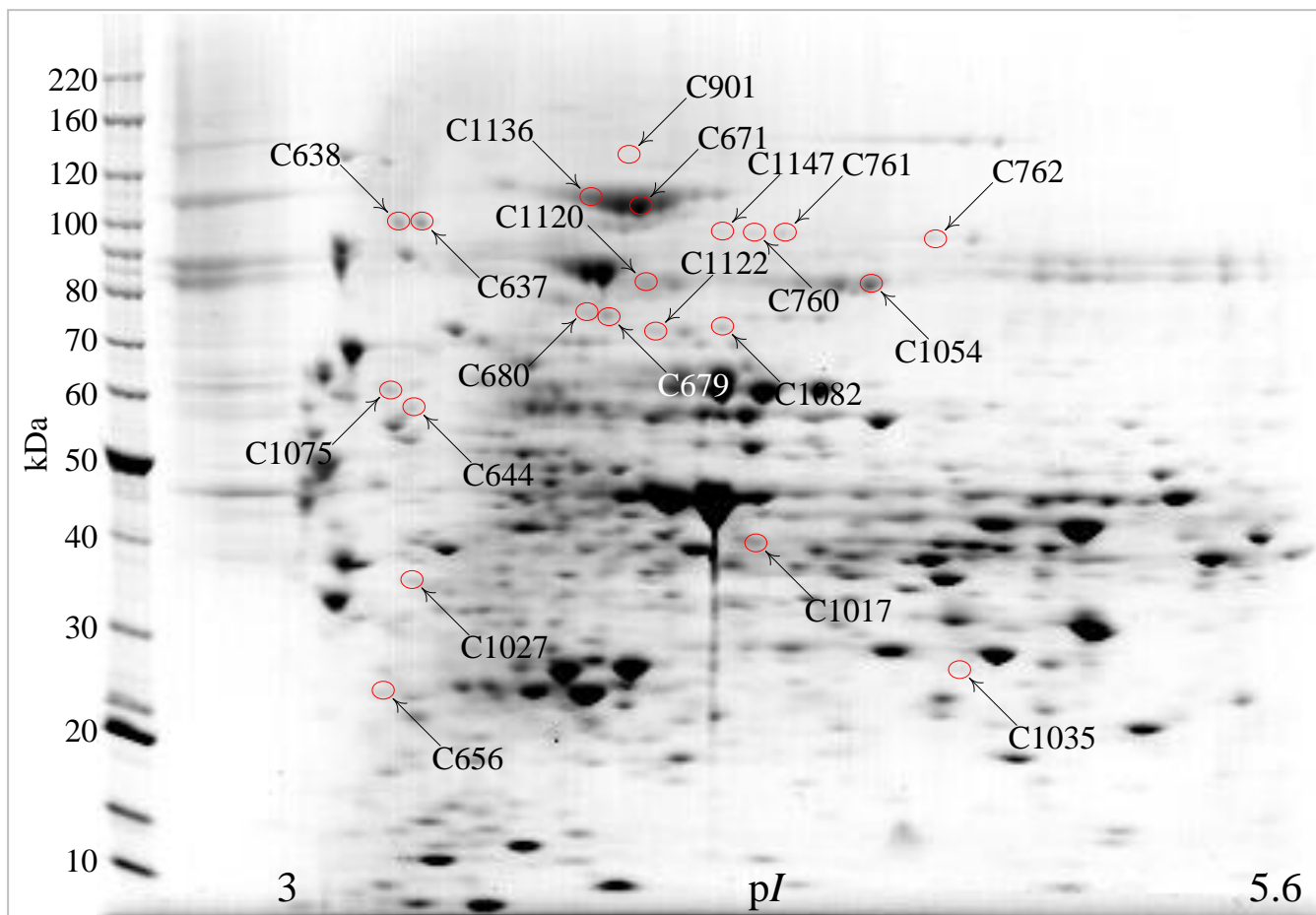


Figure 4.5. Polysaccharide degrading enzymes identified in the *B. proteoclasticus* cytosol by pI 3-5.6 2-DE/MALDI-TOF MS. Shown is a colloidal CBB G-250 stained 2-DE separation of cytosolic proteins harvested from mid-log phase, xylan grown cells. Circled spots denote the the polysaccharidases present in the set of 485 proteins from the cytosolic proteome, and are summarised in Table 4.1.

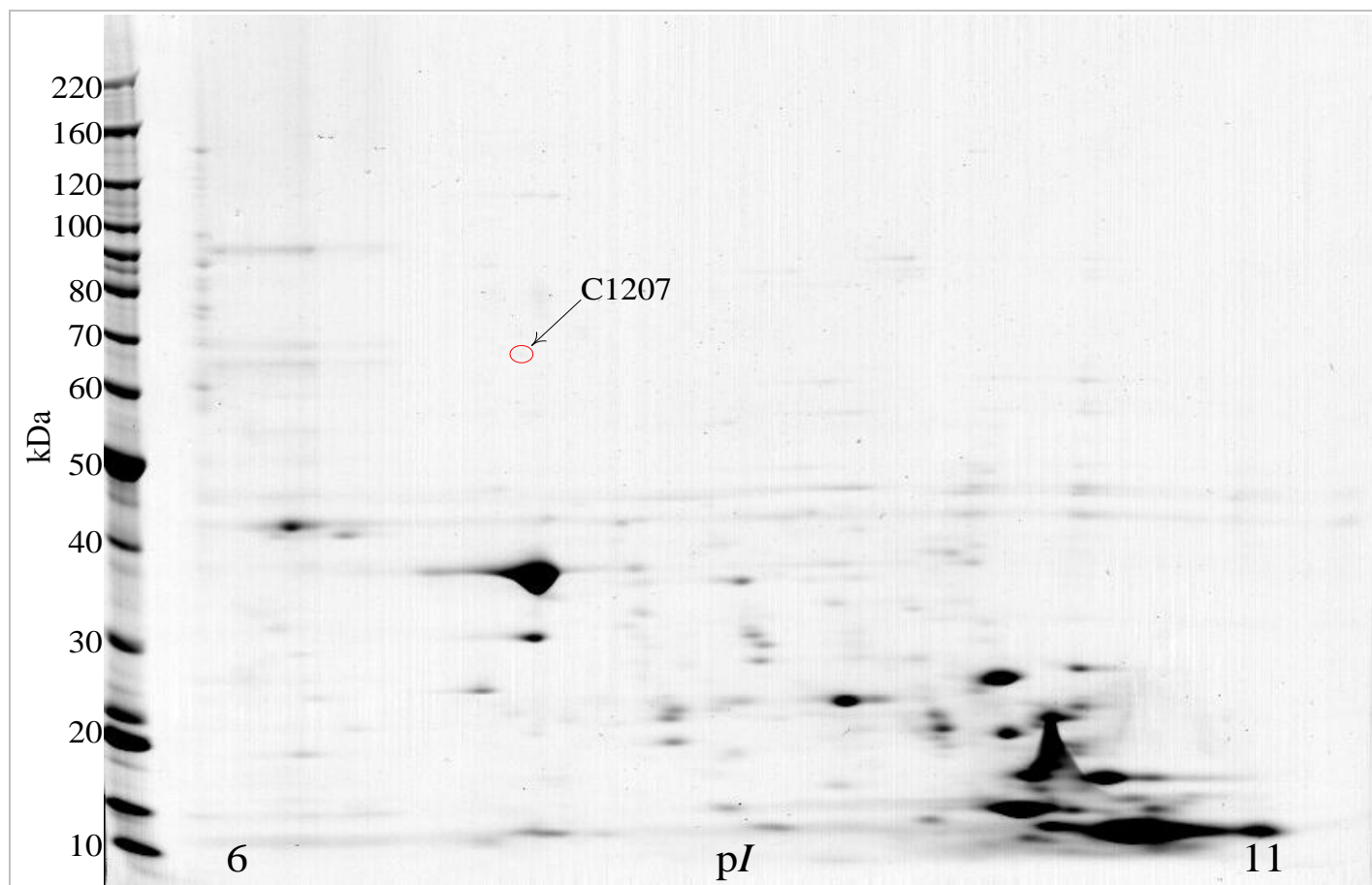


Figure 4.6. Polysaccharide degrading enzymes identified in the *B. proteoclasticus* cytosol by pI 6-11 2-DE/MALDI-TOF MS. Shown is a colloidal CBB G-250 stained 2-DE separation of cytosolic proteins harvested from mid-log phase, xylan grown cells. Circled spots are summarised in Table 4.1.

Table 4.1. Summary of the polysaccharidases identified by 2-DE/MALDI-TOF in the *B. proteoclasticus* cytosol.

Spot	Protein	Locus	EC	PHX ^a	Expect ^b	SigP ^c	pI	kDa	Pep. ^a	Cov. ^a
C1122	α -Galactosidase, Aga27A	Bpr_I0205	3.2.1.22	N	2.4e-16	N	4.6	66.8	12	18%
C1054	α -Galactosidase, Aga36C	Bpr_III065	-	N	1.2e-21	N	5.1	83.2	27	37%
C1082	α -D-Glucuronidase, Agu67A	Bpr_I0177	3.2.1.139	N	3.0e-22	N	4.9	76.1	27	37%
C901	β -Galactosidase, Bga2A	Bpr_I0279	3.2.1.23	N	1.5e-12	N	4.7	118.7	20	24%
	β -Galactosidase, Bga2B	Bpr_III209	-	N		N	4.8	91.1	11	16%
C1147	β -Glucosidase, Bgl3C	Bpr_I0138	3.2.1.21	N	1.9e-18	N	4.9	91.5	13	18%
	β -Mannosidase, Man2A	Bpr_III237	-	N		N	4.8	95.9	20	28%
C1120	β -Galactosidase, Bga35B	Bpr_I2006	3.2.1.23	N	1.9e-27	N	4.9	83.2	17	24%
	β -Xylosidase, Xyl3A	Bpr_I0184	3.2.1.37	Y		N	4.8	78.2	25	37%
C671	β -Glucosidase, Bgl3B	Bpr_I0847	-	N	1.2e-30	N	4.7	103.7	22	24%
C1136	β -Glucosidase, Bgl3B	Bpr_I0847	-	N	2.4e-14	N	4.7	103.7	26	25%
C760	β -Glucosidase, Bgl3C	Bpr_I0138	3.2.1.21	N	1.5e-12	N	4.9	91.5	17	21%
C761	β -Glucosidase, Bgl3C	Bpr_I0138	3.2.1.21	N	6.1e-13	N	4.9	91.5	18	21%
C762	Cellobiose phosphorylase, Cbp94A	Bpr_I2447	2.4.1.20	Y	9.6e-11	N	5.1	91.5	15	23%
C1035	Feruloyl esterase, Est1E	Bpr_I2870	-	N	2.4e-12	N	5.2	27.8	13	61%
C1017	Acetyl-xylan esterase, Est2A	Bpr_I2939	-	N	3.0e-07	N	4.8	42.4	13	33%

Table continues

Table 4.1 continued.

Spot	Protein	Locus	EC	PHX ^a	Expect ^b	SigP ^c	pI	kDa	Pep. ^a	Cov. ^a
C656	Acetyl-xylan esterase	Bpr_I0174	-	N	1.5e-09	N	4.6	75.1	16	22%
C644	Glycoside hydrolase family 30, GH30A	Bpr_I2937	-	N	7.6e-13	Y	4.3	67.0	15	25%
C679	Glycoside hydrolase family 31, GH31C	Bpr_I1974	3.2.1.-	Y	3.0e-15	N	4.7	78.2	21	27%
C680	Glycoside hydrolase family 31, GH31C	Bpr_I1974	3.2.1.-	Y	7.6e-18	N	4.7	78.2	25	35%
C1207	Glycogen debranching enzyme, Glx2	Bpr_I1494	3.2.1.-	N	8.2e-05	N	6.8	76.2	8	10%
C637	Pullulanase, Pul13A	Bpr_III161	3.2.1.41	N	1.5e-09	Y	4.4	99.7	20	22%
C638	Pullulanase, Pul13A	Bpr_III161	3.2.1.41	N	1.8e-06	Y	4.4	99.7	11	11%
C1075	Xylosidase/arabinofuranosidase, Xsa43A	Bpr_I0302	3.2.1.55	N	7.6e-11	Y	4.3	57.5	17	33%
C1027	Xylosidase/arabinofuranosidase, Xsa43E	Bpr_I2319	3.2.1.37	N	3.8e-08	N	4.3	35.3	11	46%

^a PHX, predicted highly expressed gene; Pep, the number of matched peptides masses/searched peptides masses; Cov, Protein sequence coverage of the matched peptides.

^b Expectation score is the statistical probability of the top ranked protein match being a false positive identification, $p < 0.05$.

^c Secretory signal-peptides were predicted using SignalP (Ver. 3.0) (www.cbs.dtu.dk/services/SignalP/) and LipoP (Ver. 1.0) (<http://www.cbs.dtu.dk/services/LipoP/>).

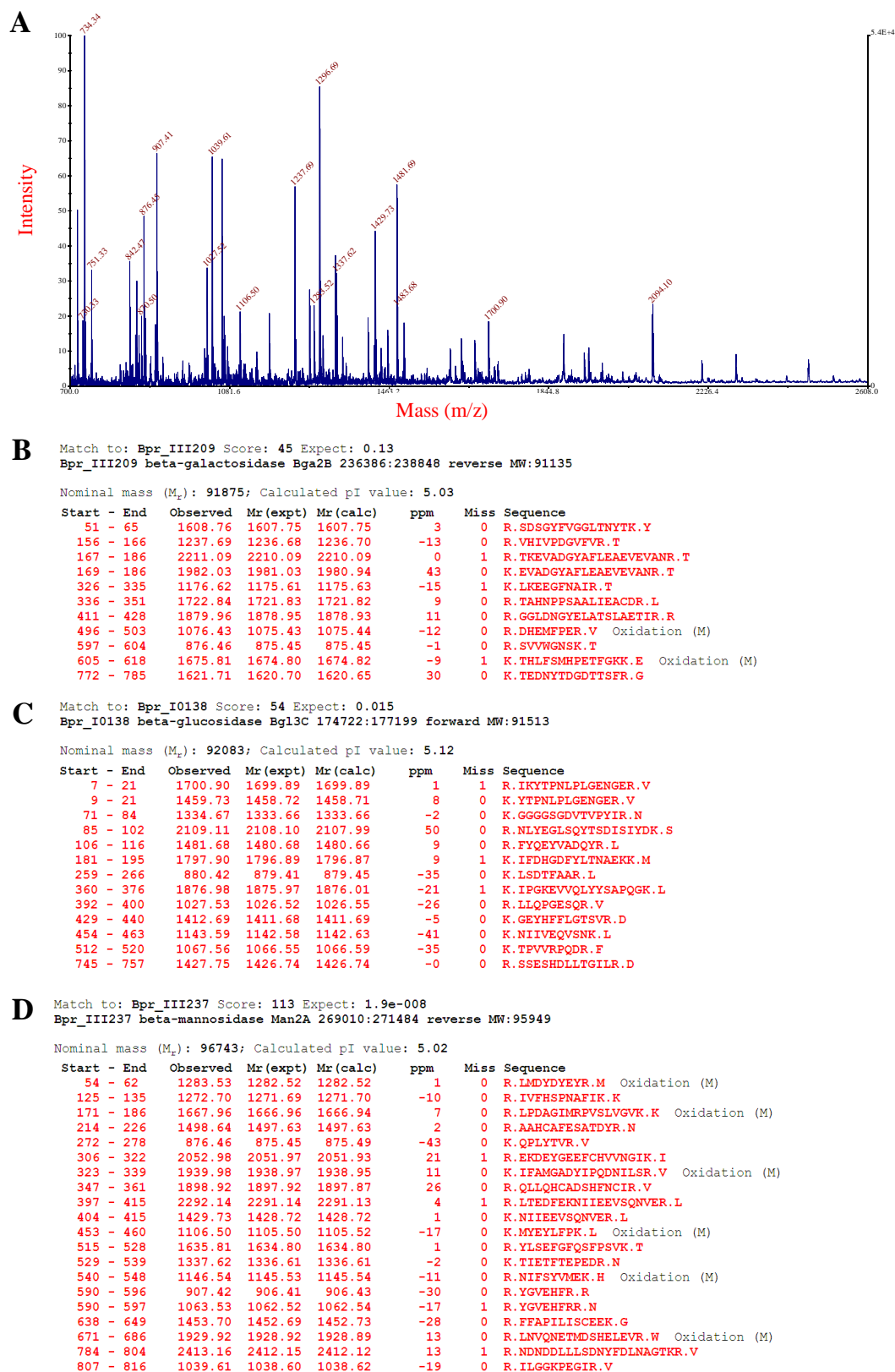
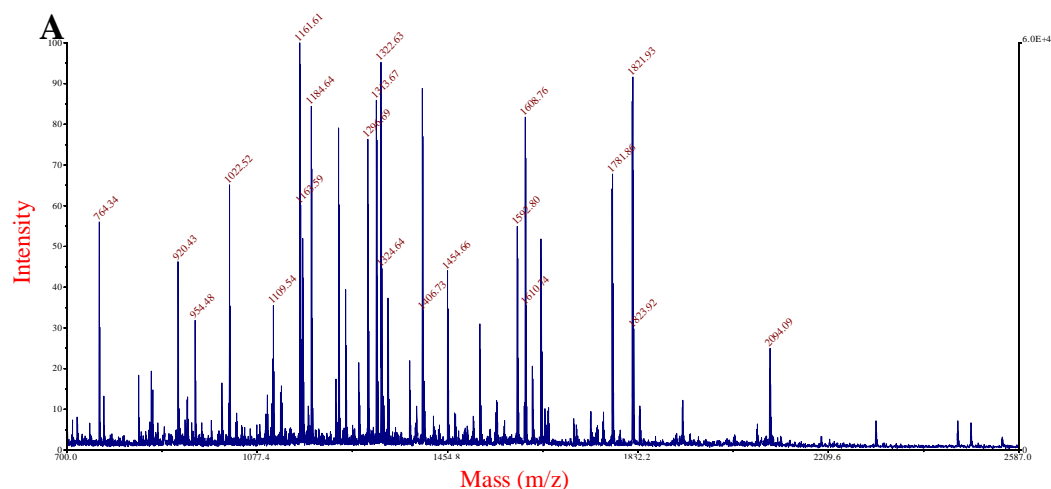


Figure 4.7. MS identification of Bga2B, Bgl3C, and Man2A. MALDI-TOF peptide mass fingerprint derived from spot C1147 (A) and MASCOT sequence coverage summary of the identified Bga2B (B), Bgl3C (C), and Man2A (D). Note that all peptide masses except m/z 876.46 were unique to the respective enzymes.



B

Match to: Bpr_I2006 Score: 101 Expect: 3e-007
Bpr_I2006 beta-galactosidase Bga35A 2445000:2447216 reverse MW:83162

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss Sequence
5 - 13	1022.52	1021.52	1021.55	-31	1 K.VMKITSEK.V
14 - 27	1518.83	1517.82	1517.82	2	1 K.VGFDNVSLVDGKR.W
93 - 101	1082.58	1081.58	1081.60	-17	1 K.KVGMYSILR.I Oxidation (M)
94 - 101	954.48	953.47	953.50	-29	0 K.KVGMYSILR.I Oxidation (M)
132 - 142	1313.67	1312.66	1312.65	6	0 R.SNAPQYLEHVR.N
233 - 246	1639.85	1638.85	1638.83	13	0 R.LTEIEPSGNYIFTR.E
288 - 303	1709.87	1708.86	1708.85	9	0 R.RPIATGLDTAAMTK.I 2 Oxidation (M)
326 - 333	947.49	946.48	946.51	-27	0 K.LTTLESR.A
334 - 353	2303.10	2302.09	2302.05	17	0 R.ATGYFNDLPEYSYDFNAPLR.E
354 - 363	1233.53	1232.52	1232.51	9	0 R.EFGQMEGTYSR.E Oxidation (M)
416 - 429	1780.86	1779.85	1779.82	15	0 R.EYNTGYLFVNNYQR.R
499 - 509	1379.63	1378.62	1378.60	11	0 R.DAYVFYTDTER.E
524 - 529	700.44	699.43	699.46	-45	0 R.IIVLSR.E
524 - 537	1551.85	1550.85	1550.88	-20	1 R.IIVLSREEAVHASK.V
577 - 587	1232.62	1231.62	1231.64	-22	0 R.VYPELAGELNKF
621 - 634	1592.80	1591.79	1591.78	7	0 K.VSFELTGNENDVLR.Y
729 - 736	939.47	938.47	938.51	-43	0 R.LSTQYSIK.I

C

Match to: Bpr_I0184 Score: 196 Expect: 9.6e-017
Bpr_I0184 beta-xylosidase Xyl3A 235721:237850 forward MW:78182

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss Sequence
28 - 35	920.43	919.42	919.44	-21	0 R.YDAPAIK.L
97 - 103	938.46	937.45	937.45	1	1 R.EDRDIYK.G
123 - 136	1608.76	1607.75	1607.72	17	0 R.GHETGSDPFLTSR.L
160 - 171	1322.63	1321.63	1321.62	8	0 K.HFAVHSGPEGER.H
172 - 177	764.34	763.33	763.37	-41	0 R.HFFDAK.A
182 - 196	1737.89	1736.88	1736.89	-2	0 K.DLEEVLPFAFEALVK.E
197 - 209	1454.66	1453.66	1453.65	5	0 K.EAEVAVMGAYNR.T Oxidation (M)
210 - 227	2092.02	2091.02	2090.95	30	0 R.TNGECCANKPLAVDTLR.G Oxidation (M)
252 - 261	1034.49	1033.48	1033.49	-12	0 K.VTSSPEESAK.L
345 - 357	1426.77	1425.76	1425.78	-13	1 K.KDGLLPINKDSIK.T
358 - 369	1184.64	1183.63	1183.63	-2	0 K.TIGVVGPNANSR.L
370 - 382	1404.74	1403.73	1403.72	9	0 R.LSLIGNVHTSSR.Y
383 - 393	1278.68	1277.67	1277.69	-12	0 R.YITVLEGIQDK.V
383 - 399	1920.02	1919.01	1919.00	5	1 R.YITVLEGIQDKVGDVVR.V
465 - 474	1167.67	1166.66	1166.68	-14	0 K.INLNPLSQK.Q
502 - 518	1820.93	1819.92	1819.89	18	0 K.AQDEANAVLQAFYFGAR.G
519 - 536	1704.85	1703.84	1703.84	5	0 R.GGADVADILFGDVSPSGK.L
537 - 543	867.46	866.46	866.49	-39	0 K.LPVTFFK.S
544 - 552	1007.46	1006.46	1006.46	-3	0 K.SADDLPDFK.D
544 - 557	1647.73	1646.72	1646.71	3	1 K.SADDLPDFKDYSMK.N Oxidation (M)
643 - 653	1178.62	1177.61	1177.63	-18	1 K.RVHVPAAGGTR.V
644 - 653	1022.52	1021.52	1021.53	-14	0 R.VHVPAAGGTR.V
662 - 672	1238.57	1237.56	1237.56	4	0 K.AFTSVNEEGER.A
681 - 691	1161.61	1160.60	1160.59	8	0 R.LYAGTQPGAR.S
700 - 709	1165.59	1164.58	1164.60	-13	0 K.CITVDFAIAR.-

Figure 4.8. MS identification of Bga35B and Xyl3A. MALDI-TOF peptide mass fingerprint derived from spot C1120 (A) and MASCOT sequence coverage summary of the identified Bga35B (B) and Xyl3A (C). Note that all peptide masses except m/z 1022.52 were unique to the respective enzymes.

4.2.3 Gel-free analysis of the *B. proteoclasticus* cytosolic proteome

Proteins extracted from the cytosol of xylan and xylose grown, mid-log phase harvested *B. proteoclasticus* cells were analysed by 1-D LC-MS/MS. Aliquots of the

three biological replicate samples of each growth condition that were analysed by 2-DE MALDI-TOF (Table 5.1) were purified by phenol/methanol/ammonium acetate precipitation, and in-solution trypsin digested. Each peptide mixture was concentrated using a C18 pipette tip, then bound to an HPLC-C18 reverse phase column, and eluted using a 77-minute LC gradient at a 200 nL.min⁻¹ flow rate. Peptide analysis was achieved using an LTQ MS/MS (ThermoFinnigan, San Jose, CA) operating in data dependent acquisition mode. Each biological replicate was analysed five times to give 15 replicate raw data files per growth substrate.

4.2.3.1 1-D LC-MS/MS protein identification and control of false positive identification rate

Each 1-D LC-MS/MS raw data file was searched in-house against the *B. proteoclasticus* protein sequence database using BioWorks (Ver. 3.1) (Thermo Finnigan, San Jose, CA). The peptide false discovery rate (FDR), defined as the proportion of significant matches that are null (Weatherly *et al.*, 2005), was determined by calculating the ratio of the number of unique peptides matched to the reverse-, and forward protein sequences, respectively (Elias *et al.*, 2005). BioWorks search parameters were empirically optimised to obtain a peptide FDR of less than 1% for each 1-D LC-MS/MS raw data file. Following peptide to protein assignment, the protein expectation value (less than 1.0e-4), cross correlation score (greater than 10), and minimum number of matched peptides per protein (minimum 2) were used as filters to obtain an average protein FDR rate across all technical replicate search results of below 1%. Protein FDR was calculated as the ratio of the number of protein identifications obtained from the reverse and forward protein sequences, respectively. The mean peptide and protein FDR values for each of the six biological replicate 1-D LC-MS/MS BioWorks results files is summarised in Table 4.2.

Proteins were included in the BioWorks identification dataset only if they were identified with statistical significance in at least two of the three biological replicate samples. Three-hundred and twenty-one proteins were identified in one or both culture conditions using the BioWorks search algorithm. Of these, 80% were identified in all three biological replicate samples of at least one growth condition, which signified a low level of analytical variation between the replicate 1-D LC-MS/MS analyses.

Table 4.2. Mean peptide and protein false positive rates for the BioWorks and Scaffold2™ analysis of the 1-D LC-MS/MS raw data files.

Sample name	BioWorks		Scaffold2™
	Mean peptide FDR	Mean protein FDR	Mean protein FDR
Xylan Rep_021107	0.45±0.14	0.53±0.64	0.52±0.01
Xylose Rep_021107	0.37±0.26	0.38±0.62	0.46±0.01
Xylan Rep_150108	0.45±0.24	0.92±0.32	0.52±0.01
Xylose Rep_150108	0.21±0.09	0.38±0.40	0.45±0.02
Xylan Rep_280508	0.43±0.17	0.10±0.23	0.49±0.04
Xylose Rep_280508	0.28±0.11	0.28±0.41	0.45±0.02

Scaffold2™ (Ver. Scaffold_2_01_02) (Proteome Software, Portland, OR) was also used to analyse the MS/MS based peptide and protein identifications. The five technical replicate raw data files for each biological replicate were uploaded to the Scaffold2™ software, condensed into a single file, and used to interrogate the *B. proteoclasticus* protein sequence database. Peptide identifications were accepted if they were established at greater than 95% probability. Protein identification was accepted only if it was established at greater than 95% probability and contained at least two top ranking peptide hits. Peptide and protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al.*, 2003). The mean protein FDR for each biological replicate set of Scaffold2™ output files are shown in Table 4.2, and demonstrate that the mean FDR values for each biological replicate are well below the 1% level. As with the BioWorks analysis, proteins were included in the Scaffold2™ dataset only if they were identified with statistical significance in at least two of the three biological replicate samples. Scaffold2™ identified 237 proteins in the cytosolic proteome of cells grown in the presence of xylan or xylose, of which 77% were identified in all three biological replicates in at least one growth condition.

In total, 329 non-redundant proteins were identified in the cytosol of cells grown in the presence of xylan or xylose, which included 50 proteins involved in carbohydrate metabolism. The importance of using multiple search algorithms to analyse MS/MS data is demonstrated by the fact that only 70% of the 329 1-D LC-MS/MS identified proteins were present in both datasets (Figure 4.9). Furthermore, only 35 of the 50 carbohydrate active proteins were identified by both search algorithms. Notably, the BioWorks search algorithm identified 93 proteins that were not identified by

Scaffold2™. In contrast, eight proteins were uniquely identified by the Scaffold2™ algorithm.

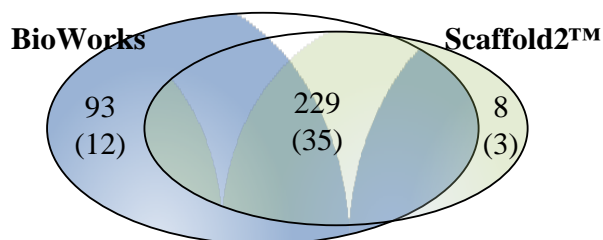


Figure 4.9. Protein identification overlap between BioWorks and Scaffold2™ analysis of the mid-log phase, cytosolic *B. proteoclasticus* proteome. Bracketed values denote the number of proteins predicted to be involved in carbohydrate metabolism.

4.2.3.2 Polysaccharide degrading enzymes identified by 1-D LC-MS/MS

The 1-D LC-MS/MS dataset contained 20 polysaccharide-degrading enzymes, eight of which had not been previously identified by the gel-based analyses (Table 4.3; the eight polysaccharidases unique to the 1-D LC-MS/MS dataset are shown in bold). Strikingly similar to the pattern observed in the gel-based analyses, all 20 1-D LC-MS/MS identified enzymes had a predicted *pI* value within a narrow range between 4.1 and 6.0. As predicted by the search algorithm protein identification overlap summarised in Figure 4.9, 19 of the 20 identified enzymes were identified using BioWorks, 13 of which were also identified by the Scaffold2™ algorithm. The GH78 α -L-rhamnosidase (Bpr_I1686) was the only enzyme identified uniquely by the Scaffold2™ search algorithm. In this case, the protein identification probability assigned by Scaffold2™ was 99.9%. The poorest false positive expectation value for the identification of each of the 19 other enzymes was almost two orders of magnitude lower than the cut-off score that yielded a protein FDR below 1% for the complete dataset. The false positive expectation value for the majority of the enzymes identified by BioWorks was at least six orders of magnitude below the cut-off score.

Table 4.3. Summary of polysaccharide degrading enzymes identified in the mid-log phase harvested *B. proteoclasticus* cytosolic proteome by 1-D LC-MS/MS.^a

Protein	Locus	EC	PHX ^b	Expect.	SigP ^c	pI	kDa	Algorithm	Pep. ^b	% Cov. ^b
α -Galactosidase, Aga36C	Bpr_III065	-	N	9.3e-05	N	5.1	83.2	BioW	7	4
α -D-Glucuronidase, Agu67A	Bpr_I0177	3.2.1.139	N	4.9e-11	N	4.9	76.1	BioW/Sc	22	13.9
α-Amylase, Amy13G	Bpr_I0729	-	N	1.7e-11	N	4.7	60.3	BioW	3	3.4
α-L-Arabinofuranosidase, Arf51A	Bpr_I0329	3.2.1.55	N	1.1e-15	N	5.2	57.0	BioW/Sc	14	14.6
β -Galactosidase, Bga2A	Bpr_I0279	3.2.1.23	N	1.5e-12	N	4.7	118.7	BioW/Sc	8	6.1
β -Galactosidase, Bga35B	Bpr_I2006	3.2.1.23	N	3.1e-10	N	4.9	83.2	BioW/Sc	4	3.8
β-Glucosidase, Bgl3A	Bpr_I0693	-	Y	2.5e-06	Y	4.1	115.6	BioW	23	3.9
β -Glucosidase, Bgl3B	Bpr_I0847	-	N	2.6e-11	N	4.7	103.7	BioW/Sc	7	6.7
β -Glucosidase, Bgl3C	Bpr_I0138	3.2.1.21	N	8.1e-07	N	4.9	91.5	BioW/Sc	2	5.5
Cellobiose phosphorylase, Cbp94A	Bpr_I2447	2.4.1.20	Y	1.1e-16	N	5.1	91.5	BioW/Sc	6	11
Cellodextrinase, Cel9B	Bpr_I1593	3.2.1.4	N	1.0e-30	N	4.6	61.0	BioW/Sc	22	18.3
Feruloyl esterase, Est1E	Bpr_I2870	-	N	1.3e-10	N	5.2	27.8	BioW/Sc	5	12.5
Glycoside hydrolase family 31, GH31C	Bpr_I1974	3.2.1.-	Y	1.0e-11	N	4.7	78.2	BioW/Sc	32	22.1
Glycogen phosphorylase, Glgp2	Bpr_I2847	2.4.1.1	N	1.8e-10	N	6.0	94.8	BioW/Sc	10	10.4
β -Mannosidase, Man2A	Bpr_III237	-	N	3.1e-11	N	4.8	95.9	BioW	7	4.1
Pullulanase, Pul13A	Bpr_III161	3.2.1.41	N	1.1e-10	Y	4.4	99.7	BioW	7	6.3
α-L-Rhamnosidase, Rha78A	Bpr_I1686	3.2.1.40	N	99.9	N	4.8	85.0	Sc	3	9.3

Table continues

Table 4.3 continued.

Protein	Locus	EC	PHX ^b	Expect.	SigP ^c	pI	kDa	Algorithm	Pep. ^b	% Cov. ^b
Xylosidase/arabinofuranosidase and esterase, Xsa43H	Bpr_I0301	-	N	1.6e-08	N	4.7	107.9	BioW/Sc	7	3.1
β -Xylosidase Xyl3A	Bpr_I0184	3.2.1.37	Y	8.9e-15	N	4.8	78.2	BioW/Sc	42	23.1
Endo-1,4-β-xylanase and esterase Xyn10D	Bpr_I1083	3.2.1.8	N	3.6e-10	N	5.0	79.7	BioW	5	7.4

^a Proteins shown in bold were uniquely identified by 1D LC-MS/MS. Where possible, the expectation values, and the number of top ranking matched peptides and percentage sequence coverage from the BioWorks (Ver. 3.1) are given. Scaffold2™ values denote protein identification probability (section 4.2.3.1)

^b PHX, predicted highly expressed gene; Pep, the number of matched peptides masses/searched peptides masses; Cov, Protein sequence coverage of the matched peptides.

^c Secretory signal-peptides were predicted using SignalP (Ver. 3.0) (www.cbs.dtu.dk/services/SignalP/) and LipoP (Ver. 1.0) (<http://www.cbs.dtu.dk/services/LipoP/>).

4.2.4 Bioinformatics of the identified *B. proteoclasticus* cytosolic proteome

Using the combination of gel-based and gel-free protein separation techniques, coupled to two mass spectrometry techniques, 28 polysaccharidases were identified in the cytosol of *B. proteoclasticus* cells grown in the presence of xylan or xylose. Secretory signal peptides were predicted by SignalP 3.0 (www.cbs.dtu.dk/services/SignalP/), LipoP 1.0 (<http://www.cbs.dtu.dk/services/LipoP/>), and pattern searching for Gram-positive bacteria (Bendtsen *et al.*, 2004, Juncker *et al.*, 2003, Sutcliffe and Harrington, 2002). Membrane-spanning domains were predicted using the TMHMM 2.0 (www.cbs.dtu.dk/services/TMHMM-2.0/) and SOSUI/G (<http://bp.nuap.nagoya-u.ac.jp/sosui/sosuiG/sosuiGsubmit.html>) utilities (Hirokawa *et al.*, 1998, Krogh *et al.*, 2001). Pfam analysis was used to examine the identity and location of carbohydrate active domains within each polysaccharidase. Where significant Pfam matches were not found, BLASTp and Tigrfam alignments were examined for significant matches.

4.2.4.1 Subcellular distribution of the polysaccharidases identified in the *B. proteoclasticus* cytosol

Among the set of 28 polysaccharidases identified in the *B. proteoclasticus* cytosol (summarised in Table 4.1 and Table 4.3) the β -glucosidase Bgl3A (Bpr_I0693), glycoside hydrolase family 30 Gh30A (Bpr_I2937), pullulanase Pul13A (Bpr_III161), and the xylosidase/arabinofuranosidase Xsa43A (Bpr_I0302) were each predicted to contain an N-terminal secretory signal peptide (Figure 4.11). Bgl3A and Pul13A also contained a single transmembrane helix within the C-terminal region, beginning at residues 1038 and 880 respectively, and it was therefore probable that Bgl3A and Pul13A were membrane-anchored proteins that had their N-terminal portions including the catalytic domains located outside the cell. GH30A and Xsa43A both contained lipobox motifs at their N-terminus, and were therefore likely to be lipoproteins that were tethered to the external surface of the cell wall. The four secretory signal peptide-containing enzymes were classified into four distinct GH families. Bgl3A, the GH30 family glycosyl hydrolase, and the xylosidase/arabinofuranosidase Xsa43A contained catalytic domains that were similar to domains found in other enzymes of the same GH families, but the pullulanase Pul13A contained a catalytic domain dissimilar to those found in any other identified enzymes.

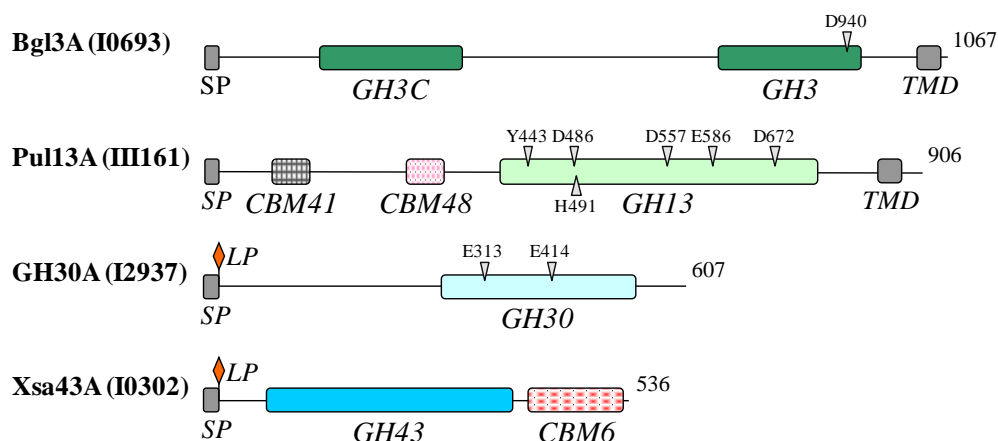


Figure 4.10. Functional domains of predicted secreted polysaccharidases identified in the *B. proteoclasticus* cytosol. Predicted GH domain active sites are shown where possible. Residue length is indicated to the right of each enzyme. CBM, carbohydrate binding module; LP, lipobox motif; PUD, bacterial pullulanase-associated domain; SP, secretory signal peptide; TMD, transmembrane domain.

4.2.4.2 Functional domains in the cytosolic polysaccharidases

Considering the 24 enzymes identified in the *B. proteoclasticus* cytosol that were predicted to be retained in the cytosol, collectively they represented 14 GH and two CE families (Figure 4.11 and Figure 4.12). They included three members each of the GH2 and GH3 families. The GH13 and GH43 families were each represented by two enzymes, and a single example of an additional 11 GH families was identified. Several carbohydrate esterases, a glycogen debranching enzyme Glx2 (Bpr_I1494) and a glycogen phosphorylase (Bpr_I2447) classified within the GT35 family were also identified. Furthermore, the CE2 domain containing acetyl-xylan esterase (Bpr_I0174) identified in spot C656 (Figure 4.5) was one of the five polysaccharidases within the predicted *B. proteoclasticus* proteome that are unmatched to a CAZy Database family.

Several enzymes contained more than one catalytic domain; including the four GH3 family enzymes that were notable in that they all contained catalytic domains matching to the same two Pfam families. The β -galactosidase Bga2A (Bpr_I0279) was distinct in that it was the only enzyme identified in the *B. proteoclasticus* that contained three catalytic domains. A distinguishing feature of the endo-1,4- β -xylanase and esterase Xyn10D (Bpr_I1083) and the xylosidase/arabinofuranosidase and esterase Xsa43H (Bpr_I0301) within the set of enzymes identified in the cytosol was that the two catalytic domains in each enzyme were predicted to be active upon different substrates. Of the remaining 18 enzymes, each contained a single catalytic

domain. The two CE2 domain-containing esterases contained a catalytic module that matched to the same Pfam family, as did the two α -galactosidases.

Several enzymes contained additional domains that may play a role in carbohydrate breakdown. The membrane anchored GH13 pullulanase and secreted GH43 xylosidase/arabinofuranosidase contained family 48 and family 6 CBMs, respectively. CBM48 modules are often found in combination with GH13 catalytic domains, and mediate enzyme attachment to α -1,4 linked glucose monomers, while CBM6 domains have been shown to target the parent enzyme to cellulose, xylan, and mixed linkage glucans. Furthermore, the GH13 pullulanase contained an N-terminal CBM41 pullulanase-associated domain that may also possess the ability to adsorb to carbohydrates. A CBMX domain and a CelD domain were found within the N-terminal region of the GH94 cellobiose phosphorylase and the GH9 cellodextrinase, respectively. Both domains are often associated with these types of enzymes, and both may play a role in enzyme/substrate attachment. The acetyl xylan esterase (Bpr_I0174) that was not classified to a CAZy family contained domains of unknown function (DUF) on either side of the esterase catalytic domain that both matched to the same protein family. In bacteria, DUFs are often found within sialic acid-specific 9-O-acetylerases or acetylxylan esterases.

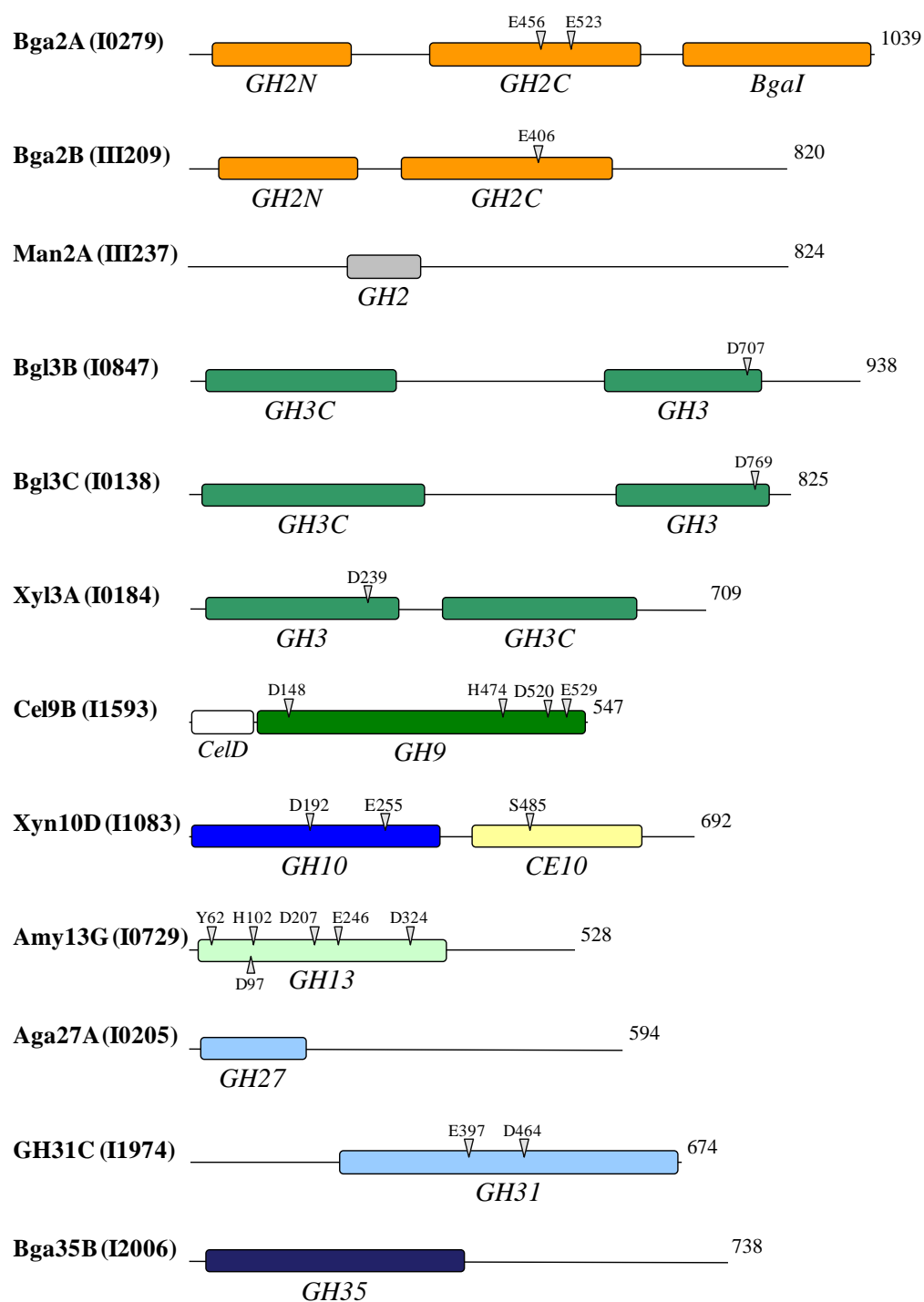


Figure 4.11. Functional domains of polysaccharidases identified in the *B. proteoclasticus* cytosol. Predicted GH domain active sites are shown where possible. Residue length is indicated to the right of each enzyme.

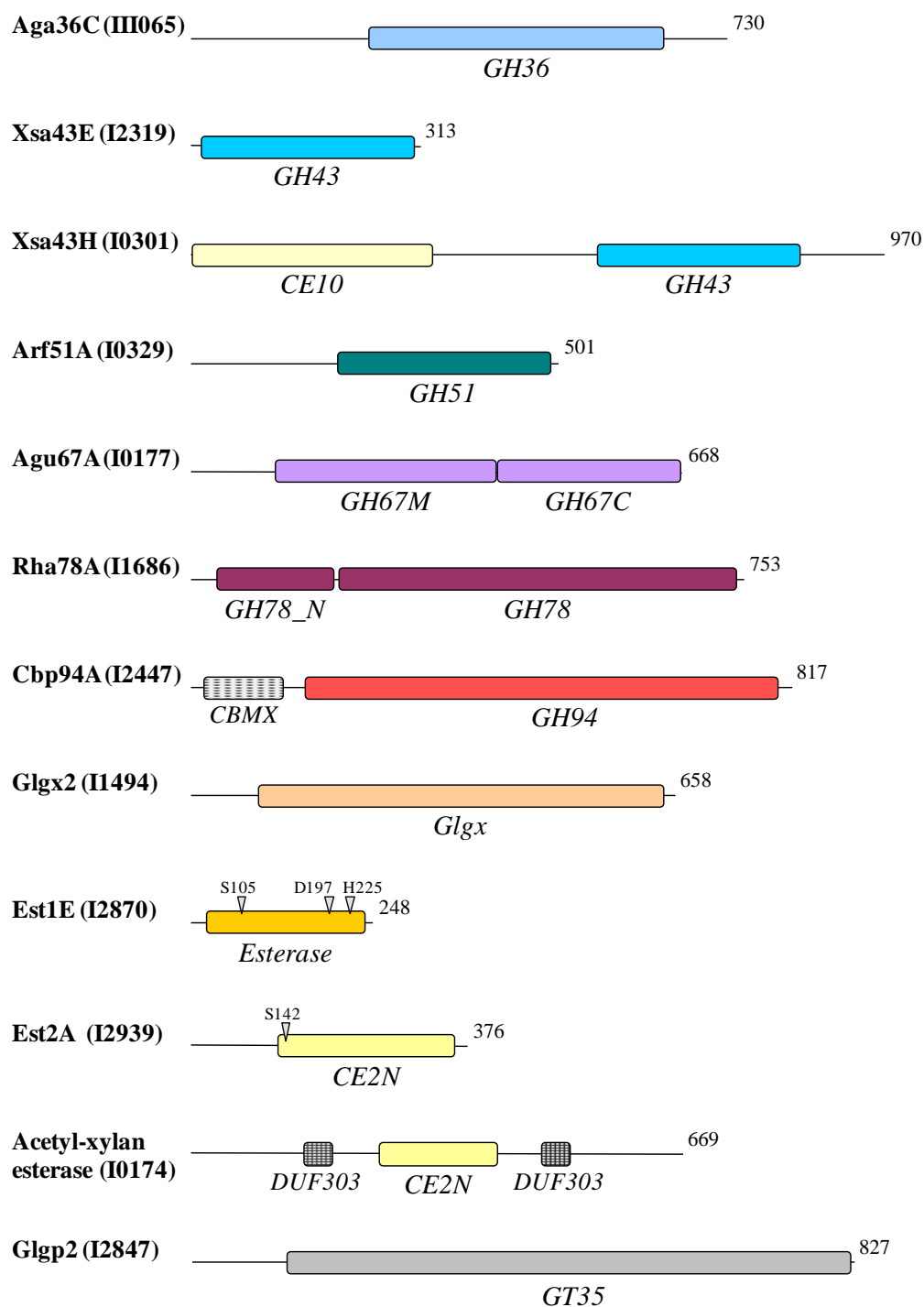


Figure 4.12. Functional domains of polysaccharidases identified in the *B. proteoclasticus* cytosol. Predicted GH domain active sites are shown where possible. Residue length is indicated to the right of each enzyme. CBM, Carbohydrate-binding module; DUF, Domain of unknown function.

4.2.4.3 BLASTp analysis of the polysaccharidases identified in the cytosolic *B. proteoclasticus* proteome

BLASTp analysis of each of 28 enzymes identified in the *B. proteoclasticus* cytosol demonstrated that all but the α -galactosidase Aga27A (Bpr_I0205) had a top ranking match to a carbohydrate active protein produced by an organism of the Clostridiales order (Table 4.4). Several enzymes including the GH9 cellodextrinase (Bpr_I1593), and GH10 endo-1,4- β -xylanase and esterase (Bpr_I1083) were at least 96% identical to homologues produced by *B. fibrisolvens* H17c(SA) (Berger *et al.*, 1990, Lin and Thomson, 1991a), and the CE1 feruloyl esterase (Bpr_I2870) was 68% identical to a cinnamoyl ester hydrolase produced by *B. fibrisolvens* E14 (Dalrymple *et al.*, 1996). Percentage identities between the *B. proteoclasticus* polysaccharidase and the top ranking BLASTp homologue ranged from 99% in the case of the GH9 cellodextrinase, to 34% for the β -glucosidase Bgl3A.

4.2.4.4 The predicted function of all identified cytosolic proteins

The predicted function of all 395 proteins identified in the *B. proteoclasticus* cytosol is summarised in Figure 4.13. The combination of gel-based and gel-free analyses of the *B. proteoclasticus* cytosol was successful in identifying a disproportionately large number of proteins involved in carbohydrate metabolism, which was almost 3-fold greater than the number predicted by the analysis of the theoretical total proteome (Appendix A, Table A.1), and signifies the utility of a multifaceted proteomic approach for examination of the fibre degrading enzyme system of *B. proteoclasticus*.

The majority of the 36 non-polysaccharide degrading proteins predicted to be involved in carbohydrate metabolism were components of the glycolytic or pentose phosphate pathways. The second most abundant functional category comprised proteins involved in protein synthesis, and significant proportions of the identified proteome were involved in amino acid biosynthesis, energy metabolism, and transmembrane transport. In particular, the 29 transporter proteins identified were comprised of 20 involved in carbohydrate transport, five that facilitated amino acid transport, and four that were not classified. Approximately 17% (71 proteins) of all identified cytosolic proteins were either hypothetical proteins, or proteins of unknown function.

Table 4.4. BLASTp analysis of the 28 polysaccharide degrading enzymes identified in the mid-log and stationary phase *B. proteoclasticus* cytosolic proteomes by 2-DE/MALDI-TOF.

<i>B. proteoclasticus</i> protein	Locus	Homologue ^a	Organism ^b	gi	Expect ^c	Identity	Region ^d
α -L-Arabinofuranosidase, Arf51A	Bpr_I0329	α -N-arabinofuranosidase 2	<i>Bryantella formatexigens</i> DSM 14469	255279950	1e-180	67%	1-497
α -Galactosidase, Aga27A	Bpr_I0205	Unknown	<i>Picea sitchensis</i>	148905920	3e-30	41%	28-220
α -Galactosidase, Aga36C	Bpr_III065	α -Galactosidase	<i>Roseburia intestinalis</i> M50/1	291537311	1e-180	61%	1-729
α -D-Glucuronidase, Agu67A	Bpr_I0177	α -Glucuronidase	<i>Clostridium</i> <i>phytofermentans</i> ISDg	160881284	1e-180	56%	3-668
α -Amylase, Amy13G	Bpr_I0729	α -Glucosidase	<i>Subdoligranulum</i> <i>variabile</i> DSM 15176	261366636	1e-174	58%	1-523
β -Galactosidase, Bga2A	Bpr_I0279	β -Galactosidase/ β -glucuronidase	<i>Ruminococcus obeum</i> A2-162	295110864	1e-180	53%	1-1039
β -Galactosidase, Bga2B	Bpr_III209	β -Galactosidase	<i>Bryantella formatexigens</i> DSM 14469	255280767	1e-180	41%	1-810
β -Galactosidase, Bga35B	Bpr_I2006	Galactosidase	<i>Cellulosilyticum</i> <i>ruminicola</i>	280977827	1e-180	52%	12-733
β -Glucosidase, Bgl3A	Bpr_I0693	β -Glucosidase-related glycosidases	<i>Ruminococcus obeum</i> A2-162	295108656	4e-160	34%	2-1066
β -Glucosidase, Bgl3B	Bpr_I0847	Glycoside hydrolase family 3 domain- containing protein sp. gb	<i>Ruminococcus</i> sp. 5_1_39B_FAA	253579190	1e-180	52%	6-924
β -Glucosidase, Bgl3C	Bpr_I0138	β -Glucosidase (EC 3.2.1.21)	<i>Butyrivibrio fibrisolvens</i>	144162	1e-180	98%	1-780

Table continues

Table 4.4 continued.

<i>B. proteoclasticus</i> protein	Locus	Homologue ^a	Organism ^b	gi	Expect ^c	Identity	Region ^d
Cellobiose phosphorylase, Cbp94A	Bpr_I2447	Cellobiose phosphorylase	<i>Butyrivibrio fibrisolvens</i> 16/4	291518248	1e-180	76%	1-816
Cellodextrinase, Cel9B	Bpr_I1593	Cellodextrinase	<i>Butyrivibrio fibrisolvens</i>	580767	1e-180	98%	1-547
Feruloyl esterase, Est1E	Bpr_I2870	CinI	<i>Butyrivibrio fibrisolvens</i>	1622732	9e-90	64%	1-248
Acetyl-xylan esterase, Est2A	Bpr_I2939	GDSL family lipase	<i>Clostridium cellulovorans</i> 743B	242260714	3e-101	51%	4-372
Acetyl-xylan esterase	Bpr_I0174	Protein of unknown function DUF303 acetylsterase putative	<i>Anaerocellum thermophilum</i> DSM 6725	222530666	1e-105	36%	4-655
Glycoside hydrolase family 30, GH30A	Bpr_I2937	Xylosidase/arabinofuranosidase Xsa43J	<i>Butyrivibrio proteoclasticus</i> B316	302672287	2e-174	64%	169-606
Glycoside hydrolase family 31, GH31C	Bpr_I1974	Glucosidase	<i>Cellulosilyticum ruminicola</i>	280977791	1e-180	72%	1-674
Glycogen phosphorylase, Glgp2	Bpr_I2847	Glycogen/starch/alpha-glucan phosphorylases	<i>Ruminococcus obeum</i> A2-162	295108762	1e-180	73%	8-824
Glycogen debranching enzyme, Glgx2	Bpr_I1494	Type II secretory pathway, pullulanase PulA and related glycosidases	<i>Ruminococcus obeum</i> A2-162	295110262	5e-115	37%	1-607
β -Mannosidase, Man2A	Bpr_III237	Putative β -mannosidase protein	<i>Bryantella formatexigens</i> DSM 14469	255281532	1e-180	57%	1-820
Pullulanase, Pul13A	Bpr_III161	Pullulanase	<i>Butyrivibrio crossotus</i> DSM 2876	260438049	1e-180	46%	1-869

Table continues

Table 4.4 continued.

<i>B. proteoclasticus</i> protein	Locus	Homologue ^a	Organism ^b	gi	Expect ^c	Identity	Region ^d
α -L-Rhamnosidase, Rha78A	Bpr_I1686	α -L-rhamnosidase	<i>Subdoligranulum</i> <i>variabile</i> DSM 15176	282600761	1e-180	48%	18-753
Xylosidase/arabinofuranosidase, Xsa43A	Bpr_I0302	Endo-1,4- β -xylanase D	<i>Roseburia intestinalis</i> L1-82	240144521	6e-173	55%	2-533
Xylosidase/arabinofuranosidase, Xsa43E	Bpr_I2319	Carbohydrate-binding family 6 protein	<i>Clostridium</i> <i>thermocellum</i> ATCC 27405	125974681	3e-107	65%	11-303
Xylosidase/arabinofuranosidase and esterase, Xsa43H	Bpr_I0301	Xylosidase/arabinosidase	<i>Roseburia intestinalis</i> L1-82	240144576	1e-180	70%	506- 968
β -Xylosidase, Xyl3A	Bpr_I0184	Glycoside hydrolase, family 3 domain- containing protein	<i>Ruminococcus</i> sp. 5_1_39B_FAA	253579611	1e-180	63%	3-707
Endo-1,4- β -xylanase and esterase Xyn10D	Bpr_I1083	β -1,4-D-xylanase	<i>Butyrivibrio fibrisolvens</i>	48963	1e-180	96%	1-635

^a BLASTp top ranked homologous protein.

^b Organism that expresses the BLASTp top ranked homologous protein.

^c Expectation score is the statistical probability of the top ranked protein match being a false positive identification, $p < 0.05$.

^d Values denote the first and last residues of the homologous region of the identified in *B. proteoclasticus* protein.

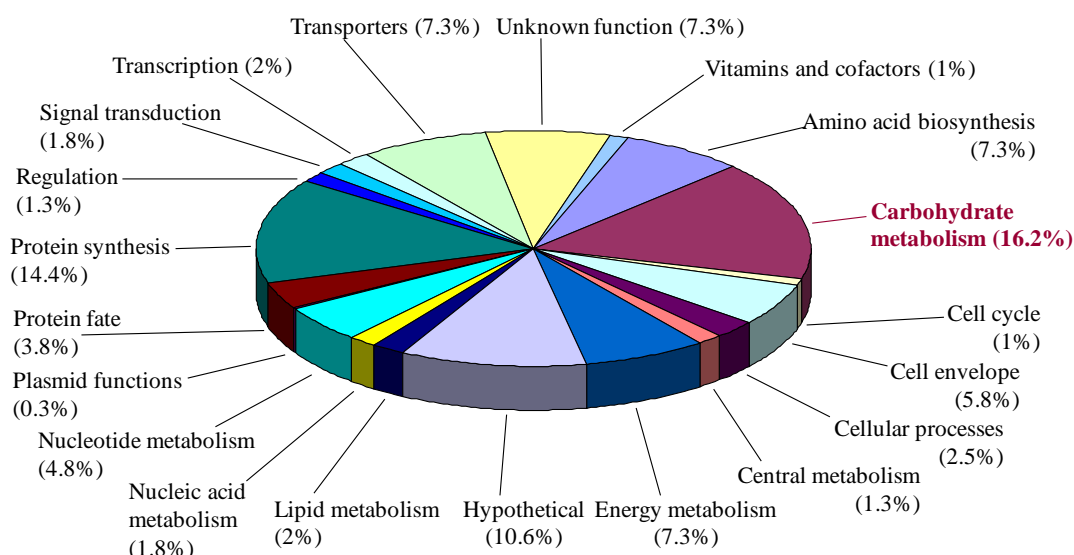


Figure 4.13. Protein function summary of all proteins identified in the *B. proteoclasticus* cytosol. Data compiled with the assistance of Dr. Bill Kelly and Dr. Sinead Leahy.

The 367 non-polysaccharide degrading enzymes proteins identified in the *B. proteoclasticus* cytosolic proteome are summarised in Appendix B, Table B.2 and Table B.3. The predicted *pI* values of the set of proteins range between 3.7 and 11.9, with a median value of 4.9. Two-hundred and sixty-six proteins had a predicted *pI* less than 5.6, 34 were between 5.6 and 6.5, and the remaining 67 had a *pI* value greater than 6.5. The smallest and largest proteins had a predicted size of 5.9 kDa and 195.3 kDa, respectively. SignalP analysis showed that 33 proteins contained an N-terminal secretory signal peptide, more than half of which were substrate-binding proteins associated with ABC-transporter systems. Notably, almost one third of the predicted secreted proteins were either hypothetical or proteins with no predicted known function. Six proteins identified in the *B. proteoclasticus* cytosol contained between five and nine TMDs. Strikingly all six were identified by 1-D LC-MS/MS analysis only (Appendix B, Table B.3). Three of these transmembrane proteins were permease proteins associated with ABC transport systems; one was a PTS system II ABC fructose-specific family protein, and the remaining two were proteins of unknown function.

4.2.4.5 Comparative analysis of the gel-based and gel-free datasets

Using two proteomic techniques to analyse the *B. proteoclasticus* cytosolic proteome offered the opportunity to compare and contrast the protein identification datasets obtained by each. The advantage of using a multi-technique approach to achieve improved proteome coverage is shown clearly in Figure 4.14.

Of the 395 proteins identified by either proteomic technique, only 40% were common to both. The gel-free analysis identified almost 50% more proteins than was identified by 2-DE MALDI-TOF, and increased the cytosolic proteome coverage by 77%. Furthermore, using both analytical techniques increased the number of proteins predicted to be involved in carbohydrate metabolism by more than 30%, and the number of identified proteins predicted to be involved in polysaccharide degradation and assimilation was increased by 40%, relative to the number achieved by either technique alone.

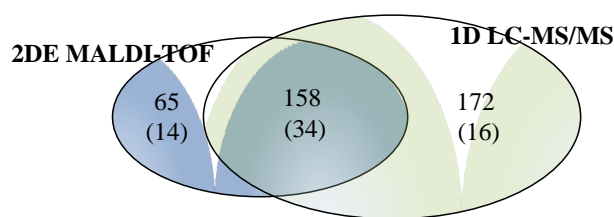


Figure 4.14. Protein identification overlap between the 2-DE MALDI-TOF and 1-D LC-MS/MS analyses. Bracketed values denote the number of proteins predicted to be involved in polysaccharide degradation and assimilation.

A two-dimensional plot of the proteins identified by 2-DE only, 1-D LC-MS/MS only, and those common to both datasets was constructed (Figure 4.15), which showed that 83% of all proteins identified in the *B. proteoclasticus* cytosol had a theoretical *pI* value below 6.5. In spite of the gel-based analyses focusing predominantly on the *pI* 3-6.5 region, 43% of all proteins that had a predicted *pI* value below 6.5 were identified by 1-D LC-MS/MS only. A small cluster of proteins with a *pI* value greater than 9.5 was identified by 1-D LC-MS/MS only, and was dominated by low molecular weight ribosomal proteins.

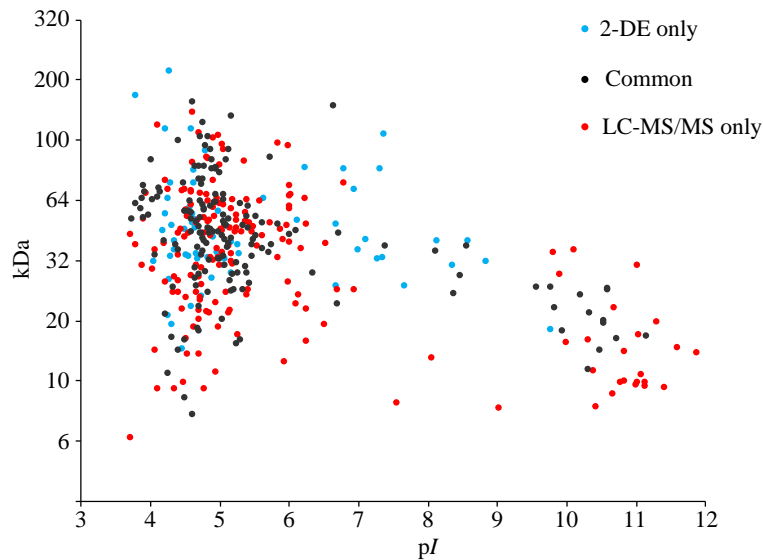


Figure 4.15. Two-dimensional plot of all proteins identified by 2-DE and/or 1-D LC-MS/MS analysis of the mid-log phase harvested *B. proteoclasticus* cytosol. The y-axis is presented in logarithmic scale to represent separation of proteins by 2-DE.

The predicted acidic skew was observed in the identified cytosolic proteome. A negative correlation between pI value and maximum protein size was also observed under experimental conditions ($r = -0.92$). Although the pI distribution of the 1-D LC-MS/MS identified proteins was comparable to that seen in the 2-DE analyses, the size distribution was noticeably different. To investigate whether the gel-free technique was biased towards the identification of lower molecular weight proteins, the mean molecular weights for the proteins identified by 2-DE only, 1-D LC-MS/MS only, and those common to both datasets were calculated and plotted (Figure 4.16). The sizes of the proteins identified by 1-D LC-MS/MS was significantly smaller than the sizes of the proteins identified by either 2-DE only, or those common to both techniques ($p < 0.01$). No significant difference in protein size was found between the sets of proteins identified by 2-DE only and those common to both techniques.

Analysis of the protein identification frequency within the gel-based and gel-free identified datasets, as a function of predicted biological activity (Figure 4.17), showed that 1-DE LC-MS/MS identified a much greater number of proteins involved in amino acid biosynthesis, lipid metabolism, protein synthesis, transcription, and transmembrane transport. Furthermore, proteins with unknown function were almost five-fold more prevalent in the 1-D LC-MS/MS dataset. Proteins involved in carbohydrate metabolism were one of the few categories not preferentially identified by either technique.

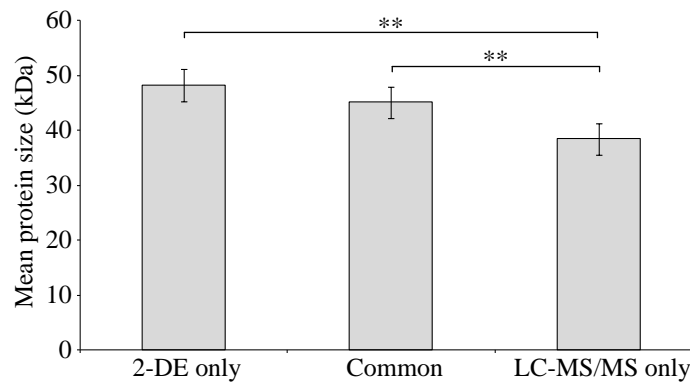


Figure 4.16. Analysis of the mean protein size for the subsets of cytosolic proteins identified by one or both proteomic techniques. Data shows that mean protein size of proteins identified by 1-D LC-MS/MS only was significantly smaller than proteins identified by 2-DE MALDI-TOF only, or by both techniques. Error bars are one standard error of the mean. Double asterisks denote significant Students *t*-tests ($p < 0.01$).

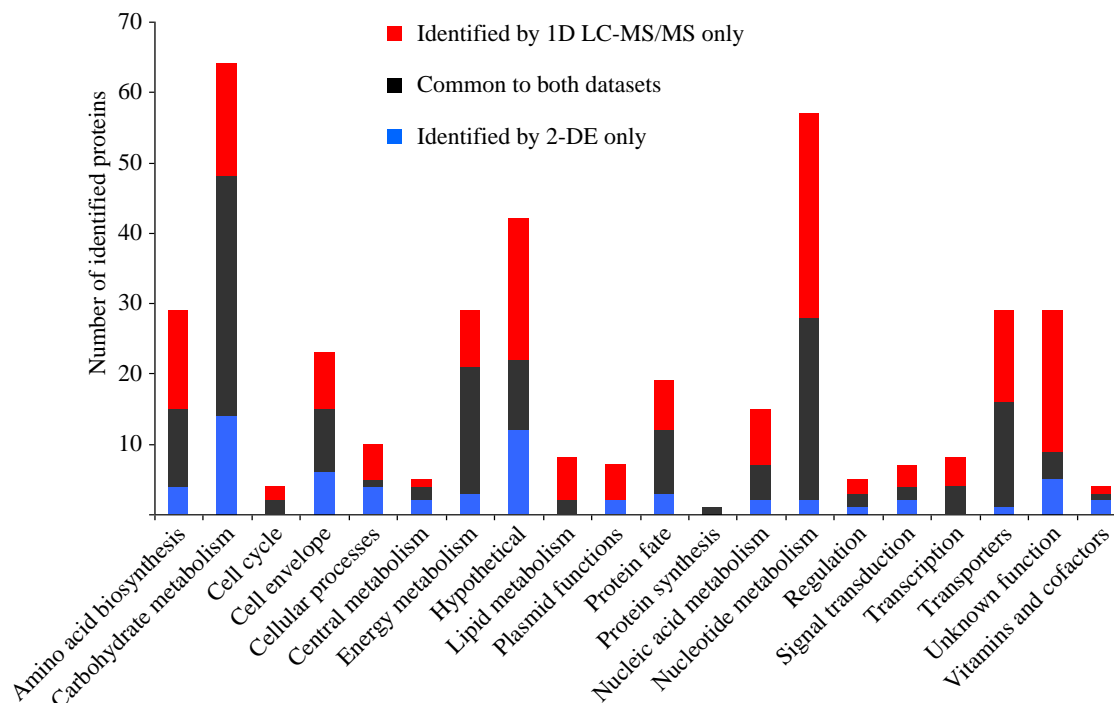


Figure 4.17. Summary of the number of proteins identified in the cytosolic proteome by 2-DE MALDI-TOF only, 1-D LC-MS/MS only, and common to both techniques, grouped according to predicted biological function.

4.2.4.6 Codon adaptation index, protein *pI*, and protein identification

The positive correlation between codon adaptation index (CAI) (Sharp and Li, 1987) and relative protein abundance in bacteria (Futcher *et al.*, 1999, Goetz and Fuglsang, 2005, Ishihama *et al.*, 2008, Jansen *et al.*, 2003, Kanaya *et al.*, 1999) was used to examine whether there was a relationship between CAI and *pI* value, and whether predicted relative protein abundance may have contributed to the acidic pattern observed in the 2-DE analyses of the *B. proteoclasticus* cytosolic proteome. A CAI value for each predicted cytosolic protein was calculated using 40 highly expressed *B. proteoclasticus* reference genes, and were used as an estimator of relative protein abundance. Ten percent of the cytosolic proteome had a CAI value greater than 0.595 and were consequently defined as predicted highly abundant proteins. The set of predicted highly expressed proteins was dominated by proteins involved in carbohydrate metabolism, energy metabolism, and protein synthesis. Notably, almost 11% of the predicted highly abundant proteins were hypothetical proteins.

Ranking the predicted cytosolic proteins in descending order of CAI value, and using the mean of 1200 protein spots routinely visualised in the 2-DE-based cytosolic separations as a guide, a CAI value of 0.454 was predicted to be the threshold for likely protein spot detection in the 2-DE separations. A scatter plot of the CAI value as a function of *pI* value for each predicted cytosolic protein was then constructed to obtain a visual representation of the correlation between predicted relative protein abundance and *pI* value (Figure 4.18).

Of the 1200 predicted cytosolic proteins that had a CAI value greater than 0.454, 85% had a *pI* value less than 7 (Figure 4.18, spots falling above the dotted line). The 15% of proteins above the CAI detection threshold that had a theoretical *pI* value greater than 7 were dominated by ribosomal proteins. Conversely, examination of the 641 predicted cytosolic proteins that had a predicted *pI* value greater than 7, 71% of the proteins had a CAI value below the predicted threshold (0.454) for likely detection by 2-DE, and were dominated by hypothetical proteins and proteins of unknown function (53%), proteins involved in exopolysaccharide synthesis (12%) and nucleic acid metabolism (8%). Examination of the CAI values of the identified cytosolic proteins showed that 86% had a CAI value greater than the 0.454 threshold for likely detection by 2-DE (Figure 4.18, red spots falling above the dotted line). Furthermore, the CAI value distribution of the identified cytosolic proteins differed markedly to that of the complete predicted cytosolic proteome (Figure 4.19).

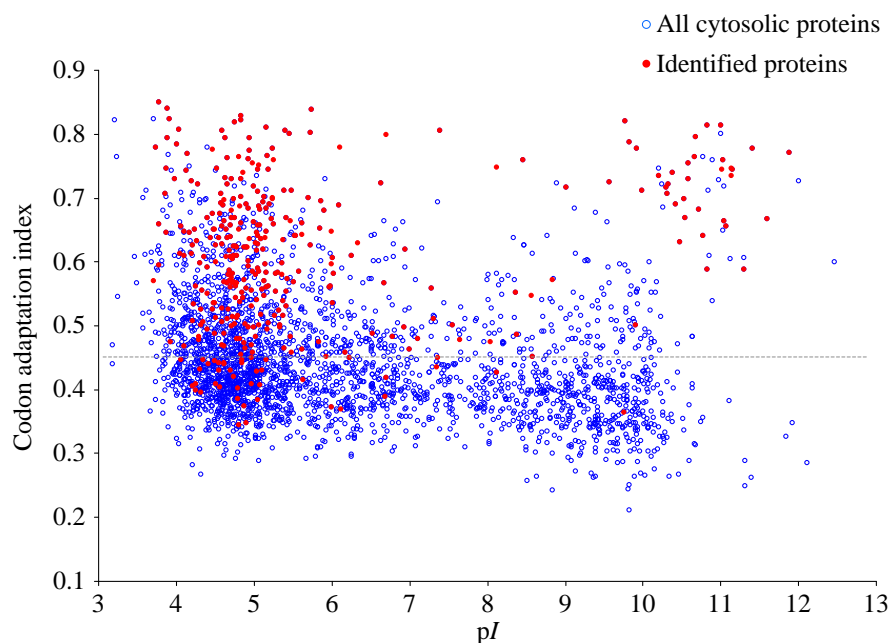


Figure 4.18. Scatter plot of the relationship between cytosolic protein codon adaptation index and pI value. Red spots represent cytosolic proteins identified in the cytosolic proteome, and blue open circles represent all unidentified cytosolic proteins.

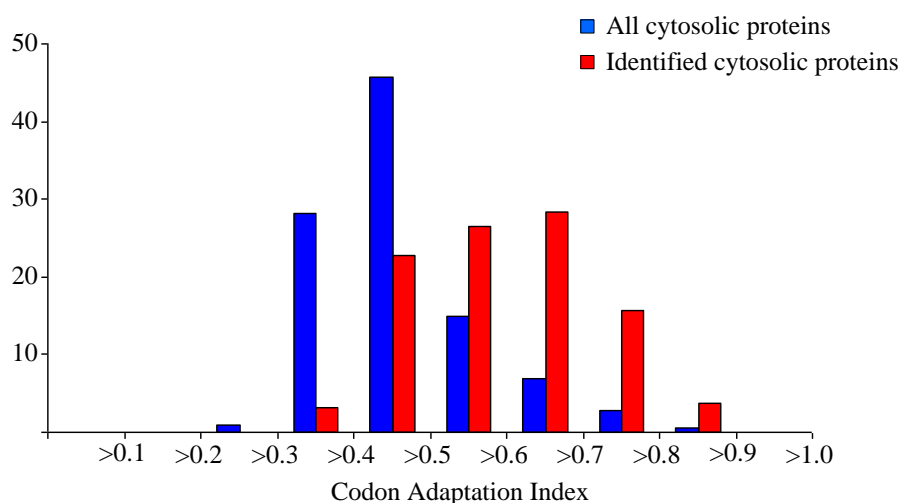


Figure 4.19. The frequency distribution of *B. proteoclasticus* cytosolic proteins as a function of codon adaptation index. Comparison of the frequency distribution of all cytosolic proteins with the set of identified cytosolic proteins shown.

4.3 Summary

The combination of gel-based and gel-free analysis of the cytosolic *B. proteoclasticus* proteome successfully identified 395 proteins, which included 24 cytosolic polysaccharidases, as well as four secreted enzymes that were not detected in the culture medium. This analysis raised the tally of polysaccharidases proven to be synthesised by cells grown in xylan containing culture medium to 36, which is 26% of the full catalogue of polysaccharidases encoded by the *B. proteoclasticus* genome. Present within the set of *B. proteoclasticus* cytosolic polysaccharidases are several classes of side chain removing enzymes that target substituent groups that the analysis of the *B. proteoclasticus* culture indicates may not be removed in the extracellular space. In particular, enzymes such as the α -L-arabinofuranosidase Arf51A, the α -D-glucuronidase Agu67A, and the feruloyl esterase Est1E contained GH domains that are predicted to target arabinofuranose, glucuronic acid, and ferulic acid side groups were all identified. Many of the identified cytosolic polysaccharidases were predicted to be components of a PUL, and two-thirds of the *B. proteoclasticus* PULs contained at least one identified protein. Collectively this data supports the likelihood that PULs are important to *B. proteoclasticus* mediated fibre degradation.

Supplementing the gel-based analysis of the *B. proteoclasticus* cytosolic proteome with the gel-free analysis increased the number of protein identifications more than 75% and made an important contribution to the examination of *B. proteoclasticus* hemicellulose degradation. Notably, the gel-free analysis increased the total number of identified polysaccharidases by almost 50%, and identified five polysaccharidases predicted to be involved in hemicellulose degradation, that were not detected in the gel-based analysis.

Chapter 5

The effect of growth substrate on the *B. proteoclasticus* cytosolic proteome

5.1 Introduction

The combination of gel-based and gel-free proteomics identified an array of cytosolic polysaccharidases that will enable *B. proteoclasticus* cells to utilise the assimilated oligomers of extracellular GAX degradation. Collectively, 14 GH and two CE families were represented including three members of both the GH2 and GH3 families, two examples of each of the GH13 and GH43 families, and an unclassified CE2 family acetyl-xylan esterase. Differential expression proteomics was therefore used to examine the relative abundance of each of these polysaccharidases, so as to gain a clearer insight into the relative importance of each of the GH families, and in particular those containing several enzymes, to the overall process of GAX degradation by *B. proteoclasticus*.

To examine the effect of cultivation in the presence of simple and complex plant carbohydrates on cytosolic protein abundance patterns *B. proteoclasticus* cells were grown *in vitro*, in modified M704 culture medium supplemented with either 0.1% xylan or 0.5% xylose, or in unsupplemented modified M704 medium. The effect of cultivation in each of these conditions was analysed in cells harvested at mid-log and stationary phase using 2-DE. Differential protein abundance patterns in cells harvested at mid-log phase were also examined using gel-free, 1-DE LC-MS/MS.

5.2 Statistical analysis of differential protein abundance patterns

5.2.1 Narrow pI range analysis of the *B. proteoclasticus* cytosolic proteome

Three biological replicate samples for each growth condition were used to examine the substrate dependent protein abundance profiles of cells harvested at mid-log and stationary phase (Table 5.1).

Each biological replicate was analysed three times using pI 3-5.6 IPG strips, and each gel was loaded with 100 µg of total protein. Due to limited sample amounts some biological replicates were analysed twice using pI 5.3-6.5 IPG strips as indicated (Table 5.1). Only the stationary phase harvested samples were analysed using pI 6-11 IPG strips, and each strip was loaded with 300 µg of total protein in an attempt to resolve low abundance proteins in this region. Limited sample dictated that each biological replicate was only analysed once using pI 6-11 IPG strips.

Table 5.1. Summary of the biological replicate samples and pI ranges used to examine the cytosolic proteome of mid-log phase harvested, *B. proteoclasticus* cells. ^a

Growth phase	Substrate	pI range	Biological replicate 1	Biological replicate 2	Biological replicate 3
Mid-log	Xylan	3-5.6	Rep_021107	Rep_150108	Rep_280508
		5.3-6.5	Rep_260606	Rep_021107	Rep_150108
	Xylose	3-5.6	Rep_021107	Rep_150108	Rep_280508
		5.3-6.5	Rep_260606	Rep_021107	Rep_150108
	No added carbon	3-5.6	Rep_021107	Rep_150108	Rep_280508
		5.3-6.5	Rep_260606	Rep_021107	Rep_150108
Stationary	Xylan	3-5.6	Rep_180407	Rep_060607	Rep_070707
		5.3-6.5	Rep_180407	Rep_060607	Rep_070707
		6-11	Rep_060607	Rep_070707	
	Xylose	3-5.6	Rep_180407	Rep_060607	Rep_070707
		5.3-6.5	Rep_180407	Rep_060607	Rep_070707
		6-11	Rep_060607	Rep_070707	
	No added carbon	3-5.6	Rep_180407	Rep_060607	Rep_070707
		5.3-6.5	Rep_180407	Rep_060607	Rep_070707
		6-11	Rep_060607	Rep_070707	

^a pI 3-5.6 and pI 5.3-6.5 IPG strips were loaded with 100 µg total protein per strip. pI 6-11 IPG strips were loaded with 300 µg total protein per strip. Biological replicates ($n = 3$) of all pI 3-5.6 and pI 5.3-6.5 2-DE separations were analysed using three technical replicates per sample.

5.2.2 2-DE spot volume normalisation

Normalisation of raw spot volumes prior to comparative examination of protein abundance profiles between samples of interest is important for quantitative 2-DE analysis. Small variations in total protein load, IEF, first-to-second-dimension transfer, gel staining, and image acquisition can cause spot volume variation between technical and biological replicate gels. Raw spot volumes for each 2-DE separation were extracted from Image Master™ 2D Platinum. The total spot volume for each separation was calculated, and each spot volume was normalised against the amount of total protein per gel. Normalised spot volumes were then subjected to log-ratio transformation (Aitchison, 1982, Beddek *et al.*, 2008). To assess the effect of normalisation, box-and-whisker plots of the \log_{10} transformed spot volume values before and after normalisation were constructed for all matched protein spots (minimum seven of nine matched spots) from the mid-log phase, xylan grown cytosolic pI 3.-5.6 gel set. Without normalisation (Figure 5.1A) considerable variation exists in the median, upper and lower quartiles and minimum values across the nine replicate gels. Comparison of Figure 5.1A and Figure 5.1B shows that normalising the raw spot volume values against the total detectable protein in each 2-DE separation reduced analytical variation between replicate gels.

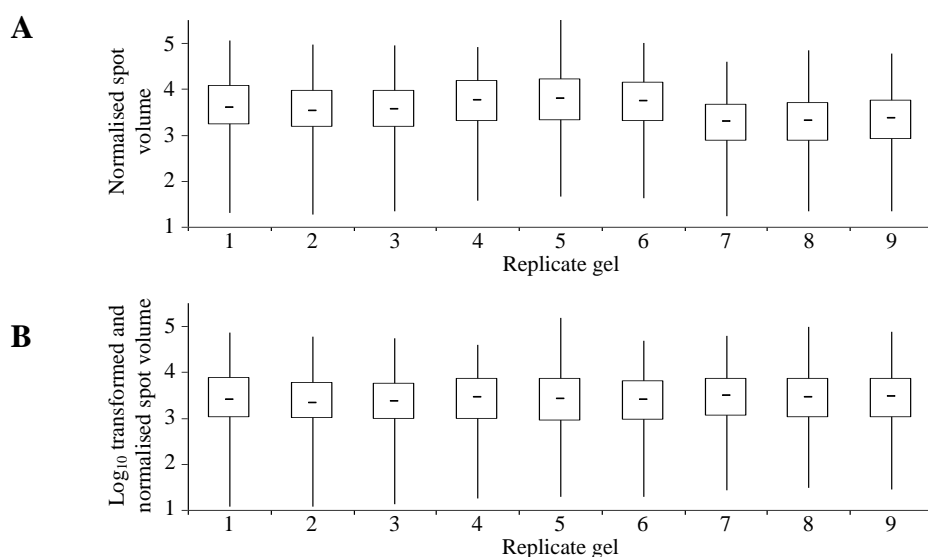


Figure 5.1. The efficiency of normalisation in correcting for differences in mean spot volume values due to analytical variation. Box-and-whisker plots are shown for the \log_{10} transformed spot volumes (A) and the \log_{10} transformed normalised spot volumes (B) for all matched spot groups (minimum eight spots from a possible nine) from the *B. proteoclasticus* mid-log phase, xylan grown cytosolic replicate gel set.

5.2.3 Distribution of missing values in 2-DE

The occurrence of missing values (spots missing in one or more replicate gels) is a common problem in 2-DE proteomic analyses, and causes problems for statistical analysis of quantitative 2-DE experiments because the power of the parametric tests used to assess statistical significance of abundance changes is reduced when comparing groups of unequal sample size (Albrecht *et al.*, 2010). Prior to performing statistical testing, the presence of missing values should be examined and dealt with in an appropriate manner. To investigate the likely cause of the majority of missing values in the 2-DE data set the relationship between mean spot volume and the rate of missing values per spot group was examined (Figure 5.2). The plot shows that spot groups with one or more missing values had a mean normalised spot volume approximately five-fold less than those with no missing values. The relationship between mean spot volume and missing value rate suggested that the majority of missing values were most likely a result of spots being below the limit of detection, and is in agreement with the findings of Grove *et al.* (Grove *et al.*, 2006). Consequently, missing data values were imputed with the minimum spot volume value detected from each 2-DE image giving rise to the missing value, which is similar to the method used by Meleth *et al.* (2005).

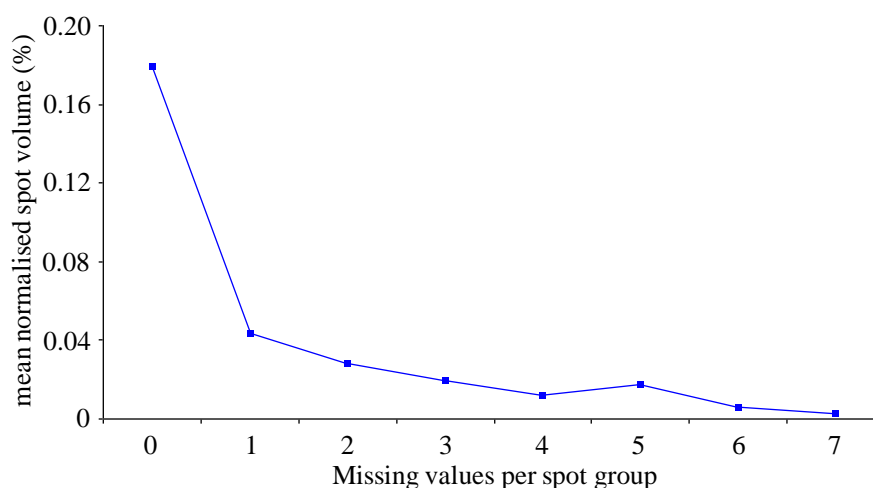


Figure 5.2. The relationship between the number of missing values per matched spot group, and the mean normalised spot volume. Matched spot groups from the *B. proteoclasticus* xylan grown, mid-log phase harvested cytosolic gel set were grouped according to the number of missing values per spot group.

5.2.4 Missing value substitution rate in 2-DE

To investigate the missing value rate across all matched spot groups in the pI 3-5.6 gel sets, the number of missing values as a proportion of the total number of data points was calculated at each missing value level. Figure 5.3 demonstrates that imputing data values into matched spot groups in the pI 3-5.6 gel sets containing no more than two missing value would result in a final data substitution rate of 5.6%, which is less than that previously reported for proteomic analyses (Ahmad *et al.*, 2006, Chang *et al.*, 2004), and is also less than that advised for microarray data (Ahmad *et al.*, 2006). For statistical analysis of protein abundance profiles, data values were therefore imputed into spot groups containing no more than two missing values in the pI 3-5.6 gel sets, and spot groups containing three or more missing values were removed from the analysis.

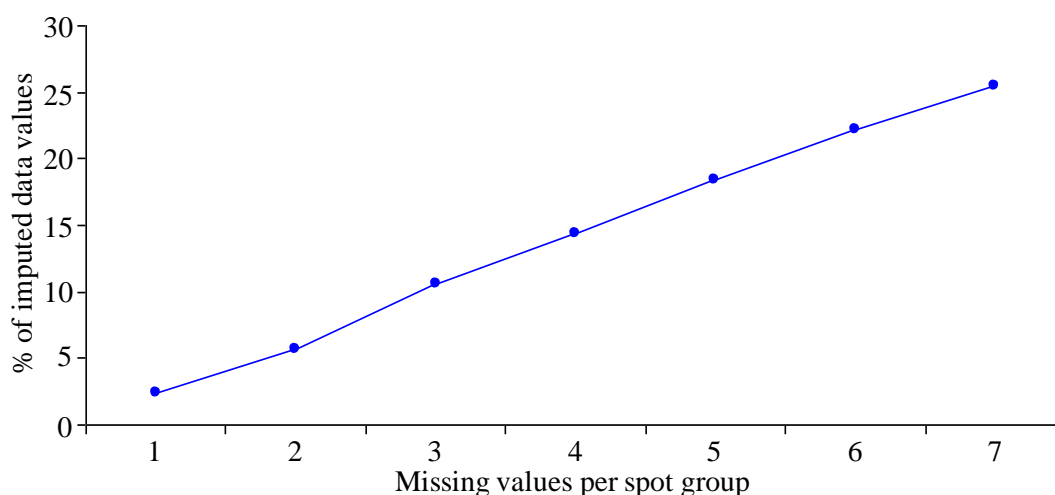


Figure 5.3. Analysis of the proportion of imputed data values as a function of missing value frequency per spot group.

Data values were also substituted into mid-log phase pI 5.3-6.5 spot groups that contained no more than two missing values, and matched spot groups containing three or more missing values were deleted. No data values were substituted into stationary phase pI 5.3-6.5 spot groups because the gel sets comprised six replicate gels and imputing values into spots groups with at least one missing value would have lead to an imputation rate greater than 10%. All spot groups with one or more missing values were removed from the analysis. Respectively, a total of 492 and 459 individual data points were imputed into the mid-log and stationary phase cytosolic pI 3-5.6 datasets, and 133 data points were imputed into the mid-log pI 5.3-6.5 dataset. The total number

of matched spot groups suitable for statistical analysis of protein abundance profiles on account of the number of missing values was consistent across each replicate gel set (Table 5.2). The larger number of total matched spot groups present in the mid-log phase gel sets is due to the larger number of mean total spots detected in the mid-log phase gel set.

Table 5.2. Total numbers of matched spot groups per cytosolic pI 3-5.6 and pI 5.3-6.5 replicate gel sets submitted for statistical analysis of protein abundance profiles.

Sample	Mid-log		Stationary	
	pI 3-5.6	pI 5.3-6.5	pI 3-5.6	pI 5.3-6.5
Xylan	540	75	464	79
Xylose	484	86	399	100
No added carbon	464	127	399	105

5.2.5 2-DE data transformation

Parametric testing was used where possible to examine the significance of differential abundance patterns of proteins matched between growth conditions. The validity of these tests is dependent on a number of assumptions being satisfied, and it was therefore necessary to ensure that these assumptions were satisfied prior to conducting statistical analysis of differential protein abundance profiles. Parametric tests assume that the experimental data being tested is obtained from normally distributed populations, and that no significant correlation exists between the mean and variance of each set of data. Transformation is often required to improve the distribution characteristics of the data in order to satisfy both assumptions. In proteomics, \log_{10} transformation has often been shown to stabilise variance versus mean dependency. The effect of \log_{10} transformation was examined using scatter plots of the variance versus mean dependency of spot groups in the stationary phase harvested, xylan grown pI 3-5.6 data set. The high positive correlation of the untransformed data (Figure 5.4A) is reduced by \log_{10} transformation (Figure 5.4B). For data sets of these sizes, small correlation coefficients can be significant; therefore a more effective transformation was sought in an attempt to reduce the variance versus mean correlation further. After iterative testing, a power^{1/5} transformation was optimal for the 2-DE data set (Figure 5.4C). Pearson correlation coefficients of the spot group variance versus mean relationship were calculated for each data set before and after power^{1/5} transformation (Table 5.3). In each data set no significant correlation was

found after transformation, and the assumption of variance versus mean independence was deemed to be satisfied.

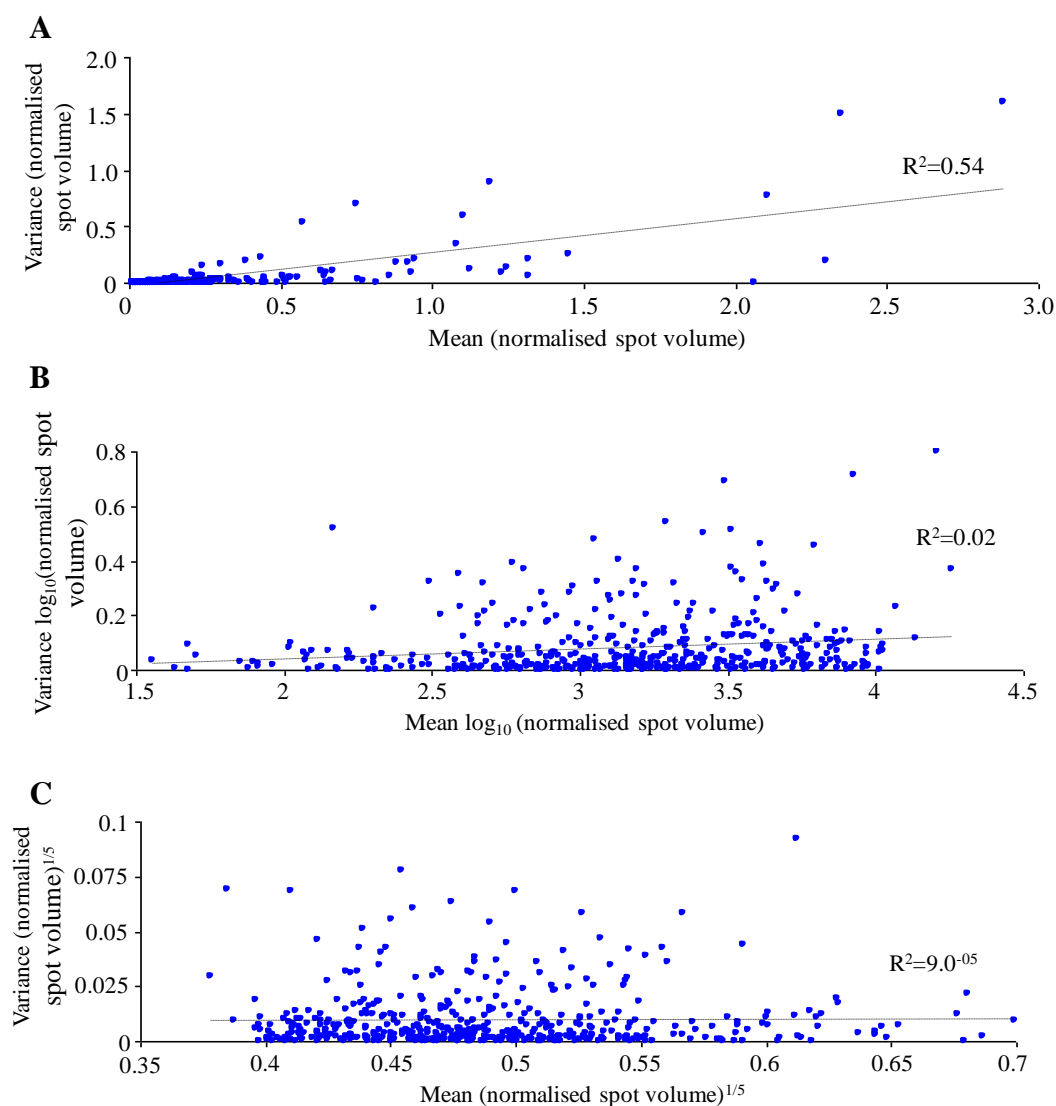


Figure 5.4. The efficiency of different transformations in correcting for variance versus mean dependency. Variance versus mean scatter plots of the stationary phase harvested, xylan grown 2-DE data set show the normalised data exhibited a strong variance versus mean dependency before transformation (A), which was reduced by \log_{10} transformation (B), and reduced further by power $^{1/5}$ transformation (C).

Table 5.3. Effect of power^{1/5} transformation on the variance versus mean dependency of spot groups in the 2-DE data sets.^a

Sample	Mid-log phase		Stationary phase	
	Normalised ^b	Normalised ^{1/5c}	Normalised	Normalised ^{1/5}
Xylan	0.788	-0.109	0.735	0.010
Xylose	0.479	-0.088	0.758	0.094
No added carbon	0.884	-0.111	0.690	-0.005

^a Values are the Pearson correlation coefficient for mean versus variance before and after transformation, for all matched spot groups in the mid-log and stationary harvested cytosolic proteome separations. Spot groups with at least eight data values were included in the analysis.

^b Pearson correlation coefficient for mean against variance of normalised spot volumes.

^c Pearson correlation coefficient for mean against variance of log₁₀ transformed normalised spot volumes.

Prior to statistical testing the normality of every spot group after power^{1/5} transformation was tested using the Shapiro-Wilk test (Shapiro and Wilk, 1965b), which is a goodness-of-fit test used to assess whether a sample is drawn from a normal distribution and was developed for small sample sizes. Spot groups with a significance score less than 0.05 were considered to exhibit a non-normal distribution. In these cases the Mann-Whitney U test (Mann and Whitney, 1947), which does not assume population normality, was used to assess statistical significance of protein abundance changes. The proportion of spot groups that failed the normality test in each power^{1/5} transformed data set is summarised in Table 5.4.

Table 5.4. Normality of matched spot groups after power^{1/5} transformation using the Shapiro-Wilk goodness-of-fit test.^a

	Mid-log phase		Stationary phase	
	pI 3-5.6	pI 5.3-6.5	pI 3-5.6	pI 5.3-6.5
Xylan	7.2	18.8	9.1	6.0
Xylose	6.6	18.1	9.8	10.5
No added carbon	11.9	31.0	12.5	22.2

^a Values are the percentage of spot groups that failed the normality test for each data set.

5.2.6 Multiple hypothesis testing and controlling the false discovery rate

Two-tailed Student's *t*-tests were performed for each spot group matched between xylan and xylose, or xylan and no added carbon growth conditions, in the mid-log and stationary phase gels sets. The mean normalised spot volume for each biological replicate gel set was calculated using the technical replicate gels. These mean values

were then used to calculate statistical significance of protein abundance changes between growth conditions.

When conducting a large number of simultaneous hypothesis tests, the multiple hypothesis testing problem is concerned with controlling the Type-I error rate (the rate of false positive outcomes), while maximising the power of the test (the ability to detect a difference when one truly exists). The family-wise-error-rate (FWER), or the probability of generating one or more false positives out of all the hypotheses tested is a commonly controlled parameter, using a method such as the Bonferroni-Holm correction (Holm, 1979). The FWER is extremely conservative and can lead to a significant loss of statistical power. It has been suggested that a more suitable approach for discovery based proteomics experiments is to control the false discovery rate (FDR) (Benjamini and Hochberg, 1995), which is the proportion of false positives among the abundance changes identified as being significant. A *q*-value (Storey, 2002) (<http://www.genomine.org/qvalue/>) is calculated for each *p*-value, which can then be used to establish a *q*-value threshold for the individual experiment, and consequently obtain a balance between the FDR and power of an analysis (Storey and Tibshirani, 2003). The cumulative frequency of significant tests and the expected number of false positives at *q*-value thresholds ranging between 0.02 and 0.1 is summarised in Table 5.5.

Table 5.5. Estimation of the differential abundance false discovery rate.^a

<i>q</i> -value threshold	FDR	Number of significant spots ^b		Estimated number of false positives ^c	
		Mid-log	Stationary	Mid-log	Stationary
0.02	2%	5	2	0	0
0.04	4%	9	5	0	0
0.06	6%	20	8	1	0
0.08	8%	30	13	2	1
0.1	10%	41	20	4	2

^a The *q*-value was used to estimate the number of false positive significant differential expression results at the designated *q*-value thresholds, in the mid-log and stationary phase cytosolic 2-DE analyses.

^b Cumulative values are given for the combined total of significant spots between the xylan and xylose and/or xylan and no added carbon growth conditions.

^c The cumulative number of estimated false positive differentially expressed results is calculated as FDR multiplied by the total number of significant spots.

Plotting the cumulative frequency of significant tests and expected false positives as a function of percentage FDR allowed guided selection of the most appropriate q -value cut-off for the cytosolic 2-DE analysis. At a 5% FDR, a near maximal separation between the total number of significant tests and the expected number of false discoveries occurred (Figure 5.5). At this level, the number of expected false discoveries was less than one. Selecting a 6% FDR would have resulted in the number of expected false discoveries rising to two. Therefore, calling significant all tests with a q -value below 0.05 maximised the balance between the statistical power of the 2-DE analyses, and the total number of expected false discoveries.

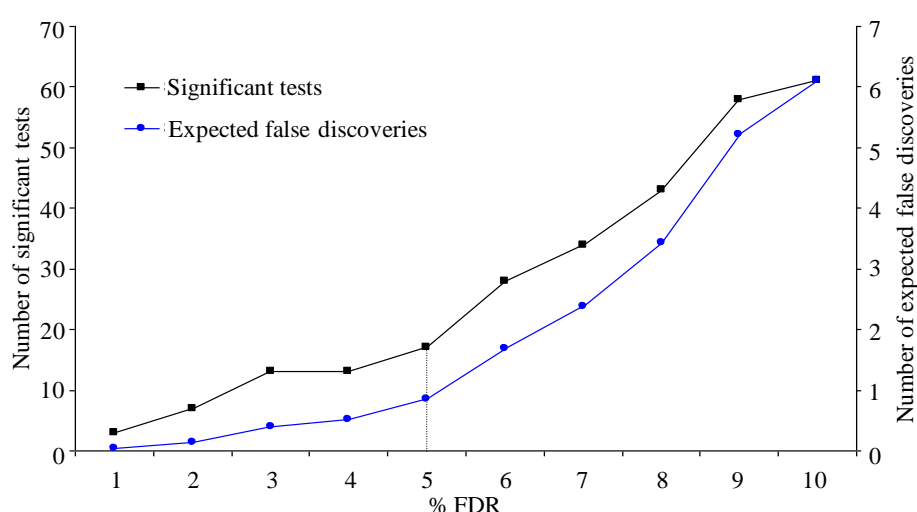


Figure 5.5. The cumulative frequency of 2-DE significant tests and expected false discoveries as a function of percentage FDR. At 5% FDR (dotted line) the significant tests (black line, primary y-axis), and the expected number of false discoveries (blue line, secondary axis) were 17 and less than 1, respectively. All tests with a q -value below 0.05 were designated significant in the cytosolic 2-DE analyses.

5.3 Relative abundance analysis of the mid-log phase cytosolic proteome by 2-DE MALDI-TOF

The outcome of the differential abundance analysis of the mid-log and stationary phase *B. proteoclasticus* cytosolic proteomes are summarised in Table 5.6. During mid-log or stationary phase growth, a total of 17 spot group pairs were at least 2-fold more abundant in the xylan growth condition, relative to either the xylose or no added carbon growth conditions. In addition, between two and 15 spots were uniquely detected in one of the three growth conditions at mid-log or stationary phase, many of which were of very low abundance.

Table 5.6. Summary of the differential abundance analysis of the mid-log and stationary phase *B. proteoclasticus* cytosolic proteomes at a 5% false discovery rate.^a

	Mid-log			Stationary		
	Matched groups	Increased abundance	Decreased abundance	Matched groups	Increased abundance	Decreased abundance
Xylan/Xylose	429	3 (3)	6 (6)	361	1 (1)	1 (1)
Xylan/No added carbon	430	1 (1)	1 (1)	349	4 (4)	0
	Unique			Unique		
Xylan		15 (7)			14 (5)	
Xylose		12(2)			2(0)	
No added carbon		3(0)			4(0)	

^a The upper portion of the table shows the total number of differentially expressed spots (FDR = 5%) in the xylan growth condition, relative to the xylose and/or no added carbon growth conditions for each growth phase. In the lower portion, the number of spots unique to each growth phase is shown. Bracketed values denote the number of identified proteins for each group of differentially expressed proteins.

In the *B. proteoclasticus* mid-log phase proteome the abundance of 10 spots was significantly different between xylan and xylose grown cells at the 5% FDR (Figure 5.6 and Figure 5.7). Four were more abundant in the xylan growth condition relative to xylose, and the remaining six were detected at lower abundance in xylan-grown cells. Proteins were identified in all 10 spots of interest, and are summarised in Table 5.7. The theoretical pI and size of each differentially abundant protein matched closely with the observed position on the 2-DE map.

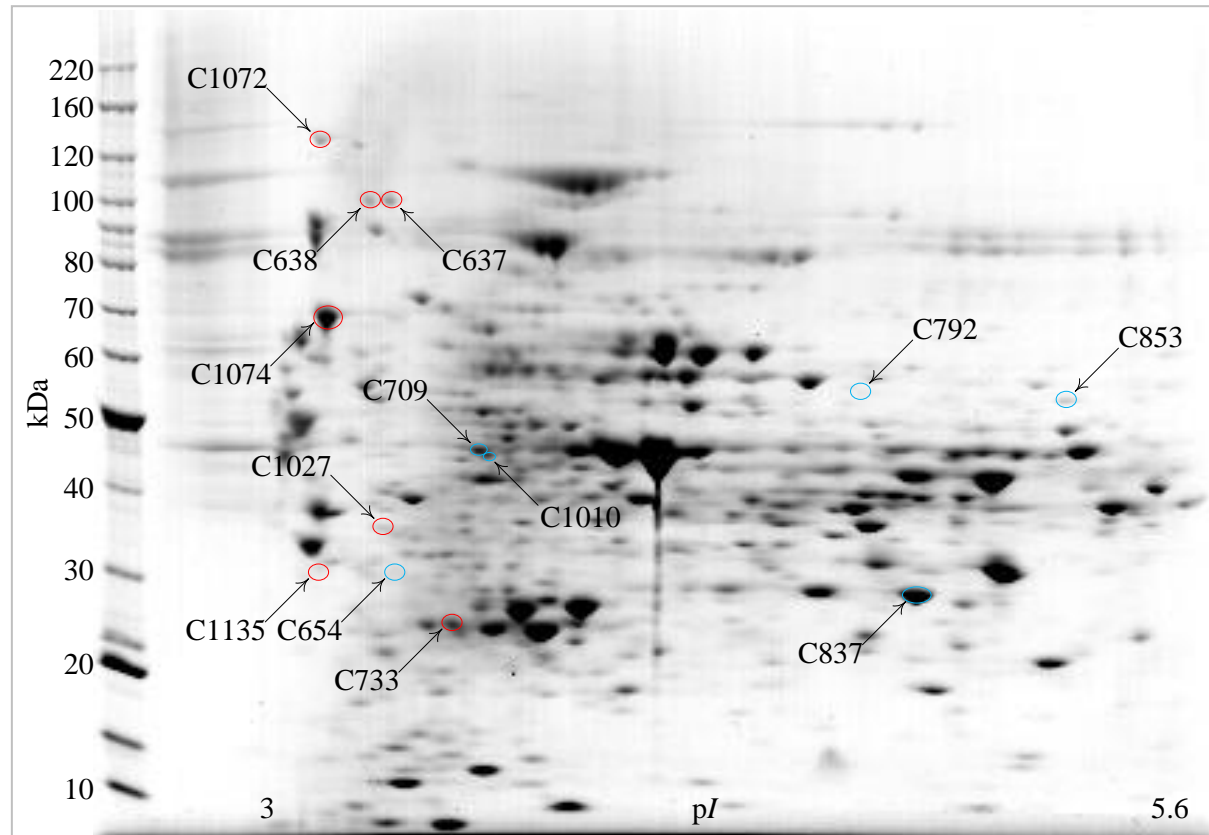


Figure 5.6. Differentially abundant spots between the mid-log phase harvested xylan grown cytosolic proteome and the alternate two growth conditions. A representative pI 3-5.6 2-DE colloidal CBB G-250 stained reference gel of cytosolic proteins harvested from mid-log phase, xylan grown *B. proteoclasticus* cells is shown. Red and blue circles denote spots with increased and decreased abundance respectively in the xylan growth condition. Circled spots are summarised in Table 5.7.

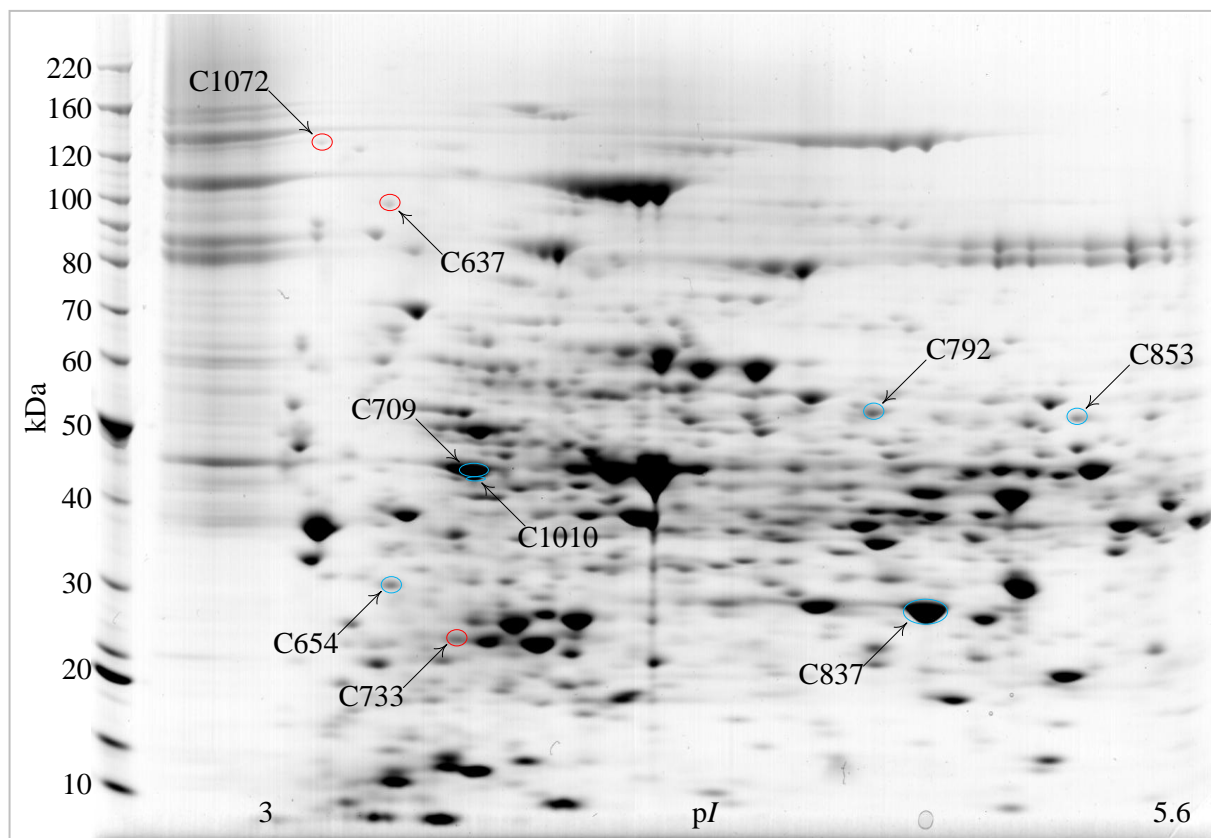


Figure 5.7. Differentially abundant spots between the mid-log phase harvested xylan and xylose grown cytosolic proteomes. A representative pI 3-5.6 2-DE colloidal CBB G-250 stained reference gel of cytosolic proteins harvested from mid-log phase, xylose grown *B. proteoclasticus* cells is shown. Red and blue circles denote spots with increased and decreased abundance respectively in the xylan growth condition. Circled spots are summarised in Table 5.7.

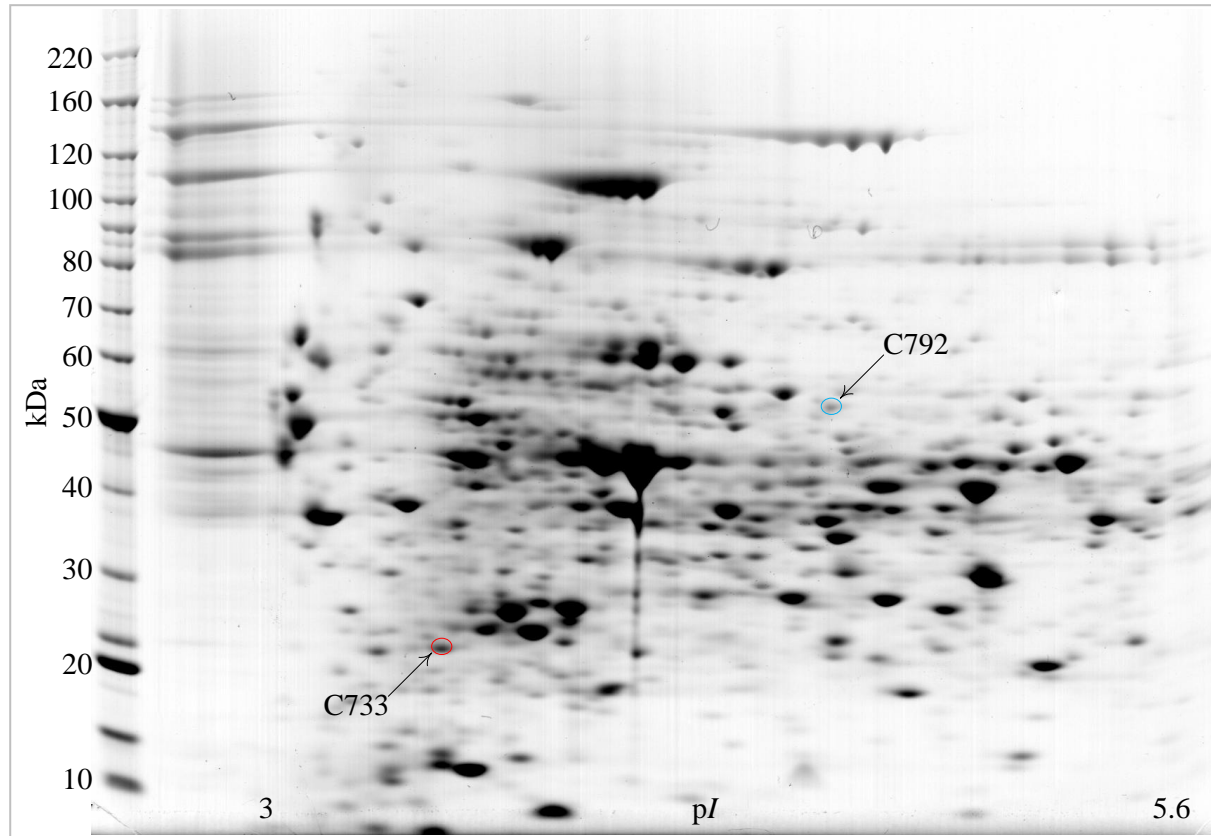


Figure 5.8. Differentially abundant spots between the mid-log phase harvested xylan and no added carbon grown cytosolic proteomes. A representative pI 3-5.6 2-DE colloidal CBB G-250 stained reference gel of cytosolic proteins harvested from mid-log phase, *B. proteoclasticus* cells grown in unsupplemented medium is shown. Red and blue circles denote spots with increased and decreased abundance respectively in the xylan growth condition. Circled spots are summarised in Table 5.7.

Table 5.7. Summary of the differentially abundant proteins between xylan and xylose, or xylan and no added carbon growth conditions, identified in the mid-log phase cytosolic proteome.

Spot	Protein	Locus	Function ^a	Expect.	pI	kDa	Pep.	Cov.	Fold-change ^b	p value/q value
C853	Adenylosuccinate lyase, PurB	Bpr_I2212	O	1.2e-10	5.4	53.7	16	28%	-3.1±0.6	0.001/0.020
C1135	Amino acid ABC transporter substrate-binding protein	Bpr_I1826	T	5.8e-04	4.0	31.6	7	21%	Xylan	n/a
C1010	Anti-sigma factor antagonist/phosphotransferase domain-containing protein	Bpr_I0249	R	3.0e-30	4.6	49.8	20	48%	-12.0±0.7	0.001/0.029
	Ribosomal protein S1, RpsA	Bpr_I2035	Q		4.6	41.9	19	42%		
C654	DNA-directed RNA polymerase α subunit, RpoA	Bpr_I0623	W	3.0e-17	4.4	35.1	16	57%	-8.7±4.3	0.002/0.049
C837	Fructose-1,6-bisphosphate aldolase, FbaA	Bpr_I2903	C	3.0e-13	5.1	30.5	14	49%	-2.1±0.2	0.002/0.048
C733	IMP cyclohydrolase, PurO	Bpr_I0731	O	1.5e-18	4.7	32.2	10	39%	2.3±0.5	0.001/0.029
	Translation elongation factor Tu, TufA	Bpr_I2364	Q		4.82	43.6	14	50%		
C792	NADPH-dependent glutamate synthase, GltA3	Bpr_I1306	A	1.9e-04	5.0	49.3	8	20%	-15.8±6.3	0.000/0.001
C637	Pullulanase, Pul13A	Bpr_III161	C	1.5e-09	4.4	99.7	20	22%	2.8±0.3	0.000/0.017
C638	Pullulanase, Pul13A	Bpr_III161	C	1.8e-06	4.4	99.7	11	11%	Xylan	n/a
C709	Ribosomal protein S1, RpsA	Bpr_I2035	Q	9.6e-19	4.6	41.9	19	42%	-4.1±1.1	0.001/0.020
C1072	Serine protease subtilisin family	Bpr_I2629	P	3.8e-08	3.8	153.3	10	9%	5.8±1.3	0.000/0.017
C1074	Sugar ABC transporter substrate-binding protein	Bpr_I0182	T	1.5e-08	4.12	63.3	14	31%	Xylan	n/a
C1027	Xylosidase/arabinofuranosidase, Xsa43E	Bpr_I2319	C	3.8e-08	4.3	35.3	11	46%	Xylan	n/a

^a A, Amino acid biosynthesis; C, Carbohydrate metabolism; O, Nucleotide metabolism; P, Protein fate; Q, Protein synthesis; R, Regulation; T, Transporters; W, Transcription.

^b Fold-change was calculated as the ratio of the mean normalised protein spot volumes in the xylan/xylose or xylan/ no added carbon growth conditions as stated. Positive values denote proteins with increased mean normalised spot volume in xylan grown cells. Xylan / xylose denotes the protein was uniquely detected in culture medium harvested from xylan / xylose grown cells respectively.

Spots C733 and C1010 each contained two proteins. In both spots, the detected proteins were identified with no overlap between the subsets of peptide masses matched to each protein, and the false positive identification score for each pair was $1.5\text{e-}18$ and $3.0\text{e-}30$ respectively. In the remaining eight differentially abundant spots a single protein was identified. Two of the nine differentially abundant spots were also significantly different between the xylan and no added carbon growth conditions (Figure 5.6 and Figure 5.8). The direction of change between the xylan and xylose or xylan and no added carbon growth condition was the same in both these cases. At the 5% FDR, no spots were significantly different between the xylan and no added carbon growth conditions only.

In addition to the differentially abundant spot pairs, 15 spots detected in the xylan growth condition were undetected in both the xylose and no added carbon conditions, and proteins were identified in seven of them. It is possible that each of these spots is uniquely produced by cells grown in the presence of xylan, or that the corresponding spots in the xylose and/or no added carbon conditions are below the limit of detection using colloidal CBB G-250 total protein staining. The identity of three of these proteins was of particular interest. The amino acid ABC transporter substrate-binding protein (Bpr_I1826) and the xylosidase/arabinofuranosidase Xsa43E (Bpr_I2319) detected in spots C1135 and C1027 respectively, were the only proteins unique to the xylan growth condition at either the mid-log or stationary growth phase that were not identified in any other spots in any other growth condition or growth phase (Figure 5.6 and Table 5.7). Both proteins were detected in all nine replicate gels of the xylan growth condition. The SBP product of Bpr_I0182 was the only protein identified in the differentially or uniquely abundant spots in the mid-log phase *B. proteoclasticus* cytosol that was a component of a polysaccharide utilisation locus.

The remaining eight protein spots unique to the xylan growth condition were unidentified due to poor quality PMF, which was a consequence of their low abundance. The majority of the protein spots uniquely detected in either the xylose or no added carbon growth conditions were unidentified for the same reason, and those that were identified were not unique to the growth condition in which they were detected.

The most strongly up- and down regulated proteins in the mid-log phase proteome of xylan grown cells relative to either alternate condition were the subtilisin family serine protease (Bpr_I2629) (spot C1072) and the glutamate synthase (NADPH) GltA

(Bpr_I1306) (spot C792), respectively. Proteins identified in seven of the differentially abundant spots were predicted to be involved in cellular metabolism. In particular, the pullulanase Pul13A (Bpr_III161) and xylosidase/arabinofuranosidase Xsa43E (Bpr_I2319) which were identified in spots C637 and C638, and C1027 respectively were predicted to be involved in the degradation of plant carbohydrates.

A complex protein spot pattern of the spots identified as the 99.7 kDa pullulanase, Pul13A (Bpr_III161) was seen between the mid-log (Figure 5.9 A) and stationary phase (Figure 5.9) time points.

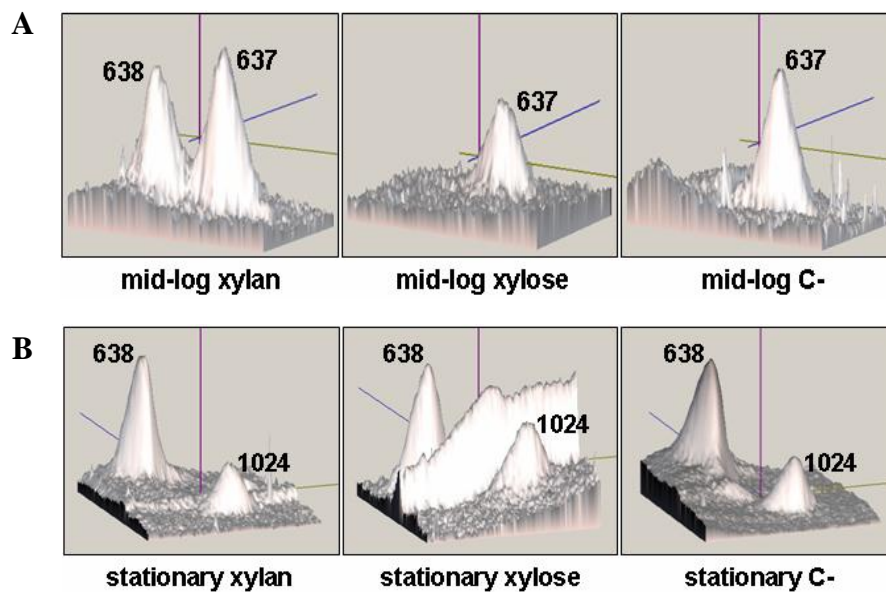


Figure 5.9. Effect of growth substrate and growth phase on pullulanase, Pul13A (Bpr_III161) abundance. (A) Corresponding regions of the mid-log phase xylan, xylose and no added carbon reference gels showing the abundance profiles of spots C637 and C638, which were both identified as pullulanase, Pul13A (Bpr_III161). (B) Corresponding regions of the stationary phase xylan, xylose and no added carbon reference gels showing the abundance profiles of spots C638 and C1024.

During mid-log phase, spot C637 was up-regulated 2.8-fold ($p < 0.00$) in the xylan growth condition relative to xylose (Figure 5.9A and B). A closely neighbouring spot (spot C638) was also identified as pullulanase, Pul13A (Bpr_III161) and was detected as being unique to the mid-log phase xylan growth condition (Figure 5.9B). Both were located at approximately pI 4.2 and 100 kDa, which corresponded closely with the theoretical values for the enzyme. During stationary phase, spot C637 became undetectable in all three culture conditions, but a new spot not detected during mid-log

phase, and approximately 5 kDa smaller than spot C637 was also identified as pullulanase, Pul13A (Bpr_III161) (Figure 5.9C, spot C1024). Manual analysis of the PMF derived from spots C637 (mid-log phase) and C1024 (stationary phase) showed two peptides matched to the C-terminal region of spot C637 that were undetected in the PMF of the spot C1024, suggesting a possible C-terminal truncation of the protein detected in spot C1024 during stationary phase. Spot C1024 was detected during stationary phase at equivalent abundance in all culture conditions. Similarly, the abundance of spot C638 during stationary phase was equivalent in all three culture conditions.

5.4 Relative abundance analysis of the stationary phase cytosolic proteome by 2-DE

In the stationary phase harvested *B. proteoclasticus* cytosolic proteome, the abundance of two spots was significantly different in xylan-grown cells relative to xylose-grown cells at the 5% FDR (Figure 5.10 and Figure 5.11, spot C601, spot C1057, and spot C1074). An additional four spots were differentially abundant between the xylan and no added carbon growth conditions, none of which changed significantly between the xylan and xylose growth states (Figure 5.10 and Figure 5.12, spot C704, spot C784, spot C785, and spot C1120). Five of these six differentially abundant proteins were detected at greater abundance in xylan-grown cells, relative to either the xylose or no added carbon culture conditions. Proteins were identified in all six spots of interest and are summarised in Table 5.8

. In addition, 14 spots were detected in the xylan growth condition that were undetected in both the xylose and no added carbon conditions, five of which were identified. Of these, the sugar ABC transporter substrate-binding protein (Brp_I0182, spot C1074), and the amino acid ABC transporter substrate-binding proteins (Bpr_I1826, spot C1135) were the only proteins not detected in any other spots in any other stationary phase growth condition. Collectively, five of the total of 11 proteins identified in the eight differentially abundant protein spots during stationary phase growth were predicted to be involved in carbohydrate metabolism and transport. BLASTp analysis of each of the non-polysaccharide degrading differentially abundant proteins during stationary phase growth is summarised in Table 5.10 and Appendix B, Table B.6.

MALDI-TOF MS showed that three of the eight differentially abundant spots of interest each contained two proteins (spot C704, spot C785, and spot C1120), while in

the remaining five spots a single protein was identified. Spot C1120 that contained the β -galactosidase Bga35A (Bpr_I2006) and β -xylosidase Xyl3A (Bpr_I0184) was 4.3-fold more abundant in xylan grown cells during stationary phase, when compared to cells grown in the no added carbon condition. Compared to xylose grown cells, spot C1120 was also 6.1-fold more abundant in xylan growth condition, but at the 5% FDR the q -value of 0.063 fell marginally outside the 5% level of significance.

Two ABC transporter SBPs were differentially abundant between the xylan and two alternate growth conditions. The SBP product of Bpr_I0182 was detected at high abundance in xylan grown cells, could not be detected at all in xylose grown cells, and was weakly abundant in the cytosol of cells grown in unsupplemented culture medium. The reverse was true for the xylose SBP product of Bpr_I1173, which was 3.9-fold less abundant in xylan grown cells relative to xylose. The abundance pattern of both SBPs is in good agreement with the pattern detected in the *B. proteoclasticus* culture medium during mid-log phase. Surprisingly, by 2-DE analysis the SBP product of Bpr_I0182 was the only protein that was detected as being differentially abundant between the xylan and alternate two growth conditions at both mid-log and stationary phase time points.

Strikingly, two proteins whose genes are located at adjacent loci (Bpr_I2455 and I2456) were up-regulated in xylan grown cells. In particular, the U62 family peptidase identified in spot C784 (Bpr_I2456) was 35.9-fold more abundant in xylan grown cells relative to cells grown in unsupplemented medium, and was the most strongly up-regulated protein in the xylan grown stationary phase harvested cytosolic proteome. This peptidase was up to 51% identical to many Clostridial zinc-dependent proteases and U62 family peptidases that modulate DNA gyrase activity. Although locus Bpr_I2455 is currently annotated as encoding a hypothetical protein (Table 5.8), BLASTp analysis demonstrated that the translated protein sequence is also most homologous to several Clostridial U62 family peptidases (Appendix B, Table B.6).

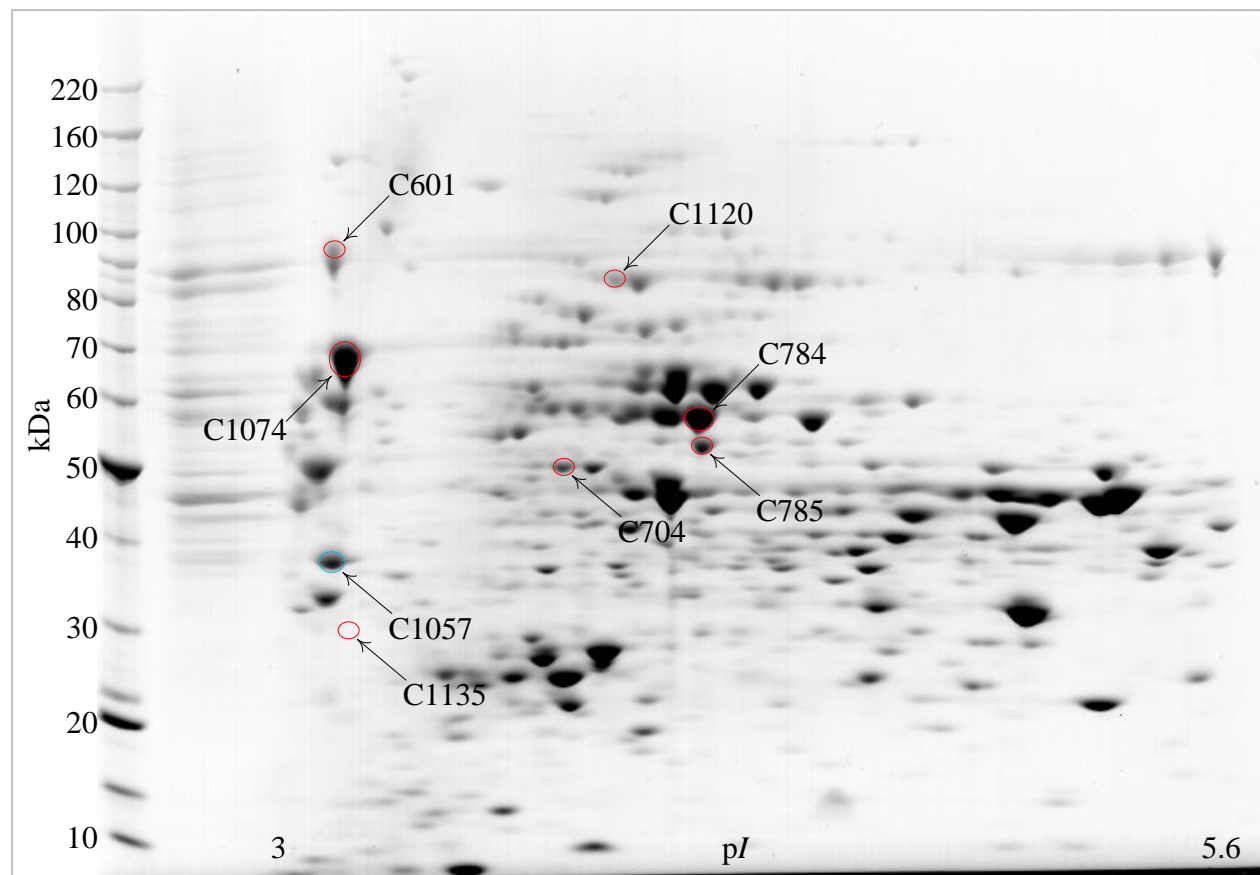


Figure 5.10. Differentially abundant spots between the stationary phase harvested xylan grown cytosolic proteome and the alternate two growth conditions. A representative pI 3-5.6 2-DE colloidal CBB G-250 stained reference gel of cytosolic proteins harvested from stationary phase, xylan grown *B. proteoclasticus* cells is shown. Red and blue circles denote spots with increased and decreased abundance respectively in the xylan growth condition. Circled spots are summarised in Table 5.8.

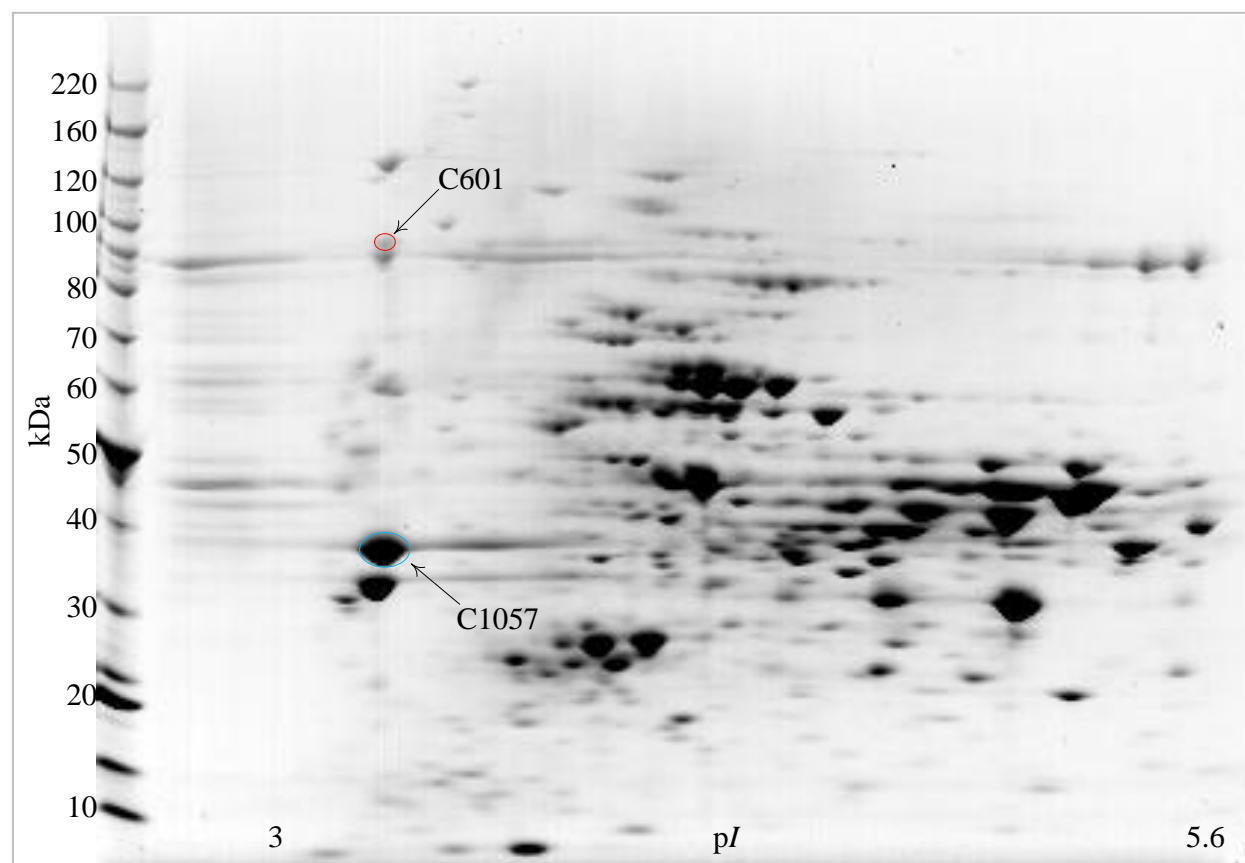


Figure 5.11. Differentially abundant spots between the stationary phase harvested xylan and xylose grown cytosolic proteomes. A representative pI 3-5.6 2-DE colloidal CBB G-250 stained reference gel of cytosolic proteins harvested from stationary phase, xylose grown *B. proteoclasticus* cells is shown. Red and blue circles denote spots with increased and decreased abundance respectively in the xylan growth condition. Circled spots are summarised in Table 5.8.

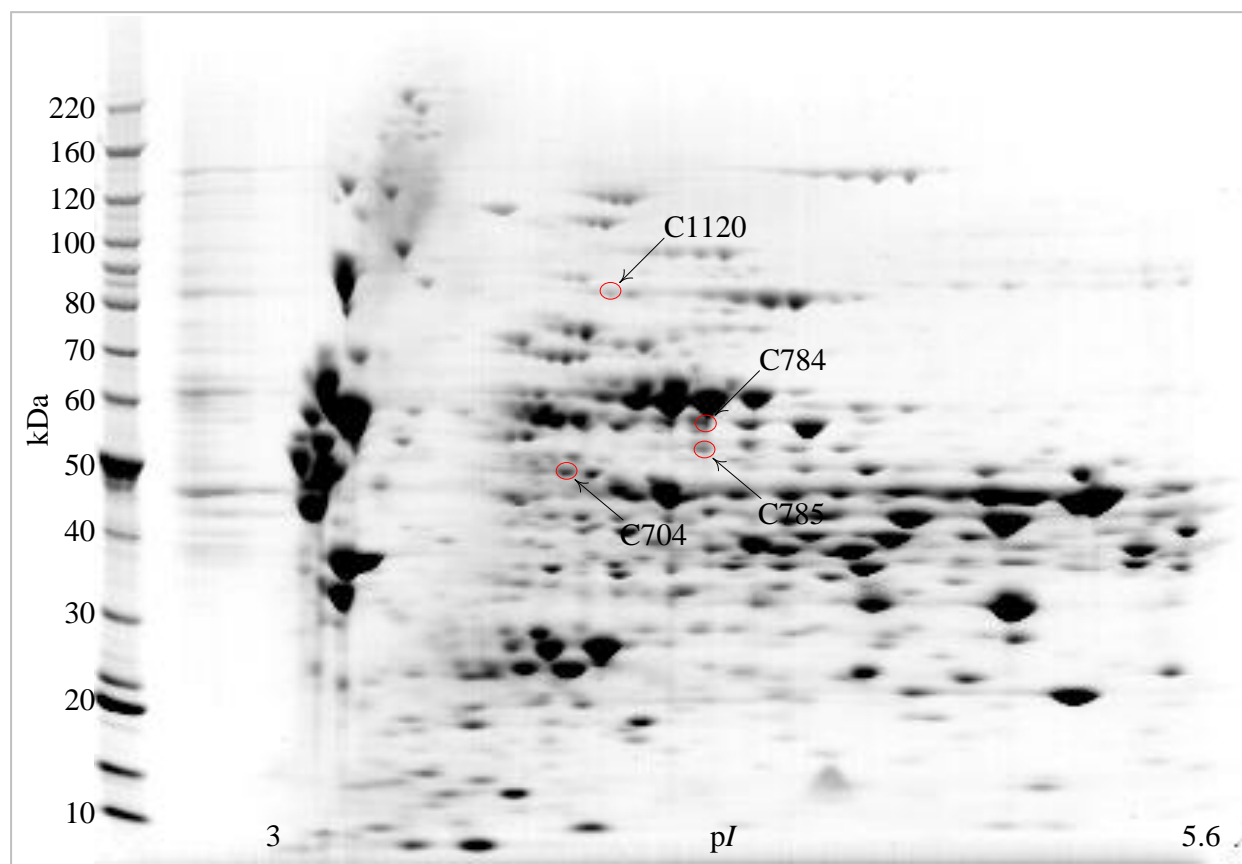


Figure 5.12. Differentially abundant spots between the stationary phase harvested xylan and no added carbon grown cytosolic proteomes. A representative pI 3-5.6 2-DE colloidal CBB G-250 stained reference gel of cytosolic proteins harvested from stationary phase, no added carbon grown *B. proteoclasticus* cells is shown. Red and blue circles denote spots with increased and decreased abundance respectively in the xylan growth condition. Circled spots are summarised in Table 5.8.

Table 5.8. Summary of the differentially abundant proteins between xylan and xylose, or xylan and no added carbon growth conditions, identified in the stationary phase cytosolic proteome.

Spot	Protein	Locus	Function ^a	PUL	Expect	pI	kDa	Pep.	Cov.	Fold-change ^b	p value/q value
C1135	Amino acid ABC transporter substrate-binding protein	Bpr_I1826	T	-	5.8e-04	4.0	31.6	7	21%	Xylan	n/a
C785	Aminotransferase DegT/DnrJ/EryC1/StrS family	Bpr_I2311	I	-	1.5e-10	5.0	51.0	11	23%	4.2±1.1	0.000/0.013
	Xylulokinase, XylB	Bpr_I0173	C	3		4.9	53.7	12	26%		
C1120	β-Galactosidase, Bga35B	Bpr_I2006	C	20	1.9e-27	4.9	83.2	25	37%	4.5±0.8	0.000/0.013
	β-Xylosidase, Xyl3A	Bpr_I0184	C	3		4.8	78.2	17	24%		
C704	Hypothetical protein	Bpr_I2455	H	-	7.6e-22	4.7	47.7	16	42%	2.3±0.1	0.001/0.020
	Phosphoribosylamine-glycine ligase, PurD	Bpr_I0870	O	-		4.7	46.3	12	29%		
C601	Oligopeptide ABC transporter substrate-binding protein, OppA1	Bpr_I1276	T	-	1.9e-13	4.0	83.1	17	23%	2.3±0.5	0.001/0.046
C784	Peptidase U62 family	Bpr_I2456	P	-	6.1e-09	4.8	52.3	15	33%	35.9±7.2	0.001/0.023
C1074	Sugar ABC transporter substrate-binding protein	Bpr_I0182	T	3	7.6e-07	4.12	63.3	13	27%	Xylan	n/a
C1057	Xylose ABC transporter substrate-binding protein	Bpr_I1173	T	-	9.6e-12	4.2	38.4	14	57%	-3.9±1.1	0.000/0.046

^a C, Carbohydrate metabolism; H, Hypothetical; I, Cell envelope; O, Nucleotide metabolism; P, Protein fate; T, Transporters.

^b Calculated as the ratio of the mean normalised protein spot volumes in the xylan/xylose or xylan/ no added carbon growth conditions as stated. Positive values denote proteins with increased mean normalised spot volume in xylan grown cells. “Xylan” denotes the protein was uniquely detected in culture medium harvested from xylan grown cells.

5.5 Relative abundance analysis of the mid-log phase cytosolic proteome by 1-D LC-MS/MS

Gel-free quantitative proteomic analysis utilises a variety of automatically observed 1-D LC-MS/MS parameters such as the peptide count (Gao *et al.*, 2003), spectrum count (Liu *et al.*, 2004), and sequence coverage (Florens *et al.*, 2002), to measure changes in relative protein abundance. Scaffold2™ was used to examine the differential abundance profiles of 1-D LC-MS/MS identified proteins using spectral counting. Protein quantitation by spectral counting operates on the assumption that for MS/MS data obtained under equivalent conditions (e.g. equal protein load, unbiased sampling) the ratio of MS/MS spectra assigned to the same protein in two samples is directly proportional to the relative quantitative abundance of that protein. The Scaffold2™ output files containing the quantitative spectrum counts for each 1-D LC-MS/MS replicate analysis were extracted and filtered prior to protein abundance profiling. Proteins detected in less than 12 of the 15 technical replicate samples were removed from the analysis. Furthermore, spectral counting of low abundance proteins may produce statistically unreliable results due to the low number of counts obtained from one or both samples (Old *et al.*, 2005), therefore proteins with fewer than 45 total spectrum counts per protein per sample (an average of at least three counts per technical replicate), were filtered from the dataset. Of the 281 protein pairs identified in both xylan and xylose growth conditions by Scaffold2™, 108 fulfilled this criteria and were therefore considered suitable for statistical analysis of their differential abundance profiles. Prior to testing, each quantitative spectrum count was normalised against the total spectrum counts per sample, and \log_{10} transformed as recommended by Carvalho *et al.* (2008). A scatter plot of spectrum count variance as a function of spectrum count mean for each protein in the xylan grown data set showed a significant variance versus mean dependency remained after \log_{10} transformation (Figure 5.13 A). A more effective transformation was therefore sought, and after iterative testing a $1/3$ power transformation was optimal for the 1-D LC-MS/MS dataset (Figure 5.13 B).

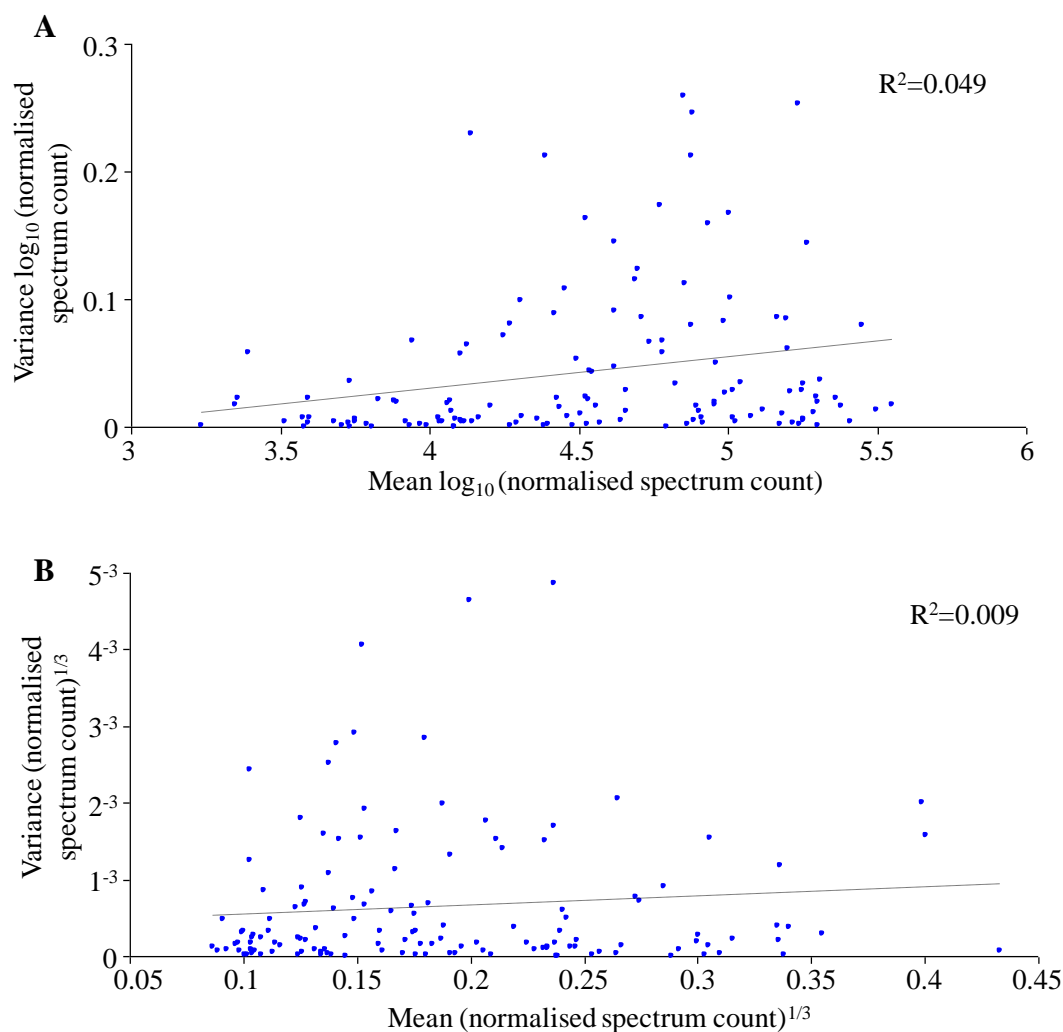


Figure 5.13. The transformation efficiency in correcting for variance versus mean dependency in the LC-MS/MS data set by \log_{10} transformation. Variance versus mean scatter plots of the mid-log phase harvested, xylan grown LC-MS/MS data set after \log_{10} transformation (A) shows a moderate correlation between the two variables still exists. After $1/3$ power transformation (B) a low correlation exists between the two variables.

The mean of the $1/3$ power transformed normalised spectral count was then calculated for each of the three biological replicate samples, and two tailed Student's *t*-tests were performed for each protein pair that passed the Shapiro-Wilk test of normality. As for the 2-DE analysis, a *q*-value was then calculated for each test, and used to determine the FDR threshold. Plotting the number of significant tests and the number of expected false positives as a function of percentage FDR (Figure 5.14) showed that the most suitable balance between statistical power and the incidence of false discoveries occurred at 5% FDR. At this level, the relative abundance of nine

proteins was significantly different between the xylan and xylose growth conditions (Table 5.9), and the number expected of false positive discoveries was less than 1.

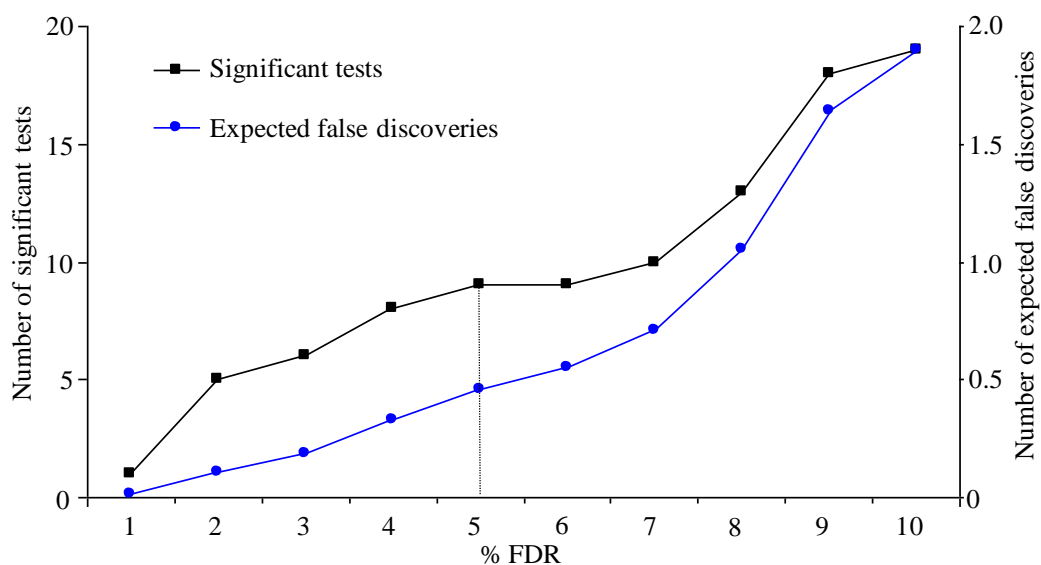


Figure 5.14. Cumulative frequency of 1-DE LC-MS/MS significant tests and expected number of false discoveries as a function of percentage FDR. A 5% FDR was chosen for the cytosolic 1-DE LC-MS/MS analysis (dotted line), which resulted in nine significant tests (black line, primary y-axis), and less than 0.5 expected false discoveries (blue line, secondary axis).

Table 5.9. Summary of proteins differentially or uniquely abundant between xylan and xylose growth conditions, identified in the mid-log phase cytosolic proteome by 1-D LC-MS/MS.

Protein	Locus	Function ^a	PUL	Expect.	pI	kDa	Unique pep. ^b	Cov. ^c	Fold change ^d	p value/q value
ATP synthase F1 β subunit, AtpD2	Bpr_I1162	E	-	3.0e-08	4.5	50.9	2	6.3%	Xylose	n/a
β -Xylosidase, Xyl3A ^e	Bpr_I0184	C	3	8.9e-15	4.8	78.2	17	36%	Xylan	n/a
Bmp family protein ^f	Bpr_I1560	U	-	1.5e-11	3.7	40.7	3/3	18%/18%	2.2 \pm 0.2	0.005/0.020
DNA-binding protein ^f	Bpr_I1236	N	-	4.0e-11	11.1	9.6	6/7	53%/57%	-6.3 \pm 5.2	0.005/0.020
DNA-directed RNA polymerase α subunit, RpoA	Bpr_I0623	W	-	3.3e-15	4.4	35.1	16/12	61%/49%	-2.1 \pm 0.4	0.025/0.050
Fatty acid/phospholipid synthesis protein, PlsX ^f	Bpr_I1576	L	-	1.6e-11	6.2	35.8	2	15%	Xylose	n/a
Orotate phosphoribosyltransferase, PyrE1 ^f	Bpr_I0863	O	-	1.1e-15	4.7	25.4	4/4	28%/36%	-3.0 \pm 1.0	0.004/0.020
PTS system HPr phosphocarrier ^f	Bpr_I2105	T	21	3.6e-14	4.3	9.3	3/4	55%/69%	-2.1 \pm 0.8	0.013/0.039
Sugar ABC transporter substrate-binding protein ^f	Bpr_I0117	T	2	1.0e-30	6.2	57.5	7	22%	Xylan	n/a
Sugar ABC transporter substrate-binding protein	Bpr_I1667	T	-	1.0e-30	3.9	47.8	11/8	39%/28%	2.9 \pm 0.1	0.000/0.000
Sugar ABC transporter substrate-binding protein	Bpr_I1720	T	16	1.1e-15	4.0	49.3	10/5	38%/18%	2.0 \pm 0.2	0.008/0.026
Sugar ABC transporter substrate-binding protein	Bpr_I2443	T	25	1.0e-30	4.0	57.0	13/5	50%/21%	7.5 \pm 3.6	0.018/0.041
Sugar ABC transporter substrate-binding protein ^f	Bpr_I2264	T	23	1.1e-16	3.9	47.5	7/5	29%/21%	6.8 \pm 1.0	0.001/0.006

Table continues

Table 5.9 continued.

Protein	Locus	Function ^a	PUL	Expect.	pI	kDa	Unique pep. ^b	Cov. ^c	Fold change ^d	<i>p</i> value/ <i>q</i> value
Xylose ABC transporter ATP-binding protein ^f	Bpr_I1174	T	10	2.7e-13	5.2	68.4	14	43%	Xylose	n/a

^a C, Carbohydrate metabolism; E, Energy metabolism; L, Lipid metabolism; N, Nucleic acid metabolism; O, Nucleotide metabolism; T, Transporters; U, Unknown function; W, Transcription.

^b Values denote the number of unique peptides matched to Scaffold2™ identified proteins in xylan/xylose samples.

^c Values denote the maximum percent sequence coverage of Scaffold2™ identified proteins in xylan/xylose samples.

^d Calculated as mean ratio of normalised spectrum counts obtained from xylan/xylose biological replicate analyses. Xylan / xylose denotes the protein was uniquely detected in culture medium harvested from xylan / xylose grown cells respectively.

^e Differentially expressed in the 2-DE MALDI-TOF analyses.

^f Not identified in the 2-DE MALDI-TOF analyses

Five of the nine proteins were more abundant in the xylan growth condition relative to growth on xylose, and the abundance fold changes ranged between 2.0 ± 0.2 and 7.5 ± 3.6 . Notably, four of the five xylan stimulated proteins were substrate-binding components of ABC transporter systems (Figure 5.15). In addition, five proteins were detected in cells grown in only one of the two culture conditions (Table 5.9). Three proteins were detected only in cells grown in xylose supplemented culture medium, and included the ABC transporter ATP binding protein (Bpr_I1174) that is a component of a xylose specific ABC transporter system. This ATP binding protein is coded for by a gene located immediately downstream of the gene that encodes the differentially abundant xylose ABC transporter substrate-binding protein (Bpr_I1173). As demonstrated by 2-DE MALDI-TOF, the abundance of the substrate-binding protein was 3.9-fold greater in xylose grown cells relative to the xylan growth condition during stationary phase growth.

The two proteins detected only in the xylan stimulated growth condition were both involved in carbohydrate metabolism and transport, and included the β -xylosidase Xyl3A (Bpr_I0184). This enzyme was one of two differentially abundant proteins detected by both the 1-D LC-MS/MS and 2-DE MALDI-TOF techniques. The direction of the abundance change was the same in both proteomic techniques. It is noteworthy that the β -xylosidase Xyl3A (Bpr_I0184) was identified in the 2-DE MALDI-TOF analyses in a single protein spot that also contained β -galactosidase Bga35B (Bpr_I2006). Surprisingly, eight of the 14 proteins found as being differentially abundant or unique to either growth substrate by the 1-D LC-MS/MS were not identified by 2-DE MALDI-TOF in any growth condition at either time point. Three of these eight proteins were uniquely detected in one or other growth condition. Two of the remaining five proteins were more abundant in the xylan growth condition, where the absolute fold changes ranged between 2.1 and 6.8. BLASTp analysis of the differentially abundant proteins detected by 1-D LC-MS/MS is summarised in Appendix B, Table B.7.

5.6 Bioinformatics of the differentially abundant sugar substrate-binding proteins identified in the *B. proteoclasticus* cytosol

The combination of 2-DE MALDI-TOF and 1-DE LC-MS/MS analysis of the *B. proteoclasticus* cytosol showed that the relative abundance patterns of seven SBP and one ATP-binding protein components of sugar ABC transporter systems were

significantly different between the xylan and alternate growth conditions Table 5.7, Table 5.8, and Table 5.9).

During stationary phase the substrate-binding protein containing the PBP1_xylose (Cd01538) domain (Bpr_I1173) constituted almost 3% of the total detectable protein in cells grown in the presence of xylose as the sole supplementary carbon source, and was 3.9 ± 1.1 -fold more abundant relative to cells grown in xylan containing culture medium (Table 5.9). A similar abundance pattern was observed during mid-log phase, but the large variance in spot abundance in the xylose growth condition contributed to the mean 5.7 ± 3.1 -fold change not being statistically significant at the 5% FDR. The Periplasmic Binding Protein Type-I domain comprising almost the full length of the protein was 82% identical to the corresponding domain of a putative periplasmic xylose binding protein produced by the cellulolytic anaerobe *Clostridium phytofermentans* ISDg.

The abundance of both proteins containing amino acid binding domains was significantly greater in the xylan growth condition relative to xylose. Indeed the SBP-Bac_3 domain containing protein (Brp_I1826) was undetectable in either the xylose or no added carbon growth conditions at both time points. Furthermore, spot C1060 which was located approximately 5 kDa below spot C601 (Figure 5.6) and also identified as the oligopeptide binding protein OppA1 (Bpr_I1276) was 4.9 ± 2.4 -fold more abundant in xylan grown cells relative to xylose. The statistical significance of this change was marginally outside the 5% FDR confidence level. The functional domains of both amino acid binding proteins were homologous to substrate-binding proteins produced by proteolytic, pathogenic *Streptococcus* sp.

Collectively, the set of 12 differentially abundant ABC transporter proteins that were comprised of 11 SBPs and one ATP-binding protein (Figure 3.26 and Figure 5.15) displayed a distinctive pattern of relative abundance change. Regardless of the sub-cellular compartment in which each ABC transporter associated sugar SBP was identified, the common feature of all nine SBPs that were more abundant in the xylan growth condition was the presence of a CUT1 family substrate-binding domain, and eight of the nine contained a SBP_Bac_1 (PF 01547) domain. Conversely, the two SBPs and that were less abundant in xylan grown cells contained a Peripla_BP_1 domain (PF00532) which is classified as a CUT2 family module.

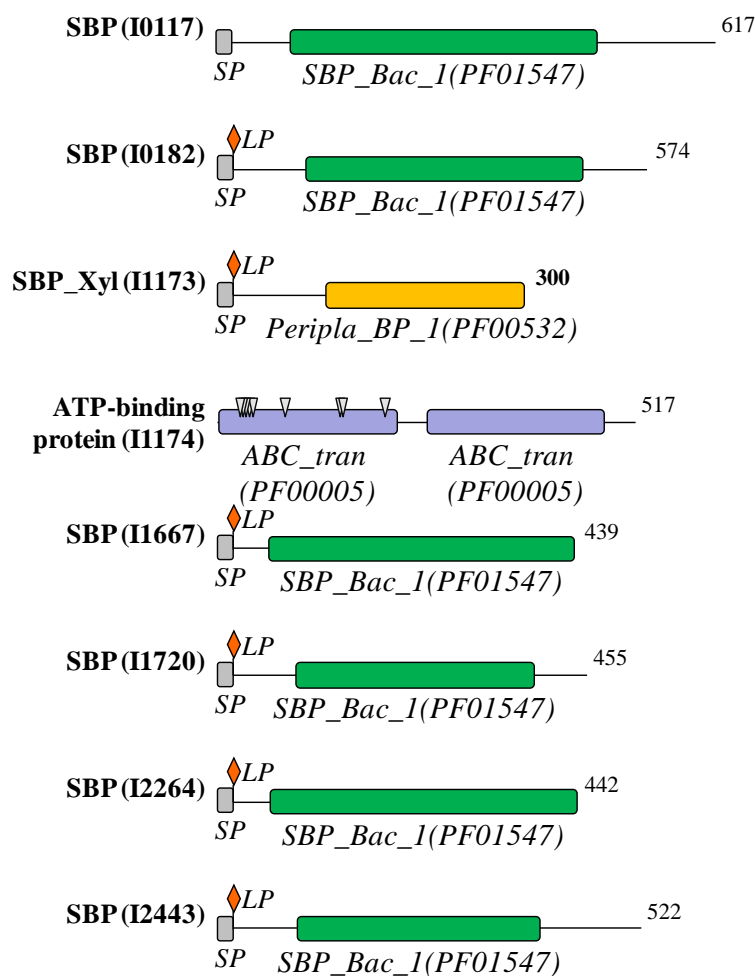


Figure 5.15. Functional domains of differentially abundant ABC transporter system proteins identified in the *B. proteoclasticus* cytosol. Values to the right of each depicted protein denote number of amino acid residues. LP, lipobox motif; SP, secretory signal peptide. Domains shown in green are associated with proteins more abundant in the xylan growth conditions. The nine conserved ATP-binding sites within the N-terminal ABC transporter ATP-binding domain (PF00005) are shown in Bpr_I1174. SP, secretory signal peptide.

The degree of primary sequence homology between each of the differentially abundant SBP_Bac_1 domain containing proteins was examined using CLUSTALW multiple sequence alignment. A neighbour-joining phylogram using each of the differentially abundant SBP protein sequences supported the distinct pattern of protein abundance change (Figure 5.16). Two clusters were formed by the nine SBP_Bac_1 (Figure 5.16, green box) and two Peripla_BP_1 domain containing proteins respectively. Nonetheless, the mean branch length of 0.42586 for the set of SBP_Bac_1 containing SBPs indicated only moderate homology between this group

of proteins, which was confirmed by examination of the CLUSTALW multiple sequence alignment (Figure 5.17 and Figure 5.18). The primary sequence divergence between the *B. proteoclasticus* differentially abundant SBP-Bac_1 proteins is reflected in the fact that each protein has a highest scoring BLASTp homologue produced by a different bacterium (Table 3.5 and Table 5.10).

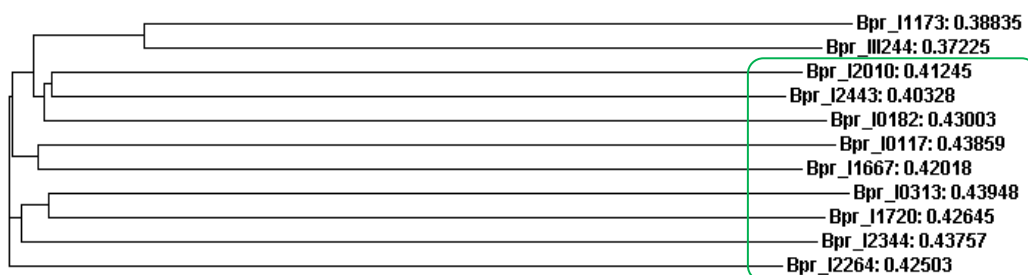


Figure 5.16. CLUSTALW neighbour-joining phylogram of the differentially abundant SBPs identified in the *B. proteoclasticus* proteome. The green box denotes the set of SBPs that were significantly more abundant in the culture medium or cytosol of *B. proteoclasticus* cells grown in the presence of xylan. Values to the right of the SBP locus number denote branch distances. Score type = proportional variability, where 1 = 100%.

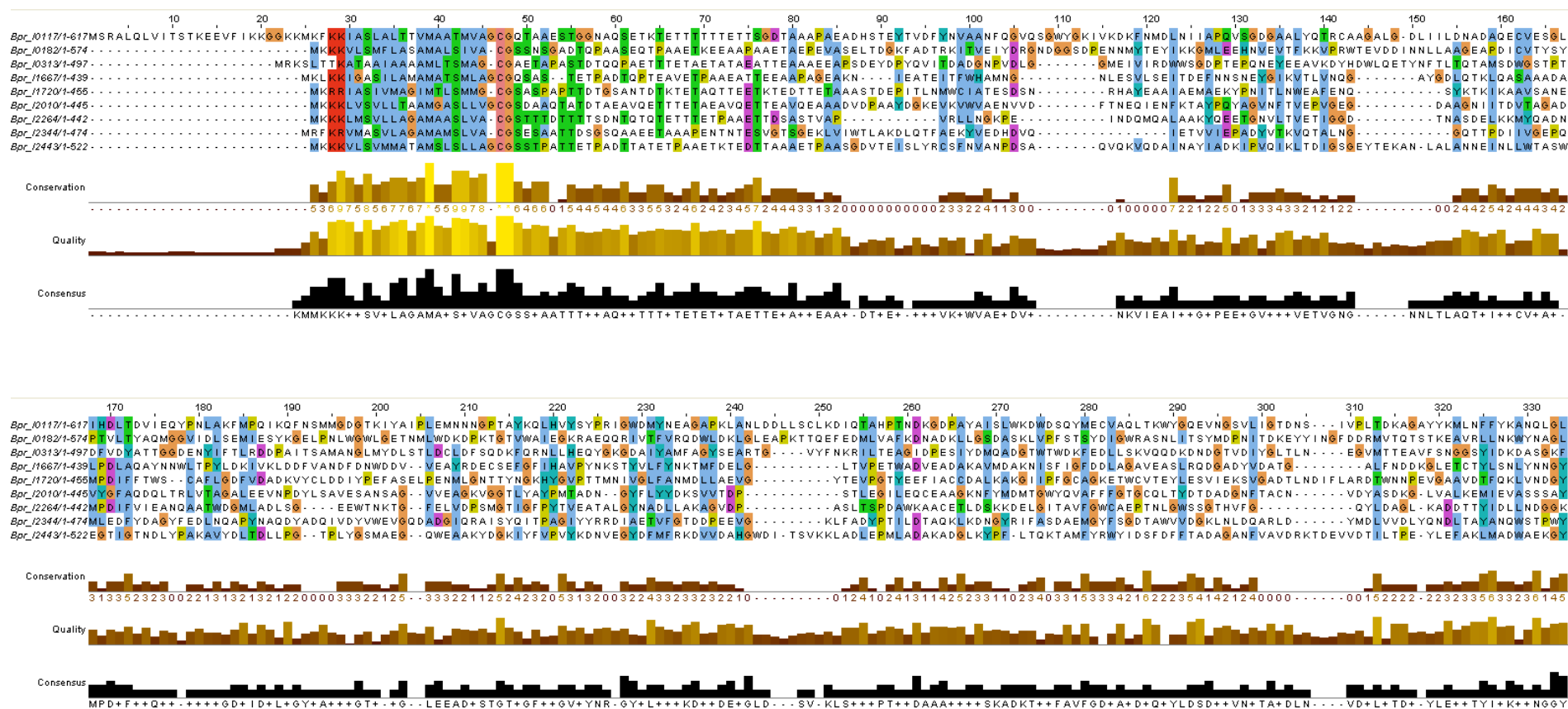


Figure 5.17. CLUSTALW multiple sequence alignment of the N-terminal half of the differentially abundant SBP_Bac_1 domain containing SBPs identified in the *B. proteoclasticus* proteome. Proteins are arranged by *B. proteoclasticus* gene locus.

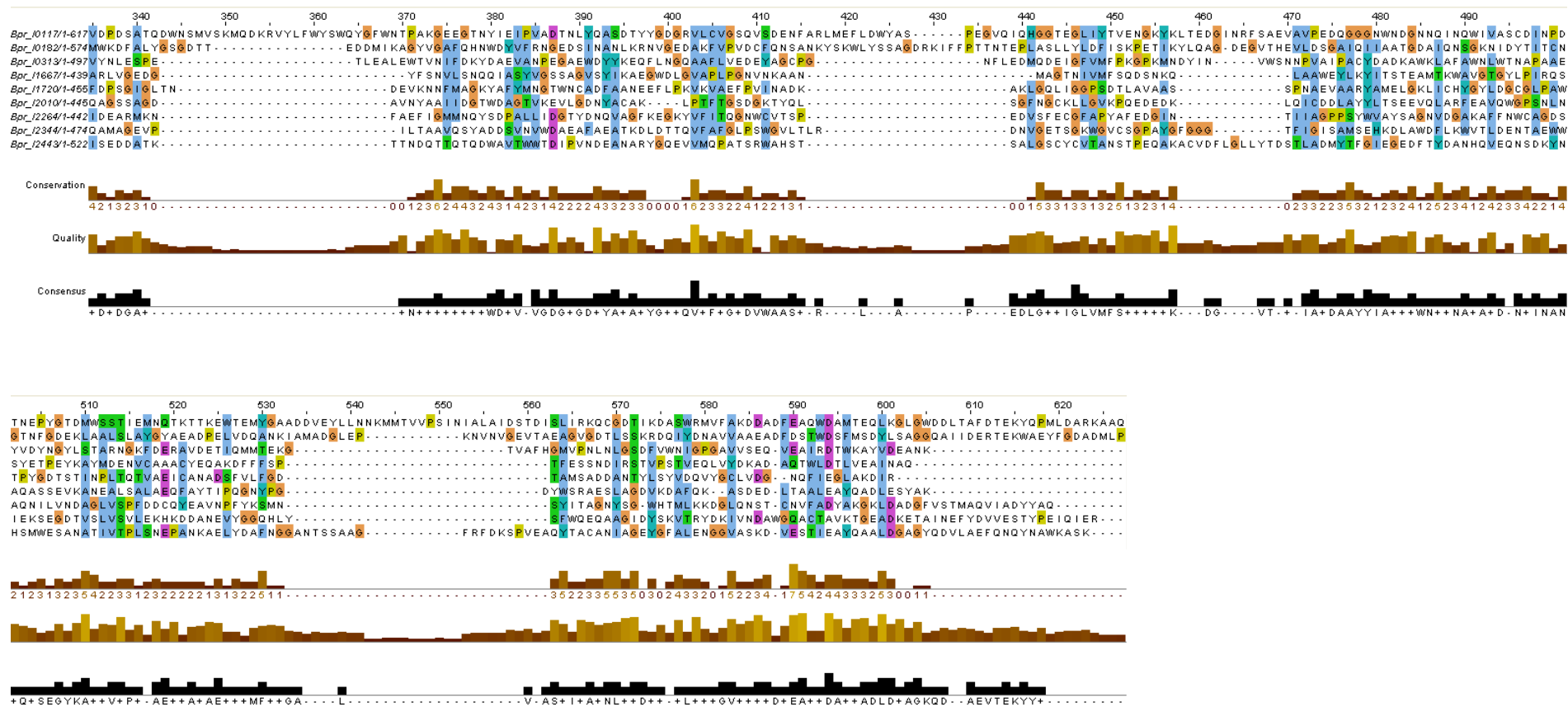


Figure 5.18. CLUSTALW multiple sequence alignment of the C-terminal half of the differentially abundant SBP_Bac_1 domain containing SBPs identified in the *B. proteoclasticus* proteome. Proteins are arranged by *B. proteoclasticus* gene locus.

Table 5.10. BLASTp analysis of the differentially abundant SBPs identified in the *B. proteoclasticus* cytosol.^a

<i>B. proteoclasticus</i> protein	Locus	Homologue ^b	Organism ^c	gi	Expect	Identity	Region
Sugar ABC transporter SBP	Bpr_I0117	Putative bacterial extracellular solute-binding protein	<i>Roseburia intestinalis</i> L1-82	240144557	1e-166	48%	25--617
Sugar ABC transporter SBP	Bpr_I0182	ABC-type sugar transport system, periplasmic component	<i>Butyrivibrio fibrisolvens</i> 16/4	291518643	1e-180	63%	1-568
Xylose ABC transporter SBP	Bpr_I1173	Putative solute-binding component of ABC transporter	<i>Clostridium phytofermentans</i> ISDg	160879730	4e-161	80%	19-359
Sugar ABC transporter SBP	Bpr_I1667	Bacterial extracellular solute-binding protein	<i>Turicibacter</i> sp. PC909	293376212	6e-58	34%	3-419
Sugar ABC transporter SBP	Bpr_I1720	Extracellular solute-binding protein family 1	<i>Acetivibrio cellulolyticus</i> CD2	302592083	5e-69	34%	1-455
Sugar ABC transporter SBP	Bpr_I2264	Extracellular solute-binding protein, family 1	<i>Ruminococcus</i> sp. 5_1_39B_FAA	253579691	6e-120	49%	1-441
Sugar ABC transporter SBP	Bpr_I2443	Extracellular solute-binding protein family 1	<i>Clostridium lentocellum</i> DSM 5427	296441832	1e-40	30%	1-522

^a Differentially abundant SBPs identified in the cytosol of mid-log or stationary phase harvested cells by a combination of 2-DE MALDI-TOF and 1-DE LC-MS/MS.

^b BLASTp top ranked homologous protein.

^c Organism that expresses the BLASTp top ranked homologous protein.

5.7 Summary

Using the combination of gel-based and gel-free differential expression proteomics has confirmed that the relative abundance of several *B. proteoclasticus* cytosolic polysaccharidases, and a significant number of substrate-binding proteins, is modulated in response to growth substrate. Similar to relative abundance changes detected in the extracellular proteome, the relative abundance changes detected in the *B. proteoclasticus* cytosol accurately reflect the growth conditions used. During mid-log phase, growth in the presence of long-chain GAX relative to cells grown on xylose elicited the increased abundance of the β -xylosidase Xyl3A, the xylosidase/arabinofuranosidase Xsa43E, which both target xylooligosaccharides, as well as the pullulanase Pul13A, which targets starch. During stationary phase, Xyl3A was also detected in one spot that was more than four-fold more abundant in xylan grown cells. The relative abundance changes detected in this analysis imply that within the consortia of enzymes detected in the cytosol, Xyl3A and Xsa43E may play a particularly important role in the cytosolic degradation of assimilated xylooligosaccharides.

The relative abundance of seven ABC-transporter SBPs detected in the *B. proteoclasticus* cytosol was also significantly different between growth substrates. This result raised the total number of differentially abundant ABC-transporter associated proteins within the *B. proteoclasticus* proteome to 11, nine of which were more abundant in cells grown on xylan. These changes imply that *B. proteoclasticus* modulates the set of SBPs in response to growth substrate, which may provide an ecological advantage to *B. proteoclasticus* by allowing cells to rapidly assimilate the products of its extracellular polysaccharide-degrading activity rather than having them lost to other members of the complex rumen microbial ecosystem.

Chapter 6

Discussion

6.1 Overview

B. proteoclasticus falls within the *Butyrivibrio/Pseudobutyrvibrio* assemblage and was the first member of the genus to have its genome fully sequenced (Kelly *et al.*, 2010). *Butyrivibrio/Pseudobutyrvibrio* species are metabolically versatile gut bacteria, but their primary role in the rumen is thought to be the degradation of plant hemicelluloses. Using a combination of gel-based and gel-free proteomics, 416 non-redundant proteins present in the secreted, cell-associated, or cytosolic sub-cellular compartments were detected, which represent 14% of the *B. proteoclasticus* extracellular and cytosolic theoretical proteins. Present in this set of proteins were 12 predicted extracellular, and 24 predicted cytosolic polysaccharidases. Collectively, these enzymes represent 26% of the total predicted fibrolytic potential of the bacterium. An additional 59 identified proteins were predicted to be involved in the assimilation and further metabolism of polysaccharide degradation products. This study has confirmed the hypothesis that *B. proteoclasticus* produces an extensive array of polysaccharidases that enable it to degrade and assimilate GAX, and that the abundance of many of these proteins is sensitive to growth substrate. The results of the study support the notion that *B. proteoclasticus* is an important contributor to degradation and utilisation of the types of hemicelluloses that are abundant in ruminant forages. Furthermore, this study demonstrated that the abundance of *B. proteoclasticus* secreted and cytosolic fibrolytic enzymes and proteins that are likely to be involved in the translocation of hemicellulose degradation products are responsive to the substrates that cells encounter in the external environment.

An overview of the predicted catalytic activity of the set of 36 identified polysaccharidases identified in the *B. proteoclasticus* proteome predicts that these enzymes are likely to be active upon a range of plant-derived oligosaccharides (Table 6.1 and Table 6.2). Furthermore, the experimentally determined catalytic activities for homologues of the majority of the identified *B. proteoclasticus* polysaccharidases support these predictions. Clustering of the 36 polysaccharidases by function signifies that 11 and 18 distinct hydrolytic activities are predicted to be present within the *B. proteoclasticus* extracellular and cytosolic proteomes respectively. These predictions do not include the three identified *B. proteoclasticus* polysaccharidases whose GH domains are currently unclassified.

Table 6.1. Predicted catalytic activity of the identified extracellular *B. proteoclasticus* polysaccharidases.^a

Protein	Locus	Substrate	Reaction catalysed ^b
α -Amylase, Amy13A	Bpr_I1087	Starch and glycogen	Endohydrolysis of (1→4)- α -D-glucosidic linkages.
β -Glucosidase, Bgl3A	Bpr_I0693	β -D-Glucosides	Hydrolysis of terminal, non-reducing β -D-glucosyl residues with release of β -D-glucose.
Endo-1,4- β -glucanase, Cel5C	Bpr_I1710	Cellulose cereal β -D-glucans	Endohydrolysis of (1→4)- β -D-glucosidic linkages.
Glycoside hydrolase family 30, GH30A	Bpr_I2937	Unknown	Unknown.
Endo-1,4- β -xylanase and xylosidase, Mxy10-43A	Bpr_I0737	α -L-Arabinofuranose containing xylans	Endohydrolysis of (1→4)- β -D-xylosidic linkages, and hydrolysis of terminal, non-reducing D-xylose residues.
Endo-1,4-β-xylanase, Xyn10B Endo-1,4- β -xylanase, Xyn10C	Bpr_I0026 Bpr_I1008	Xylan	Endohydrolysis of (1→4)-β-D-xylosidic linkages.
Pectate lyase, Pel1A	Bpr_I2372	Pectate	Cleavage of (1→4)- α -D-galacturonan. Favours pectate over the methyl ester pectin.
Pectin methylesterase, Pme8B	Bpr_I2473	Pectin	Hydrolysis of pectin yielding pectate and methanol.
Pullulanase, Pul13A	Bpr_III161	Pullulan, amylopectin and glycogen	Hydrolysis of (1→6)-α linkages.
Xylosidase/arabinofuranosidase, Xsa43A	Bpr_I0302	α -L-Arabinofuranosides	Hydrolysis of terminal, non-reducing D-xylose or α -L-arabinofuranoside residues.
Xylosidase/arabinofuranosidase, Xsa43J	Bpr_I2935	Xylooligosaccharides and α-L-arabinofuranose containing xylans	Hydrolysis of (1→4)-β-D-linkages, releasing D-xylose residues from the non-reducing termini/Hydrolysis of terminal non-reducing α-L-arabinofuranoside residues.

^a Bold text denotes enzymes that were more abundant in the culture medium of cells grown on xylan, relative to those grown on xylose or no added carbon.

^b Swiss Institute of Bioinformatics Enzyme nomenclature (<http://us.expasy.org/enzyme/>).

Table 6.2. Predicted catalytic activity of the identified cytosolic *B. proteoclasticus* polysaccharidases.^a

Protein	Locus	Substrate	Reaction catalysed ^b
α -Galactosidase, Aga27A	Bpr_I0205	α -D-Galactosides	Hydrolysis of terminal, non-reducing α -D-galactose residues.
α -Galactosidase, Aga36C	Bpr_III065		
α -D-Glucuronidase, Agu67A	Bpr_I0177	Glucuronoxylans	Hydrolysis of glucuronic acid substituted xylooligosaccharides.
α -Amylase, Amy13G	Bpr_I0729	Starch and glycogen	Endohydrolysis of (1 \rightarrow 6)- α -D-glucosidic linkages.
α -L-Arabinofuranosidase, Arf51A	Bpr_I0329	α -L-Arabinofuranosides	Hydrolysis of terminal, non-reducing α -L-arabinofuranoside residues.
β -Galactosidase, Bga2A	Bpr_I0279	β-D-Galactosides	Hydrolysis of terminal, non-reducing β-D-galactose residues.
β -Galactosidase, Bga2B	Bpr_III209		
β-Galactosidase, Bga35B	Bpr_I2006		
β -Glucosidase, Bgl3B	Bpr_I0847	β -D-Glucosides	Hydrolysis of terminal, non-reducing β -D-glucosyl residues with release of β -D-glucose.
β -Glucosidase, Bgl3C	Bpr_I0138		
Cellobiose phosphorylase, Cbp94A	Bpr_I2447	Cellobiose	Hydrolysis of cellobiose.
Cellodextrinase, Cel9B	Bpr_I1593	Cellulose, lichenin and cereal β -D-glucans	Endohydrolysis of (1 \rightarrow 4)- β -D-glucosidic linkages.
Feruloyl esterase, Est1E	Bpr_I2870	Esterified oligosaccharides	Deferuloylation of esterified oligosaccharides.
Acetyl-xylan esterase, Est2A	Bpr_I2939	Acetylated xylans and xylo-oligosaccharides	Deacetylation of xylans and xylo-oligosaccharides.
Acetyl-xylan esterase	Bpr_I0174	Unknown	Unknown.
Glycoside hydrolase family 31, GH31C	Bpr_I1974	Unknown	Unknown.
Glycogen phosphorylase, Glp2	Bpr_I2847	1,4- α -D-Glucans	Exohydrolysis and phosphorylation of (1 \rightarrow 4)- α -D-glucan.

Table continues

Table 6.2 continued.

Protein	Locus	Substrate	Reaction catalysed ^b
Glycogen debranching enzyme, Glgx2	Bpr_I1494	1,6- α -D-Glucans	Deglycosylation of (1→6)- α -linked glucose monomers.
β -Mannosidase, Man2A	Bpr_III237	β -D-Mannosides	Hydrolysis of terminal, non-reducing β -D-mannose residues.
α -L-Rhamnosidase, Rha78A	Bpr_I1686	α -L-Rhamnosides	Hydrolysis of terminal, non-reducing α -L-rhamnose residues.
Xylosidase/arabinofuranosidase, Xsa43E	Bpr_I2319	α-L-Arabinofuranosides	Hydrolysis of terminal, non-reducing D-xylose or α-L-arabinofuranoside residues.
Xylosidase/arabinofuranosidase and esterase, Xsa43H	Bpr_I0301	α -L-Arabinosides and triacylglycerols	Hydrolysis of terminal, non-reducing D-xylose or α -L-arabinofuranoside residues/hydrolysis of triacylglycerols with release of a diacylglycerol and a carboxylate.
β-Xylosidase, Xyl3A	Bpr_I0184	1,4-β-D-Xylans	Hydrolysis of terminal, non-reducing D-xylose residues.
Endo-1,4- β -xylanase and esterase, Xyn10D	Bpr_I1083	1,4- β -D-Xylans	Endohydrolysis of (1→4)- β -D-xylosidic linkages.

^a Bold text denotes enzymes that were more abundant in the cytosol of cells grown on xylan, relative to those grown on xylose or no added carbon.

^b Swiss Institute of Bioinformatics Enzyme nomenclature (<http://us.expasy.org/enzyme/>).

The functional diversity within the set of polysaccharidases synthesised by *B. proteoclasticus* grown on xylan most likely reflects the chemical complexity of the polysaccharide. Hemicellulose, of which GAX is the predominant form found in ruminant forage crops (Ebringerova *et al.*, 2005, Ishii, 1997), is a chemically heterogeneous and extensively branched polymer of pentoses, hexoses, and sugar acids that are linked to each other and to surrounding structural polymers by a variety of glycosidic linkages (Kabel *et al.*, 2007, Linder *et al.*, 2003). Efficient GAX degradation requires the cooperative action of a variety of enzymes, including endoxylanases, β -xylosidases, α -D-glucuronidases α -L-arabinofuranosidases, acetylxylan esterases, and ferulic acid esterases. Examples of each of these classes of hemicellulolytic enzymes were identified in the *B. proteoclasticus* proteome. Liberation of hemicellulose from the surrounding polymers also requires the presence of several non-xylanolytic polysaccharidases, and several of these are also present in the *B. proteoclasticus* proteome.

6.1.1 Proposed mechanism of GAX degradation by *B. proteoclasticus*

A distinct pattern of fibrolytic capability has emerged from the analysis of the extracellular proteome of *B. proteoclasticus*. The results suggest that *B. proteoclasticus* attacks primarily the xylan backbone and main substituent groups of hemicellulose in the extracellular space, assimilates the variable length substituted or un-substituted xylooligosaccharides, and performs the final stages of degradation within the cell. The secreted enzymes that are predicted to degrade GAX contain only GH10 or GH43 domains that target (1 \rightarrow 4)- β -D-linkages within the xylan backbone, or between the backbone and arabinose side chain groups. Polysaccharidases necessary for the removal of common GAX substituent groups such as glucuronic acid and acetate groups, or the hydrolysis of interchain linkages such as ferulic acid esterases were not detected in the *B. proteoclasticus* extracellular proteome. The hypothesis that *B. proteoclasticus* does not perform complete degradation of complex substituted GAX in the extracellular space implies that cells must be capable of transporting a variety of substituted xylooligomers across the cell wall. This scenario is supported by the number of the sugar specific ABC transporter SBPs that were present in the *B. proteoclasticus* secreted proteome. Furthermore, prior to their utilisation within the cytosol these assimilated substituted xylooligomers must be degraded to their constituent monomers, which is likely to be achieved by the activity of several classes of side chain removing enzymes and β -xylosidases identified in the *B. proteoclasticus*

cytosol. This hypothesis is in general consistent with the transcriptome data obtained by microarray analysis of the *B. proteoclasticus* genes involved in hemicellulose degradation (Kong, 2007). Comparison of cells grown in xylan or xylose supplemented culture medium showed that several secreted endoxylanases, an endocellulase and a feruloyl esterase were up-regulated in cells grown on xylan. In contrast, none of the up-regulated β -xylosidases were secreted enzymes. The proposed mechanism of GAX hydrolysis is supported by the observations of Cotta and Zeltwanger (1995) who demonstrated that the rapid hydrolysis of oat-spelt xylan by enzymes derived from *B. fibrisolvens* H17c caused the accumulation of an extracellular pool of soluble, low molecular weight xylooligosaccharides that contained significant amounts of xylose, xylobiose and xylotriose, but low amounts of glucuronic acids.

A schematic representation of *B. proteoclasticus* mediated extracellular hemicellulose degradation, and carbohydrate uptake based on the data obtained in this project is presented in Figure 6.1 and Figure 6.2. A schematic representation summarising the degradation of internalised xylooligosaccharides is given in Figure 6.3. The enzymatic degradation of pectins is also depicted in view of the fact that these complex polysaccharides can contribute up to 5% of the cell walls of grasses, and are implicated in cross-linking surrounding structural cell wall polymers.

It should be noted that four of the 12 identified polysaccharidases that are predicted to be secreted enzymes were identified only in the cytosol. Although the presence of a Type-I or Type-II N-terminal secretory signal peptide was identified with high statistical probability in each of these enzymes, it is not possible to conclude with certainty that these enzymes are secreted *in vivo*. Nonetheless, for the purposes of clarity these enzymes are discussed based on the assumption that they act on insoluble polysaccharides in the extracellular space.

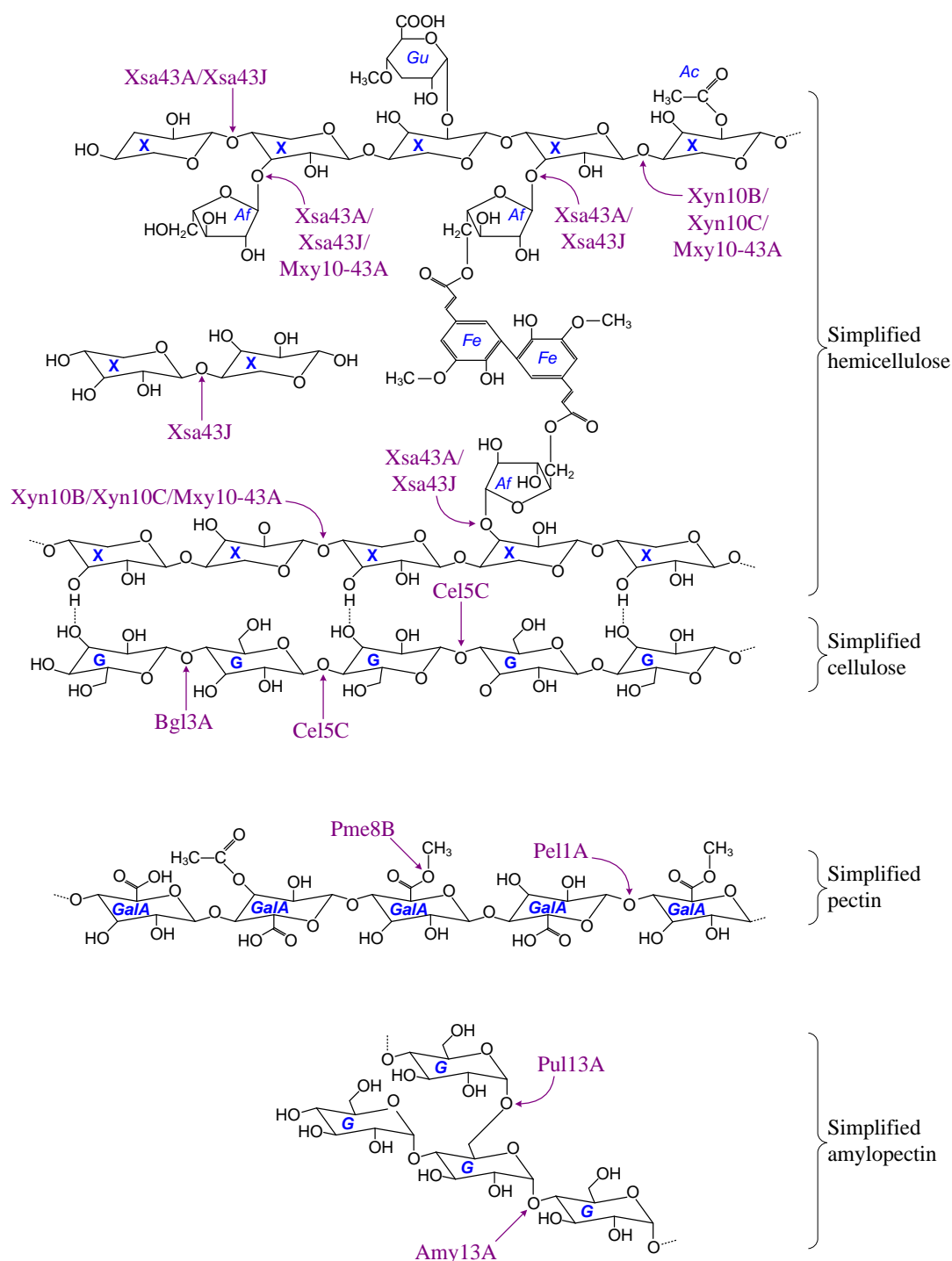


Figure 6.1. A simplified schematic representation of extracellular plant polysaccharide metabolism by *B. proteoclasticus*. The simplified structure of xylan and cellulose (A), pectin (B) and amylopectin (C) showing the predicted sites of hydrolysis by the identified *B. proteoclasticus* secreted polysaccharide-degrading enzymes. Dashed lines denote intra- and inter-polymer hydrogen bonds. Ac., O-acetyl group; Af, α -L-arabinofuranose; Fe, ferulic acid; G, glucose; GalA, D-galacturonic acid; Gu, 4-O-methyl-D-glucuronic acid; X, xylopyranose (xylose).

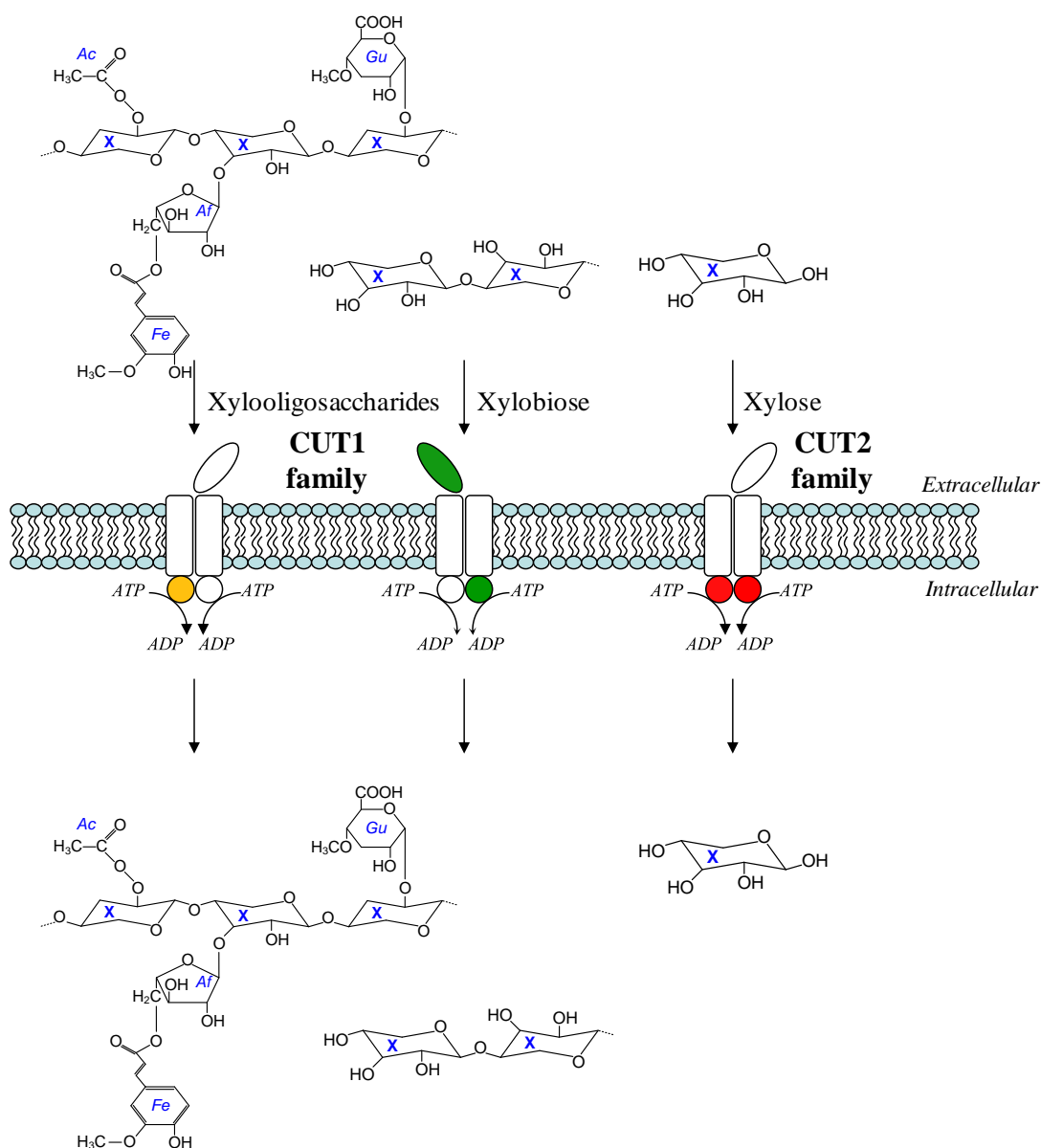


Figure 6.2. A simplified schematic representation of the ATP-driven transport systems likely to mediate the assimilation of xylan derived soluble sugars. Ac., O-acetyl group; Af, α-L-arabinofuranose; CUT1 and CUT2, Carbohydrate Uptake Transporter family 1 and family 2 respectively; Fe, ferulic acid; Gu, 4-O-methyl-D-glucuronic acid; X, xylopyranose (xylose).

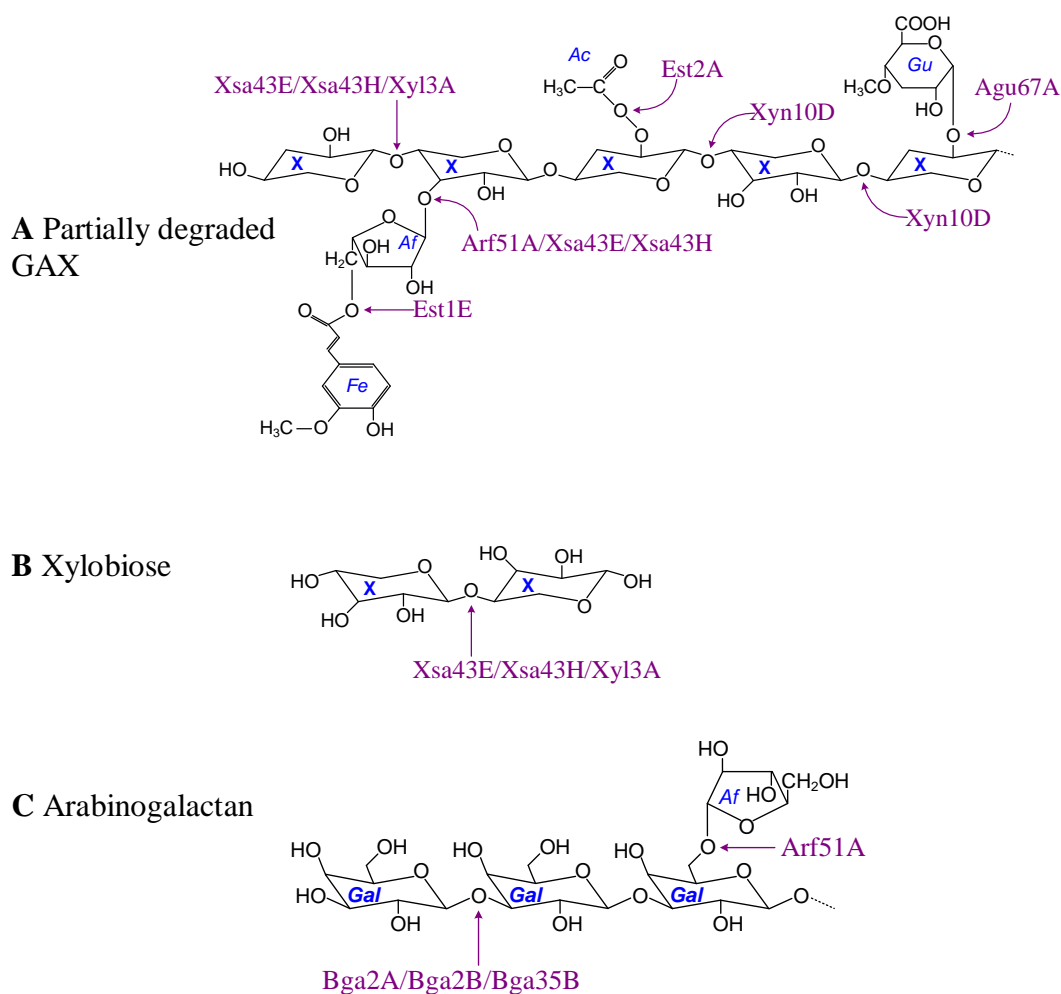


Figure 6.3. A simplified schematic representation of intracellular xylooligosaccharide degradation by *B. proteoclasticus*. Simplified structure of partially degraded GAX (A), xylobiose (B), and arabinogalactan (C) showing the predicted sites of hydrolysis by the identified *B. proteoclasticus* cytosolic polysaccharidases Ac., O-acetyl group; Af, α-L-arabinofuranose; Fe, ferulic acid; G, glucose; Gal, D-galacturonic acid, Gu, 4-O-methyl-D-glucuronic acid; X, xylopyranose (xylose).

6.2 A proteomic view of extracellular fibre degradation by *B. proteoclasticus*

6.2.1 The role of secreted xylanolytic enzymes

The size and insolubility of forage derived polysaccharides dictates that they are degraded in the extracellular environment before the released soluble oligo- and monosaccharides can be assimilated and utilised by microbial cells. This requirement necessitates the secretion of a variety of polysaccharide degrading enzymes across the bacterial cell wall, where they may be covalently or non-covalently tethered to the external cell-surface, or released into the rumen fluid. Assimilation of the released soluble sugars necessitates their rapid uptake across the bacterial cell wall, which in Gram-positive bacteria is mediated by a variety of extracellular substrate-binding proteins linked to dedicated sugar transport systems. An examination of the secreted component of the *B. proteoclasticus* proteome was an important step for characterisation of the fibre degrading enzyme system of the bacterium.

6.2.1.1 Endoxylanase Xyn10B

The observed differential abundance pattern of Xyn10B in response to growth on xylan and xylose, together with the prediction that the 137 kDa endoxylanase is the product of a highly expressed gene, suggests that it is the primary xylanolytic enzyme in the extracellular polysaccharide-degrading system of *B. proteoclasticus*. GH10 endoxylanases are important for xylan breakdown due to their catalytic versatility, wide substrate specificity, and ability to hydrolyse heavily substituted xylooligosaccharides (Biely *et al.*, 1997). The GH10 and CBM9 domains in Xyn10B both display homology to the equivalent domains in XynA of *Eubacterium ruminantium* (Q47871_9FIRM), and the XynA precursor of *Thermoanaerobacterium thermosulfurigenes* (Q60046_THETU). The *E. ruminantium* XynA exhibited high endoxylanase activity against oat-spelt xylan and xylooligosaccharides when expressed in *E. coli*, and the primary degradation products were xylobiose and xylotriose (Taguchi *et al.*, 2004b). It is plausible that *B. proteoclasticus* Xyn10B is an endo-1,4- β -xylanase capable of liberating variable length xylooligomers from hemicellulose (Figure 6.1 A).

The relative abundance pattern of Xyn10B in xylan compared to xylose grown cells is similar to the xylan stimulated differential abundance pattern of XynA from *C. cellulovorans* (Han *et al.*, 2004), and implies that Xyn10B abundance is sensitive to the available extracellular carbon source. Hespell *et al.* (1987) observed a significant

increase in total secreted endoxylanase activity in several *B. fibrisolvens* strains grown in the xylan containing medium compared to growth on non-xylan carbohydrates, which is consistent with our findings. Little is known of the mechanisms controlling polysaccharidase expression in rumen bacteria. The expression of a *Prevotella bryantii* B(1)4 xylanase gene cluster containing an extracellular endoxylanase is induced by medium to large xylooligosaccharides, and is controlled by the multi-domain regulatory protein XynR, which operates via a two-component mechanism (Miyazaki *et al.*, 2003, Miyazaki *et al.*, 2005). The gene encoding Xyn10B does not appear to be clustered within a polysaccharide utilisation locus, and no corresponding transcriptional regulator is located near the *xyn10B* locus in the *B. proteoclasticus* genome. Therefore, the mechanism by which Xyn10B abundance is increased in the presence of xylan, as well as the identity of the inducer molecules remains to be elucidated.

6.2.1.2 Xylosidase/arabinofuranosidase Xsa43J and glycoside hydrolase family 30 GH30A

The GH43 family includes β -xylosidases, α -L-arabinofuranosidases, and endoxylanases, and enzymes belonging to this family are necessary for the efficient degradation of hemicellulose (Sorensen *et al.*, 2007). In particular, α -L-arabinofuranosidases are important for their ability to cleave arabinose side chains that participate in inter- and intra-polymer cross-linking within the plant cell wall (Grabber *et al.*, 2000). The constitutively expressed xylosidase/arabinofuranosidase Xsa43J contains an N-terminal GH30 domain within the N-terminal region and a C-terminal GH43 catalytic domain, and in the CAZy database is classified in the GH30 family. Currently, the lack of homology of either GH domain to enzymes in the NCBI database with characterised functions makes it difficult to accurately predict the catalytic potential of this enzyme, and suggests Xsa43J possesses novel β -xylosidase or α -L-arabinofuranosidase activity in the rumen ecosystem. It is notable that the GH30 domain detected in the glycoside hydrolase family 30 Gh30A (Bpr_I2937), which is found two CDS downstream of Xsa43J, is most similar to the GH30 domain in Xsa43J.

Xsa43J also contains multiple C-terminal Type-I CWBs, two Big4 (Bacterial Ig-like domain-group 4) domains, and an uncharacterised sugar-binding domain suggesting that the enzyme may interact with insoluble polysaccharides *in vivo*. Xsa43J is one of three polysaccharidases encoded by the *B. proteoclasticus* genome

that is predicted to possess at least one Big4 module, which are present in a variety of secreted bacterial GH (Hung and Lee, 1998, Kataeva *et al.*, 2004), but are relatively poorly characterised. A Big4 domain in the *C. thermocellum* CbhA may improve the stability of the neighbouring catalytic domain, and is vital to enzyme function as deletion of the domain completely abolished CbhA catalytic activity (Kataeva *et al.*, 2004). The combination of catalytic and binding domains detected in Xsa43J suggests that the enzyme may attack glucuronoxylan, but functional analysis of catalytic and non-catalytic modules of Xsa43J will be important in assessing the contribution of this enzyme to *B. proteoclasticus* mediated hemicellulose degradation. This will also assist in determining the functional role of the glycoside hydrolase family 30 Gh30A.

6.2.1.3 Xylosidase/arabinofuranosidase Xsa43A

The xylosidase/arabinofuranosidase, Xsa43A (Bpr_I0302) is a 57.5 kDa enzyme that contains an N-terminal secretory signal peptide, a central GH43 catalytic domain, and a C-terminal CBM6 domain. The GH43 of Xsa43A is 44% identical to that of the XynD secreted by *Bacillus subtilis* subsp. *subtilis* ATCC 6051 (Bourgois *et al.*, 2007). Expression of recombinant XynD in *E. coli* produced an enzyme that was highly active upon wheat-bran arabinoxyloligosaccharides and intact arabinoxylan. The enzyme specifically hydrolysed (1→2)- α -linked and (1→3)- α -linked L-arabinofuranosyl groups from xylooligosaccharides of varying lengths, and showed a strong preference for mono-substituted backbone residues.

Xsa43A was one of only two identified polysaccharidases that contained a CBM6 domain. The CBM6 family contains CBM modules that are approximately 120 residues long, and different members of the family are capable of binding crystalline, and amorphous cellulose, as well as xylan. A CBM6 from *Clostridium cellulolyticum* (CcCBM6) recognised xylose either as a monosaccharide, at the non-reducing end of xylooligosaccharides, or within the side chain components of xyloglucan (Abbott *et al.*, 2009), while several family 6 modules recognise (1→3)- β -D-linkages at the non-reducing end of β -glucans (Correia *et al.*, 2009, van Bueren *et al.*, 2005). Furthermore, a CBM6 produced by *Cellvibrio mixtus* endoglucanase 5A showed a particularly diverse binding repertoire, in that it was capable of attachment to β -(1→3),(1→4)-mixed linkage glucans and barley β -glucan, cello-oligosaccharides, crystalline cellulose, and xylooligosaccharides (Henshaw *et al.*, 2004, Pires *et al.*, 2004). The promiscuous substrate recognition may be explained by the presence of multiple binding sites within the secondary structure of some CBM6 domains.

Family 6 modules appear to have co-evolved with their associated catalytic domains to acquire the same substrate specificity (Michel *et al.*, 2009). Xsa43A was not identified in the set of proteins that adsorbed to Avicel *in vitro*. Nonetheless, if the substrate preference and hydrolytic activity of the *B. subtilis* XynD homologue is reflected in the hydrolytic activity of the *B. proteoclasticus* Xsa43A, it is plausible that Xsa43A is secreted into the extracellular environment where it acts on intact or degraded arabinose containing xylans. It is possible that Xsa43A augments the activity of Xsa43J to hydrolyse α -linked L-arabinofuranosyl groups from the xylan backbone prior to xylooligosaccharide assimilation.

6.2.1.4 Endo-1,4- β -xylanase and xylosidase Mxy10-43A, and endo-1,4- β -xylanase Xyn10C

The two endoxylanases identified in the Avicel affinity assay but not by 2-DE were predicted to be functionally similar enzymes that may play an important role in extracellular GAX degradation. The functional domain arrangement of Xyn10C is replicated in the C-terminal region of Mxy10-43A. This similarity is supported by the fact that the closest BLASTp homologue to Xyn10C currently in the NCBI database is Mxy10-43A, which suggests that a prior gene duplication event may have occurred.

The GH10 domain of both enzymes is similar to the catalytic domain in the 47 kDa GH10 *B. fibrisolvens* 49 XynA (76% identity). When expressed in *E. coli* JM83 XynA exhibited endoxylanase activity, but no arabinosidase, β -xylosidase or cellulase activity (Mannarelli *et al.*, 1990), which is a hydrolytic pattern characteristic of true endoxylanases (Reilly, 1981). Based on this similarity and the apparent low abundance of Xyn10C and Mxy10-43A in the culture medium, it is plausible to expect that both enzymes augment the activity of the highly abundant Xyn10B by similarly attacking internal β -1,4 glycosidic linkages within the xylan backbone of hemicellulose.

Mxy10-43A contains an additional N-terminal GH43 domain and two centrally located CBM6 domains that are not present in Xyn10C. The GH10/GH43 combination is unique within the Pfam database, suggesting that Mxy10-43A represents a newly identified enzyme within the rumen microbial ecosystem. The GH43 domain of Mxy10-43A is 45% identical to that found in the 64 kDa GH43 XynD produced by *Paenibacillus polymyxa*. Expression and purification of the *xynD* gene product in *B. subtilis* produced an enzyme that specifically hydrolysed α -(1 \rightarrow 3) linkages between arabinose and xylose, and released only arabinose from oat-spelt and wheat flour xylans (Morales *et al.*, 1995). The enzyme did not hydrolyse the backbone of

xylooligosaccharides. The purified XynD did not require the presence of additional xylan backbone degrading enzymes for maximal activity, but enhanced the ability of *B. subtilis* endoxylanases to attack the xylan backbone.

This data supports the likelihood that Mxy10-43A possesses both endoxylanase and arabinosidase activity, and can attack xylose/arabinose side chain linkages as well at the backbone of xylooligosaccharide and xylan. The potential synergism between these domains may enhance Mxy10-43A fibrolytic potential. Studies have demonstrated up to 20-fold positive synergistic affects between endoxylanase and arabinofuranosidase activities (Greve *et al.*, 1984, Lee and Forsberg, 1987), possibly due to a reduction of arabinose induced stearic hindrance (Brillouet, 1987).

6.2.2 The possible role of secreted non-xylanolytic enzymes

In addition to the five xylanolytic enzymes identified in the *B. proteoclasticus* culture medium, the presence of an endoglucanase, a pectin methylesterase, and a pectate lyase was also confirmed. *B. proteoclasticus* is incapable of growth on either cellulose or galacturonic acid *in vitro* (Attwood *et al.*, 1996) therefore the secretion of these three non-xylanolytic polysaccharidases suggests they act synergistically with the xylanolytic enzymes to enhance the rate and extent of plant cell wall degradation (Yu *et al.*, 2003).

6.2.2.1 Endo-1,4- β -glucanase Cel5C, pectate lyase Pel1A, and pectin methylesterase Pme8B

Extracellular degradation of cellulose requires the presence of endo- β -1,4-glucanases and cellodextrinases that cleave internal and external (1 \rightarrow 4)- β -D-glycosidic linkages respectively, to produce soluble cellodextrins that can be assimilated and degraded to glucose. The full length sequence of Cel5C is 99% identical to that of the product of *end1* expressed by *B. fibrisolvens* H17c(SA) (Berger *et al.*, 1989), which is now classified as *B. proteoclasticus* (Moon *et al.*, 2008). When cloned and expressed in *E. coli* End1 was active against several (1 \rightarrow 4)- β -D-glycosidic linkage containing glucans, but did not hydrolyse (1 \rightarrow 3)- β -D-linkages, and was not active upon synthetic cellobiose substrates (Berger *et al.*, 1989). It was concluded that End1 was a true endoglucanase hydrolysing internal glycosidic bonds within cellulose. The presence of the endocellulase Cel5C in the *B. proteoclasticus* culture medium suggests that cells are capable of hydrolysing cellulose to cellotriose and longer cellodextrins, but the absence of extracellular cellodextrinases may prevent further degradation.

B. proteoclasticus cells grown *in vitro* are able to utilise glucose and cellobiose, but not cellulose or dextrin (Attwood *et al.*, 1996). The outer parts of the cellulose microfibrils are often more amorphous than the core regions (O’Sullivan, 1997), and it is possible that Cel5C targets these ‘weaker’ regions but is unable to attack crystalline cellulose directly. Although one predicted extracellular β -glucosidase (Bgl3A) was identified in the cytosolic compartments, this enzyme has a C-terminal TMD indicating that it has limited access to cellodextrins in the external environment and may be insufficient to liberate enough glucose to support rapid growth. Cel5C was the only member of PUL16 that was identified. PUL16 encodes two cytosolic polysaccharidases that are annotated as glycoside hydrolase family 2 Gh2F (Bpr_I1699) and glycoside hydrolase family 32 Gh32A (Bpr_I1704), as well as several ABC-transporter proteins. Functional analysis of these proteins may assist in determining the role of Cel5C in the degradation of forage material by *B. proteoclasticus*.

Taken together, this data is consistent with the possibility that *B. proteoclasticus* cells lack the transport machinery necessary for the assimilation of cellotriose and longer cellodextrins, which supports the proposed synergistic role for Cel5C in lignocellulose degradation. Evidence for a possible synergistic role for Cel5C in hemicellulose degradation is provided by Murashima *et al.*, (2003) who demonstrated that extent of corn cell wall degradation after 15 h incubation with a mixture containing xylanase and cellulase enzymes was 61% greater than the degradation achieved by xylanase alone. Furthermore, the quantity of both xylo- and cellooligosaccharides present after the incubation was significantly greater when xylanase and cellulase enzymes were present, which suggested that the activity of each enzyme complemented that of the other. That Cel5C is down-regulated during mid-log and stationary phase growth in the presence of xylan suggests that accumulation of xylan breakdown products in the culture medium may repress Cel5C abundance, although further analysis using additional substrates such as short chain xylooligosaccharides will be needed to elucidate how abundance of this enzyme is controlled.

The presence of pectin in forage plant cell walls restricts the access of xylanases and cellulases to their substrates and hinders lignocellulose degradation (Benshalom, 1986). Removal of esterified methyl groups from pectin by Pme8B is likely to increase the susceptibility of the pectin backbone to degradation by Pel1A, and improve the

activity of other polysaccharidases. Although Pel1A was not detected in the culture medium of xylan-grown cells during stationary phase, the low abundance in the culture medium of xylose-grown means we cannot discount the possibility that Pel1A is constitutively expressed. Similar abundance of PelA was observed in the supernatant of *C. cellulovorans* cells grown on a variety of plant polysaccharides suggesting that it was constitutively expressed (Han *et al.*, 2004).

6.2.2.1.1 Pullulanase Pul13A

Pul13A is the only pullulanase encoded by the *B. proteoclasticus* genome, and was detected in significantly greater abundance in the cytosol of xylan grown cells at mid-log phase growth. During mid-log phase Pul13A was detected in two spots located at approximately pI 4.2 and 100 kDa, which corresponded closely with the theoretical values for this enzyme. One of these spots was unique to the xylan growth condition. During stationary phase, the differentially abundant spot became undetectable in all three culture conditions, but a new spot approximately 5 kDa smaller became apparent, and was present in all three growth conditions at comparable abundance. The presence of xylan in the culture medium therefore appears to stimulate the up-regulation of pullulanase Pul13A (Bpr_III161) during mid-log phase growth, but during stationary phase the enzyme may be constitutively expressed.

The genomes of many Gram-positive, polysaccharolytic bacteria encode type I pullulanases (Ben Messaoud *et al.*, 2002, Bertoldo *et al.*, 2004, Doman-Pytka and Bardowski, 2004, Malle *et al.*, 2006), and pullulanase activity has been detected in polysaccharolytic gut isolates including *B. fibrisolvens* 16/4 (Anderson, 1995, Ramsay *et al.*, 2006). Nonetheless, Pul13A is the first pullulanase shown to be endogenously synthesised by a rumen bacterium.

Pullulanases (EC 3.2.1.41) are enzymes capable of hydrolysing α -1,6 glycosidic linkages in polymers such as starch, pullulan, and other branched oligosaccharides. On the basis of substrate specificity and product profile, pullulanases are classified into two groups. Pullulanase type I enzymes (also called debranching enzymes) specifically hydrolyse α -1,6 glycosidic bonds in polymers such as amylopectin, and belong to the family 13 glycosyl hydrolases. Pullulanase type II enzymes (also called amylopullulanases) hydrolyse both α -1,4 and α -1,6 linkages in α -glucans such as amylose, and fall mostly within the GH57 family. The action of type I and type II pullulanases on starch generates α -1,4 linked linear oligosaccharides, and sugars such as glucose, maltose, and maltotriose, respectively. Pul13A is 46% identical to a

pullulanase encoded by the recently sequenced *Butyrivibrio crossotus* DSM 2876 genome, and 38% identical to several pullulanase type I enzymes produced *Bacillus* species. It is therefore likely that Pul13A specifically hydrolyses α -1,6 glycosidic linkages in glucooligosaccharides.

Starch is the main component of the endosperm tissue within the seeds of flowering plants, and is an important energy storage polysaccharide and source of nutrition to the developing embryo during the early stages of plant development. Starch consists of approximately 25% amylose and 75% amylopectin, but these proportions vary depending on plant type. Amylose is a water soluble, linear polymer of α -1,4 linked D-glucose units, while amylopectin is a highly branched polymer of α -1,4-D-glucose main chains with α -1,6 linked branch points occurring every 24 to 30 glucose residues. Starch is present in large amounts in many nutritionally important plants including wheat, maize, barley, sorghum, and oats, but is not a major component of primary or secondary plant cell walls.

The relative abundance pattern of Pul13A is most likely explained by the presence of starch in the xylan supplemented culture medium. The oat-spelt xylan used in this project was selected because it was considered most similar to GAX that is likely to be present in ruminant forage crops. An unavoidable contaminant of the oat-spelt xylan is glucooligosaccharides that may be present in quantities as high as 15% of the dry mass. This possibility is supported by the induction pattern of *Bacillus cereus* FDTA-13 pullulanase activity in response to growth on branched and linear polysaccharides (Nair *et al.*, 2007). High pullulanase activity was detected only in medium containing branched polysaccharides, and medium containing soluble starch, maltodextrin, and amylopectin elicited particularly high enzyme titres. Pullulanase activity was low in the culture medium containing glucose or amylose, and was not detected in the absence of carbon source, which indicated that pullulanase activity was inducible and responsive to the presence of α -1,6 linkage containing branched polysaccharides only.

6.2.2.1.2 α -Amylase Amy13A

The N-terminal portion of *B. proteoclasticus* Amy13A, spanning residues 1-975 contains the GH13 catalytic domains and is 99% homologous to the full length sequence of the α -amylase AmyA synthesised by *B. fibrisolvens* H17c(SA) (Rumbak *et al.*, 1991), now *B. proteoclasticus* (Moon *et al.*, 2008). Furthermore, this N-terminal portion is also 50% and 45% identical to predicted α -amylases encoded within the recently sequenced genomes of *B. fibrisolvens* 16/4 and *B. crossotus*. When the

B. fibrisolvens H17c(SA) *amy13A* was cloned and expressed in *E. coli* C600, a 107 kDa enzyme exhibiting α -amylolytic activity was synthesised (Rumbak *et al.*, 1991). Regardless of whether the *B. fibrisolvens amy13A* transfected cells were grown in the presence of maltose, glucose, or starch, almost 90% of the α -amylolytic activity was associated with the periplasm, with the remainder distributed evenly between the cytosol and secreted sub-cellular compartments. The amylolytic activity had a pH optimum of 6.8, and was stable between pH 5.5 and 7.5. The optimum temperature for enzyme activity was 50°C, although above 45°C the enzyme became unstable and was completely inactivated at 52°C. In all conditions examined the α -amylolytic was dependent on the presence of CaCl₂ in both the growth medium and assay buffers.

When crude enzyme extracts harvested from *B. fibrisolvens amy13A* transfected *E. coli* cells were incubated with maltohexose, the major degradation product detected was maltotriose, suggesting that the enzyme was an endoamylase (Rumbak *et al.*, 1991). This hypothesis was confirmed by analysis of starch, amylose, and amylopectin degradation, where the initial hydrolysis product was maltotriose, with small amounts of maltose (G2) and maltotetraose (G4). After 24 h incubation only a portion of the maltotriose was degraded to maltose and glucose. Furthermore, no activity was detected against maltose when provided as a substrate. The enzyme appeared unable to hydrolyse branched chain maltooligosaccharides, suggesting it was not active upon α -1,6 branch points, which is a characteristic of endoamylases.

6.2.3 Extracellular proteins may be localised to the external cell surface *in vivo*

Fibrolitic rumen bacteria such as *R. flavefaciens* and *R. albus*, and several non-rumen fibrolitic *Clostridial* species incorporate many of their secreted glycosyl hydrolases into a cell wall associated multienzyme complex termed the cellulosome (Bayer *et al.*, 1985, Doi and Tamaru, 2001, Kim *et al.*, 2001, Rincon *et al.*, 2003). *B. proteoclasticus* secreted enzymes lack the scaffoldin/cohesin domains that are necessary for cellulosome assembly (Carvalho *et al.*, 2007, Park *et al.*, 2001, Rincon *et al.*, 2003), and the set of identified secreted polysaccharides appear to use several approaches to tether extracellular enzymes to the cell wall. A feature of Amy13A, Pel1A, Pme8B, Xsa43J, and Xyn10B is the presence of multiple CWBD1 domains at their C-termini (Kelly *et al.*, 2010). Cell wall binding (CWB) domains are found in many extracellular GH where they tether the enzyme to the external surface of the bacterial cell wall and are important for polysaccharide degradation (Brechtel *et al.*, 1999, Burchhardt *et al.*, 1994, Fuchs *et al.*, 2003, Matuschek *et al.*, 1994, Matuschek

et al., 1996, May *et al.*, 2006, Sahm *et al.*, 1996). Attachment is via a covalent linkage to choline residues of cell wall associated teichuronic acids, or non-covalent interaction with pyruvylated cell wall polymers (Mesnage *et al.*, 2000). It is likely that the five secreted *B. proteoclasticus* polysaccharidases are non-covalently attached to the external surface of the cell wall *in vivo*. An additional five predicted secreted proteins that also contained CWBD1 domains were identified, which implies that CWBD1 modules play an important role in the attachment of secreted proteins to the external cell surface of *B. proteoclasticus*. The presence of these enzymes in the culture media may be explained by the activity of extracellular proteases, as demonstrated in a number of *Bacillus* strains (Antelmann *et al.*, 2001, Voigt *et al.*, 2006), or the ionic composition of the culture medium (Yother and White, 1994).

6.2.4 Carbohydrate-binding module containing proteins, cellulose adsorption, and substrate adherence

An intriguing feature of the set of identified secreted proteins was the presence of the two carbohydrate-binding proteins encoded by the *B. proteoclasticus* genome sequence (products of loci Bpr_I0736 and Bpr_I1599). Both proteins contained tandem CBM2a domains that were homologous to those present in the secreted *B. proteoclasticus* endocellulase Cel5C. Furthermore, the gene product of locus Bpr_I1599 also contained tandem CBM6 domains (PF03422) located towards the N-terminus. Neither protein contained an identifiable cell wall binding domain, suggesting they are released into the extracellular space rather than being retained on the external cell surface, and do not therefore mediate bacterial cell attachment to plant polysaccharides. CBMs are known to possess additional non-hydrolytic functions such as polysaccharide disruption and surface modification. The CBM2 of the *C. fimi* endoglucanase A was shown to disrupt the structure of cellulose fibres, causing the release of small cellulosic particles (Din *et al.*, 1991). The secretion of both carbohydrate-binding proteins at equivalent levels by cells grown in the presence of xylan and xylose suggests they play a non-hydrolytic role in fibre-degradation and may target non-hemicellulosic substrates such as crystalline or amorphous cellulose. It is possible that the secreted carbohydrate-binding proteins also act synergistically with the secreted polysaccharide-degrading enzymes. Molecular and biochemical examination of each of these secreted proteins will be important in understanding their roles *in vivo*, and their relative contributions to the overall process of plant cell wall degradation by *B. proteoclasticus*.

Most of the identified extracellular polysaccharidases and both the secreted non-catalytic carbohydrate-binding proteins contained one or more carbohydrate binding modules. CBMs are well characterised domains that target the associated catalytic domain or domains to specific regions of the polysaccharide, which dramatically increases enzyme concentration at the site of activity, and improves enzyme substrate interaction and catalysis (Black *et al.*, 1996, Bolam *et al.*, 1998). Cellulose adsorption assays have been used to examine the substrate specificity and to affinity purify CBM containing polysaccharidases from the culture medium of fibrolytic bacteria (Devillard *et al.*, 2004, Han *et al.*, 2004, Shoseyov and Doi, 1990),

Adsorption to Avicel (crystalline cellulose) was used in this project to examine the binding potential of *B. proteoclasticus* proteins present in the growth medium of xylan grown, stationary phase harvested cells. Fourteen well resolved 1-DE bands could be detected after the Avicel adsorption assay, 10 of which were identified and found to be a mixture of four secreted polysaccharidases, one carbohydrate binding protein, and several ABC transporter SBPs. The 1-DE separation of extracellular proteins that adsorbed to crystalline cellulose showed that Xyn10C was detected in one band at approximately 105 kDa, which is similar to the predicted size of the full length enzyme, and in a second band of approximately 100 kDa. The peptide sequence coverage data indicated the lower size band may have been due to an N-terminal truncation. Mxy10-43A was present in one band of approximately 75 kDa that was likely to be a C-terminal fragment. The full-length Xyn10C and the C-terminal fragment of Mxy10-43A had estimated *pI* values of 5.5 and 4.9 respectively, which together with their observed sizes suggest that they resolve to a region of the 2-DE separation that contained few detectable proteins. Therefore, the most likely explanation for the fact that neither polysaccharidase was detected in the 2-DE separations of proteins harvested from culture medium of xylan grown cells is that the abundance was below the limit of detection by colloidal CBB G-250 staining.

A feature common to three of the four polysaccharidases and the carbohydrate binding protein was the presence of at least one CBM2a module. The dissimilar enzyme was the CBM9 containing Xyn10B. The CBM2 family is currently the largest within the CAZy Database and is subdivided in families 2a and 2b on the basis of primary sequence within the substrate-binding region (Tomme *et al.*, 1995). An important functional difference between families 2a and 2b is that family 2a CBMs bind cellulose, irrespective of the type of enzyme from which they originate, while

family 2b CBMs bind xylan. Intriguingly, no CBM2b containing polysaccharidases were identified in the *B. proteoclasticus* genome. The family 2a CBM found within the *Cellulomonas fimi* Xyn10A had a high affinity for crystalline cellulose, but also bound to amorphous regions of the polymer (McLean *et al.*, 2000), while the CBM2a module of the *C. fimi* CenA was recently shown to adsorb to crystalline cellulose (Jing *et al.*, 2009), using a methodology similar to that used in this project.

CBM9 domains such as the one present in Xyn10B are found only in xylanases, and some possess promiscuous substrate specificity. For instance, the CBM9-2 domain in the *Thermotoga maritima* Xyn10A was able to bind to crystalline and amorphous cellulose, xylan, and a range of soluble cello- and xylooligomers (Boraston *et al.*, 2001). A unique property of the CBM9 modules is their ability to bind specifically to the non-reducing ends of cellulose and soluble polysaccharides (Boraston *et al.*, 2001). The CBM9 domain in *E. ruminantium* XynA that is homologous to the *B. proteoclasticus* Xyn10B domain adsorbed to acid swollen cellulose, and to a lesser degree xylan (Taguchi *et al.*, 2004b).

The discovery that several secreted *B. proteoclasticus* polysaccharidases and one carbohydrate binding module are able to adsorb specifically to Avicel *in vitro* is in agreement with the previously reported substrate specificities of their associated CBMs and implies that they primarily target cellulose containing polymers *in vivo*. Biochemical characterisation of their precise substrate specificities will be essential to defining the contribution of each of these enzymes to the fibre degrading system of *B. proteoclasticus*.

6.2.4.1 Cell adherence to plant polysaccharides

Adherence to plant material is crucial for efficient degradation of plant cell wall polysaccharides (Jindou *et al.*, 2006, Jun *et al.*, 2007, Miron *et al.*, 1998, Mosoni and Gaillard-Martinie, 2001), and the fibre adherent bacteria account for approximately 90% of rumen cellulase and xylanase activity (Akin, 1980, Craig *et al.*, 1987, Minato *et al.*, 1993). With the exception of Xyn10B, (Bpr_I0026), a feature of the set of polysaccharidases and carbohydrate binding proteins that adsorbed to Avicel was that none of the remaining enzymes contain cell wall binding domains and would therefore not be expected to be tethered to the external cell surface. In contrast, the four polysaccharidases identified by 2-DE that contained no CBMs, and therefore did not bind to Avicel *in vitro*, all possessed multiple cell wall binding domains located at their C-termini. This CBM/CWB domain mutual exclusivity implies that none of these

extracellular polysaccharidases play a role in bacterial cell attachment to insoluble polysaccharides *in vivo*.

The *B. proteoclasticus* genome sequence indicates that the cell has several potential mechanism for adhering to plant cell walls, which include flagella, pili, cell surface proteins, and exopolysaccharides (Kelly *et al.*, 2010). One of four cell surface proteins containing *Listeria/Bacteroides* repeats (TIGRfam accession number TIGR02543) that are encoded in the *B. proteoclasticus* genome was identified in the culture medium of stationary phase harvested xylan and xylose grown cells (Bpr_I2508), and many proteins involved in exopolysaccharide synthesis were also identified. Therefore it is possible that *B. proteoclasticus* may use a variety of cell surface proteins as well as exopolysaccharide as a means of cell attachment to polysaccharide, although further analysis is required as no conclusions can be made based on the proteomic data alone.

6.2.5 Proteolytic enzymes in the *B. proteoclasticus* proteome

B. proteoclasticus was originally isolated from a screen for proteolytic rumen bacteria (Attwood and Reilly, 1995), and the cells were found to possess high cell-associated serine-type proteolytic activity (Attwood *et al.*, 1996). In this project the secreted subtilisin family serine protease (Bpr_I2629) was detected in several abundant protein spots in culture medium and cytosol, and was the only secreted enzyme of the type identified. The protease was one of the most abundant proteins detected in the secreted fraction under all growth conditions. It is possible that this protein accounted for the proteolytic activity documented in the original study (Attwood *et al.*, 1996). Furthermore, in the mid-log phase harvested *B. proteoclasticus* cytosolic fraction, the serine protease was the most strongly up-regulated protein in the proteome of xylan grown cells, although this pattern was not reflected in the cytosol of stationary phase harvested cells, or in the secreted fraction.

Despite the 153 kDa subtilisin family serine protease (Bpr_I2629) not being directly involved in forage degradation and fermentation, reports suggest that bacterial extracellular proteases may play an auxiliary role in plant polysaccharide degradation in the rumen (Colombatto and Beauchemin, 2008, Eun and Beauchemin, 2005). Treatment of ruminant feed with a commercially prepared protease mixture resulted in a significantly increased *in vivo* fibre digestibility, and concomitant elevation in rumen xylanase and endoglucanase activities (Eun and Beauchemin, 2005). Biochemical characterisation demonstrated that it was a subtilisin-like serine protease that accelerated the rumen fluid induced disappearance of alfalfa hay hemicellulose

(Colombatto and Beauchemin, 2008), but had negligible fibrolytic activity. Cell wall structural proteins are implicated in mediating polysaccharide/lignin cross-links (Showalter, 1993) and it was hypothesised that cell wall proteolysis allowed rumen fibrolytic microbes better access to plant polysaccharides. It is possible that the secreted *B. proteoclasticus* serine protease may play a similar role, targeting cell wall structural proteins and acting synergistically with the secreted polysaccharidases to enhance the rate and extent of polysaccharide degradation *in vivo*. Examination of the effect of the serine protease on plant fibre, in isolation or in combination with other *B. proteoclasticus* fibrolytic enzymes will be important in understanding the role of the 153 kDa subtilisin family serine protease in the extracellular environment.

6.3 A proteomic view of sugar assimilation and cytosolic degradation by *B. proteoclasticus*

6.3.1 Polysaccharide utilisation loci

PULs such as the Sus PUL of the human gut bacterium *B. thetaiotaomicron* (Bjursell *et al.*, 2006, Martens *et al.*, 2009a, Reeves *et al.*, 1997) are co-ordinately regulated cell envelope systems that typically comprise polysaccharide degrading enzymes and cell associated transmembrane transporter proteins, often together with signal transducers, transcriptional regulators, and auxiliary proteins necessary for polysaccharide utilisation. PULs have been identified in several Gram-positive (Berg Miller *et al.*, 2009) and Gram-negative bacteria. In *B. thetaiotaomicron* a subset of the 88 putative PULs were shown to be transcriptionally regulated in response to the presence of gut mucin O- and N-glycans (Bjursell *et al.*, 2006, Reeves *et al.*, 1997). Almost half the genes in the *B. proteoclasticus* genome that encode lignocellulose degrading enzymes, and many of the sugar specific ABC transporter associated proteins are clustered within one of 36 PULs. In complex microbial ecosystems such as the rumen where microbes are exposed to a dynamic and diverse range of poly- and oligosaccharides, such genetic arrangement presumably allows the rapid and coordinated control of enzyme production, substrate transport and intracellular metabolism (Kelly *et al.*, 2010).

The set of 416 proteins identified in the secreted, cell-associated, or cytosolic *B. proteoclasticus* proteomes contained 51 proteins (12% of the identified proteome) that were predicted to be components of a polysaccharide utilisation locus. Furthermore, proteomic analysis of the *B. proteoclasticus* membrane fraction performed by Dr. Judy Bond confirmed the presence of 10 additional PUL proteins,

nine of which were not detected in this project. Collectively, these 60 proteins were components of 24 of the 36 *B. proteoclasticus* PULs (Table 6.3). Despite the fact that none of the 36 PULs were fully represented in the complete *B. proteoclasticus* proteomic dataset, the proportion of PULs containing at least one identified protein together with the prevalence of differentially abundant polysaccharidases and transport proteins strongly supports the notion that PULs are an important component of the fibre degradation and assimilation system of *B. proteoclasticus*.

The 17 polysaccharidases within the set of PUL proteins represented 13 different GH families, and also included the acetyl-xylan esterase unmatched to a CAZy family. Notably, only one of the 20 sugar specific ABC transporter proteins identified in *B. proteoclasticus* proteome (Bpr_I1667) was not a predicted PUL component. Of the set of 46 differentially abundant proteins detected in all subcellular fractions at either mid-log or stationary phase, 17 were PUL components that were associated with 13 of the 36 PULs. Of the 17 differentially abundant PUL proteins, 12 were more abundant in the cytosol of cells grown in the presence of xylan, relative to either of the alternate growth substrates. These results are similar to those obtained from transcriptional analysis of mucosal glucan usage by *B. thetaiotaomicron* (Martens *et al.*, 2008), where it was found that 44% of all differentially regulated genes were PUL components, and these were distributed across 43 of the 88 PULs. It is possible that the differences in PUL protein response between the *B. thetaiotaomicron* study and this project are in part due to the transcript versus proteomic profiling, together with the undoubted chemical complexity and heterogeneity of gut mucins relative to oat-spelt xylan.

Table 6.3. Predicted polysaccharide utilisation proteins identified in the *B. proteoclasticus* proteome.^a

PUL	Protein	Locus	Function ^b	Size ^c
2	Sugar ABC transporter substrate-binding protein	Bpr_I0117	T	5
3	Xylulokinase, XylB	Bpr_I0173	C	23
	Acetyl-xylan esterase	Bpr_I0174	C	
	α -D-Glucuronidase, Agu67A	Bpr_I0177	C	
	Sugar ABC transporter permease protein	Bpr_I0180	T	
	Sugar ABC transporter substrate-binding protein	Bpr_I0182	T	
	β-Xylosidase, Xyl3A	Bpr_I0184	C	
	L-Fucose isomerase related protein	Bpr_I0185	C	
	Hypothetical protein	Bpr_I0188	H	
	UTP-glucose-1-phosphate uridylyltransferase, GalU	Bpr_I0191	C	
	UDP-galactose 4-epimerase, GalE	Bpr_I0192	C	
4	Sugar ABC transporter substrate-binding protein	Bpr_I0237	T	6
5	Xylosidase/arabinofuranosidase and esterase, Xsa43H	Bpr_I0301	C	15
	Xylosidase/arabinofuranosidase, Xsa43A	Bpr_I0302	C	
	Sugar ABC transporter substrate-binding protein ^d	Bpr_I0305	T	
	NHL repeat-containing protein ^d	Bpr_I0308	U	
	Hypothetical protein ^d	Bpr_I0310	H	
	Sugar ABC transporter substrate-binding protein	Bpr_I0313	T	
7	β -Glucosidase, Bgl3A	Bpr_I0693	C	16
9	5-Keto 4-deoxyuronate isomerase, KduI	Bpr_I0929	C	10
	2-Deoxy-D-gluconate 3-dehydrogenase, KduD	Bpr_I0930	C	
	2-Keto-3-deoxygluconate 6-phosphate aldolase/2-keto-4-hydroxyglutarate aldolase	Bpr_I0931	C	
	2-Dehydro-3-deoxygluconokinase, KdgK	Bpr_I0932	C	
	Sugar ABC transporter substrate-binding protein	Bpr_I0937	T	
10	Endo-1,4- β -xylanase and esterase, Xyn10D	Bpr_I1083	C	4
11	Xylose ABC transporter substrate-binding protein	Bpr_I1173	T	8
	Xylose ABC transporter ATP-binding protein	Bpr_I1174	T	
12	Two-component system response regulator ^d	Bpr_I1234	S	6
	Chemotaxis protein CheW ^d	Bpr_I1237	B	
14	Sugar ABC transporter substrate-binding protein	Bpr_I1589	T	13
	Glucuronate isomerase, UxaC	Bpr_I1591	C	
	Cellodextrinase, Cel9B	Bpr_I1593	C	
15	α -L-Rhamnosidase, Rha78A	Bpr_I1686	C	11
16	Endo-1,4-β-glucanase, Cel5C	Bpr_I1710	C	12
17	Sugar ABC transporter substrate-binding protein	Bpr_I1720	T	5
19	Aldose 1-epimerase family protein	Bpr_I1782	C	7

Table continues

Table 6.3 continued.

PUL	Protein	Locus	Function ^b	Size ^c
21	β-Galactosidase, Bga35B	Bpr_I2006	C	8
	Sugar ABC transporter substrate-binding protein	Bpr_I2010	T	
22	Transcriptional regulator DeoR family ^d	Bpr_I2102	R	12
	1-Phosphofructokinase, PfkB	Bpr_I2103	C	
	PTS system IIABC fructose-specific family	Bpr_I2104	T	
	PTS system HPr phosphocarrier	Bpr_I2105	T	
	PTS system I PEP-phosphotransferase	Bpr_I2106	T	
23	Sugar ABC transporter substrate-binding protein ^d	Bpr_I2116	T	7
	Methyl-accepting chemotaxis protein McpL ^d	Bpr_I2117	B	
	Glycogen phosphorylase GlgP1 ^d	Bpr_I2118	C	
24	Sugar ABC transporter substrate-binding protein	Bpr_I2264	T	6
25	Sugar ABC transporter substrate-binding protein	Bpr_I2344	T	6
26	Sugar ABC transporter substrate-binding protein	Bpr_I2443	T	5
	Sugar ABC transporter permease protein	Bpr_I2444	T	
	Cellobiose phosphorylase, Cbp94A	Bpr_I2447	C	
29	Feruloyl esterase, Est1E	Bpr_I2870	C	4
31	Xylosidase/arabinofuranosidase, Xsa43J	Bpr_I2935	C	5
	Glycoside hydrolase family 30, Gh30A	Bpr_I2937	C	
	Oxidoreductase aldo/keto reductase family	Bpr_I2938	U	
	Acetyl-xylan esterase Est2A	Bpr_I2939	C	
34	β-Galactosidase, Bga2B	Bpr_III209	C	5
36	β-Mannosidase, Man2A	Bpr_III237	C	14
	HTH domain-containing protein	Bpr_III238	R	
	Sugar ABC transporter substrate-binding protein	Bpr_III244	T	

^a Bold text denotes enzymes that were more abundant in cells grown on xylan, relative to those grown on xylose or no added carbon.

^b B, Cellular processes; C, Carbohydrate metabolism; H, Hypothetical; R, Regulation; S, Signal transduction; T, Transporters; U, Unknown function.

^c The number of genes predicted to contribute to the PUL.

^d Tabulated data includes proteins identified in the *B. proteoclasticus* membrane proteome analysis performed by Dr. Judy Bond.

6.3.2 Polysaccharide utilisation locus 3

When *B. proteoclasticus* is grown in xylan containing culture medium, the hydrolytic activity of the secreted GH10 endoxylanase Xyn10B is expected to liberate variable length xylooligomers that may be substituted with (1→2)-α-linked and (1→3)-α-linked L-arabinofuranosyl groups (Vietor *et al.*, 1992), O-acetyl, and (1→2)-α-D-glucuronic acid or 4-O-methyl-D-glucuronic acid groups (Borneman *et al.*, 1990, Hartley *et al.*, 1990b, Mueller-Harvey *et al.*, 1986). The α-L-arabinofuranosyl side groups may also be esterified at the O-2 and/or O-5 positions with phenolic polymers

such as ferulic acid (Kulkarni *et al.*, 1999). Subsequent assimilation and utilisation of the substituted oligosaccharides requires the expression of genes encoding transporters, α -D-glucuronidase, and enzymes that convert xylooligosaccharides to xylose. PUL 3 appears to harbour most of the metabolic machinery necessary to affect the extensive degradation and subsequent utilisation of substituted oligosaccharides (Figure 6.4). PUL3 comprises three identified polysaccharidases, two ABC transporter system proteins, and the first two key xylose metabolism genes (L-fucose isomerase and XylB) (Kelly *et al.*, 2010), together with an unidentified α -D-glucuronidase (Gh115A), feruloyl esterase (Est1C), several AraC transcriptional regulators.

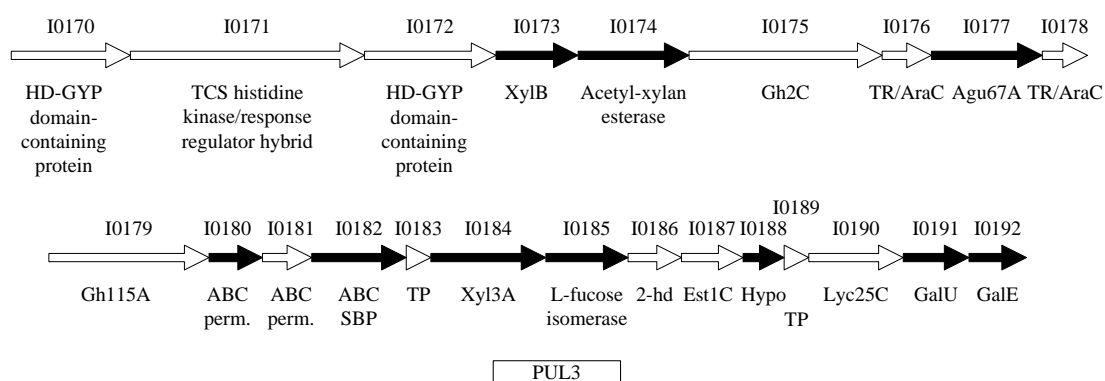


Figure 6.4. Organisation of the *B. proteoclasticus* PUL 3. Arrow size indicates the relative length of each protein. Filled and empty arrows indicate identified and unidentified proteins respectively. 2-hd, 2-hydroxyacid dehydrogenase; Hypo, hypothetical protein; perm., permease; SBP, substrate-binding protein; TCS, Two-component system; TP, transposase; TR/AraC, transcriptional regulator AraC family.

6.3.2.1 β -xylosidase Xyl3A

During stationary phase, C1120 was 4.3-fold more abundant in xylan grown cells relative to cells grown in the no added carbon condition. Furthermore, a 6.1-fold abundance increase was detected cells grown in xylan relative to xylose, although at a 5% FDR this increase was only marginally above the level of statistical significance ($q = 0.63$). The MALDI-TOF MS data indicated that spot C1120 contained the β -galactosidase Bga35A (Bpr_I2006) and the β -xylosidase Xyl3A (Bpr_I0184). The two proteins have very similar sizes and pI values, and the likelihood that the dual protein identification was a false positive result was $1.9e-27$. Taking the 2-DE protein abundance data alone, it is not possible to conclude whether one or both enzymes contributed to the differential abundance pattern.

1-D LC-MS/MS analysis of the cytosolic fraction showed that the differential abundance pattern of spot C1120 was most likely due to a difference in relative amounts of β -xylosidase Xyl3A. Xyl3A was identified with high statistical probability, a large number of matched peptides, and high protein sequence coverage in all three biological replicate fractions extracted from mid-log phase harvested xylan grown cells (Table 5.9). Furthermore, when proteins identified in the xylan grown cytosolic fraction were ranked according to relative protein abundance, Xyl3A was within the top 15% most abundant proteins. In contrast, Xyl3A was not detected in any replicate fraction harvested from xylose grown cells. The β -galactosidase Bga35A was undetected in both cytosolic fractions by 1-DE LC-MS/MS. Taking the 2-DE and LC-MS/MS datasets together, it is likely that the relative abundance of β -xylosidase Xyl3A was modulated by growth substrate, and positively regulated by the presence of xylan in the extracellular environment. This conclusion is supported by the results of the *B. proteoclasticus* microarray analysis that detected a >10-fold up-regulation of *xyl3A* during growth on xylan (Kong, 2007).

Xyl3A comprises two non-homologous GH3 functional domains located towards the N- (residues 30-273, pfam00933) and C-termini (residues 340-579, pfam01915) of the 709 amino acid long, full-length protein, and is between 54% and 63% identical to several GH family 3 domain-containing enzymes expressed by *Firmicute* phylum. Despite the fact that the GH3 family is one of the largest in the CAZy database (currently greater than 2000 entries) none of close homologues of Xyl3A have been examined experimentally. The strongest homologue for which there is biochemical and molecular biology data is the β -xylosidase B (Bxl3B) produced by *Clostridium stercorarium* NCIMB 11754 (Q8GJ42_CLOSR, 53% identical and 71% similar across the full-length protein sequence) (Adelsberger *et al.*, 2004). Bxl3B is a 715 amino acid long enzyme that has a predicted *pI* and size of 5.38 and 79.3 kDa respectively, which are similar to the corresponding values of *B. proteoclasticus* Xyl3A (*pI* 4.8 and 78.2 kDa). Bxl3B encodes a predicted cytosolic enzyme that contains a single GH3 catalytic module spanning residues 2-715. Recombinant Bxl3B was expressed in *E. coli* and isolated from the cell lysate as an 80 kDa band that had an activity towards pNP- β -xyloside¹ of 440 U.mg⁻¹ (Adelsberger *et al.*, 2004). At an optimal temperature of 50-55°C and an optimal pH of 6.0, Bxl3B rapidly and completely hydrolysed

¹ pNP- β -xyloside is a synthetic, chromogenic, *p*-Nitrophenol conjugated oligoxyloside used to assay for β -xylosidase activity in vitro, by measurement of liberated *p*-Nitrophenol.

xylobiose, and released xylose from the non-reducing end of xylotriose. The enzyme liberated only small amounts of xylose from pNP- α -arabinofuranoside, and did not hydrolyse xylose from intact arabinoxylan, or arabinose substituted xylooligosaccharides. Furthermore, Bxl3B showed very low or no detectable activity towards substrates such as pNP- β -glucoside, pNP- α -mannoside, carboxymethylcellulose, and barley β -glucan. This data suggests that the *B. proteoclasticus* Xyl3A may act exclusively on the non-reducing ends of short chain xylooligosaccharides.

The differential abundance pattern of protein spot C1120, containing *B. proteoclasticus* Xyl3A, after growth on xylan and xylose is similar to the induction pattern observed for *C. stercoreum* Bxl3B (Adelsberger *et al.*, 2004) and several other Gram-positive and Gram-negative bacteria. Compared to xylose grown cultures, a 2.3-fold increase in the transcription of the *C. stercoreum* *bxl3B* was detected in cells grown in arabinoxylan, which was accompanied by a concomitant 2.2-fold increase in cytosolic *C. stercoreum* β -xylosidase activity (Adelsberger *et al.*, 2004). The transcription of all other *C. stercoreum* β -xylosidase genes assayed was unchanged between xylan and xylose grown cells, indicating the enhanced cytosolic β -xylosidase activity was likely due to an increased abundance of Bxl3B. In addition, relative to xylose grown cells, a two-fold-reduction in cytosolic β -xylosidase activity was observed when *C. stercoreum* cells were grown in glucose containing medium, which suggests *bxl3B* was subject to glucose repression. Consequently, Bxl3B was proposed to play a direct role in arabinoxylan hydrolysis by the degradation of xylooligosaccharides. Cytosolic β -xylosidase activity in several other Gram-positive xylanolytic bacteria including *Thermoanaerobacterium saccharolyticum* and *Bacillus stearothermophilus* has also been shown to be stimulated by growth on xylan or long-chain xylo-oligosaccharides (Cho and Choi, 1998, Lee *et al.*, 1993). Furthermore, the β -xylosidase gene *xynB* of the Gram-negative *P. bryantii*, together with its cytosolic transcription product are positively regulated in the presence of medium to large xylooligosaccharides (Miyazaki *et al.*, 2003, Miyazaki *et al.*, 2005).

6.3.2.2 α -D-Glucuronidase Agu67A

The α -D-glucuronidase, Agu67A (Bpr_I0177) is a 76.1 kDa enzyme containing a C-terminal GH67 domain, and was identified with very high statistical confidence in the *B. proteoclasticus* cytosol using 2-DE and 1-D LC-MS/MS techniques. α -D-

Glucuronidases (EC 3.2.1.139) cleave the α -1,2-glycosidic linkage between 4-O-methyl-D-glucuronic acid and the xylan backbone, and are necessary for the complete degradation of hemicelluloses such as GAX. Currently, the structural and functional properties of nine bacterial GH67 enzymes have been examined, including the AguA enzymes produced by *Geobacillus stearothermophilus* T-1 (Golan *et al.*, 2004, Shallom *et al.*, 2004) and *Paenibacillus* sp strain JDR-2, which are 51% and 56% identical to the *B. proteoclasticus* α -D-glucuronidase respectively. The *Geobacillus stearothermophilus* T-1 AguA released 4-O-methyl-D-glucuronic acid from synthetic 4-O-methyl-D-glucuronoyl oligosaccharides, and was distinctive in that it possessed a homodimeric organisation that was 240-fold more active than a monomeric mutant counterpart (Shallom *et al.*, 2004). Subtle conformation changes within the substrate-binding region induced by dimer formation were predicted to account for the marked difference in catalytic activity. In the 2-DE separations of cytosolic proteins Agu67A was detected at a size that corresponded with the monomeric form of the enzyme, which may have been due to the reducing conditions used in the analysis. Non-reducing 2-DE in conjunction with enzyme activity assays may assist in determining whether Agu67A exists as a dimer *in vivo*.

The *Paenibacillus* sp strain JDR-2 homologue of the *B. proteoclasticus* Agu67A is a 77.8 kDa enzyme encoded as part of an eight component aldouronate utilization gene cluster that also encodes a cytosolic GH10 endoxylanase, a GH43 β -xylosidase/ α -arabinofuranosidase, and several ABC transporter components and signal transduction regulatory proteins (Chow *et al.*, 2007). The gene cluster operates as a regulon, and expression of these genes as well as a downstream, cell-associated GH10 endoxylanase is induced by growth on methylglucuronoxylan, and repressed by glucose. When expressed in *E. coli* the *Paenibacillus* enzyme hydrolysed the α -1,2-glycosidic bond between the 4-O-methylglucuronic acid residues and the xylan backbone in methylglucuronoxylan (Nong *et al.*, 2005). *Paenibacillus* AguA contains the highly conserved catalytic aspartate and glutamate residues characteristic of GH67 α -D-glucuronidases, which are homologues of Asp355 and Glu383 in the *B. proteoclasticus*. Taken together, this data provides good evidence that the *B. proteoclasticus* Agu67A is likely to remove 4-O-methylglucuronic acid residues from the backbone of assimilated methylglucuronoxyloligosaccharides.

6.3.3 Polysaccharide utilisation locus 5

PUL5 is a 15 component gene cluster that encodes a GH5 endoglucanase/xylanase Cel5A (Bpr_I0299), a GH10 endoxylanase (Bpr_I0304), and the GH43 xylosidase/arabinofuranosidase Xsa43A (Bpr_I0302), all of which are secreted enzymes (Figure 6.5). A second GH43 xylosidase/arabinofuranosidase Xsa43H (Bpr_I0301) is clustered in PUL5 and is localised to the cytosol (Kelly *et al.*, 2010). Both GH43 xylosidase/arabinofuranosidases were identified in the *B. proteoclasticus* cytosol, although the secreted Xsa43A could not be detected at either time point in the *B. proteoclasticus* culture medium. The low relative abundance of Xsa43A in the cytosol might explain why the detection of the secreted form was not possible by 2-DE. An ABC transporter protein SBP belonging to PUL5 was also present in the *B. proteoclasticus* proteome, which together with the identification of three additional PUL5 proteins in the membrane proteome analysis (Dr. Judy Bond, personal communication) made PUL5 the second most prevalent PUL in the *B. proteoclasticus* proteome.

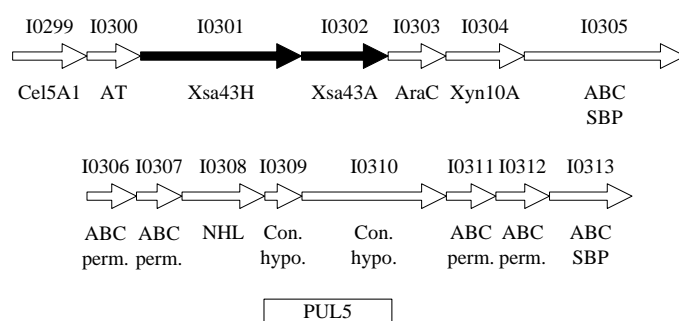


Figure 6.5. Organisation of the *B. proteoclasticus* PUL 5. Arrow size indicates the relative length of each protein. Filled and empty arrows indicate identified and unidentified proteins respectively. AraC, transcriptional regulator AraC family; AT, acetyltransferase GNAT family; Con. hypo, conserved hypothetical protein; NHL, NHL repeat-containing protein; perm., permease; SBP, substrate-binding protein.

6.3.3.1 Xylosidase/arabinofuranosidase Xsa43H

Xsa43H (Bpr_I0301) is encoded by the gene found immediately upstream of the extracellular Xsa43A, and is a 107.9 kDa cytosolic protein that contains an N-terminal family 10 carbohydrate esterase domain in association with a C-terminal GH43 domain. The enzyme was identified by 1-D LC-MS/MS only, where seven distinct peptides were matched to 3.1% of the full-length protein sequence. The C-terminal, GH43 domain of Xsa43H was at least 65% identical to several *Roseburia* and

Clostridium species. A *C. stercorarium* homologue (XylA) was particularly interesting in that at an optimal pH of 7.0 it exhibited xylosidase and arabinofuranosidase activity apparently within a single GH43 domain (Sakka *et al.*, 1993). XylA was incapable of degrading xylan, which was in keeping with the cytosolic location and predicted role in the degradation of arabinofuranosyl xylooligosaccharides, and contrasts with the predicted activity of the *B. proteoclasticus* Xsa43A. If the *B. proteoclasticus* Xsa43H GH43 domain has similar catalytic activity to that of *C. stercorarium* XylA, it is possible that in the cytosol Xsa43H acts synergistically with the PUL3 associated Xyl3A to affect the removal of arabinofuranose substituents from assimilated xylooligosaccharides, prior to subsequent hydrolysis of the xylooligomers to monomeric xylose.

The N-terminal CE10 domain of Xsa43H is one of two CE10 domains present in polysaccharidases detected in the *B. proteoclasticus* cytosol, the other being the endoxylanase and esterase Xyn10D (Bpr_I1083). Examination of the BLASTp homologues of the Xsa43H CE10 domain provides few clues to the specific function of this domain. Nonetheless, no identifiable BLASTp homologue within the NCBI database possessed a similar domain arrangement, which implies that Xsa43H might be a novel enzyme that possesses unique catalytic activity within the rumen microbiome, and a good candidate for further biochemical and biophysical analysis.

6.3.4 Polysaccharide utilisation locus 10

PUL10 is a small cluster of genes that contains the α -glucuronidase Gh115B (Bpr_I1081), the endo-1,4- β -xylanase and esterase Xyn10D (Bpr_I1083), and the pectin methylesterase Pme8A (Bpr_I1084), all of which are predicted cytosolic enzymes (Figure 6.6). In addition, PUL10 contains a single secreted hypothetical protein (Bpr_I1082). The locus does not appear to encode any other proteins, and only the endo-1,4- β -xylanase and esterase Xyn10D was identified in the *B. proteoclasticus* proteome.

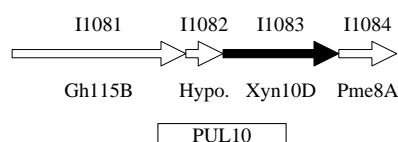


Figure 6.6. Organisation of the *B. proteoclasticus* PUL 10. Arrow size indicates the relative length of each protein. Filled and empty arrows indicate identified and unidentified proteins respectively. Hypo, hypothetical protein.

6.3.4.1 Endo-1,4- β -xylanase and esterase Xyn10D

The endo-1,4- β -xylanase and esterase, Xyn10D (Bpr_I1083) is a 79.7 kDa cytosolic enzyme that was one of eight polysaccharide degrading enzymes identified by 1-D LC-MS/MS only. The relative abundance of the enzyme was not influenced by the growth conditions used in this project. Xyn10D contains an N-terminal GH10 catalytic domain, and a C-terminal CE10 domain, and the full-length protein is 97% identical to the 73 kDa product of the *xynB* gene found in *B. fibrisolvens* H17c(SA), now *B. proteoclasticus* (Lin and Thomson, 1991a, Moon *et al.*, 2008). When the *B. fibrisolvens xynB* gene was expressed in *E. coli* up to 94% of the detected endoxylanase activity was retained in the cytosol, which is in agreement with the predicted subcellular location of the *B. proteoclasticus* Xyn10D. It was concluded that the *B. fibrisolvens* Xyn10D was able to cleave the backbone of xylooligomers close to sites of arabinose substitution. These substitutions can block cleavage of the xylan backbone by other endoxylanases, which makes Xyn10D a key enzyme for the cytosolic degradation of high arabinose content hemicelluloses commonly found in forage material. Lin and Thomson (1991a) found no evidence for the presence of the CE10 domain in *xynB*, which might be explained by the fact that the cloned enzyme is missing a portion of the C-terminus that is present in the predicted *B. proteoclasticus* enzyme (635 vs. 692 amino acids respectively). The homology of *B. proteoclasticus* Xyn10D to other bacterial polysaccharide degrading enzymes is considerably lower than to the *B. fibrisolvens xynB* gene product. As a result, information describing the possible catalytic function of the CE10 domain in *B. proteoclasticus* is not available. The CE10 classification within the CAZy database contains an assortment of catalytic activities, including several that are active upon non-carbohydrate substrates. Bifunctional polysaccharide degrading enzymes can act synergistically upon hemicellulose derived substrates (Levasseur *et al.*, 2005), therefore if the catalytic activity of the *B. proteoclasticus* Xyn10D is similar to that of the *B. fibrisolvens* homologue, Xyn10D may play an important role in the cytosolic degradation of internalised esterified xylooligomers, and is an excellent candidate for future biochemical and biophysical characterisation. Furthermore, no single enzyme in the CAZy database contains catalytic domains homologous those found in the *B. proteoclasticus* Xyn10D, which implies that the GH10/CE10 is a novel catalytic combination.

6.3.5 Polysaccharide utilisation locus 29

In grasses, dimerisation of ferulic acid residues attached to arabinosyl side chains extensively cross-link neighbouring hemicellulose backbone chains (Harris and Hartley, 1976); (Ralph *et al.*, 1994), and arabinoxylan-bound ferulic acids may also form ester-lignin bonds with surrounding lignin (Jacquet *et al.*, 1995, Scalbert *et al.*, 1985). Collectively these covalent cross-links create a formidable structural barrier to degradation of the lignocellulose in the rumen, and are a determining factor in the degradability of high fibre forage material, by restricting the accessibility of polysaccharide degrading enzymes to their substrate. Polysaccharidases that target these linkages act synergistically with xylanases to hydrolyse the hemicellulose component of forage material (Hashimoto and Nakata, 2003), and are likely to be particularly important targets for improving microbial mediated rumen forage digestion. PUL 29 (Figure 6.7) is a four-component gene cluster that encodes a single feruloyl esterase that is predicted to target feruloylated xylooligosaccharides.

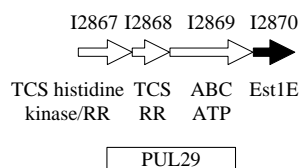


Figure 6.7. Organisation of the *B. proteoclasticus* PUL 29. Arrow size indicates the relative length of each protein. Filled and empty arrows indicate identified and unidentified proteins respectively. ATP, ATPase; RR, response regulator; TCS, Two-component system.

6.3.5.1 Feruloyl esterase Est1E

The feruloyl esterase Est1E (Bpr_I2870) is a 27.8 kDa cytosolic enzyme that was detected at equivalent abundance in the cytosol of *B. proteoclasticus* cells grown in xylan, xylose, and unsupplemented culture medium. Although Est1E is one of four feruloyl esterases encoded by the *B. proteoclasticus* genome (two cytosolic and two secreted), it was the only enzyme of its type identified in this study. Est1E is also the only polysaccharidase clustered in PUL29, which suggests that PUL29 may mediate the degradation and assimilation of feruloyl xylooligosaccharides. The structure of Est1E was recently determined to 1.6 Å resolution in the apo- and ligand-bound forms (Goldstone *et al.*, 2010) and was the first *B. proteoclasticus* polysaccharidase to be structurally characterised. A lid domain within the active site that is novel within the PDB database was discovered, which indicated a dynamic mechanism for substrate-

binding and release. When incubated with esterified synthetic substrates or purified rye-grass hemicellulose, Est1E released ferulic and *p*-coumaric acids. Furthermore, a variety of ester linkages were hydrolysed and phenolic compounds released when Est1E was incubated with birch-wood xylan, suggesting that the enzyme has wide substrate specificity. Est1E is 64% identical to CinI produced by *B. fibrisolvens* E14 which also released ferulic acid from wheat bran extracted ferulated arabinofuranose (Dalrymple *et al.*, 1996).

Information describing the induction patterns and the nature of inducer molecules of bacterial feruloyl esterases is limited. Examination of the Gram-positive soil bacterium *Streptomyces avermitilis* has demonstrated that feruloyl esterase production is induced by the presence of oat-spelt xylan in the culture medium (Garcia *et al.*, 1998). The expression of a second *B. fibrisolvens* E14 esterase gene that is similar to CinI is also induced by the presence of intact hemicelluloses but not hemicellulose derived simple sugars (Dalrymple and Swadling, 1997). In view of the fact that Est1E was present at equivalent abundance in the cytosol of cells grown in all three conditions, oat-spelt xylan or its degradation products do not appear to be an inducer molecule for *B. proteoclasticus* Est1E. Previous reports analysed the expression and synthesis of secreted feruloyl esterases, whereas the *B. proteoclasticus* Est1E is predicted to be a cytosolic enzyme. Est1E is the only polysaccharidase component of PUL29, which also contains an unidentified ABC transporter ATP-binding protein (Bpr_I2869) and two two-component system regulatory proteins (Bpr_I2867 and Bpr_I2868). Additional experiments will be needed to identify the nature of the inducer molecules. Nonetheless, identification of feruloyl esterase Est1E in the *B. proteoclasticus* cytosol, together with the novel structure and wide substrate specificity implies that it is an important contributor to hemicellulose degradation by *B. proteoclasticus*.

6.3.6 ABC transporter mediated uptake of xylooligosaccharides

The tally and differential abundance patterns of the ABC transporter substrate-binding proteins (SBPs) detected in the *B. proteoclasticus* proteome signifies the importance that SBP dependent ATP-driven active transport of oligo- and monosaccharides plays in *B. proteoclasticus* growth. In Gram-positive bacteria, SBPs are often lipoproteins that are tethered to the external surface of the cell wall or fused directly to the membrane transporter (Davidson *et al.*, 2008, van der Heide and Poolman, 2002). Lipoproteins in *B. subtilis* are translocated by the general secretory

(Sec) pathway proteins, lipid modified by the diacylglycerol transferase Lgt, and cleaved by the type II signal peptidase LspA (Leskela *et al.*, 1999). Homologues of each of these proteins are encoded by the *B. proteoclasticus* genome (Kelly *et al.*, 2010).

The proportion of SBPs identified in this study is comparable to that identified in proteomic analyses of other highly fibrolytic Gram-positive bacteria such as *Bacillus licheniformis* and *B. subtilis* (Bunai *et al.*, 2004, Voigt *et al.*, 2006). Furthermore, the abundance of many of these proteins is regulated depending on the growth substrate encountered, which is likely to provide an advantage within the complex rumen microbial ecosystem allowing *B. proteoclasticus* to efficiently utilise the products of its polysaccharide-degrading activity. Within the secreted, cell-associated and cytosolic fractions, 14 sugar specific ABC transporter SBPs were identified, 11 of which were differentially abundant between the xylan and xylose, or xylan and no added carbon growth conditions.

The common feature of the nine SBPs that were more abundant in the xylan growth condition was the presence of a Carbohydrate Uptake Transporter 1 (CUT1) family substrate-binding domain. The CUT1 family comprises a large and diverse collection of SBP-dependent ABC transporter systems that are dominated by those that mediate the uptake of di- and oligosaccharides. The maltose/maltodextrin transporter of *E. coli* represents the paradigm for CUT1 family and ABC transporters in general (Boos and Shuman, 1998). Substrate recognition is mediated primarily by the high-affinity maltose/maltodextrin SBP MalE, which contains two opposing lobes that move towards each other and trap the substrate in the binding pocket. Interaction with the membrane spanning domains triggers ATP-hydrolysis and substrate translocation. The expression of all components of the maltose/maltodextrin transport system of *E. coli* is controlled by the positive transcriptional regulator MalT, which is activated by the presence of maltotriose (Dardonville and Raibaud, 1990). The aldouronate transporter (*LplABC*) of *Paenibacillus* sp strain JDR-2 (Chow *et al.*, 2007) is also a CUT1 family member, and is part of the aldouronate utilization gene cluster that contains the homologue of the *B. proteoclasticus* polysaccharidase Agu67A. The *Paenibacillus* gene cluster is strongly induced by methylglucouronoxylan, only weakly by xylose, and is repressed by glucose (Chow *et al.*, 2007). It appears likely that the *B. proteoclasticus* CUT1 family SBP-dependent ABC transporters are sensitive to

extracellular oligosaccharides in a manner that is similar to the well characterised members of the CUT1 family.

The 2-DE protein abundance pattern of the SBP product of the PUL3 gene Bpr_I0182 implies that this SBP might play a particularly important role in the assimilation of GAX derived oligosaccharides, which supports the hypothesis that PUL3 is an integral component of the fibre degrading enzyme system of *B. proteoclasticus*. The 63.5 kDa ABC transporter substrate-binding protein was the only protein of its type identified in PUL3 (Kelly *et al.*, 2010), and was the most abundant protein detected in both the mid-log and stationary phase harvested xylan-grown culture medium. In both the culture medium and cytosol, the SBP was at least 5-fold more abundant in xylan grown cells relative to those grown on xylose, at both mid-log and stationary phase. This differential abundance pattern was the largest consistent abundance change among the SBPs identified in the *B. proteoclasticus* proteome. Furthermore, the product of Bpr_I0182 was detected with high probability by 1-D LC-MS/MS as the 11th most abundant protein in the cytosol of xylan grown cells. The protein was detected at very low relative abundance in the cytosol of xylose grown cells, and in less biological replicates that necessary to permit reliable statistical analysis of the relative abundance change. Nonetheless, this data supports that obtained by 2-DE. The substrate-binding protein was 63% identical to the recently released sequence of the *B. fibrisolvens* 16/4 ABC transporter system SBP (gi 291518643) and 34% identical to several family 1 extracellular solute binding protein (SPBbac1, Pfam 01547) secreted by *Geobacillus* and *Paenibacillus* species.

6.3.6.1 Polysaccharide utilisation locus 11

Using proteomics in isolation it is not possible to predict with certainty the substrate specificity of the differentially abundant SBPs. Although many show a moderate to high identity with SBPs produced by other Gram-positive bacteria, very few of the homologues have been experimentally characterised. Structural and biochemical examination of the differentially abundant SBPs, as well as the chemical nature of the inducing molecules will be necessary in determining the precise role and importance of these proteins in affecting the assimilation of xylooligosaccharides by *B. proteoclasticus*.

PUL11 is a small gene cluster that contains three sugar specific substrate-binding proteins, ATP binding and permease proteins, a pair of two-component signal transduction proteins, and an EAL-domain containing protein (Figure 6.8) (Kelly *et*

al., 2010). PUL 11 contains no identifiable polysaccharide degrading proteins and no regulatory proteins. The fact that the identified SBP and ATP-binding protein are both significantly more abundant in cells growth on xylose relative to xylan suggests a role for PUL11 in the uptake of xylose monomers from the extracellular environment.

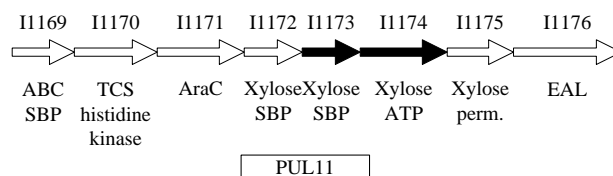


Figure 6.8. Organisation of the *B. proteoclasticus* PUL 11. Arrows indicate the length of each protein. Filled and empty arrows indicate identified and unidentified proteins respectively. AraC, transcriptional regulator AraC family; ATP, ATPase; EAL, EAL domain-containing protein; perm, permease; SBP, substrate-binding protein; TCS, Two-component system.

The xylose ABC transporter SBP (Bpr_I1173) and the xylose ABC transporter ATP-binding protein (Bpr_I1174) were the only components of PUL11 identified. The xylose ABC transporter SBP was identified as being between 3.3-fold and 4.8-fold more abundant in the secreted and cytosolic compartments of xylose grown cells respectively. The ATP-binding protein was detected uniquely in the cytosol of xylose grown cells by 1-D LC/MS, which indicated that when xylose is actively utilised by the bacterium the Bpr_I1174 product is present in the cytosol at low concentration.

In addition to being 80% identical to a putative solute-binding component of ABC transporter synthesised by *Clostridium phytofermentans* ISDg, the *B. proteoclasticus* xylose ABC transporter SBP (Bpr_I1173) was 68% identical to the recently sequenced *B. fibrisolvens* and *Ruminococcus* sp CUT2 family substrate-binding proteins (GenBank CBK73023 and CBL20170.1 respectively), and the *B. proteoclasticus* ATP-binding protein (Bpr_I1174) is 63% identical to ATP-binding proteins encoded by the same organisms. Taken together this data implies that the *B. proteoclasticus* PUL11 SBP and ATP-binding proteins form part of a CUT2 family ABC transporter system that is specific for xylose assimilation. The CUT2 family is a large collection of microbial ABC transporter systems that almost exclusively assimilate monosaccharides (Schneider, 2001). The ribose and galactose transporters of *E. coli* and *S. typhimurium* respectively are the most well characterised CUT2 family members, and the *E. coli* and *Thermoanaerobacter ethanolicus* high-affinity transport

systems that are essential for xylose uptake also fall within the CUT2 family. The *xylF* genes encoding the D-xylose binding proteins found in the xylose uptake systems of the *E. coli* and *T. ethanolicus* systems were induced (up to 70-fold in *E. coli* K-12) by the presence of D-xylose in the culture medium (Erbeznik *et al.*, 1998, Song and Park, 1997).

Bpr_I0117 is a component of PUL2 and was one of only two PUL associated SBPs encoded by the *B. proteoclasticus* genome that did not contain a Type-II secretory signal peptide and lipobox motif, suggesting that it is not tethered to the external cell surface *in vivo*. PUL2 is a small gene cluster that includes an uncharacterised xylosidase enzyme. Together this data suggests that PUL2 may be functionally distinct from the majority of other 35 PULs encoded by the *B. proteoclasticus* genome.

6.4 Non-PUL polysaccharidases and their role in hemicellulose degradation

At least two polysaccharide degrading enzymes synthesised by *B. proteoclasticus* that are not part of a detectable PUL are either homologous to well characterised bacterial enzymes, or contain GH domains, that provide evidence that they are likely to be important contributors to hemicellulose degradation in the rumen.

6.4.1 α -L-Arabinofuranosidase Arf51A

The α -L-arabinofuranosidase, Arf51A (Bpr_I0329) is a 57 kDa cytosolic enzyme that contains a C-terminal GH51 catalytic domain and is likely to play a key role in the removal of arabinofuranose side chains from internalised substituted xylooligosaccharides that have not been hydrolysed by extracellular arabinosidases such as Mxy10-43A and Xsa43A. The enzyme is at least 60% identical to α -L-arabinofuranosidases produced by several *Firmicute* species including ArfB of *C. stercorarium* and ArfA of *C. cellulovorans* (Kosugi *et al.*, 2002, Schwarz *et al.*, 1995, Zverlov *et al.*, 1998), and is 57% identical to the recently characterised Abf2 of *B. subtilis* (Inacio *et al.*, 2008). All three Arf51A homologues preferentially hydrolysed arabinose containing polysaccharides such as wheat arabinoxylan over other hemicellulosic substrates. The catalytic activity of the *C. stercorarium* ArfB was distinctive in that it was highly specific for the hydrolysis of non-reducing α -L-arabinosyl residues and was restricted to the furanose form of arabinose (as opposed to the pyranose form) such as that present in arabinoxylan and GAX (Schwarz *et al.*, 1995). Arabinofuranose liberation from intact hemicellulose could be achieved prior to

hydrolysis of the main xylan backbone, and produced linear (1→4)-β-D-xylooligosaccharides that could then be degraded further by the action of endoxylanases. Moreover, the presence of arabinofuranose side-chains in hemicellulosic substrates blocked the activity of *C. stercorarium* endoxylanases, and complete hemicellulose degradation could only be achieved by the subsequent addition of ArfB (Schwarz *et al.*, 1995).

The *B. subtilis* Abf2 displayed a less fastidious substrate specificity than that of the *C. stercorarium* arabinofuranosidase, but assay of wheat hemicellulose hydrolysis products indicated Abf2 preferentially attacked (1→2)-α- and (1→3)-α-linked L-arabinofuranosyl units (Inacio *et al.*, 2008). The expression of Abf2 was responsive to the presence of both arabinose and arabinan in a temporal manner with a peak during early post-exponential phase (Raposo *et al.*, 2004).

B. subtilis expresses a second cytosolic GH51 α-L-arabinofuranosidase (AbfA) that has similar physicochemical properties to the *B. subtilis* Abf2, but is only 23% identical and has a catalytic preference for (1→5)-linkages in linear α-(1→5)-L-arabinan and α-(1→5)-linked arabinoxylooligomers over the (1→2)-α- or (1→3)-α-linkages in arabinoxylan (Inacio *et al.*, 2008). The catalytic disparity between the two enzymes suggests *B. subtilis* is equipped to accomplish optimal utilisation of arabinose containing polysaccharides such as arabinooligomers and glucuronoarabinoxylan. BLASTp searches of the *B. proteoclasticus* genome verified that the *B. subtilis* AbfA was 60% identical to the only other cytosolic GH51 arabinofuranosidase Arf51B, which is located at Bpr_I0017. Arf51B has very similar pI, size, and solubility characteristics to Arf51A, but was not identified in this project. Nonetheless, Arf51B was identified with high statistical probability in a 1-DE LC-MS/MS analysis of the *B. proteoclasticus* membrane proteome (Dr. Judy Bond, personal communication). It appears therefore that *B. proteoclasticus* may utilise a similar cytosolic arabinofuranosidase enzyme system to that of *B. subtilis*, employing two catalytically divergent α-L-arabinofuranosidases to hydrolyse the variety of α-linkages found in arabinose containing hemicelluloses. It is likely that *B. proteoclasticus* makes a significant contribution to the conversion of α-L-arabinosyl containing hemicelluloses to utilisable energy in the rumen.

6.4.2 Xylosidase/arabinofuranosidase Xsa43E

The 35.3 kDa product of the *Xsa43E* gene (Bpr_I2319) was identified in a single spot in the 2-DE separation of protein extracted from mid-log phase harvested cells

grown in the presence of xylan. The enzyme could not be detected in any other mid-log phase harvested growth condition, and it could not be detected in any growth condition during stationary phase. Xsa43E contained a single GH43 catalytic domain that spanned 92% of the full length protein sequence (residues 13-300), and was 65% and 58% identical to catalytic domains of GH43 family enzymes produced by *Clostridium thermocellum* ATCC 27405 (YP_001038591) and *Clostridium acetobutylicum* ATCC 824 (NP_149278) respectively. Both homologues, as well as several more distantly related enzymes, were predicted to be secreted proteins that also contained CBM6 domains within their C-terminal regions. No domains corresponding to these regions were detected in the *B. proteoclasticus* Xsa43E protein sequence. Furthermore, none of the close BLASTp homologues of Xsa43E have been experimentally examined. It is therefore not possible to determine the specific activity or substrates of Xsa43E. The GH43 domain of Xsa43E is no greater than 33% identical to the GH43 domains in the identified *B. proteoclasticus* Xsa43A, Xsa43J, or Xsa43H enzymes. This observation implies that the substrate specificity might be different to any of these enzymes, and as a whole this set of β -xylosidases may be capable of hydrolysing glycosidic linkages at varied positions along the xylan backbone and between substituent groups. Biochemical examination of Xsa43E will be a necessary and important step in advancing the understanding of hemicellulose degradation by *B. proteoclasticus*. Xsa43E has been cloned and expressed, and is currently undergoing structural analysis (Dr. W. J. Kelly, personal communication).

6.5 Additional enzymes identified in the *B. proteoclasticus* proteome

The presence of a variety of other hemicellulosic and non-hemicellulosic enzymes in the *B. proteoclasticus* proteome when cells are grown in xylan containing culture medium implies that they also make an important contribution to lignocellulose degradation. Xyloglucan is a hemicellulose that is present at low levels in the primary cell walls of plants used as forage crops, but is prevalent in the primary cell wall of many dicotyledonous plants. The xyloglucan backbone consists of a repeating pattern of four (1 \rightarrow 4)- β -D-glucose monomers that are substituted with (1 \rightarrow 6)- α -xylopyranose residues. In Type-I cell walls the xylose residues are often substituted further with (1 \rightarrow 2)- α -D-galactose and (1 \rightarrow 2)- α -L-fucose monomers. Mannan based xylans, including glucomannan (a heteropolymer of D-glucose and D-mannose) and galactoglucomannan (a heteropolymer of D-galactose, D-glucose, and D-mannose)

comprise up to 10% of Type-I cell walls. Mixed linkage glucans can comprise up to 30% of the primary cell wall of Type-II plants depending on the developmental stage of the plant. Examples of several of the types of enzymes necessary for the degradation of these polysaccharides are present in the *B. proteoclasticus* proteome, including the β -mannosidase Man2A (Bpr_III237), and two α -galactosidases, Aga27A (Bpr_I0205) and Aga36C (Bpr_III065). The predicted secreted β -glucosidase Bgl3A (Bpr_I0693) and the two cytosolic β -glucosidases Bgl3B (Bpr_I0847) and Bgl3C (Bpr_I0138), were all detected in the cytosol. Bgl3C is homologous to a *B. fibrisolvens* H17c(SA), now *B. proteoclasticus* (Moon *et al.*, 2008), β -glucosidase that degrades celooligosaccharides to glucose (Lin *et al.*, 1990). The *B. proteoclasticus* Bgl3C may therefore perform the same function upon celooligosaccharides transported across the bacterial cell wall. In at least one case, β -glucosidases have been shown to be important for the hydrolysis of rice bran derived hemicellulose (Harada *et al.*, 2005). The collection of *B. proteoclasticus* β -glucosidases may therefore act in a similar manner upon the main chains of both xyloglucans and mixed linkage glucans. Several pectin degrading enzymes, including the α -amylase Amy13G and the α -L-rhamnosidase Rha78A were detected in the *B. proteoclasticus* cytosol, which implies that pectin degradation products may be utilised by *B. proteoclasticus*, consistent with the presence of the pectate lyase and pectin methylesterase enzymes in the culture medium. None of these cytosolic enzymes were differentially abundant in *B. proteoclasticus* cells grown in the presence of xylan, xylose, or unsupplemented culture medium, suggesting that they are not transcriptionally or post-transcriptionally responsive to these growth substrates. The presence of this diverse set of enzymes implies that *B. proteoclasticus* is capable of degrading the majority of hemicellulosic and non-hemicellulosic polymers encountered in the rumen, and together with the previously determined substrate utilisation profiles (Attwood *et al.*, 1996) implies that a variety of catabolic pathways might be active in *B. proteoclasticus* cells grown *in vitro*.

6.6 A proteomic view of *B. proteoclasticus* central metabolism

6.6.1 Central metabolism

Within the *B. proteoclasticus* proteome, components of the catabolic pathways necessary for the utilisation of D-arabinose, L-fucose, D-fructose, D-galactose, D-glucose, L-rhamnose, and D-xylose were all identified, as were pathways for the catabolism of galacturonates and glucuronates (Figure 6.9). The first two reactions in

the conversion of D-xylose to xylulose 5-phosphate are catalysed by xylose isomerase and xylulokinase respectively. The *B. proteoclasticus* genome encodes at least three proteins that may possess xylose isomerase activity, and following the reconstruction of the xylose and xyloside utilisation pathways in 24 *Firmicute* genomes (Gu *et al.*, 2010) it is predicted that the product of the PUL3 clustered Bpr_I0185 is the most likely candidate (Dr. W. J. Kelly, personal communication). The product of Bpr_I0185 was identified in the cytosol of xylan and xylose grown *B. proteoclasticus* cells at equivalent abundance at mid-log and stationary phase. The *B. proteoclasticus* genome encodes a single xylulokinase, XylB (Bpr_I0173) that is also a PUL3 component, which is consistent with the fact in *Firmicutes* xylulokinase encoding genes are frequently clustered with other enzymes involved in xyloside utilisation (Gu *et al.*, 2010). XylB was 4.2-fold more abundant in the cytosol of xylan grown cells during stationary phase, which is consistent with the increased abundance of several other PUL3 proteins in the *B. proteoclasticus* proteome. In *C. acetobutylicum* the specific activity of xylulokinase may be a rate limiting step in xylose utilisation (Gu *et al.*). If this is the case for *B. proteoclasticus*, then co-regulation of the *xylB* gene and other cytosolic xylooligosaccharide utilisation enzymes such as the β -xylosidase Xyl3A and α -D-glucuronidase Agu67A may maximise growth and fermentation when xylan is present in the external environment. Taken together, this data supports the hypothesis that PUL3 is an important for xylan degradation and utilisation by *B. proteoclasticus*.

Galacturonate and glucuronate catabolism proceeds via the common intermediate 2-keto-3-deoxygluconate (KDG), which is ultimately converted to 2-keto-3-deoxygluconate phosphate (KDGP) by the activity of 2-dehydro-3-deoxygluconokinase. KDGP is then converted to pyruvate and glyceraldehyde-3-phosphate (GAP) by 2-keto-3-deoxygluconate 6-phosphate aldolase. The metabolic footprint of *B. proteoclasticus* grown on xylan, pectin and glucose (Villas-Boas *et al.*, 2006) supports the possibility that galacturonates and glucuronates released by xylan and pectin degradation are rapidly consumed by *B. proteoclasticus*, and are a major source of pyruvate for central metabolism. This hypothesis is consistent with the presence of at least five pectin degrading enzymes in the *B. proteoclasticus* proteome. The proteomic data obtained in this project supports this argument, and is strengthened by the fact that all except one of the genes involved in glucuronate metabolism are clustered in PULs and are likely to be coordinated with other gene products involved in polysaccharide breakdown (Kelly *et al.*, 2010).

A feature of the genome sequence is the apparent absence of a gene coding for enolase (phosphopyruvate hydratase, EC4.2.1.11), which is normally a well conserved enzyme that converts 2-phosphoglycerate to phosphoenolpyruvate in the penultimate step of the Embden-Meyerhof pathway (Kelly *et al.*, 2010). At least one example of all other types of enzymes required for the glycolytic conversion of glucose to pyruvate was identified in the *B. proteoclasticus* proteome. Potential alternatives to the Embden-Meyerhof pathway are the Entner-Doudoroff pathway and the methylglyoxal shunt. The Entner-Doudoroff pathway is common in Gram-negative bacteria, but very few Gram-positive bacteria utilise the pathway, and the *B. proteoclasticus* genome sequence suggests that the Entner-Doudoroff pathway is not utilised (Kelly *et al.*, 2010). The methylglyoxal shunt has been proposed as a strategy allowing rumen bacteria to dispose of excess carbohydrate and decrease ATP production (Russell, 1998). The conversion of dihydroxyacetone phosphate (DHAP) to pyruvate proceeds via methylglyoxal and D-lactate (Cooper, 1984), and enzymes catalysing each step in the pathway are encoded in the *B. proteoclasticus* genome. None of the components of the methylglyoxal shunt were identified in this proteomic analysis which makes it difficult to speculate how *B. proteoclasticus* achieves the conversion of monosaccharides such as D-glucose and L-rhamnose to phosphoenolpyruvate.

Figure 6.9. A proteomic view of *B. proteoclasticus* central metabolism. Identified proteins, and proteins with increased and decreased abundance in xylan grown cells relative to xylose are shown in green, red, and blue respectively. Figure adapted from Kelly *et al.* (2010).

6.6.2 Butyrate synthesis by *B. proteoclasticus*

The main end-products of the fermentative metabolism of *B. proteoclasticus* are butyrate and formate, as well as low amounts of acetate and propionate (Attwood *et al.*, 1996). Butyrate synthesis is achieved by the conversion of acetyl-CoA to butyryl-CoA via four sequential reactions, which are catalysed by thiolase, β -hydroxybutyryl-CoA dehydrogenase (BHBD), crotonase, and butyryl-CoA dehydrogenase (BCD) respectively (Miller and Jenesel, 1979). *Butyrivibrio* species that use butyrate kinase to catalyse the conversion of butyryl-CoA to butyrate in the final stage of butyrate synthesis pathway (Type I producers) produce more butyrate than species using an alternate pathway (Type II producers), and are important contributors to the ruminant energy supply (Shane *et al.*, 1969). For this reason, Type I butyrate producers are considered potential targets for the positive manipulation of butyrate production in the rumen of forage fed animals (Asanuma *et al.*, 2003, Asanuma *et al.*, 2005). Phylogenetically, *B. proteoclasticus* clusters with a group of closely related *Butyrivibrio* strains that are without exception positive for the butyrate kinase gene (Paillard *et al.*, 2007a). All components of the butyrate synthesis pathway, including the butyrate kinase (Bpr_I2323), were identified in the *B. proteoclasticus* proteome regardless of growth substrate. The substrates tested did not cause detectable changes in the abundance of any members of the pathway. Consequently, the butyrate pathway of *B. proteoclasticus* appears to be constitutively active in cells that are growing in variable external environments, and may be a convenient potential target for the positive manipulation of short-chain volatile fatty acid synthesis and supply to the ruminant. These results support the likelihood that *B. proteoclasticus* makes an important contribution to ruminant energy supply.

6.7 Features of the complete identified protein dataset

6.7.1 The *B. proteoclasticus* acidic proteome signature

The predicted *B. proteoclasticus* proteome ranges from pI 3.2 to 13.4, and 3.4 kDa to 656 kDa (Kelly *et al.*, 2010). The theoretical 2-DE map of the *B. proteoclasticus* proteome showed a bimodal pI distribution and a moderate acidic skew, with 73% of the proteins predicted to have a pI value below 7. The 2-DE examination of the secreted and cytosolic proteomes supported the theoretical predictions. In particular, the majority of *B. proteoclasticus* secreted proteins were strongly acidic. Knight *et al.* (2004) revealed a correlation between microbial protein pI distribution and ecological niche. Strongly acidic pI distributions have been observed in the secreted and whole

cell 2-DE patterns of gastrointestinal (Lee *et al.*, 2008, Sanchez *et al.*, 2008) and non-polysaccharide-degrading rumen bacteria (Lee *et al.*, 2006). The *B. proteoclasticus* *pI* distribution raises the possibility that the gastrointestinal environment may exert a selective pressure for acidic secreted proteins. An acidic *pI* is likely to ensure efficient protein function even during periods of rumen acidosis, which are especially common when fermentation rates are elevated (Ash and Dobson, 1963).

6.7.2 The importance of a multiplex approach for improved proteome coverage

2-DE analysis of the *B. proteoclasticus* cytosolic proteome produced a protein dataset containing 223 non-redundant entries, which included 48 proteins involved in polysaccharide degradation and assimilation. Augmenting the 2-DE analysis with a gel-free 1-D LC-MS/MS examination improved the depth of coverage by more than 75%, and added significantly to the total number of identified polysaccharidases and proteins involved in polysaccharide assimilation. Notably, five of the eight polysaccharidases identified only by 1-D LC-MS/MS were predicted to be involved in hemicellulose degradation. The 1-D LC-MS/MS analysis also made an important contribution in terms of the number of ABC transporter associated proteins mediating soluble carbohydrate assimilation, and non-GH proteins involved in carbohydrate metabolism, which helped elucidate the number of potentially active catabolic pathways. Overall, the 1-D LC-MS/MS of the cytosolic proteome made an important contribution to the examination of hemicellulose degradation and utilisation by *B. proteoclasticus*.

The low overlap between the 2-DE and 1-D LS-MS/MS cytosolic protein datasets is a function of the sample preparation requirements (for each method of separation), aspects of the separation technologies and to a lesser degree the MS instrumentation. Various types of proteins are often difficult to detect by 2-DE. For instance, low abundance proteins, or proteins that by chance may resolve to a 2-DE position similar to more high abundance proteins, and may therefore be difficult to identify by MALDI TOF. This is likely to be the major factor contributing to the low number of proteins identified by 2-DE/MALDI-TOF, and supports the use of narrow range IPG strips. Hydrophobic proteins or proteins that may have low solubility in 2-DE buffer are also difficult to detect using a 2-DE/MALDI-TOF approach. Proteins that digest into a small number of large peptides may be difficult to elute from a 2-DE gel plug. Furthermore, because most of the proteins separated 2-DE were identified by MALDI-TOF (single MS in contrast to MS/MS), proteins that digest into a small number of

large peptides, or a large number of small peptides may also be difficult to identify. Using LC-MS/MS as the means of protein identification circumvents many of these problems, including the possibility of high and low abundance proteins co-resolving, because protein digestion occurs prior to peptide separation.

The acidic skew observed in the 2-DE separation of *B. proteoclasticus* cytosol was not predicted by the theoretical 2-DE analysis. Nonetheless the 2-DE spot patterns were supported by the plot of the theoretical 2-DE distribution of the cytosolic proteins identified by 1D LC-MS/MS. The predicted *pI* and size of the proteins detected by 1-D LC-MS/MS were consistent with the gel positions. The large majority of all proteins identified by 1D LC-MS/MS also had a *pI* value less than 6.5, and the majority of those that had a *pI* greater than 6.5 were small proteins which can sometimes be difficult to detect by 2-DE because they are close to the dye front. The CAI (Sharp and Li, 1987) is a measure of the level of synonymous codon usage bias, which in bacterial genomes is positively correlated with relative protein abundance (Futcher *et al.*, 1999, Goetz and Fuglsang, 2005, Ishihama *et al.*, 2008, Jansen *et al.*, 2003, Kanaya *et al.*, 1999). Examination of protein *pI* as a function of CAI predicted that the majority of the cytosolic proteins that had a theoretical *pI* value greater than 7 were present in the *B. proteoclasticus* cytosolic proteome at low abundance and may have been below the detection threshold of CBB-G250. It appears likely therefore that the acidic *pI* pattern detected in the project is a reasonable representation of the *B. proteoclasticus* proteome when cells are grown in the conditions used in these experiments.

It is difficult to explain the low agreement between the sets of mid-log phase harvested cytosolic proteins that were detected as being differentially abundant by 2-DE and 1D LC/MSMS. Fifty percent of the proteins detected as being differentially abundant by 2-DE were present in the 1D LC-MS/MS dataset but were not determined to be differentially abundant at the 5% FDR. Conversely, only 20% of the 1D LC-MS/MS differentially abundant proteins were also present in the unchanging 2-DE dataset. It should be noted that only the β -xylosidase Xyl3A was present in both 2-DE and 1D LC-MS/MS datasets but was determined to be differentially abundant by 1D LC-MS/MS only.

6.8 Future directions

Examination of the catalytic properties of the polysaccharidases and transporter proteins identified in this project should be an important component of subsequent analyses aimed at understanding the hemicellulose degrading enzyme system of *B. proteoclasticus*. These polysaccharidases include the set of secreted enzymes identified in cells grown in xylan supplemented culture medium. The presence of 36 PUL within the *B. proteoclasticus* genome together with the proportion of PUL proteins present in the identified proteome signifies their importance to hemicellulose degradation and assimilation by *B. proteoclasticus*. The cytosolic PUL associated polysaccharidases that target the main chain and common substituent groups of xylooligosaccharides, which include each of the enzymes discussed above, will also be important targets for future analysis. This work has already begun with the characterisation of the PUL29 associated ferulic acid esterase Est1E, which has uncovered novel structural biology (Goldstone *et al.*, 2010), and the cloning and characterisation of several xylosidases and esterases that target hemicelluloses (Dr. W. J. Kelly, personal communication).

The model of glucuronoarabinoxylan degradation by *B. proteoclasticus* presented is in the context of the growth substrates used. It is likely that growth in the presence of xylans of different origins, such as wheat-bran, birchwood, or larchwood xylan, will stimulate the production of different sets of polysaccharidases that include enzymes not identified so far. Different substrates are also likely to elicit changes in the relative abundance of some of these polysaccharidases. These differences are evident when comparing the proteomic and transcriptomic data (Kong, 2007) from *B. proteoclasticus* grown on oat-spelt or wheat-bran xylan. Growth on wheat-bran xylan elicited a 15-fold up-regulation of a secreted feruloyl esterase transcript, the protein product of which was unidentified in the proteomic data. In contrast, oat-spelt xylan contains almost no alkali extractable ferulic acid, and at least 10-fold less ferulic acid than wheat-bran xylan (Faulds *et al.*, 1997, Garcia *et al.*, 1998). It will therefore be important to investigate the proteomic profiles of hemicellulose degrading enzymes after cultivation of *B. proteoclasticus* cells on xylans of different origins with different chemical and structural properties.

Several factors can contribute to false positive protein abundance changes in 2DE and label free proteomics analyses, and as a consequence the differential protein abundance changes detected in this project must be validated by alternative methods.

Validation of proteomic data is often performed using 1-DE or 2-DE Western Blotting. There are currently few commercially available antibodies with which this goal could be achieved, and antibody production is time consuming and expensive. An alternative is to use a chemical labelling-based LC-MS/MS approach such as iTRAQ (isobaric tags for relative and absolute quantification) (Ross *et al.*, 2004). Unlike previous labelled protein approaches for quantitative proteomics, iTRAQ quantitation is performed at the peptide level, which means that multiple peptides derived from the same protein may be detected, therefore giving multiple quantitation measurements per protein and increasing the confidence of protein identification and quantitation. LC-MS/MS coupled with iTRAQ has recently been used to analyse the secretomes of several lignocellulose degrading microbes, where relative abundance changes in several glycoside hydrolases including endoxylanases, α -L-arabinofuranosidases, and acetyl xylan esterases were detected between cells grown in differing carbohydrate containing media (Adav *et al.*, 2010a, Adav *et al.*, 2010b, Evans *et al.*, 2007).

The likely importance of PUL associated proteins suggests that future examination and characterisation of PUL activity, sugar and linkage specificity, and regulation will be necessary to determine the contributions made by each PUL. Cultivation and proteomic examination of *B. proteoclasticus* cells on a wider variety of hemicellulosic substrates than was used in this study will be necessary to achieve this goal. An interesting feature of the PUL system of *B. thetaiotaomicron* is the coordinated regulation of PUL expression and capsular polysaccharide biosynthesis in cells degrading and utilising gut O-mucins (Martens *et al.*, 2009b). It was proposed that the coordinated regulation optimises the miscibility of target glycans and other nutrients within the capsular layer, and ensures that the chemical nature of capsular polysaccharides does not negatively influence the interaction between cell-surface bound SBPs and their target sugars. A feature of *B. proteoclasticus* cells grown in liquid media is the production of large amounts of exopolysaccharide, which is also visible on the surface of cells grown on plant material (Kelly *et al.*, 2010). *B. proteoclasticus* has 363 genes that encode proteins predicted to be involved in exopolysaccharide synthesis, and these genes are clustered into 13 exopolysaccharide loci. Although there was no evidence for the differential abundance of exopolysaccharide loci proteins during this analysis, it will be important to analyse the proteomic profiles of exopolysaccharide proteins in cells grown on a range of substrates. It is possible that growth on hemicellulosic substrates not used in this

project may elicit protein abundance changes that shed light on the exopolysaccharide production and regulation.

The rumen is a vast source of untapped fibrolytic potential, and potentially novel fibrolytic enzymes that are likely to have uses outside the rumen. Bacterial polysaccharide degrading enzymes including arabinosidases and feruloyl esterases, examples of which are present in the *B. proteoclasticus* proteome, have industrial and biotechnological applications (Topakas *et al.*, 2007). The most demanding and potentially rewarding application is the development of efficient processes for the solubilisation of lignocellulose material to serve as a renewable energy and carbon source (Mielenz, 2001). Bioethanol is currently the most commonly used renewable fuel and is produced predominantly from corn grain starch, but further production increases will have to be met by other sources because of corn grain supply limitations. These factors call for the exploitation of lignocellulose feedstock, such as agricultural residues and dedicated crops. Bioethanol production begins with biomass thermochemical pre-treatment that reduces particle size, disrupts the cell wall, and improves enzyme access to the cell wall polysaccharides. The exposed polysaccharides are enzymatically or chemically hydrolysed to release soluble sugars, which can be fermented by bacteria, yeast, or filamentous fungi. Since the rate-limiting steps in the conversion of lignocellulose to ethanol are hemicellulose solubilisation and cellulose hydrolysis (Demain *et al.*, 2005), cellulases and hemicellulases are vital components of the production process. While cellulases have been the targets of successful genetic engineering efforts to improve performance and decrease the amount of enzyme needed to degrade lignocellulosic biomass (Bower, 2005, Day, 2003, Lange *et al.*, 2002, Teter *et al.*, 2004, Wu *et al.*, 2003), hemicellulases and hemicellulose engineering for bioethanol production have to date received little attention. The set of hemicellulose degrading enzymes identified in this project, and in particular those that contain multiple catalytic domains such as the endo-1,4- β -xylanase and xylosidase Mxy10-43A, are excellent candidates for assessment of their utility in the process of biofuel production.

6.9 Conclusion

The research presented in this thesis is biologically significant in several ways. Current understanding of the metabolic processes that underpin the microbial mediated conversion of polysaccharides to utilisable energy in the rumen, especially in regard to

the suite of enzymes that are synthesised in response to conditions commonly encountered in the rumen of pasture grazing ruminants, is currently limited. The results of this project provide significant insight into a xylan degrading enzyme system that has evolved to enable *B. proteoclasticus* to efficiently degrade and utilise glucuronoarabinoxylan, and extend our understanding of the types of enzymes that are likely to play important roles in hemicellulose degradation in the rumen. The close genetic relationship between *B. proteoclasticus* and several members of the *Butyrivibrio/Pseudobutyrvibrio* genus suggests that other prominent hemicellulose degrading members of this genus may utilise similar sets of enzymes. Examination of the protein components of the fibrolytic system of hemicellulose degrading rumen microbes provides a clearer understanding of the metabolic processes involved in plant cell-wall degradation by the fibrolytic rumen microbiota. In addition, this project has demonstrated that the relative abundance of several secreted and cytosolic polysaccharide degrading enzymes, as well as substrate-binding proteins that are likely to mediate the assimilation of soluble oligosaccharides, is modulated in response to the growth conditions encountered by the bacterium. It is therefore plausible to expect that *B. proteoclasticus* possesses the sensory mechanisms that enable cells to monitor the polysaccharide status of the external environment, and respond to that status in a manner that is likely to maximise the utilisation of available growth substrate.

The use of exogenous fibrolytic enzymes as feed additives, either by forage pretreatment, or fed directly to ruminants, is an emerging technology that holds promise as a means of enhancing forage utilisation and improving ruminant productivity (Beauchemin *et al.*, 2003, Yang and Xie, 2010). Application of fibrolytic enzymes to rumen forage prior to consumption has resulted in increased voluntary intake, milk production, and average daily weight gain as a consequence of increased forage digestibility (Beauchemin *et al.*, 1995, Cruywagen and van Zyl, 2008, Kung *et al.*, 2000, Pinos-Rodriguez *et al.*, 2002, Schingoethe *et al.*, 1999, Titi and Lubbadah, 2004, Yang *et al.*, 2000). Direct-fed enzyme cocktails also increase forage degradation *in vivo*, and stimulate the growth of cellulolytic bacteria in forage-fed sheep (Giraldo *et al.*, 2008). Nonetheless, results are inconsistent and not always positive (Peters *et al.*, 2010, ZoBell *et al.*, 2000). Commercially prepared fibrolytic enzymes for ruminant applications are cocktails of crude enzyme extracts that usually contain specified levels of several defined enzyme activities such as xylanase or cellulase, and are assessed primarily on their capacity to degrade plant cell walls *in vitro*

(Colombatto *et al.*, 2003, Yang and Xie, 2010). Although the majority of enzyme cocktails contain a combination of xylanase, cellulase, and β -glucanase activities (Colombatto *et al.*, 2003, Yang and Xie, 2010), little attention has been paid to specific formulations or the suitability of these formulations for their intended purposes. In particular, it has been suggested that side-chain hydrolyzing enzymes such as ferulic acid esterases, acetyl-xylan esterases and arabinofuranosidases should be incorporated (Yang and Xie, 2010). *B. proteoclasticus* enzymes could be considered good candidates for product development. The results presented in this thesis provide a clearer understanding of the classes of enzymes and metabolic processes that are utilized by prominent hemicellulose degrading microbes, and may assist in achieving the more rational and targeted design of multi-enzyme products aimed at enhancing the cell wall degradation of rumen forages.

This project is the first global proteomic analysis of the fibre degrading enzyme system of a polysaccharolytic rumen bacterium. Prior to commencing this project the feasibility of applying proteomic technologies to an objective of this nature was unknown. The results demonstrate that proteomics is a valuable tool for the examination of polysaccharidases synthesised by a prominent fibre degrading rumen bacterium. Furthermore, this project has demonstrated that combining gel-based and gel-free approaches significantly improves the depth of proteome coverage over a single technique. The results of this project will be a valuable benchmark for analyses of this type in the future.

The proteomic data and the proposed model of hemicellulose degradation and assimilation elucidated in this project, taken together with information derived from the *B. proteoclasticus* genome sequence project (Kelly *et al.*, 2010), will help to provide a platform for the development of new approaches and opportunities to improve forage degradation, manipulate rumen function, and maximise the productivity of forage fed ruminants with a view to improving the conversion of plant biomass into milk and meat for human consumption.

References

- Abbott, D. W., Ficko-Blean, E., van Bueren, A. L., Rogowski, A., Cartmell, A., Coutinho, P. M., Henrissat, B., Gilbert, H. J. and Boraston, A. B. (2009). Analysis of the structural and functional diversity of plant cell wall specific Family 6 carbohydrate binding modules. *Biochemistry* **48** (43): 10395-10404.
- Adav, S. S., Li, A. A., Manavalan, A., Punt, P. and Sze, S. K. (2010a). Quantitative iTRAQ secretome analysis of *Aspergillus niger* reveals novel hydrolytic enzymes. *Journal of Proteome Research* **9** (8): 3932-3940.
- Adav, S. S., Ng, C. S., Arulmani, M. and Sze, S. K. (2010b). Quantitative iTRAQ secretome analysis of cellulolytic *Thermobifida fusca*. *Journal of Proteome Research* **9** (6): 3016-3024.
- Adelsberger, H., Hertel, C., Glawischnig, E., Zverlov, V. V. and Schwarz, W. H. (2004). Enzyme system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the *in vivo* system from recombinant enzymes. *Microbiology-Sgm* **150**: 2257-2266.
- Ahmad, N., Zhang, J., Brown, P. J., James, D., Birch, J. R., Racher, A. J. and Smales, C. M. (2006). On the statistical analysis of the GS-NS0 cell proteome: Imputation, clustering and variability testing. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1764** (7): 1179-1187.
- Aitchison, J. (1982). The statistical-analysis of compositional data. *Journal of the Royal Statistical Society, Series B-Methodological* **44** (2): 139-177.
- Akin, D. E. (1980). Evaluation by electron microscopy and anaerobic culture of types of rumen bacteria associated with digestion of forage cell walls. *Applied and Environmental Microbiology* **39** (1): 242-252.
- Akin, D. E., Amesgottfred, N., Hartley, R. D., Fulcher, R. G. and Rigsby, L. L. (1990a). Microspectrophotometry of phenolic compounds in bermudagrass cell-walls in relation to rumen microbial digestion. *Crop Science* **30** (2): 396-401.
- Akin, D. E., Hartley, R. D., Morrison, W. H. and Himmelsbach, D. S. (1990b). Diazonium compounds localize grass cell-wall phenolics - relation to wall digestibility. *Crop Science* **30** (5): 985-989.
- Akin, D. E., Ljungdahl, L. G., Wilson, J. R. and Harris, P. J. (1990c). *Microbial and plant opportunities to improve lignocellulose utilization by ruminants*. New York, Elsevier Science Publishing Co.
- Albrecht, D., Kniemeyer, O., Brakhage, A. A. and Guthke, R. (2010). Missing values in gel-based proteomics. *Proteomics* **10** (6): 1202-1211.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215** (3): 403-410.
- Ames, N. P., Hartley, R. D., Akin, D. E., Wilson, J. R., Harbers, L. M. and Reffner, J. A. (1992). Distribution of aromatic compounds in coastal bermudagrass cell-walls using ultraviolet absorption scanning microspectrophotometry. *Food Structure* **11** (1): 25-32.
- Anderson, K. L. (1995). Biochemical analysis of starch degradation by *Ruminobacter amylophilus*-70. *Applied and Environmental Microbiology* **61** (4): 1488-1491.
- Anderson, N. L., Esquerblasco, R., Hofmann, J. P. and Anderson, N. G. (1991). A 2-dimensional gel database of rat-liver proteins useful in gene-regulation and drug effects studies. *Electrophoresis* **12** (11): 907-930.
- Andrewartha, K. A., Phillips, D. R. and Stone, B. A. (1979). Solution properties of wheat-flour arabinoxylans and enzymically modified arabinoxylans. *Carbohydrate Research* **77**: 191-204.

- Antelmann, H., Tjalsma, H., Voigt, B., Ohlmeier, S., Bron, S., van Dijk, J. M. and Hecker, M. (2001). A proteomic view on genome-based signal peptide predictions. *Genome Research* **11** (9): 1484-1502.
- Antonopoulos, D. A., Russell, W. M. and White, B. A. (2003). Phylogenetic reconstruction of Gram-positive organisms based on comparative sequence analysis of molecular chaperones from the ruminal microorganism *Ruminococcus flavefaciens* FD-1. *FEMS Microbiology Letters* **227** (1): 1-7.
- Asanuma, N., Kawato, M., Ohkawara, S. and Hino, T. (2003). Characterization and transcription of the genes encoding enzymes involved in butyrate production in *Butyrivibrio fibrisolvens*. *Current Microbiology* **47** (3): 203-207.
- Asanuma, N., Ishiwata, M., Yoshii, T., Kikuchi, M., Nishina, Y. and Hino, T. (2005). Characterization and transcription of the genes involved in butyrate production in *Butyrivibrio fibrisolvens* Type-I and II strains. *Current Microbiology* **51** (2): 91-94.
- Ash, R. W. and Dobson, A. (1963). Effect of absorption on acidity of rumen contents. *Journal of Physiology-London* **169** (1): 39-61.
- Aspinall, G. O. (1980). Chemistry of cell wall polysaccharides. Carbohydrates: structure and function. . *The biochemistry of plants (a comprehensive treatise)*. Eds. Preiss, J. New York., Academic Press. **3**: 473-500.
- Atalla, R. H., Hackney, J. M., Uhlin, I. and Thompson, N. S. (1993). Hemicelluloses as structure regulators in the aggregation of native cellulose. *International Journal of Biological Macromolecules* **15** (2): 109-112.
- Attwood, G. T. and Reilly, K. (1995). Identification of proteolytic rumen bacteria isolated from New Zealand cattle. *Journal of Applied Bacteriology* **79** (1): 22-29.
- Attwood, G. T., Reilly, K. and Patel, B. K. C. (1996). *Clostridium proteoclasticum* sp nov, a novel proteolytic bacterium from the bovine rumen. *International Journal of Systematic Bacteriology* **46** (3): 753-758.
- Attwood, G. T., Kelly, W. J., Altermann, E. H., Moon, C. D., Leahy, S. and Cookson, A. L. (2008). Application of rumen microbial genome information to livestock systems in the postgenomic era. *Australian Journal of Experimental Agriculture* **48** (6-7): 695-700.
- Aurilia, V., Martin, J. C., McCrae, S. I., Scott, K. P., Rincon, M. T. and Flint, H. J. (2000). Three multidomain esterases from the cellulolytic rumen anaerobe *Ruminococcus flavefaciens* 17 that carry divergent dockerin sequences. *Microbiology-Uk* **146**: 1391-1397.
- Barr, B. K., Hsieh, Y. L., Ganem, B. and Wilson, D. B. (1996). Identification of two functionally different classes of exocellulases. *Biochemistry* **35** (2): 586-592.
- Barrett, A. J. (1999). Enzyme nomenclature. Recommendations 1992. Supplement 5: Corrections and additions (1997). *European Journal of Biochemistry* **264**: 610-650.
- Barriere, Y., Guillet, C., Goffner, D. and Pichon, M. (2003). Genetic variation and breeding strategies for improved cell wall digestibility in annual forage crops. A review. *Animal Research* **52** (3): 193-228.
- Barriere, Y., Ralph, J., Mechin, V., Guillaumie, S., Grabber, J. H., Argillier, O., Chabbert, B. and Lapierre, C. (2004). Genetic and molecular basis of grass cell wall biosynthesis and degradability. II. Lessons from brown-midrib mutants. *Comptes Rendus Biologies* **327** (9-10): 847-860.
- Bayer, E. A., Setter, E. and Lamed, R. (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. *Journal of Bacteriology* **163** (2): 552-559.
- Beauchemin, K. A., Rode, L. M. and Sewalt, V. J. H. (1995). Fibrolytic enzymes increase fiber digestibility and growth rate of steers fed dry forages. *Canadian Journal of Animal Science* **75** (4): 641-644.

- Beauchemin, K. A., Colombato, D., Morgavi, D. P. and Yang, W. Z. (2003). Use of exogenous fibrolytic enzymes to improve feed utilization by ruminants. *Journal of Animal Science* **81**: E37-47.
- Beddek, A. J., Rawson, P., Peng, L., Snell, R., Lehnert, K., Ward, H. E. and Jordan, T. W. (2008). Profiling the metabolic proteome of bovine mammary tissue. *Proteomics* **8** (7): 1502-1515.
- Ben Messaoud, E., Ben Ammar, Y., Mellouli, L. and Bejar, S. (2002). Thermostable pullulanase Type-I from new isolated *Bacillus thermoleovorans* US105: cloning, sequencing and expression of the gene in *E. coli*. *Enzyme and Microbial Technology* **31** (6): 827-832.
- Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. *Journal of Molecular Biology* **340** (4): 783-795.
- Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57** (1): 289-300.
- Benshalom, N. (1986). Hindrance of hemicellulose and cellulose hydrolysis by pectic substances. *Journal of Food Science* **51** (3): 720-725.
- Berg Miller, M. E., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., Coutinho, P. M., Borovok, I., Jindou, S., Lamed, R., Flint, H. J., Bayer, E. A. and White, B. A. (2009). Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS One* **4** (8): e6650.
- Berger, E., Jones, W. A., Jones, D. T. and Woods, D. R. (1989). Cloning and sequencing of an endoglucanase (*end1*) gene from *Butyrivibrio fibrisolvens* H17c. *Molecular and General Genetics* **219** (1-2): 193-198.
- Berger, E., Jones, W. A., Jones, D. T. and Woods, D. R. (1990). Sequencing and expression of a cellodextrinase (*ced1*) gene from *Butyrivibrio fibrisolvens* H17c cloned in *Escherichia coli*. *Molecular & General Genetics* **223** (2): 310-318.
- Berkelman, T. (2006). Sensitivity and protein-to-protein consistency of Flamingo (TM) Fluorescent Gel Stain compared to other fluorescent stains. *Molecular & Cellular Proteomics* **5** (10): S47-S47.
- Bertoldo, C., Armbricht, M., Becker, F., Schafer, T., Antranikian, G. and Liebl, W. (2004). Cloning, sequencing, and characterization of thermoalkalstable Type-I pullulanase from *Anaerobranca gottschalkii*. *Applied and Environmental Microbiology* **70** (6): 3407-3416.
- Biely, P., Vrsanska, M., Tenkanen, M. and D., K. (1997). Endo-beta-1,4-xylanase families: differences in catalytic properties. *Journal of Bacteriology* **57** (1-3): 151-66.
- Bjellqvist, B., Ek, K., Righetti, P. G., Gianazza, E., Gorg, A., Westermeier, R. and Postel, W. (1982). Isoelectric focusing in immobilized pH gradients - principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods* **6** (4): 317-339.
- Bjursell, M. K., Martens, E. C. and Gordon, J. I. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *Journal of Biological Chemistry* **281** (47): 36269-36279.
- Black, G. W., Rixon, J. E., Clarke, J. H., Hazlewood, G. P., Theodorou, M. K., Morris, P. and Gilbert, H. J. (1996). Evidence that linker sequences and cellulose-binding domains enhance the activity of hemicellulases against complex substrates. *Biochemical Journal* **319**: 515-520.

- Blair, B. G. and Anderson, K. L. (1999). Cellulose-inducible ultrastructural protuberances and cellulose-affinity proteins of *Eubacterium cellulosolvens*. *Anaerobe* **5** (5): 547-554.
- Blouzard, J. C., Bourgeois, C., de Philip, P., Valette, O., Belaich, A., Tardif, C., Belaich, J. P. and Pages, S. (2007). Enzyme diversity of the cellulolytic system produced by *Clostridium cellulolyticum* explored by two-dimensional analysis: Identification of seven genes encoding new dockerin-containing proteins. *Journal of Bacteriology* **189** (6): 2300-2309.
- Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P. and Gilbert, H. J. (1998). *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochemical Journal* **331**: 775-781.
- Bonhomme, A. (1990). Rumen ciliates - Their metabolism and relationships with bacteria and their hosts. *Animal Feed Science and Technology* **30** (3-4): 203-266.
- Boos, W. and Shuman, H. (1998). Maltose/maltodextrin system of *Escherichia coli*: Transport, metabolism, and regulation. *Microbiology and Molecular Biology Reviews* **62** (1): 204-229.
- Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A. J. and Kilburn, D. G. (2001). Binding specificity and thermodynamics of a Family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A. *Biochemistry* **40** (21): 6240-6247.
- Borneman, W. S., Hartley, R. D., Himmelsbach, D. S. and Ljungdahl, L. G. (1990). Assay for trans-para-coumaroyl esterase using a specific substrate from plant-cell walls. *Analytical Biochemistry* **190** (1): 129-133.
- Bourgeois, T. M., Van Craeyveld, V., Van Campenhout, S., Courtin, C. M., Delcour, J. A., Robben, J. and Volckaert, G. (2007). Recombinant expression and characterization of XynD from *Bacillus subtilis* subsp. *subtilis* ATCC 6051: a GH 43 arabinoxylan arabinofuranohydrolase. *Applied Microbiology and Biotechnology* **75** (6): 1309-1317.
- Bower, B. S. (2005). Fusion proteins of an exocellobiohydrolase and an endoglucanase for use in the saccharification of cellulose and hemicellulose. Genencor International, IU, USA. 2005093073.
- Brechtel, E., Matuschek, M., Hellberg, A., Egelseer, E. M., Schmid, R. and Bahl, H. (1999). Cell wall of *Thermoanaerobacterium thermosulfurigenes* EM1: isolation of its components and attachment of the xylanase XynA. *Archives of Microbiology* **171** (3): 159-165.
- Brillouet, J. M. (1987). Mode of action of xylanase-II from *Polyporus tulipiferae* on oat-spelt arabinoglucuronoxylan and larchwood 4-O-methylglucuronoxylan. *Carbohydrate Research* **159** (1): 165-170.
- Brillouet, J. M. and Joseleau, J. P. (1987). Investigation of the structure of a heteroxylan from the outer pericarp (beeswing bran) of wheat kernel. *Carbohydrate Research* **159** (1): 109-126.
- Brown, R. M. and Saxena, I. M. (2000). Cellulose biosynthesis: A model for understanding the assembly of biopolymers. *Plant Physiology and Biochemistry* **38** (1-2): 57-67.
- Brulc, J. M., Antonopoulos, D. A., Miller, M. E. B., Wilson, M. K., Yannarell, A. C., Dinsdale, E. A., Edwards, R. E., Frank, E. D., Emerson, J. B., Wacklin, P., Coutinho, P. M., Henrissat, B., Nelson, K. E. and White, B. A. (2009). Gene-centric metagenomics of the fiber-adherent bovine rumen microbiome reveals forage specific glycoside hydrolases. *Proceedings of the National Academy of Sciences of the United States of America* **106** (6): 1948-1953.

- Bryant, M. P. and Doetsch, R. N. (1955). Factors necessary for the growth of *Bacteroides succinogenes* in the volatile acid fraction of rumen fluid. *Journal of Dairy Science* **38** (4): 340-350.
- Buanafina, M. M. D., Langdon, T., Hauck, B., Dalton, S. J. and Morris, P. (2006). Manipulating the phenolic acid content and digestibility of Italian ryegrass (*Lolium multiflorum*) by vacuolar-targeted expression of a fungal ferulic acid esterase. *Applied Biochemistry and Biotechnology* **130** (1-3): 416-426.
- Buanafina, M. M. D. O., Langdon, T., Hauck, B., Dalton, S. and Morris, P. (2008). Expression of a fungal ferulic acid esterase increases cell wall digestibility of tall fescue (*Festuca arundinacea*). *Plant Biotechnology Journal* **6** (3): 264-280.
- Bunai, K., Ariga, M., Inoue, T., Nozaki, M., Ogane, S., Kakeshita, H., Nemoto, T., Nakanishi, H. and Yamane, K. (2004). Profiling and comprehensive expression analysis of ABC transporter solute-binding proteins of *Bacillus subtilis* membrane based on a proteomic approach. *Electrophoresis* **25** (1): 141-155.
- Burchhardt, G., Sahm, K., Matuschek, M., Muller, H., Haeckel, K., Spreinat, A., Antranikian, G., Steinborn, G. and Bahl, H. (1994). Molecular analysis of the pullulanase-encoding gene of *Clostridium thermosulfurogenes em1*. *ECB6: Proceedings of the 6th European Congress on Biotechnology, Parts I and II* **9**: 341-344.
- Burr, S. J. and Fry, S. C. (2009). Extracellular cross-linking of maize arabinoxylans by oxidation of feruloyl esters to form oligoferuloyl esters and ether-like bonds. *Plant Journal* **58** (4): 554-567.
- Campbell, M. M. and Sederoff, R. R. (1996). Variation in lignin content and composition - Mechanism of control and implications for the genetic improvement of plants. *Plant Physiology* **110** (1): 3-13.
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research* **37**: D233-D238.
- Carpentier, S. C., Witters, E., Laukens, K., Deckers, P., Swennen, R. and Panis, B. (2005). Preparation of protein extracts from recalcitrant plant tissues: An evaluation of different methods for two-dimensional gel electrophoresis analysis. *Proteomics* **5** (10): 2497-2507.
- Carpita, N. C. (1989). Pectic polysaccharides of maize coleoptiles and proso millet cells in liquid culture. *Phytochemistry* **28** (1): 121-125.
- Carpita, N. C. and Gibeaut, D. M. (1993). Structural models of primary-cell walls in flowering plants - Consistency of molecular-structure with the physical properties of the walls during growth. *Plant Journal* **3** (1): 1-30.
- Carpita, N. C. (1996). Structure and biogenesis of the cell walls of grasses. *Annual Review of Plant Physiology and Plant Molecular Biology* **47**: 445-476.
- Carpita, N. C., Defernez, M., Findlay, K., Wells, B., Shoue, D. A., Catchpole, G., Wilson, R. H. and McCann, M. C. (2001). Cell-wall architecture of the elongating maize coleoptile. *Plant Physiology* **127** (2): 551-565.
- Carvalho, A. L., Dias, F. M. V., Nagy, T., Prates, J. A. M., Proctor, M. R., Smith, N., Bayer, E. A., Davies, G. J., Ferreira, L. M. A., Romao, M. J., Fontes, C. and Gilbert, H. J. (2007). Evidence for a dual binding mode of dockerin modules to cohesins. *Proceedings of the National Academy of Sciences of the United States of America* **104** (9): 3089-3094.
- Carvalho, P. C., Hewel, J., Barbosa, V. C. and Yates, J. R. (2008). Identifying differences in protein expression levels by spectral counting and feature selection. *Genetics and Molecular Research* **7** (2): 342-356.

- Cassab, G. I. and Varner, J. E. (1987). Immunocytochemical localization of extensin in developing soybean seed coats by immunogold silver staining and by tissue printing on nitrocellulose paper. *Journal of Cell Biology* **105** (6): 2581-2588.
- Chabannes, M., Ruel, K., Yoshinaga, A., Chabbert, B., Jauneau, A., Joseleau, J. P. and Boudet, A. M. (2001). *In situ* analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels. *Plant Journal* **28** (3): 271-282.
- Chamrad, D. C., Korting, G., Stuhler, K., Meyer, H. E., Klose, J. and Bluggel, M. (2004). Evaluation of algorithms for protein identification from sequence databases using mass spectrometry data. *Proteomics*. **4** (3): 619-28.
- Chang, J., van Remmen, H., Ward, W. F., Regnier, F. E., Richardson, A. and Cornell, J. (2004). Processing of data generated by 2-dimensional gel electrophoresis for statistical analysis: Missing data, normalization, and statistics. *Journal of Proteome Research* **3**: 1210-1218.
- Chelius, D. and Bondarenko, P. V. (2002). Quantitative profiling of proteins in complex mixtures using liquid chromatography and mass spectrometry. *Journal of Proteome Research* **1** (4): 317-323.
- Chen, J., Sharma, S., Quiocho, F. A. and Davidson, A. L. (2001). Trapping the transition state of an ATP-binding cassette transporter: Evidence for a concerted mechanism of maltose transport. *Proceedings of the National Academy of Sciences of the United States of America* **98** (4): 1525-1530.
- Cheng, K. J., Forsberg, C. W., Minato, H. and Costerton, J. W. (1991). Microbial ecology and physiology of feed degradation within the rumen. *Physiological Aspects of Digestion and Metabolism in Ruminants*. Eds. Tsuda, T., Sasaki, Y. and Kawashima, R. Toronto, ON, Academic Press: 595-624.
- Chitlaru, T. and Shafferman, A. (2009). Proteomic studies of *Bacillus anthracis*. *Future Microbiology* **4** (8): 983-998.
- Cho, C. W., Lee, S. H., Choi, J., Park, S. J., Ha, D. J., Kim, H. J. and Kim, C. W. (2003). Improvement of the two-dimensional gel electrophoresis analysis for the proteome study of *Halobacterium salinarum*. *Proteomics*. **3** (12): 2325-9.
- Cho, S. G. and Choi, Y. J. (1998). Regulation of beta-xylosidase (XylA) synthesis in *Bacillus stearothermophilus*. *Journal of Microbiology and Biotechnology* **8** (1): 14-20.
- Cho, S. J., Cho, K. M., Shin, E. C., Lim, W. J., Hong, S. Y., Choi, B. R., Kang, J. M., Lee, S. M., Kim, Y. H., Kim, H. and Yun, H. D. (2006). 16S rDNA analysis of bacterial diversity in three fractions of cow rumen. *Journal of Microbiology and Biotechnology* **16** (1): 92-101.
- Chow, V., Nong, G. and Preston, J. F. (2007). Structure, function, and regulation of the aldouronate utilization gene cluster from *Paenibacillus* sp strain JDR-2. *Journal of Bacteriology* **189** (24): 8863-8870.
- Chu, P. W., Yap, M. N., Wu, C. Y., Huang, C. M., Pan, F. M., Tseng, M. J. and Chen, S. T. (2000). A proteomic analysis of secreted proteins from xylan-induced *Bacillus* sp strain K-1. *Electrophoresis* **21** (9): 1740-1745.
- Clauser, K. R., Baker, P. and Burlingame, A. L. (1999). Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Analytical Chemistry* **71** (14): 2871-2882.
- Coen, J. A. and Dehority, B. A. (1970). Degradation and utilization of hemicellulose from intact forages by pure cultures of rumen bacteria. *Applied Microbiology* **20** (3): 362-368.

- Coenen, G. J., Bakx, E. J., Verhoef, R. P., Schols, H. A. and Voragen, A. G. J. (2007). Identification of the connecting linkage between homo- or xylogalacturonan and rhamnogalacturonan Type-I. *Carbohydrate Polymers* **70** (2): 224-235.
- Cohen, D. P. A., Renes, J., Bouwman, F. G., Zoetendal, E. G., Mariman, E., de Vos, W. M. and Vaughan, E. E. (2006). Proteomic analysis of log to stationary growth phase *Lactobacillus plantarum* cells and a 2-DE database. *Proteomics* **6** (24): 6485-6493.
- Colombatto, D., Morgavi, D. P., Furtado, A. F. and Beauchemin, K. A. (2003). Screening of exogenous enzymes for ruminant diets: Relationship between biochemical characteristics and in vitro ruminal degradation. *Journal of Animal Science* **81** (10): 2628-2638.
- Colombatto, D. and Beauchemin, K. A. (2008). A protease additive increases fermentation of alfalfa diets by mixed ruminal microorganisms *in vitro*. *Journal of Animal Science* **87** (3): 1097-1105.
- Cooper, R. A. (1984). Metabolism of methylglyoxal in microorganisms. *Annual Review of Microbiology* **38**: 49-68.
- Correia, M. A. S., Pires, V. M. R., Gilbert, H. J., Bolam, D. N., Fernandes, V. O., Alves, V. D., Prates, J. A. M., Ferreira, L. M. A. and Fontes, C. (2009). Family 6 carbohydrate-binding modules display multiple beta-1,3-linked glucan-specific binding interfaces. *FEMS Microbiology Letters* **300** (1): 48-57.
- Cosgrove, D. J. (2000). Expansive growth of plant cell walls. *Plant Physiology and Biochemistry* **38** (1-2): 109-124.
- Cotta, M. A. and Hespell, R. B. (1986). Proteolytic activity of the ruminal bacterium *Butyrivibrio fibrisolvens*. *Applied and Environmental Microbiology* **52** (1): 51-58.
- Cotta, M. A. (1993). Utilization of xylooligosaccharides by selected ruminal bacteria. *Applied and Environmental Microbiology* **59** (11): 3557-3563.
- Cotta, M. A. and Zeltwanger, R. L. (1995). Degradation and utilization of xylan by the ruminal bacteria *Butyrivibrio fibrisolvens* and *Selenomonas ruminantium*. *Applied and Environmental Microbiology* **61** (12): 4396-4402.
- Cousins, S. K. and Brown, R. M. (1995). Cellulose-I microfibril assembly - Computational molecular mechanics energy analysis favors bonding by van der Waals forces as the initial step in crystallization. *Polymer* **36** (20): 3885-3888.
- Craig, W. M., Broderick, G. A. and Ricker, D. B. (1987). Quantitation of microorganisms associated with the particulate phase of ruminal ingesta. *Journal of Nutrition* **117** (1): 56-62.
- Creagh, A. L., Ong, E., Jervis, E., Kilburn, D. G. and Haynes, C. A. (1996). Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proceedings of the National Academy of Sciences of the United States of America* **93** (22): 12229-12234.
- Cruywagen, C. W. and van Zyl, W. H. (2008). Effects of a fungal enzyme cocktail treatment of high and low forage diets on lamb growth. *Animal Feed Science and Technology* **145** (1-4): 151-158.
- Dabard, J., Bridonneau, C., Phillippe, C., Anglade, P., Molle, D., Nardi, M., Ladire, M., Girardin, H., Marcille, F., Gomez, A. and Fons, M. (2001). Ruminococcin A, a new lantibiotic produced by a *Ruminococcus gnavus* strain isolated from human feces. *Applied and Environmental Microbiology* **67** (9): 4111-4118.
- Dado, R. G. and Allen, M. S. (1995). Intake limitations, feeding-behavior, and rumen function of cows challenged with rumen fill from dietary fiber or inert bulk. *Journal of Dairy Science* **78** (1): 118-133.

- Dalrymple, B. P., Swadling, Y., Cybinski, D. H. and Xue, G. P. (1996). Cloning of a gene encoding cinnamoyl ester hydrolase from the ruminal bacterium *Butyrivibrio fibrisolvens* E14 by a novel method. *Fems Microbiology Letters* **143** (2-3): 115-120.
- Dalrymple, B. P. and Swadling, Y. (1997). Expression of a *Butyrivibrio fibrisolvens* E14 gene (*cinB*) encoding an enzyme with cinnamoyl ester hydrolase activity is negatively regulated by the product of an adjacent gene (*cinR*). *Microbiology-Uk* **143**: 1203-1210.
- Dardonville, B. and Raibaud, O. (1990). Characterization of *malT* mutants that constitutively activate the maltose regulon of *Escherichia coli*. *Journal of Bacteriology* **172** (4): 1846-1852.
- Dassa, E. and Bouige, P. (2001). The ABC of ABCs: a phylogenetic and functional classification of ABC systems in living organisms. *Research in Microbiology* **152** (3-4): 211-229.
- Davidson, A. L., Shuman, H. A. and Nikaido, H. (1992). Mechanism of maltose transport in *Escherichia coli* - transmembrane signaling by periplasmic binding-proteins. *Proceedings of the National Academy of Sciences of the United States of America* **89** (6): 2360-2364.
- Davidson, A. L., Dassa, E., Orelle, C. and Chen, J. (2008). Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiology and Molecular Biology Reviews* **72** (2): 317-364.
- Day, A. G. (2003). Novel variant *Hypocrea jecorina* CBH I cellulases, their production with recombinant cells, and their uses. Genencor International, IU, USA. 2004016760.
- Dea, I. C. M., Rees, D. A., Beveridge, R. J. and Richards, G. N. (1973). Aggregation with change of conformation in solutions of hemicellulose xylans. *Carbohydrate Research* **29** (2): 363-372.
- Dehority, B. A. (1966). Characterization of several bovine rumen bacteria isolated with a xylan medium. *Journal of Bacteriology* **91** (5): 1724-1729.
- Demain, A. L., Newcomb, M. and Wu, J. H. D. (2005). Cellulase, *Clostridia*, and ethanol. *Microbiology and Molecular Biology Reviews* **69** (1): 124-154.
- Deutscher, J. and Saier, M. H. (2005). Ser/Thr/Tyr protein phosphorylation in bacteria - For long time neglected, now well established. *Journal of Molecular Microbiology and Biotechnology* **9** (3-4): 125-131.
- Devillard, E., Goodheart, D. B., Karnati, S. K. R., Bayer, E. A., Lamed, R., Miron, J., Nelson, K. E. and Morrison, M. (2004). *Ruminococcus albus* 8 mutants defective in cellulose degradation are deficient in two processive endocellulases, Cel48A and Cel9B, both of which possess a novel modular architecture. *Journal of Bacteriology* **186** (1): 136-145.
- Devillard, E., Andant, N. and Wallace, R. J. (2006). Increased expression of a molecular chaperone GroEL in response to unsaturated fatty acids by the biohydrogenating ruminal bacterium, *Butyrivibrio fibrisolvens*. *FEMS Microbiology Letters* **262** (2): 244-248.
- Diez-Gonzalez, F., Bond, D. R., Jennings, E. and Russell, J. B. (1999). Alternative schemes of butyrate production in *Butyrivibrio fibrisolvens* and their relationship to acetate utilization, lactate production, and phylogeny. *Archives of Microbiology* **171** (5): 324-330.
- Din, N., Gilkes, N. R., Tekant, B., Miller, R. C., Warren, A. J. and Kilburn, D. G. (1991). Non-hydrolytic disruption of cellulose fibers by the binding domain of a bacterial cellulase. *Biotechnology* **9** (11): 1096-1099.

- Dodd, D., Kocherginskaya, S. A., Spies, M. A., Beery, K. E., Abbas, C. A., Mackie, R. I. and Cann, I. K. O. (2009). Biochemical analysis of a beta-D-xylosidase and a bifunctional xylanase-ferulic acid esterase from a xylanolytic gene cluster in *Prevotella ruminicola* 23. *Journal of Bacteriology* **191** (10): 3328-3338.
- Dodd, D., Moon, Y. H., Swaminathan, K., Mackie, R. I. and Cann, I. K. O. (2010). Transcriptomic analyses of xylan degradation by *Prevotella bryantii* and insights into energy acquisition by xylanolytic Bacteroidetes. *Journal of Biological Chemistry* **285** (39): 30261-30273.
- Doerner, K. C. and White, B. A. (1990). Assessment of the endo-1,4-beta-glucanase components of *Ruminococcus flavefaciens* FD-1. *Applied and Environmental Microbiology* **56** (6): 1844-1850.
- Doeven, M. K., Abele, R., Tampe, R. and Poolman, B. (2004). The binding specificity of OppA determines the selectivity of the oligopeptide ATP-binding cassette transporter. *Journal of Biological Chemistry* **279** (31): 32301-32307.
- Doi, R. H. and Tamaru, Y. (2001). The *Clostridium cellulovorans* cellulosome: An enzyme complex with plant cell wall degrading activity. *Chemical Record* **1** (1): 24-32.
- Doman-Pytka, M. and Bardowski, J. (2004). Pullulan degrading enzymes of bacterial origin. *Critical Reviews in Microbiology* **30** (2): 107-121.
- Duan, J. Y., Zheng, Y., Dong, Q. and Fang, J. N. (2004). Structural analysis of a pectic polysaccharide from the leaves of *Diospyros kaki*. *Phytochemistry* **65** (5): 609-615.
- Ebringerova, A., Hromadkova, Z. and Heinze, T. (2005). Hemicellulose. *Polysaccharides 1: Structure, Characterization and Use* **186**: 1-67.
- Edwards, J. E., McEwan, N. R., Travis, A. J. and Wallace, R. J. (2004). 16S rDNA library-based analysis of ruminal bacterial diversity. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **86** (3): 263-281.
- Edwards, J. E., Huws, S. A., Kim, E. J., Lee, M. R. F., Kingston-Smith, A. H. and Scollan, N. D. (2008). Advances in microbial ecosystem concepts and their consequences for ruminant agriculture. *Animal* **2** (5): 653-660.
- Elias, J. E., Haas, W., Faherty, B. K. and Gygi, S. P. (2005). Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. *Nature Methods* **2** (9): 667-675.
- Ellis, W. C. (1978). Determinants of grazed forage intake and digestibility. *Journal of Dairy Science* **61** (12): 1828-1840.
- Erbeznik, M., Strobel, H. J., Dawson, K. A. and Jones, C. R. (1998). The D-xylose binding protein, XylF, from *Thermoanaerobacter ethanolicus* 39E: Cloning, molecular analysis, and expression of the structural gene. *Journal of Bacteriology* **180** (14): 3570-3577.
- Eun, J. S. and Beauchemin, K. A. (2005). Effects of a proteolytic feed enzyme on intake, digestion, ruminal fermentation, and milk production. *Journal of Dairy Science* **88** (6): 2140-2153.
- Evans, F. F., Raftery, M. J., Egan, S. and Kjelleberg, S. (2007). Profiling the secretome of the marine bacterium *Pseudoalteromonas tunicata* using amine-specific isobaric tagging (iTRAQ). *Journal of Proteome Research* **6** (3): 967-975.
- Faulds, C. B., deVries, R. P., Kroon, P. A., Visser, J. and Williamson, G. (1997). Influence of ferulic acid on the production of feruloyl esterases by *Aspergillus niger*. *FEMS Microbiology Letters* **157** (2): 239-244.
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989). Electrospray ionization for mass-spectrometry of large biomolecules. *Science* **246** (4926): 64-71.

- Fields, M. W., Russell, J. B. and Wilson, D. B. (1998). The role of ruminal carboxymethylcellulases in the degradation of beta-glucans from cereal grain. *FEMS Microbiology Ecology* **27** (3): 261-268.
- Finn, R. D., Tate, J., Mistry, J., Coghill, P. C., Sammut, S. J., Hotz, H. R., Ceric, G., Forslund, K., Eddy, S. R., Sonnhammer, E. L. L. and Bateman, A. (2008). The Pfam protein families database. *Nucleic Acids Research* **36**: D281-D288.
- Fisher, D. S., Burns, J. C. and Pond, K. R. (1989). Kinetics of *in vitro* cell-wall disappearance and *in vivo* digestion. *Agronomy Journal* **81** (1): 25-33.
- Flint, H. J., Martin, J., McPherson, C. A., Daniel, A. S. and Zhang, J. X. (1993). A bifunctional enzyme, with separate xylanase and beta-(1,3-1,4)-glucanase domains, encoded by the *xynD* gene of *Ruminococcus flavefaciens*. *Journal of Bacteriology* **175** (10): 2943-2951.
- Flint, H. J., Whitehead, T. R., Martin, J. C. and Gasparic, A. (1997). Interrupted catalytic domain structures in xylanases from two distantly related strains of *Prevotella ruminicola*. *Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology* **1337** (2): 161-165.
- Flint, H. J. (2004). Polysaccharide breakdown by anaerobic microorganisms inhabiting the mammalian gut. *Advances in Applied Microbiology*. Eds. Amsterdam, Elsevier Publishing. **56**: 89-120.
- Florens, L., Washburn, M. P., Raine, J. D., Anthony, R. M., Grainger, M., Haynes, J. D., Moch, J. K., Muster, N., Sacci, J. B., Tabb, D. L. and al., e. (2002). A proteomic view of the *Plasmodium falciparum* life cycle. *Nature* **419** (6906): 520-526.
- Foroozandeh, A. D., Rezaeian, M., Balaly, G. R. and Alikhani, M. (2009). Relative contributions of ruminal bacteria, protozoa and fungi to degradation of forage fiber fractions. *Journal of Animal and Veterinary Advances* **8** (3): 603-607.
- Francis, A. W., Ruggiero, C. E., Koppisch, A. T., Dong, J. G., Song, J., Brettin, T. and Iyer, S. (2005). Proteomic analysis of *Bacillus anthracis* Sterne vegetative cells. *Biochimica Et Biophysica Acta-Proteins and Proteomics* **1748** (2): 191-200.
- Fritz, J. O., Moore, K. J. and Vogel, K. P. (1991). Ammonia-labile bonds in high-digestibility and low-digestibility strains of switchgrass. *Crop Science* **31** (6): 1566-1570.
- Fry, S. C., Willis, S. C. and Paterson, A. E. J. (2000). Intraprotoplasmic and wall-localised formation of arabinoxylan-bound diferulates and larger ferulate coupling-products in maize cell-suspension cultures. *Planta* **211** (5): 679-692.
- Fuchs, K. P., Zverlov, V. V., Velikodvorskaya, G. A., Lottspeich, F. and Schwarz, W. H. (2003). Lic16A of *Clostridium thermocellum*, a non-cellulosomal, highly complex endo-beta-1, 3-glucanase bound to the outer cell surface. *Microbiology-Sgm* **149**: 1021-1031.
- Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. and Garrels, J. I. (1999). A sampling of the yeast proteome. *Molecular and Cellular Biology* **19** (11): 7357-7368.
- Gao, J., Opiteck, G. J., Friedrichs, M. S., Dongre, A. R. and Hefta, S. A. (2003). Changes in the protein expression of yeast as a function of carbon source. *Journal of Proteome Research* **2** (6): 643-649.
- Garcia, B. L., Ball, A. S., Rodriguez, J., Perez-Leblic, M. I., Arias, M. E. and Copa-Patino, J. L. (1998). Induction of ferulic acid esterase and xylanase activities in *Streptomyces avermitilis* UAH30. *FEMS Microbiology Letters* **158** (1): 95-99.
- Gasparic, A., Marinseklogar, R., Martin, J., Wallace, R. J., Nekrep, F. V. and Flint, H. J. (1995a). Isolation of genes encoding beta-D-xylanase, beta-D-xylosidase and alpha-L-arabinofuranosidase activities from the rumen bacterium *Prevotella Ruminicola* B(1)4. *FEMS Microbiology Letters* **125** (2-3): 135-141.

- Gasparic, A., Martin, J., Daniel, A. S. and Flint, H. J. (1995b). A xylan hydrolase gene cluster in *Prevotella ruminicola* B(1)4 - Sequence relationships, synergistic interactions, and oxygen sensitivity of a novel enzyme with exoxylanase and beta-(1,4)-xylosidase activities. *Applied and Environmental Microbiology* **61** (8): 2958-2964.
- Gilois, N., Ramarao, N., Bouillaut, L., Perchat, S., Aymerich, S., Nielsen-LeRoux, C., Lereclus, D. and Gohar, M. (2007). Growth-related variations in the *Bacillus cereus* secretome. *Proteomics* **7** (10): 1719-1728.
- Giraldo, L. A., Tejido, M. L., Ranilla, M. J., Ramos, S. and Carro, M. D. (2008). Influence of direct-fed fibrolytic enzymes on diet digestibility and ruminal activity in sheep fed a grass hay-based diet. *Journal of Animal Science* **86** (7): 1617-1623.
- Gobius, K. S., Xue, G. P., Aylward, J. H., Dalrymple, B. P., Swadling, Y. J., McSweeney, C. S. and Krause, D. O. (2002). Transformation and expression of an anaerobic fungal xylanase in several strains of the rumen bacterium *Butyrivibrio fibrisolvens*. *Journal of Applied Microbiology* **93** (1): 122-133.
- Goetz, R. M. and Fuglsang, A. (2005). Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from *Escherichia coli*. *Biochemical and Biophysical Research Communications* **327** (1): 4-7.
- Golan, G., Shallom, D., Teplitsky, A., Zaide, G., Shulami, S., Baasov, T., Stojanoff, V., Thompson, A., Shoham, Y. and Shoham, G. (2004). Crystal structures of *Geobacillus stearothermophilus* alpha-glucuronidase complexed with its substrate and products - Mechanistic implications. *Journal of Biological Chemistry* **279** (4): 3014-3024.
- Goldstone, D. C., Villas-Boas, S. G., Till, M., Kelly, W. J., Attwood, G. T. and Arcus, V. L. (2010). Structural and functional characterization of a promiscuous feruloyl esterase (Est1E) from the rumen bacterium *Butyrivibrio proteoclasticus*. *Proteins* **78** (6): 1457-69.
- Gong, J. H., Egbosimba, E. E. and Forsberg, C. W. (1996). Cellulose binding proteins of *Fibrobacter succinogenes* and the possible role of a 180-kDa cellulose-binding glycoprotein in adhesion to cellulose. *Canadian Journal of Microbiology* **42** (5): 453-460.
- Grabber, J. H., Hatfield, R. D., Ralph, J., Zon, J. and Amrhein, N. (1995). Ferulate cross-linking in cell-walls isolated from maize cell-suspensions. *Phytochemistry* **40** (4): 1077-1082.
- Grabber, J. H., Hatfield, R. D. and Ralph, J. (1998a). Diferulate cross-links impede the enzymatic degradation of non-lignified maize walls. *Journal of the Science of Food and Agriculture* **77** (2): 193-200.
- Grabber, J. H., Ralph, J. and Hatfield, R. D. (1998b). Ferulate cross-links limit the enzymatic degradation of synthetically lignified primary walls of maize. *Journal of Agricultural and Food Chemistry* **46** (7): 2609-2614.
- Grabber, J. H., Ralph, J. and Hatfield, R. D. (2000). Cross-linking of maize walls by ferulate dimerization and incorporation into lignin. *Journal of Agricultural and Food Chemistry* **48** (12): 6106-6113.
- Grabber, J. H., Ralph, J. and Hatfield, R. D. (2002). Model studies of ferulate-coniferyl alcohol cross-product formation in primary maize walls: Implications for lignification in grasses. *Journal of Agricultural and Food Chemistry* **50** (21): 6008-6016.
- Grabber, J. H., Ralph, J., Lapierre, C. and Barriere, Y. (2004). Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions. Review. *C R Biol.* **327** (5): 455-65.

- Greve, L. C., Labavitch, J. M. and Hungate, R. E. (1984). Alpha-L-arabinofuranosidase from *Ruminococcus albus* 8 - Purification and possible role in hydrolysis of alfalfa cell-wall. *Applied and Environmental Microbiology* **47** (5): 1135-1140.
- Grove, H., Hollung, K., Uhlen, A. K., Martens, H. and Faergestad, E. M. (2006). Challenges related to analysis of protein spot volumes from two-dimensional gel electrophoresis as revealed by replicate gels. *Journal of Proteome Research* **5** (12): 3399-3410.
- Gu, Y., Ding, Y., Ren, C., Sun, Z., Rodionov, D. A., Zhang, W. W., Yang, S., Yang, C. and Jiang, W. H. (2010) Reconstruction of xylose utilization pathway and regulons in *Firmicutes*. *Bmc Genomics* **11** 10.1186/1471-2164-11-255.
- Guillaume, E., Berger, B., Affolter, M. and Kussmann, M. (2009). Label-free quantitative proteomics of two *Bifidobacterium longum* strains. *Journal of Proteomics* **72** (5): 771-784.
- Ha, M. A., Apperley, D. C., Evans, B. W., Huxham, M., Jardine, W. G., Vietor, R. J., Reis, D., Vian, B. and Jarvis, M. C. (1998). Fine structure in cellulose microfibrils: NMR evidence from onion and quince. *Plant Journal* **16** (2): 183-190.
- Haft, D. H., Selengut, J. D. and White, O. (2003). The TIGRFAMs database of protein families. *Nucleic Acids Research* **31** (1): 371-373.
- Hahne, H., Wolff, S., Hecker, M. and Becher, D. (2008). From complementarity to comprehensiveness - targeting the membrane proteome of growing *Bacillus subtilis* by divergent approaches. *Proteomics* **8** (19): 4123-4136.
- Halliwell, G. and Bryant, M. P. (1963). Cellulolytic activity of pure strains of bacteria from rumen of cattle. *Journal of General Microbiology* **32** (3): 441-448.
- Han, S. O., Cho, H. Y., Yukawa, H., Inui, M. and Doi, R. H. (2004). Regulation of expression of cellulosomes and noncellulosomal (hemi)cellulolytic enzymes in *Clostridium cellulovorans* during growth on different carbon sources. *Journal of Bacteriology* **186** (13): 4218-4227.
- Han, S. O., Yukawa, H., Inui, M. and Doi, R. H. (2005). Effect of carbon source on the cellulosomal subpopulations of *Clostridium cellulovorans*. *Microbiology-Sgm* **151**: 1491-1497.
- Hansmeier, N., Bartels, F. W., Ros, R., Anselmetti, D., Tauch, A., Puhler, A. and Kalinowski, J. (2004). Classification of hyper-variable *Corynebacterium glutamicum* surface-layer proteins by sequence analyses and atomic force microscopy. *Journal of Biotechnology* **112** (1-2): 177-193.
- Harada, K. M., Tanaka, K., Fukuda, Y., Hashimoto, W. and Murata, K. (2005). Degradation of rice bran hemicellulose by *Paenibacillus* sp strain HC1: gene cloning, characterization and function of beta-D-glucosidase as an enzyme involved in degradation. *Archives of Microbiology* **184** (4): 215-224.
- Harris, P. J. and Hartley, R. D. (1976). Detection of bound ferulic acid in cell-walls of *Gramineae* by ultraviolet fluorescence microscopy. *Nature* **259** (5543): 508-510.
- Harris, W. A., Janecki, D. J. and Reilly, J. P. (2002). Use of matrix clusters and trypsin autolysis fragments as mass calibrants in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Rapid Communications in Mass Spectrometry* **16** (18): 1714-1722.
- Harrison, P. W., Lower, R. P. J., Kim, N. K. D. and Young, J. P. W. (2010). Introducing the bacterial 'chromid': not a chromosome, not a plasmid. *Trends in Microbiology* **18** (4): 141-148.
- Hartley, R. D. (1972). Para-coumaric and ferulic acid components of cell-walls of ryegrass and their relationships with lignin and digestibility. *Journal of the Science of Food and Agriculture* **23** (11): 1347-1354.

- Hartley, R. D. and Ford, C. W. (1989). *Plant Cell Wall polymers: Biogenesis and Biodegradation*. Washington, DC, American Chemical Society.
- Hartley, R. D., Akin, D. E., Himmelsbach, D. S. and Beach, D. C. (1990a). Microspectrophotometry of bermudagrass (*Cynodon dactylon*) cell-walls in relation to lignification and wall biodegradability. *Journal of the Science of Food and Agriculture* **50** (2): 179-189.
- Hartley, R. D., Morrison, W. H., Himmelsbach, D. S. and Borneman, W. S. (1990b). Cross-linking of cell-wall phenolic arabinoxylans in gramineous plants. *Phytochemistry* **29** (12): 3705-3709.
- Hashimoto, T. and Nakata, Y. (2003). Synergistic degradation of arabinoxylan with alpha-L-arabinofuranosidase, xylanase and beta-xylosidase from soy sauce koji mold, *Aspergillus oryzae*, in high salt condition. *Journal of Bioscience and Bioengineering* **95** (2): 164-169.
- Hatfield, R. D., Wilson, J. R. and Mertens, D. R. (1999). Composition of cell walls isolated from cell types of grain sorghum stems. *Journal of the Science of Food and Agriculture* **79** (6): 891-899.
- Hecker, M., Antelmann, H., Buettner, K. and Bernhardt, J. (2008). Gel-based proteomics of Gram-positive bacteria: A powerful tool to address physiological questions. *Proteomics* **8** (23-24): 4958-4975.
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* **280**: 309-316.
- Henrissat, B. and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* **293**: 781-788.
- Henrissat, B. and Davies, G. (1997). Structural and sequence-based classification of glycoside hydrolases. *Current Opinion in Structural Biology* **7** (5): 637-644.
- Henrissat, B. (1998). Glycosidase families. *Biochemical Society Transactions* **26** (2): 153-156.
- Henshaw, J. L., Bolam, D. N., Pires, V. M. R., Czjzek, M., Henrissat, B., Ferreira, L. M. A., Fontes, C. and Gilbert, H. J. (2004). The Family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *Journal of Biological Chemistry* **279** (20): 21552-21559.
- Hespell, R. B., Wolf, R. and Bothast, R. J. (1987). Fermentation of xylans by *Butyrivibrio fibrisolvens* and other ruminal bacteria. *Applied and Environmental Microbiology* **53** (12): 2849-53.
- Hespell, R. B. (1991). The genera *Butyrivibrio*, *Lachnospira*, and *Roseburia*. *The procaryotes: a handbook on the biology of bacteria. Ecophysiology, isolation, identification, and applications*. Eds. Balows, A., Truper, H. G., Dworkin, M., Harder, W. and Schleifer, K. H. New York, Springer-Verlag: 2022-2033.
- Hespell, R. B. and Obryan, P. J. (1992). Purification and characterization of an alpha-L-arabinofuranosidase from *Butyrivibrio fibrisolvens* GS113. *Applied and Environmental Microbiology* **58** (4): 1082-1088.
- Hespell, R. B. and Cotta, M. A. (1995). Degradation and utilization by *Butyrivibrio fibrisolvens* H17c of xylans with different chemical and physical-properties. *Applied and Environmental Microbiology* **61** (8): 3042-3050.
- Higgins, C. F. and Ames, G. F. L. (1981). Two periplasmic transport proteins which interact with a common membrane-receptor show extensive homology - Complete nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **78** (10): 6038-6042.

- Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998). SOSUI: classification and secondary structure prediction system for membrane proteins. *Bioinformatics* **14** (4): 378-379.
- Hobson, P. N. (1989). *The Rumen Microbial Ecosystem*. London, Elsevier Applied Science.
- Hochberg, Y. (1988). A Sharper Bonferroni Procedure for Multiple Tests of Significance. *Biometrika* **75** (4): 800-802.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* **6**: 65-70.
- Hong, S. H., Kim, J. S., Lee, S. Y., In, Y. H., Choi, S. S., Rih, J. K., Kim, C. H., Jeong, H., Hur, C. G. and Kim, J. J. (2004). The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nature Biotechnology* **22** (10): 1275-1281.
- Hung, M. N. and Lee, B. H. (1998). Cloning and expression of beta-galactosidase gene from *Bifidobacterium infantis* into *Escherichia coli*. *Biotechnology Letters* **20** (7): 659-662.
- Hungate, R. E. (1957). Microorganisms in the rumen of cattle fed a constant ration. *Canadian Journal of Microbiology* **3** (2): 289-311.
- Iiyama, K., Lam, T. B. T. and Stone, B. A. (1994). Covalent Cross-Links in the Cell-Wall. *Plant Physiology* **104** (2): 315-320.
- Inacio, J. M., Correia, I. L. and de Sa-Nogueiral, I. (2008). Two distinct arabinofuranosidases contribute to arabino-oligosaccharide degradation in *Bacillus subtilis*. *Microbiology-Sgm* **154**: 2719-2729.
- Ishihama, Y., Schmidt, T., Rappsilber, J., Mann, M., Hartl, F. U., Kerner, M. J. and Frishman, D. (2008) Protein abundance profiling of the *Escherichia coli* cytosol. *Bmc Genomics* **9** 10.1186/1471-2164-9-102.
- Ishii, T. (1997). Structure and functions of feruloylated polysaccharides. *Plant Science* **127** (2): 111-127.
- Jacquet, G., Pollet, B. and Lapierre, C. (1995). New ether-linked ferulic acid-coniferyl alcohol dimers identified in grass straws. *Journal of Agricultural and Food Chemistry* **43** (10): 2746-2751.
- Janecek, S., Svensson, B. and MacGregor, E. A. (2003). Relation between domain evolution, specificity, and taxonomy of the alpha-amylase family members containing a C-terminal starch-binding domain. *European Journal of Biochemistry* **270** (4): 635-645.
- Jansen, R., Bussemaker, H. J. and Gerstein, M. (2003). Revisiting the codon adaptation index from a whole-genome perspective: analyzing the relationship between gene expression and codon occurrence in yeast using a variety of models. *Nucleic Acids Research* **31** (8): 2242-2251.
- Janssen, P. H. and Kirs, M. (2008). Structure of the archaeal community of the rumen. *Applied and Environmental Microbiology* **74** (12): 3619-3625.
- Jarvis, M. C. (1984). Structure and properties of pectin gels in plant cell-walls. *Plant Cell and Environment* **7** (3): 153-164.
- Jiang, L., He, L. and Fountoulakis, M. (2004). Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *Journal of Chromatography A* **1023** (2): 317-320.
- Jindou, S., Borovok, I., Rincon, M. T., Flint, H. J., Antonopoulos, D. A., Berg, M. E., White, B. A., Bayer, E. A. and Lamed, R. (2006). Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens*. *Journal of Bacteriology* **188** (22): 7971-7976.

- Jing, H. Q., Cockburn, D., Zhang, Q. X. and Clarke, A. J. (2009). Production and purification of the isolated family 2a carbohydrate-binding module from *Cellulomonas fimi*. *Protein Expression and Purification* **64** (1): 63-68.
- Jones, L., Ennos, A. R. and Turner, S. R. (2001). Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. *Plant Journal* **26** (2): 205-216.
- Jun, H. S., Qi, M., Gong, J., Egbosimba, E. E. and Forsberg, C. W. (2007). Outer membrane proteins of *Fibrobacter succinogenes* with potential roles in adhesion to cellulose and in cellulose digestion. *Journal of Bacteriology* **189**: 6806-6815.
- Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Science* **12** (8): 1652-1662.
- Jung, H. J. G., Ralph, J. and Hatfield, R. D. (1991). Degradability of phenolic acid hemicellulose esters - a model system. *Journal of the Science of Food and Agriculture* **56** (4): 469-478.
- Jung, H. J. G., Valdez, F. R., Hatfield, R. D. and Blanchette, R. A. (1992). Cell-wall composition and degradability of forage stems following chemical and biological delignification. *Journal of the Science of Food and Agriculture* **58** (3): 347-355.
- Jung, H. J. G. (2003). Maize stem tissues: ferulate deposition in developing internode cell walls. *Phytochemistry* **63** (5): 543-549.
- Jungblut, P. R., Schiele, F., Zimny-Arndt, U., Ackermann, R., Schmid, M., Lange, S., Stein, R. and Pleissner, K. P. (2010). *Helicobacter pylori* proteomics by 2-DE/MS, 1-DE-LC/MS and functional data mining. *Proteomics* **10** (2): 182-193.
- Kabel, M. A., van den Borne, H., Vincken, J. P., Voragen, A. G. J. and Schols, H. A. (2007). Structural differences of xylans affect their interaction with cellulose. *Carbohydrate Polymers* **69** (1): 94-105.
- Kamisaka, S., Takeda, S., Takahashi, K. and Shibata, K. (1990). Diferulic and ferulic acid in the cell-wall of *Avena* coleoptiles - Their relationships to mechanical properties of the cell-wall. *Physiologia Plantarum* **78** (1): 1-7.
- Kanaya, S., Yamada, Y., Kudo, Y. and Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene* **238** (1): 143-155.
- Karas, M. and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. *Analytical Chemistry* **60** (20): 2299-301.
- Kasperowicz, A. (1994). Comparison of utilization of pectins from various sources by pure cultures of pectinolytic rumen bacteria and mixed cultures of rumen microorganisms. *Acta Microbiologica Polonica* **43** (1): 47-56.
- Kataeva, I. A., Uversky, V. N., Brewer, J. M., Schubot, F., Rose, J. P., Wang, B. C. and Ljungdahl, L. G. (2004). Interactions between immunoglobulin-like and catalytic modules in *Clostridium thermocellum* cellulosomal cellobiohydrolase CbhA. *Protein Engineering Design & Selection* **17** (11): 759-769.
- Katsuraya, K., Okuyama, K., Hatanaka, K., Oshima, R., Sato, T. and Matsuzaki, K. (2003). Constitution of konjac glucomannan: chemical analysis and C-13 NMR spectroscopy. *Carbohydrate Polymers* **53** (2): 183-189.
- Kelly, W. J., Leahy, S., Altermann, E., Yeoman, C. J., Dunne, J. C., Kong, Z., Pacheco, D. M., Li, D., Noel, S., Moon, C. D., Cookson, A. and Attwood, G. T. (2010) The glycobiome of the rumen bacterium *Butyrivibrio proteoclasticus* B316^T highlights adaptation to a polysaccharide-rich environment. *PLoS One* **8**, (9) doi:10.1371/journal.pone.0011942.

- Kim, J. B., Olek, A. T. and Carpita, N. C. (2000). Cell-wall and membrane-associated exo-beta-D-glucanases from developing maize seedlings. *Plant Physiology* **123** (2): 471-485.
- Kim, Y. S., Singh, A. P., Wi, S. G., Myung, K. H., Karita, S. and Ohmiya, K. (2001). Cellulosome-like structures in ruminal cellulolytic bacterium *Ruminococcus albus* F-40 as revealed by electron microscopy. *Asian-Australasian Journal of Animal Sciences* **14** (10): 1429-1433.
- Klose, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis of mouse tissues - novel approach to testing for induced point mutations in mammals. *Humangenetik* **26** (3): 231-243.
- Kluskens, L. D., van Alebeek, G., Voragen, A. G. J., de Vos, W. M. and van der Oost, J. (2003). Molecular and biochemical characterization of the thermoactive family 1 pectate lyase from the hyperthermophilic bacterium *Thermotoga maritima*. *Biochemical Journal* **370**: 651-659.
- Knight, C. G., Kassen, R., Hebestreit, H. and Rainey, P. B. (2004). Global analysis of predicted proteomes: Functional adaptation of physical properties. *Proceedings of the National Academy of Sciences of the United States of America* **101** (22): 8390-8395.
- Koike, S., Pan, J., Kobayashi, Y. and Tanaka, K. (2003a). Kinetics of *in sacco* fiber attachment of representative ruminal cellulolytic bacteria monitored by competitive PCR. *Journal of Dairy Science* **86** (4): 1429-1435.
- Koike, S., Yoshitani, S., Kobayashi, Y. and Tanaka, K. (2003b). Phylogenetic analysis of fiber-associated rumen bacterial community and PCR detection of uncultured bacteria. *FEMS Microbiology Letters* **229** (1): 23-30.
- Kondo, T., Mizuno, K. and Kato, T. (1990). Cell-wall bound para-coumaric and ferulic acids in Italian ryegrass. *Canadian Journal of Plant Science* **70** (2): 495-499.
- Kong, Z. H. (2007). Microarray analysis of *Clostridium proteoclasticus* genes involved in hemicellulose degradation. M.Sc Thesis. School of Biological Sciences. University of Auckland.
- Kopečný, J., Logar, R. M. and Kobayashi, Y. (2001). Phenotypic and genetic data supporting reclassification of *Butyrivibrio fibrisolvens* isolates. *Folia Microbiologica* **46** (1): 45-48.
- Kosugi, A., Murashima, K. and Doi, R. H. (2002). Characterization of two noncellulosomal subunits, ArfA and BgaA, from *Clostridium cellulovorans* that cooperate with the cellulosome in plant cell wall degradation. *Journal of Bacteriology* **184** (24): 6859-6865.
- Kotiranta, A., Haapasalo, M., Lounatmaa, K. and Kari, K. (1995). Crystalline surface protein of *Peptostreptococcus anaerobius*. *Microbiology* **141** (5): 1065-73.
- Krause, D. O., Dalrymple, B. P., Smith, W. J., Mackie, R. I. and McSweeney, C. S. (1999). 16S rDNA sequencing of *Ruminococcus albus* and *Ruminococcus flavefaciens*: design of a signature probe and its application in adult sheep. *Microbiology-Sgm* **145**: 1797-1807.
- Krause, D. O., Bunch, R. J., Dalrymple, B. D., Gobius, K. S., Smith, W. J. M., Xue, G. P. and McSweeney, C. S. (2001). Expression of a modified *Neocallimastix patriciarum* xylanase in *Butyrivibrio fibrisolvens* digests more fibre but cannot effectively compete with highly fibrolytic bacteria in the rumen. *Journal of Applied Microbiology* **90** (3): 388-396.
- Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E. L. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305** (3): 567-580.

- Kulkarni, N., A., S. and M., R. (1999). Molecular and biotechnological aspects of xylanases. Review. *FEMS Microbiol Rev.* **23** (4): 411-56.
- Kung, L., Treacher, R. J., Nauman, G. A., Smagala, A. M., Endres, K. M. and Cohen, M. A. (2000). The effect of treating forages with fibrolytic enzymes on its nutritive value and lactation performance of dairy cows. *Journal of Dairy Science* **83** (1): 115-122.
- Kuntumalla, S., Braisted, J. C., Huang, S. T., Parmar, P. P., Clark, D. J., Alami, H., Zhang, Q. S., Donohue-Rolfe, A., Tzipori, S., Fleischmann, R. D., Peterson, S. N. and Pieper, R. (2009) Comparison of two label-free global quantitation methods, APEX and 2D gel electrophoresis, applied to the *Shigella dysenteriae* proteome. *Proteome Science* **7** 10.1186/1477-5956-7-22.
- Lam, T. B. T., Iiyama, K. and Stone, B. A. (1992). Cinnamic acid bridges between cell-wall polymers in wheat and *Phalaris* internodes. *Phytochemistry* **31** (4): 1179-1183.
- Lam, T. B. T., Kadoya, K. and Iiyama, K. (2001). Bonding of hydroxycinnamic acids to lignin: ferulic and p-coumaric acids are predominantly linked at the benzyl position of lignin, not the beta-position, in grass cell walls. *Phytochemistry* **57** (6): 987-992.
- Lam, T. B. T., Iiyama, K. and Stone, B. A. (2003). Hot alkali-labile linkages in the walls of the forage grass *Phalaris aquatica* and *Lolium perenne* and their relation to *in vitro* wall digestibility. *Phytochemistry* **64** (2): 603-607.
- Lamed, R., Setter, E. and Bayer, E. A. (1983). Characterization of a cellulose-binding, cellulase-containing complex in *Clostridium thermocellum*. *Journal of Bacteriology* **156** (2): 828-836.
- Lamport, D. T. A. (1986). Roles for peroxidases in cell-wall genesis. *Molecular and Physiological Aspects of Plant Peroxidases*. Eds. Greppin, H., Penel, C. and Gaspar, T. Geneva, University of Geneva Press: 199-208.
- Lange, L., Wu, W., Aubert, D., Landvik, S., Schnorr, K. M. and Clausen, I. G. (2002). Cloning and sequences of microbial cellobiohydrolase I for use in DNA shuffling, ethanol prodn., transgenic plants, and detergents. Novozymes, USA. 2003000941.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J. and Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* **23** (21): 2947-2948.
- Larue, R., Yu, Z. T., Parisi, V. A., Egan, A. R. and Morrison, M. (2005). Novel microbial diversity adherent to plant biomass in the herbivore gastrointestinal tract, as revealed by ribosomal intergenic spacer analysis and rrs gene sequencing. *Environmental Microbiology* **7** (4): 530-543.
- Lazaridou, A., Chornick, T. and Izydorczyk, M. S. (2008). Variations in morphology and composition of barley endosperm cell walls. *Journal of the Science of Food and Agriculture* **88** (13): 2388-2399.
- Lee, J. W., Lee, S. Y., Song, H. and Yoo, J. S. (2006). The proteome of *Mannheimia succiniciproducens*, a capnophilic rumen bacterium. *Proteomics* **6** (12): 3550-3566.
- Lee, J. W. and Lee, S. Y. (2010). Proteome-based physiological analysis of the metabolically engineered succinic acid producer *Mannheimia succiniciproducens* LPK7. *Bioprocess and Biosystems Engineering* **33** (1): 97-107.
- Lee, K., Lee, H. G., Pi, K. and Choi, Y. J. (2008). Effect of low pH on protein expression by the probiotic bacterium *Lactobacillus reuteri*. *Proteomics* **8** (8): 1624-1630.

- Lee, S. F. and Forsberg, C. W. (1987). Purification and characterization of an alpha-L-arabinofuranosidase from *Clostridium acetobutylicum* ATCC 824. *Canadian Journal of Microbiology* **33** (11): 1011-1016.
- Lee, Y. E., Lowe, S. E. and Zeikus, J. G. (1993). Regulation and characterization of xylanolytic enzymes of *Thermoanaerobacterium saccharolyticum* B6A-RI. *Applied and Environmental Microbiology* **59** (3): 763-771.
- Leskela, S., Wahlstrom, E., Kontinen, V. P. and Sarvas, M. (1999). Lipid modification of prelipoproteins is dispensable for growth but essential for efficient protein secretion in *Bacillus subtilis*: characterization of the lgt gene. *Molecular Microbiology* **31** (4): 1075-1085.
- Levasseur, A., Navarro, D., Punt, P. J., Belaich, J. P., Asther, M. and Record, E. (2005). Construction of engineered bifunctional enzymes and their overproduction in *Aspergillus niger* for improved enzymatic tools to degrade agricultural by-products. *Applied and Environmental Microbiology* **71** (12): 8132-8140.
- Levene, H. (1960). Robust tests for equality of variances. *Contributions to probability and statistics*. Eds. I. Olkin, S. G. G. W. and Hoeffding, W. G. M., and H. B. Mann. Stanford, California., Stanford University Press: 278-292.
- Li, K. C., Azadi, P., Collins, R., Tolan, J., Kim, J. S. and Eriksson, K. E. L. (2000). Relationships between activities of xylanases and xylan structures. *Enzyme and Microbial Technology* **27** (1-2): 89-94.
- Lin, L. L., Rumbak, E., Zappe, H., Thomson, J. A. and Woods, D. R. (1990). Cloning, sequencing and analysis of expression of a *Butyrivibrio fibrisolvens* gene encoding a beta-glucosidase. *Journal of General Microbiology* **136**: 1567-1576.
- Lin, L. L. and Thomson, J. A. (1991a). Cloning, sequencing and expression of a gene encoding a 73 kDa xylanase enzyme from the rumen anaerobe *Butyrivibrio fibrisolvens* H17c. *Molecular and General Genetics* **228** (1-2): 55-61.
- Lin, L. L. and Thomson, J. A. (1991b). An analysis of the extracellular xylanases and cellulases of *Butyrivibrio fibrisolvens* H17c. *FEMS Microbiology Letters* **84** (2): 197-204.
- Linder, A., Bergman, R., Bodin, A. and Gatenholm, P. (2003). Mechanism of assembly of xylan onto cellulose surfaces. *Langmuir* **19** (12): 5072-5077.
- Liu, H., Sadygov, R. G. and Yates, J. R. r. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Analytical Chemistry* **76** (14): 4193-4201.
- Lortal, S., Vanheijenoort, J., Gruber, K. and Sleytr, U. B. (1992). S-layer of *Lactobacillus helveticus* ATCC-12046 - Isolation, chemical characterization and re-formation after extraction with lithium chloride. *Journal of General Microbiology* **138**: 611-618.
- Lynd, L. R., Weimer, P. J., van Zyl, W. H. and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. Review. *Microbiol Mol Biol Rev.* **66** (3): 506-77.
- Lyons, P. C., Hipskind, J., Vincent, J. R. and Nicholson, R. L. (1993). Phenylpropanoid dissemination in maize resistant or susceptible to *Helminthosporium maydis*. *Maydica* **38** (3): 175-181.
- MacAdam, J. W. and Grabber, J. H. (2002). Relationship of growth cessation with the formation of diferulate cross-links and p-coumaroylated lignins in tall fescue leaf blades. *Planta* **215** (5): 785-793.
- Mackenzie, C. R., Bilous, D., Schneider, H. and Johnson, K. G. (1987). Induction of cellulolytic and xylanolytic enzyme-systems in *Streptomyces* spp. *Applied and Environmental Microbiology* **53** (12): 2835-2839.

- Macrae, J. C. and Ulyatt, M. J. (1974). Quantitative digestion of fresh herbage by sheep. 2. Sites of digestion of some nitrogenous constituents. *Journal of Agricultural Science* **82**: 309-319.
- Maglione, G., Russell, J. B. and Wilson, D. B. (1997). Kinetics of cellulose digestion by *Fibrobacter succinogenes* S85. *Applied and Environmental Microbiology* **63** (2): 665-669.
- Malle, D., Itoh, T., Hashimoto, W., Murata, K., Utsumi, S. and Mikami, B. (2006). Overexpression, purification and preliminary X-ray analysis of pullulanase from *Bacillus subtilis* strain 168. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications* **62**: 381-384.
- Mann, H. B. and Whitney, D. R. (1947). On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics* **18**: 50-60.
- Mannarelli, B. M., Evans, S. and Lee, D. (1990). Cloning, sequencing, and expression of a xylanase gene from the anaerobic ruminal bacterium *Butyrivibrio fibrisolvens*. *Journal of Bacteriology* **172** (8): 4247-4254.
- Marounek, M. and Duskova, D. (1999). Metabolism of pectin in rumen bacteria *Butyrivibrio fibrisolvens* and *Prevotella ruminicola*. *Letters in Applied Microbiology* **29** (6): 429-433.
- Martens, E. C., Chiang, H. C. and Gordon, J. I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host & Microbe* **4** (5): 447-457.
- Martens, E. C., Koropatkin, N. M., Smith, T. J. and Gordon, J. I. (2009a). Complex glycan catabolism by the human gut microbiota: The *Bacteroidetes* Sus-like paradigm. *Journal of Biological Chemistry* **284** (37): 24673-24677.
- Martens, E. C., Roth, R., Heuser, J. E. and Gordon, J. I. (2009b). Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *Journal of Biological Chemistry* **284** (27): 18445-18457.
- Matte, A. and Forsberg, C. W. (1992). Purification, characterization, and mode of action of endoxylanases 1 and 2 from *Fibrobacter succinogenes* S85. *Applied and Environmental Microbiology* **58** (1): 157-168.
- Matuschek, M., Burchhardt, G., Sahm, K. and Bahl, H. (1994). Pullulanase of *Thermoanaerobacterium thermosulfurigenes* Em1 (*Clostridium thermosulfurogenes*) - Molecular analysis of the gene, composite structure of the enzyme, and a common model for its attachment to the cell-surface. *Journal of Bacteriology* **176** (11): 3295-3302.
- Matuschek, M., Sahm, K., Zibat, A. and Bahl, H. (1996). Characterization of genes from *Thermoanaerobacterium thermosulfurigenes* EM1 that encode two glycosyl hydrolases with conserved S-layer-like domains. *Molecular and General Genetics* **252** (4): 493-496.
- May, A., Pusztahelyi, T., Hoffmann, N., Fischer, R. J. and Bahl, H. (2006). Mutagenesis of conserved charged amino acids in SLH domains of *Thermoanaerobacterium thermosulfurigenes* EM1 affects attachment to cell-wall sacculi. *Archives of Microbiology* **185** (4): 263-269.
- McAllister, T. A., Bae, H. D., Jones, G. A. and Cheng, K. J. (1994). Microbial attachment and feed digestion in the rumen. *Journal of Animal Science* **72** (11): 3004-3018.
- McCoy, E. C., Doyle, D., Burda, K., Corbeil, L. B. and Winter, A. J. (1975). Superficial antigens of *Campylobacter (Vibrio) Fetus* - Characterization of an anti-phagocytic component. *Infection and Immunity* **11** (3): 517-525.

- McLean, B. W., Bray, M. R., Boraston, A. B., Gilkes, N. R., Haynes, C. A. and Kilburn, D. G. (2000). Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Engineering* **13** (11): 801-809.
- McSweeney, C. S., Dulieu, A. and Bunch, R. (1998). *Butyrivibrio* spp and other xylanolytic microorganisms from the rumen have cinnamoyl esterase activity. *Anaerobe* **4** (1): 57-65.
- Meier, H. and Reid, J. S. G. (1982). *Reserve polysaccharides other than starch in higher plants*. Berlin, Springer-Verlag.
- Meleth, S., Deshane, J. and Kim, H. (2005). The case for well-conducted experiments to validate statistical protocols for 2D gels: different pre-processing = different lists of significant proteins. *Bmc Biotechnology* **5**.
- Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M. and Fouet, A. (2000). Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. *EMBO Journal* **19** (17): 4473-4484.
- Michel, G., Barbeyron, T., Kloareg, B. and Czjzek, M. (2009). The Family 6 carbohydrate-binding modules have coevolved with their appended catalytic modules toward similar substrate specificity. *Glycobiology* **19** (6): 615-623.
- Mielenz, J. R. (2001). Ethanol production from biomass: technology and commercialization status. *Current Opinion in Microbiology* **4** (3): 324-329.
- Miller, T. L. and Jenesel, S. E. (1979). Enzymology of butyrate formation by *Butyrivibrio fibrisolvens*. *Journal of Bacteriology* **138** (1): 99-104.
- Minato, H., Mitsumori, M. and Cheng, K. J. (1993). Attachment of microorganisms to solid substrate in the rumen. *Genetics, Biochemistry and Ecology of Lignocellulose Degradation*. Eds. Shimada, K. and *al., e.*, Uni Publishers, Tokyo.: 139-145.
- Miron, J. and Benghedalia, D. (1993). Digestion of cell-wall monosaccharides of ryegrass and alfalfa hays by the ruminal bacteria *Fibrobacter succinogenes* and *Butyrivibrio fibrisolvens*. *Canadian Journal of Microbiology* **39** (8): 780-786.
- Miron, J., Duncan, S. H. and Stewart, C. S. (1994). Interactions between rumen bacterial strains during the degradation and utilization of the monosaccharides of barley straw cell-walls. *Journal of Applied Bacteriology* **76** (3): 282-287.
- Miron, J., Morag, E., Bayer, E. A., Lamed, R. and Ben-Ghedalia, D. (1998). An adhesion-defective mutant of *Ruminococcus albus* SY3 is impaired in its capability to degrade cellulose. *Journal of Applied Microbiology* **84** (2): 249-254.
- Mitchell, R. A. C., Dupree, P. and Shewry, P. R. (2007). A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiology* **144** (1): 43-53.
- Miyazaki, K., Martin, J. C., Marinsek-Logar, R. and Flint, H. J. (1997). Degradation and utilization of xylans by the rumen anaerobe *Prevotella bryantii* (formerly *P. ruminicola* subsp *brevis*) B(1)4. *Anaerobe* **3** (6): 373-381.
- Miyazaki, K., Miyamoto, H., Mercer, D. K., Hirase, T., Martin, J. C., Kojima, Y. and Flint, H. J. (2003). Involvement of the multidomain regulatory protein XynR in positive control of xylanase gene expression in the ruminal anaerobe *Prevotella bryantii* B(1)4. *Journal of Bacteriology* **185** (7): 2219-2226.
- Miyazaki, K., Hirase, T., Kojima, Y. and Flint, H. J. (2005). Medium- to large-sized xylo-oligosaccharides are responsible for xylanase induction in *Prevotella bryantii* B(1)4. *Microbiology-Sgm* **151**: 4121-4125.

- Monedero, V., Yebra, M. J., Poncet, S. and Deutscher, J. (2008). Maltose transport in *Lactobacillus casei* and its regulation by inducer exclusion. *Research in Microbiology* **159** (2): 94-102.
- Moon, C. D., Pacheco, D. M., Kelly, W. J., Leahy, S. C., Li, D., Kopečný, J. and Attwood, G. T. (2008). Reclassification of *Clostridium proteoclasticum* as *Butyrivibrio proteoclasticus* comb. nov., a butyrate producing ruminal bacterium. *International Journal of Systematic and Evolutionary Microbiology*. **58** (9): 2041-2045.
- Morales, P., Sendra, J. M. and PerezGonzalez, J. A. (1995). Purification and characterization of an arabinofuranosidase from *Bacillus polymyxa* expressed in *Bacillus subtilis*. *Applied Microbiology and Biotechnology* **44** (1-2): 112-117.
- Morrison, I. M. (1991). Changes in the biodegradability of ryegrass and legume fibers by chemical and biological pretreatments. *Journal of the Science of Food and Agriculture* **54** (4): 521-533.
- Morrison, M., Nelson, K. E., Cann, I. K. O., Forsberg, C. W., Mackie, R. I., Russell, J. B., White, B. A., Amava, K., Cheng, B., Qi, M., Jun, H., Mulligan, S., Tran, K., Carty, H. A., Khouri, H., Nelson, W., Daugherty, S. and Fraser, C. M. (2003). The *Fibrobacter succinogenes* strain S85 genome sequencing project. 3rd ASM-TIGR Conference of Microbial Genomes.
- Mosoni, P. and Gaillard-Martinie, B. (2001). Characterization of a spontaneous adhesion-defective mutant of *Ruminococcus albus* strain 20. *Archives of Microbiology* **176** (1-2): 52-61.
- Mrazek, J., Tepsic, K., Avgustin, G. and Kopečný, J. (2006). Diet-dependent shifts in ruminal butyrate-producing bacteria. *Folia Microbiologica* **51** (4): 294-298.
- Mueller-Harvey, I., Hartley, R. D., Harris, P. J. and Curzon, E. H. (1986). Linkage of para-coumaroyl and feruloyl groups to cell-wall polysaccharides of barley straw. *Carbohydrate Research* **148** (1): 71-85.
- Murashima, K., Kosugi, A. and Doi, R. H. (2002a). Synergistic effects on crystalline cellulose degradation between cellulosomal cellulases from *Clostridium cellulovorans*. *Journal of Bacteriology* **184** (18): 5088-5095.
- Murashima, K., Kosugi, A. and Doi, R. H. (2002b). Determination of subunit composition of *Clostridium cellulovorans* cellulosomes that degrade plant cell walls. *Applied and Environmental Microbiology* **68** (4): 1610-1615.
- Murashima, K., Kosugi, A. and Doi, R. H. (2003). Synergistic effects of cellulosomal xylanase and cellulases from *Clostridium cellulovorans* on plant cell wall degradation. *Journal of Bacteriology* **185** (5): 1518-1524.
- Nair, S. U., Singhal, R. S. and Kamat, M. Y. (2007). Induction of pullulanase production in *Bacillus cereus* FDA-13. *Bioresource Technology* **98** (4): 856-859.
- Nakamura, A., Furuta, H., Maeda, H., Takao, T. and Nagamatsu, Y. (2002a). Analysis of the molecular construction of xylogalacturonan isolated from soluble soybean polysaccharides. *Bioscience Biotechnology and Biochemistry* **66** (5): 1155-1158.
- Nakamura, A., Furuta, H., Maeda, H., Takao, T. and Nagamatsu, Y. (2002b). Structural studies by stepwise enzymatic degradation of the main backbone of soybean soluble polysaccharides consisting of galacturonan and rhamnogalacturonan. *Bioscience Biotechnology and Biochemistry* **66** (6): 1301-1313.
- Nakamura, M., Nagamine, T., Takenaka, A., Aminov, R. I., Ogata, K., Tajima, K., Matsui, H., Benno, Y. and Itabashi, H. (2002c). Molecular cloning, nucleotide sequence and characteristics of a xylanase gene (*xynA*) from *Ruminococcus albus* 7. *Animal Science Journal* **73** (5): 347-352.

- Nesvizhskii, A. I., Keller, A., Kolker, E. and Aebersold, R. (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry* **75** (17): 4646-4658.
- Nocek, J. E. and Russell, J. B. (1988). Protein and energy as an integrated system - Relationship of ruminal protein and carbohydrate availability to microbial synthesis and milk production. *Journal of Dairy Science* **71** (8): 2070-2107.
- Nong, G., Chow, V., Rice, J. D., St. John, F. and Preston, J. F. (2005). An aldouronic acid-utilization operon in a *Paenibacillus* sp encodes an alpha-glucuronidase with activity on aldouronic acids generated by acid and enzyme mediated digestion of methylglucuronoxylan. *Abstracts of the General Meeting of the American Society for Microbiology* **105**: 433.
- Notenboom, V., Boraston, A. B., Chiu, P., Freelove, A. C. J., Kilburn, D. G. and Rose, D. R. (2001a). Recognition of cello-oligosaccharides by a Family 17 carbohydrate-binding module: An X-ray crystallographic, thermodynamic and mutagenic study. *Journal of Molecular Biology* **314** (4): 797-806.
- Notenboom, V., Boraston, A. B., Kilburn, D. G. and Rose, D. R. (2001b). Crystal structures of the Family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A in native and ligand-bound forms. *Biochemistry* **40** (21): 6248-6256.
- O' Farrell, P. H. (1975). High-resolution 2-dimensional electrophoresis of proteins. *Journal of Biological Chemistry* **250** (10): 4007-4021.
- O'Neil, M. A. and York, W. S. (2003). The composition and structure of plant primary cell walls. *The Plant Cell Wall*. Eds. Rose, J. K. C., CRC Press: 1-54.
- O'Neill, M., Albersheim, P. and Darvill, A. (1990). The polysaccharides of primary cell walls. *Methods in Plant Biochemistry*. Eds. Dey, P. M. London, Academic Press. **2**.
- O'Sullivan, A. C. (1997). Cellulose: The structure slowly unravels. *Cellulose* **4**: 173-207.
- Ohara, H., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (2000a). Characterization of the cellulolytic complex (cellulosome) from *Ruminococcus albus*. *Bioscience Biotechnology and Biochemistry* **64** (2): 254-260.
- Ohara, H., Noguchi, J., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (2000b). Sequence of *egV* and properties of EgV, a *Ruminococcus albus* endoglucanase containing a dockerin domain. *Bioscience Biotechnology and Biochemistry* **64** (1): 80-88.
- Ohlmeier S, S. C., Hecker M. (2000). Alkaline proteins of *Bacillus subtilis*: first steps towards a two-dimensional alkaline master gel. *Electrophoresis* **21** (17): 3701-9.
- Old, W. M., Meyer-Arendt, K., Aveline-Wolf, L. D., Pierce, K. G., Mendoza, A. M., Sevinsky, J. R., Resing, K. A. and Ahn, N. G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Molecular & Cellular Proteomics* **4**: 1487-1502.
- Ong, S. E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A. and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Molecular & Cellular Proteomics* **1** (5): 376-386.
- Orpin, C. G., Mathiesen, S. D., Greenwood, Y. and Blix, A. S. (1985). Seasonal changes in the ruminal microflora of the high-arctic Svalbard reindeer (*Rangifer tarandus platyrhynchus*). *Applied and Environmental Microbiology* **50** (1): 144-151.
- Orpin, C. G. and Joblin, K. N. (1997). The rumen anaerobic fungi. *The Rumen Microbial Ecosystem*. Eds. Hobson, P. N. and Stewart, C. S. London, Blackie Academic and Professional Publishers.: 140-195.

- Osborne, J. M. and Dehority, B. A. (1989). Synergism in degradation and utilization of intact forage cellulose, hemicellulose, and pectin by 3 pure cultures of ruminal bacteria. *Applied and Environmental Microbiology* **55** (9): 2247-2250.
- Paillard, D., McKain, N., Chaudhary, L. C., Walker, N. D., Pizette, F., Koppova, I., McEwan, N. R., Kopecny, J., Vercoe, P. E., Louis, P. and Wallace, R. J. (2007a). Relation between phylogenetic position, lipid metabolism and butyrate production by different *Butyrivibrio*-like bacteria from the rumen. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **91** (4): 417-422.
- Paillard, D., McKain, N., Rincon, M. T., Shingfield, K. J., Givens, D. I. and Wallace, R. J. (2007b). Quantification of ruminal *Clostridium proteoclasticum* by real-time PCR using a molecular beacon approach. *Journal of Applied Microbiology* **103** (4): 1251-1261.
- Pappin, D. J. C., Hojrup, P. and Bleasby, A. J. (1993). Rapid identification of proteins by peptide mass fingerprinting. *Current Biology* **3** (6): 327-332.
- Paradis, F. W., Zhu, H., Krell, P. J., Phillips, J. P. and Forsberg, C. W. (1993). The *xynC* gene from *Fibrobacter succinogenes* S85 codes for a xylanase with 2 similar catalytic domains. *Journal of Bacteriology* **175** (23): 7666-7672.
- Park, J. S., Matano, Y. and Doi, R. H. (2001). Cohesin-dockerin interactions of cellulosomal subunits of *Clostridium cellulovorans*. *Journal of Bacteriology* **183** (18): 5431-5435.
- Pauly, M., Albersheim, P., Darvill, A. and York, W. S. (1999). Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *Plant Journal* **20** (6): 629-639.
- Pegden, R. S., Larson, M. A., Grant, R. J. and Morrison, M. (1998). Adherence of the Gram-positive bacterium *Ruminococcus albus* to cellulose and identification of a novel form of cellulose-binding protein which belongs to the Pil family of proteins. *Journal of Bacteriology* **180** (22): 5921-5927.
- Perkins, D. N., Pappin, D. J. C., Creasy, D. M. and Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20** (18): 3551-3567.
- Peters, A., Lebzien, P., Meyer, U., Borchert, U., Bulang, M. and Flachowsky, G. (2010). Effect of exogenous fibrolytic enzymes on ruminal fermentation and nutrient digestion in dairy cows. *Archives of Animal Nutrition* **64** (3): 221-237.
- Pettipfer, G. L. and Latham, M. J. (1979). Characteristics of enzymes produced by *Ruminococcus flavefaciens* which degrade plant cell-walls. *Journal of General Microbiology* **110** (1): 21-27.
- Pinos-Rodriguez, J. M., Gonzalez, S. S., Mendoza, G. D., Barcena, R., Cobos, M. A., Hernandez, A. and Ortega, M. E. (2002). Effect of exogenous fibrolytic enzyme on ruminal fermentation and digestibility of alfalfa and rye-grass hay fed to lambs. *Journal of Animal Science* **80** (11): 3016-3020.
- Pires, V. M. R., Henshaw, J. L., Prates, J. A. M., Bolam, D. N., Ferreira, L. M. A., Fontes, C., Henrissat, B., Planas, A., Gilbert, H. J. and Czjzek, M. (2004). The crystal structure of the Family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5A in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *Journal of Biological Chemistry* **279** (20): 21560-21568.
- Pizzi, A. and Eaton, N. (1985). The structure of cellulose by conformational analysis. 2. The cellulose polymer-chain. *Journal of Macromolecular Science-Chemistry* **A22** (1): 105-137.

- Potrykus, J., White, R. L. and Bearne, S. L. (2008). Proteomic investigation of amino acid catabolism in the indigenous gut anaerobe *Fusobacterium varium*. *Proteomics* **8** (13): 2691-2703.
- Qi, X. Y., Behrens, B. X., West, P. R. and Mort, A. J. (1995). Solubilization and partial characterization of extensin fragments from cell-walls of cotton suspension-cultures - evidence for a covalent cross-link between extensin and pectin. *Plant Physiology* **108** (4): 1691-1701.
- Ralph, J., Quideau, S., Grabber, J. H. and Hatfield, R. D. (1994). Identification and synthesis of new ferulic acid dehydrodimers present in grass cell-walls. *Journal of the Chemical Society-Perkin Transactions 1* (23): 3485-3498.
- Ralph, J., Grabber, J. H. and Hatfield, R. D. (1995). Lignin-ferulate cross-links in grasses - active incorporation of ferulate polysaccharide esters into ryegrass lignins. *Carbohydrate Research* **275** (1): 167-178.
- Ralph, J., Lapierre, C., Marita, J. M., Kim, H., Lu, F. C., Hatfield, R. D., Ralph, S., Chapple, C., Franke, R., Hemm, M. R., Van Doorselaere, J., Sederoff, R. R., O'Malley, D. M., Scott, J. T., MacKay, J. J., Yahiaoui, N., Boudet, A. M., Pean, M., Pilate, G., Jouanin, L. and Boerjan, W. (2001). Elucidation of new structures in lignins of CAD- and COMT-deficient plants by NMR. *Phytochemistry* **57** (6): 993-1003.
- Ramsay, A. G., Scott, K. P., Martin, J. C., Rincon, M. T. and Flint, H. J. (2006). Cell-associated alpha-amylases of butyrate-producing *Firmicute* bacteria from the human colon. *Microbiology-Sgm* **152**: 3281-3290.
- Raposo, M. P., Inacio, J. M., Mota, L. J. and de Sa-Nogueira, I. (2004). Transcriptional regulation of genes encoding arabinan-degrading enzymes in *Bacillus subtilis*. *Journal of Bacteriology* **186** (5): 1287-1296.
- Rappsilber, J., Ryder, U., Lamond, A. I. and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Research* **12** (8): 1231-1245.
- Reeves, A. R., Wang, G. R. and Salyers, A. A. (1997). Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *Journal of Bacteriology* **179** (3): 643-649.
- Reilly, K. and Attwood, G. T. (1998). Detection of *Clostridium proteoclasticum* and closely related strains in the rumen by competitive PCR. *Applied and Environmental Microbiology* **64** (3): 907-913.
- Reilly, K., Carruthers, V. R. and Attwood, G. T. (2002). Design and use of 16S ribosomal DNA-directed primers in competitive PCRs to enumerate proteolytic bacteria in the rumen. *Microbial Ecology* **43** (2): 259-270.
- Reilly, P. J. (1981). Xylanases: Structure and Function. *Trends in the biology of fermentations*. Eds. Hollaender, A. New York, Plenum Publishing Corp.: 11-129.
- Resing, K. A. and Ahn, N. G. (2004). Protein identification by in-gel digestion and mass spectrometry. *Proteome Analysis: Interpreting the Genome*: 163-182.
- Ridley, B. L., O'Neill, M. A. and Mohnen, D. A. (2001). Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **57** (6): 929-967.
- Rincon, M. T., McCrae, S. I., Kirby, J., Scott, K. P. and Flint, H. J. (2001). EndB, a multidomain family 44 cellulase from *Ruminococcus flavefaciens* 17, binds to cellulose via a novel cellulose-binding module and to another *R. flavefaciens* protein via a dockerin domain. *Applied and Environmental Microbiology* **67** (10): 4426-4431.
- Rincon, M. T., Ding, S. Y., McCrae, S. I., Martin, J. C., Aurilia, V., Lamed, R., Shoham, Y., Bayer, E. A. and Flint, H. J. (2003). Novel organization and divergent

- dockerin specificities in the cellulosome system of *Ruminococcus flavefaciens*. *Journal of Bacteriology* **185** (3): 703-713.
- Rincon, M. T., Cepeljnik, T., Martin, J. C., Barak, Y., Lamed, R., Bayer, E. A. and Flint, H. J. (2007). A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of *Ruminococcus flavefaciens*. *Journal of Bacteriology* **189** (13): 4774-4783.
- Romaniec, M. P. M., Davidson, K., White, B. A. and Hazlewood, G. P. (1989). Cloning of *Ruminococcus albus* endo-beta-1,4-glucanase and xylanase genes. *Letters in Applied Microbiology* **9** (3): 101-104.
- Ross, P. L., Huang, Y. L. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlett-Jones, M., He, F., Jacobson, A. and Pappin, D. J. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Molecular & Cellular Proteomics* **3** (12): 1154-1169.
- Rouau, X., Cheynier, V., Surget, A., Gloux, D., Barron, C., Meudec, E., Louis-Montero, J. and Criton, M. (2003). A dehydrotrimer of ferulic acid from maize bran. *Phytochemistry* **63** (8): 899-903.
- Ruiz, L., Coute, Y., Sanchez, B., de los Reyes-Gavilan, C. G., Sanchez, J. C. and Margolles, A. (2009). The cell-envelope proteome of *Bifidobacterium longum* in an *in vitro* bile environment. *Microbiology-Sgm* **155**: 957-967.
- Rumbak, E., Rawlings, D. E., Lindsey, G. G. and Woods, D. R. (1991). Cloning, nucleotide-sequence, and enzymatic characterization of an alpha-amylase from the ruminal bacterium *Butyrivibrio fibrisolvens* H17c. *Journal of Bacteriology* **173** (13): 4203-4211.
- Russell, J. B., Oconnor, J. D., Fox, D. G., Vansoest, P. J. and Sniffen, C. J. (1992a). A net carbohydrate and protein system for evaluating cattle diets. 1. Ruminant fermentation. *Journal of Animal Science* **70** (11): 3551-3561.
- Russell, J. B. (1998). Strategies that ruminal bacteria use to handle excess carbohydrate. *Journal of Animal Science* **76** (7): 1955-1963.
- Russell, R. R. B., Aduseopoku, J., Sutcliffe, I. C., Tao, L. and Ferretti, J. J. (1992b). A binding protein-dependent transport-system in *Streptococcus mutans* responsible for multiple sugar metabolism. *Journal of Biological Chemistry* **267** (7): 4631-4637.
- Sahm, K., Matuschek, M., Muller, H., Mitchell, W. J. and Bahl, H. (1996). Molecular analysis of the amy gene locus of *Thermoanaerobacterium thermosulfurigenes* EM1 encoding starch-degrading enzymes and a binding protein-dependent maltose transport system. *Journal of Bacteriology* **178** (4): 1039-1046.
- Saier, M. H., Yen, M. R., Noto, K., Tamang, D. G. and Elkan, C. (2009). The Transporter Classification Database: recent advances. *Nucleic Acids Research* **37**: D274-D278.
- Sakai, T., Sakamoto, T., Hallaert, J. and Vandamme, E. J. (1993). Pectin, pectinase, and protopectinase - production, properties, and applications. *Advances in Applied Microbiology* **39**: 213-294.
- Sakka, K., Yoshikawa, K., Kojima, Y., Karita, S., Ohmiya, K. and Shimada, K. (1993). Nucleotide-sequence of the *Clostridium stercoarium* xylA gene encoding a bifunctional protein with beta-D-xylosidase and alpha-L-arabinofuranosidase activities, and properties of the translated product. *Bioscience Biotechnology and Biochemistry* **57** (2): 268-272.
- Sales, M., Lucas, F. and Blanchart, G. (2000). Effects of ammonia and amino acids on the growth and proteolytic activity of three species of rumen bacteria: *Prevotella*

- albensis*, *Butyrivibrio fibrisolvens*, and *Streptococcus bovis*. *Current Microbiology* **40** (6): 380-386.
- Sanchez-Herrera, L. M., Ramos-Valdivia, A. C., de la Torre, M., Salgado, L. M. and Ponce-Noyola, T. (2007). Differential expression of cellulases and xylanases by *Cellulomonas flavigena* grown on different carbon sources. *Applied Microbiology and Biotechnology* **77**: 589-595.
- Sanchez, B., Champomier-Verges, M. C., Collado, M. D., Anglade, P., Baraige, F., Sanz, Y., Reyes-Gavilan, C., Margolles, A. and Zagorec, M. (2007). Low-pH adaptation and the acid tolerance response of *Bifidobacterium longum* biotype longum. *Applied and Environmental Microbiology* **73**: 6450-6459.
- Sanchez, B., Champomier-Verges, M. C., Anglade, P., Baraige, F., de los Reyes-Gavilan, C. G., Margolles, A. and Zagorec, M. (2008). A preliminary analysis of *Bifidobacterium longum* exported proteins by two-dimensional electrophoresis. *Journal of Molecular Microbiology and Biotechnology* **14** (1-3): 74-79.
- Sato, S., Liu, F., Koc, H. and Tien, M. (2007). Expression analysis of extracellular proteins from *Phanerochaete chrysosporium* grown on different liquid and solid substrates. *Microbiology-Sgm* **153**: 3023-3033.
- Saulnier, L. and Thibault, J. F. (1999). Ferulic acid and diferulic acids as components of sugarbeet pectins and maize bran heteroxylans. *Journal of the Science of Food and Agriculture* **79** (3): 396-402.
- Scalbert, A., Monties, B., Lallemand, J. Y., Guittet, E. and Rolando, C. (1985). Ether linkage between phenolic-acids and lignin fractions from wheat straw. *Phytochemistry* **24** (6): 1359-1362.
- Schingoethe, D. J., Stegeman, G. A. and Treacher, R. J. (1999). Response of lactating dairy cows to a cellulase and xylanase enzyme mixture applied to forages at the time of feeding. *Journal of Dairy Science* **82** (5): 996-1003.
- Schmidt, F., Donahoe, S., Hagens, K., Mattow, J., Schaible, U. E., Kaufmann, S. H. E., Aebersold, R. and Jungblut, P. R. (2004). Complementary analysis of the *Mycobacterium tuberculosis* proteome by two-dimensional electrophoresis and isotope-coded affinity tag technology. *Molecular & Cellular Proteomics* **3** (1): 24-42.
- Schneider, E. and Hunke, S. (1998). ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiology Reviews* **22** (1): 1-20.
- Schneider, E. (2001). ABC transporters catalyzing carbohydrate uptake. *Research in Microbiology* **152** (3-4): 303-310.
- Schwarz, K., Fiedler, T., Fischer, R. J. and Bahl, H. (2007). A Standard Operating Procedure (SOP) for the preparation of intra- and extracellular proteins of *Clostridium acetobutylicum* for proteome analysis. *Journal of Microbiological Methods* **68** (2): 396-402.
- Schwarz, W. H., Bronnenmeier, K., Krause, B., Lottspeich, F. and Staudenbauer, W. L. (1995). Debranching of arabinoxylan - properties of the thermoactive recombinant alpha-l-arabinofuranosidase from *Clostridium-stercorarium* (Arfb). *Applied Microbiology and Biotechnology* **43** (5): 856-860.
- Schwarz, W. H., Zverlov, V. V. and Bahl, H. (2004). Extracellular glycosyl hydrolases from *Clostridia*. *Advances in Applied Microbiology* **56**: 215-261.
- Shallom, D., Golan, G., Shoham, G. and Shoham, Y. (2004). Effect of dimer dissociation on activity and thermostability of the alpha-glucuronidase from *Geobacillus stearothermophilus*: Dissecting the different oligomeric forms of Family 67 glycoside hydrolases. *Journal of Bacteriology* **186** (20): 6928-6937.

- Shane, B. S., Gouws, L. and Kistner, A. (1969). Cellulolytic bacteria occurring in rumen of sheep conditioned to low-protein Teff hay. *Journal of General Microbiology* **55**: 445-457.
- Shapiro, S. S. and Wilk, M. B. (1965a). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* **52**: 591-&.
- Shapiro, S. S. and Wilk, M. B. (1965b). An analysis of variance test for normality (complete samples). *Biometrika* **52**: 591-611.
- Sharp, P. M. and Li, W. H. (1987). The codon adaptation index - a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research* **15** (3): 1281-1295.
- Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F. and Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Research* **33** (4): 1141-1153.
- Shevchenko, A., Jensen, O. N., Podtelejnikov, A. V., Sagliocco, F., Wilm, M., Vorm, O., Mortensen, P., Shevchenko, A., Boucherie, H. and Mann, M. (1996). Linking genome and proteome by mass spectrometry: Large-scale identification of yeast proteins from two dimensional gels. *Proceedings of the National Academy of Sciences of the United States of America* **93** (25): 14440-14445.
- Shevchik, V. E. and Hugouvieux-Cotte-Pattat, N. (2003). PaeX, a second pectin acetylesterase of *Erwinia chrysanthemi* 3937. *Journal of Bacteriology* **185** (10): 3091-3100.
- Shibuya, N. and Nakane, R. (1984). Pectic polysaccharides of rice endosperm cell-walls. *Phytochemistry* **23** (7): 1425-1429.
- Shin, E. C., Cho, K. M., Lim, W. J., Hong, S. Y., An, C. L., Kim, E. J., Kim, Y. K., Choi, B. R., An, J. M., Kang, J. M., Kim, H. and Yun, H. D. (2004a). Phylogenetic analysis of protozoa in the rumen contents of cow based on the 18S rDNA sequences. *Journal of Applied Microbiology* **97** (2): 378-383.
- Shin, E. C., Choi, B. R., Lim, W. J., Hong, S. Y., An, C. L., Cho, K. M., Kim, Y. K., An, J. M., Kang, J. M., Lee, S. S., Kim, H. and Yun, H. D. (2004b). Phylogenetic analysis of archaea in three fractions of cow rumen based on the 16S rDNA sequence. *Anaerobe* **10** (6): 313-319.
- Shin, E. C., Kim, Y. K., Lim, W. J., Hong, S. Y., An, C. L., Kim, E. J., Cho, K. M., Choi, B. R., An, J. M., Kang, J. M., Jeong, Y. J. J., Kwon, E. J., Kim, H. and Yun, H. D. (2004c). Phylogenetic analysis of yeast in the rumen contents of cattle based on the 26S rDNA sequence. *Journal of Agricultural Science* **142**: 603-611.
- Shoseyov, O. and Doi, R. H. (1990). Essential 170 kDa subunit for degradation of crystalline cellulose by *Clostridium cellulovorans* cellulase. *Proceedings of the National Academy of Sciences of the United States of America* **87** (6): 2192-2195.
- Showalter, A. M. (1993). Structure and function of plant-cell wall proteins. *Plant Cell* **5** (1): 9-23.
- Shulami, S., Zaide, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L. and Shoham, Y. (2007). A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. *Applied and Environmental Microbiology* **73** (3): 874-884.
- Shuman, H. A. (1982). Active-transport of maltose in *Escherichia coli* K12 - role of the periplasmic maltose-binding protein and evidence for a substrate recognition site in the cytoplasmic membrane. *Journal of Biological Chemistry* **257** (10): 5455-5461.
- Skillman, L. C., Evans, P. N., Strompl, C. and Joblin, K. N. (2006). 16S rDNA directed PCR primers and detection of methanogens in the bovine rumen. *Letters in Applied Microbiology* **42** (3): 222-228.

- Smith, B. G. and Harris, P. J. (1999). The polysaccharide composition of Poales cell walls: *Poaceae* cell walls are not unique. *Biochemical Systematics and Ecology* **27** (1): 33-53.
- Song, S. G. and Park, C. (1997). Organization and regulation of the D-xylose operons in *Escherichia coli* K-12: XylR acts as a transcriptional activator. *Journal of Bacteriology* **179** (22): 7025-7032.
- Sorensen, H. R., Pedersen, S., Jorgensen, C. T. and Meyer, A. S. (2007). Enzymatic hydrolysis of wheat arabinoxylan by a recombinant "minimal" enzyme cocktail containing beta-xylosidase and novel endo-1,4-beta-xylanase and alpha-L-arabinofuranosidase activities. *Biotechnology Progress* **23** (1): 100-107.
- Spence, C., Wells, W. G. and Smith, C. J. (2006). Characterization of the primary starch utilization operon in the obligate anaerobe *Bacteroides fragilis*: Regulation by carbon source and oxygen. *Journal of Bacteriology* **188** (13): 4663-4672.
- Staudte, R. G., Woodward, J. R., Fincher, G. B. and Stone, B. A. (1983). Water-soluble (1,3),(1,4)-beta-D-glucans from barley (*Hordeum vulgare*) endosperm. III. Distribution of cellotriosyl and cellotetraosyl residues. *Carbohydrate Polymers* **3** (299-312).
- 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): 17-28.
- Stewart, C. S., Flint, H. J. and Bryant, M. P. (1997). The rumen bacteria. *The Rumen Microbial Ecosystem*. Eds. Hobson, P. N. and Stewart, C. S. London, Blackie Academic and Professional Publishers.: 10-72.
- Stinard, P. S. and Nevins, D. J. (1980). Distribution of non-cellulosic beta-D-glucans in grasses and other monocots. *Phytochemistry* **19** (7): 1467-1468.
- Storey, J. D. (2002). A direct approach to false discovery rates. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**: 479-498.
- Storey, J. D. and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America* **100** (16): 9440-9445.
- Sun, J. L., Sakka, K., Karita, S., Kimura, T. and Ohmiya, K. (1998). Adsorption of *Clostridium stercorarium* xylanase A to insoluble xylan and the importance of the CBDs to xylan hydrolysis. *Journal of Fermentation and Bioengineering* **85** (1): 63-68.
- Sutcliffe, I. C. and Harrington, D. J. (2002). Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes. *Microbiology-Sgm* **148** (7): 2065-2077.
- Taguchi, H., Hagiwara, D., Genma, T., Karita, S., Kimura, T., Sakka, K. and Ohmiya, K. (2004a). Cloning of the *Ruminococcus albus* *cel5D* and *cel9A* genes encoding dockerin module-containing endoglucanases and expression of Cel5D in *Escherichia coli*. *Bioscience Biotechnology and Biochemistry* **68** (7): 1557-1564.
- Taguchi, H., Koike, S., Kobayashi, Y., Cann, I. K. O. and Karita, S. (2004b). Partial characterization of structure and function of a xylanase gene from the rumen hemicellulolytic bacterium *Eubacterium ruminantium*. *Animal Science Journal* **75** (4): 325-332.
- Taguchi, H., Senoura, T., Hamada, S., Matsui, H., Kobayashi, Y., Watanabe, J., Wasaki, J. and Ito, S. (2008). Cloning and sequencing of the gene for cellobiose 2-epimerase from a ruminal strain of *Eubacterium cellulosolvens*. *FEMS Microbiology Letters* **287** (1): 34-40.

- Tajima, K., Aminov, R. I., Nagamine, T., Ogata, K., Nakamura, M., Matsui, H. and Benno, Y. (1999). Rumen bacterial diversity as determined by sequence analysis of 16S rDNA libraries. *Fems Microbiology Ecology* **29** (2): 159-169.
- Tam, L. T., Antelmann, H., Eymann, C., Albrecht, D., Bernhardt, J. and Hecker, M. (2006). Proteome signatures for stress and starvation in *Bacillus subtilis* as revealed by a 2-D gel image color coding approach. *Proteomics* **6** (16): 4565-4585.
- Tamhane, A. C. (1979). Comparison of procedures for multiple comparisons of means with unequal variances. *Journal of the American Statistical Association* **74** (366): 471-480.
- Tao, L., Sutcliffe, I. C., Russell, R. R. B. and Ferretti, J. J. (1993). Transport of sugars, including sucrose, by the msm transport-system of *Streptococcus mutans*. *Journal of Dental Research* **72** (10): 1386-1390.
- Tardy, F., Nasser, W., Robert-Baudouy, J. and Hugouvieux-Cotte-Pattat, N. (1997). Comparative analysis of the five major *Erwinia chrysanthemi* pectate lyases: Enzyme characteristics and potential inhibitors. *Journal of Bacteriology* **179** (8): 2503-2511.
- Taylor, L. E., Henrissat, B., Coutinho, P. M., Ekborg, N. A., Hutcheson, S. W. and Weiner, R. A. (2006). Complete cellulase system in the marine bacterium *Saccharophagus degradans* strain 2-40(T). *Journal of Bacteriology* **188** (11): 3849-3861.
- Teeri, T. T. (1997). Crystalline cellulose degradation: New insight into the function of cellobiohydrolases. *Trends in Biotechnology* **15** (5): 160-167.
- Teter, S., Cherry, J., Ward, C., Jones, A., Harris, P. and Yi, J. (2004). Variants of cellobiohydrolase I from *Trichoderma reesei* with improved properties. Novozymes Biotech, IU, USA. 2005048619.
- Titi, H. and Lubbadah, W. F. (2004). Effect of feeding cellulase enzyme on productive responses of pregnant and lactating ewes and goats. *Small Ruminant Research* **52** (1-2): 137-143.
- Tomme, P., Warren, R. A. and Gilkes, N. R. (1995). Cellulose hydrolysis by bacteria and fungi. *Advances in Microbial Physiology* **37**: 1-81.
- Topakas, E., Vafiadi, C. and Christakopoulos, P. (2007). Microbial production, characterization and applications of feruloyl esterases. *Process Biochemistry* **42** (4): 497-509.
- Toyoda, A. and Minato, H. (2002). Cloning, nucleotide sequence and expression of the gene encoding the cellulose-binding protein A (CBPA) of *Eubacterium cellulosolvens* 5. *FEMS Microbiology Letters* **207** (2): 141-146.
- Toyoda, A., Takano, K. and Minato, H. (2003). A possible role of cellulose-binding protein A (CBPA) in the adhesion of *Eubacterium cellulosolvens* 5 to cellulose. *Journal of General and Applied Microbiology* **49** (4): 245-250.
- Toyoda, A., Yoshimatsu, M., Takano, K. and Minato, H. (2005). Cloning, nucleotide sequence and module structure of the gene encoding the cellulose-binding protein B (CBPB) of *Eubacterium cellulosolvens* 5. *Journal of General and Applied Microbiology* **51** (4): 213-220.
- Trethewey, J. A. K., Campbell, L. M. and Harris, P. J. (2005). (1-3,1-4)-beta-D-glucans in the cell walls of the Poales (*Sensu lato*): An immunogold labeling study using a monoclonal antibody. *American Journal of Botany* **92** (10): 1660-1674.
- Ulyatt, M. J., Beever, D. E., Thomson, D. J., Evans, R. T. and Haines, M. J. (1980). Measurement of nutrient supply at pasture. *Proceedings of the Nutrition Society* **39** (3): A67-A67.
- Utt, E. A., Eddy, C. K., Keshav, K. F. and Ingram, L. O. (1991). Sequencing and expression of the *Butyrivibrio fibrisolvens* xylB gene encoding a novel bifunctional

- protein with beta-D-xylosidase and alpha-L-arabinofuranosidase activities. *Applied and Environmental Microbiology* **57** (4): 1227-1234.
- Vailhe, M. A. B., Provan, G. J., Scobbie, L., Chesson, A., Mailliot, M. P., Cornu, A. and Besle, J. M. (2000). Effect of phenolic structures on the degradability of cell walls isolated from newly extended apical internode of tall fescue (*Festuca arundinacea* Schreb.). *Journal of Agricultural and Food Chemistry* **48** (3): 618-623.
- van Bueren, A. L., Morland, C., Gilbert, H. J. and Boraston, A. B. (2005). Family 6 carbohydrate binding modules recognize the non-reducing end of beta-1,3-linked glucans by presenting a unique ligand binding surface. *Journal of Biological Chemistry* **280** (1): 530-537.
- van der Heide, T. and Poolman, B. (2002). ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Reports* **3** (10): 938-943.
- van Gylswyk, N. O. and van der Toorn, J., K (1986). Description and designation of a neotype strain of *Eubacterium cellulosolvens* (*Cillobacterium cellulosolvens* bryant, Small, Bouma and Robinson) Holdeman and Moore. *International Journal of Systematic Bacteriology* **36** (2): 275-277.
- van Gylswyk, N. O. (1990). Enumeration and presumptive identification of some functional groups of bacteria in the rumen of dairy cows fed grass silage-based diets. *Fems Microbiology Ecology* **73** (3): 243-254.
- van Soest, P. J. (1973). Uniformity and nutritive availability of cellulose. *Federation Proceedings* **32** (7): 1804-1808.
- van Soest, P. J. (1994). *Nutritional ecology of the ruminant*. Ithaca, New York, Cornell University Press.
- Varel, V. H., Yen, J. T. and Kreikemeier, K. K. (1995). Addition of cellulolytic *Clostridia* to the bovine rumen and pig intestinal tract. *Applied and Environmental Microbiology* **61** (3): 1116-1119.
- Vietor, R. J., Angelino, S. and Voragen, A. G. J. (1992). Structural features of arabinoxylans from barley and malt cell-wall material. *Journal of Cereal Science* **15** (3): 213-222.
- Villas-Boas, S. G., Noel, S., Lane, G. A., Attwood, G. and Cookson, A. (2006). Extracellular metabolomics: A metabolic footprinting approach to assess fiber degradation in complex media. *Analytical Biochemistry* **349** (2): 297-305.
- Vogel, J. (2008). Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* **11** (3): 301-307.
- Voigt, B., Schweder, T., Sibbald, M., Albrecht, D., Ehrenreich, A., Bernhardt, J., Feesche, J., Maurer, K. H., Gottschalk, G., van Dijk, J. M. and Hecker, M. (2006). The extracellular proteome of *Bacillus licheniformis* grown in different media and under different nutrient starvation conditions. *Proteomics* **6** (1): 268-281.
- Voigt, B., Hoi, L. T., Jurgen, B., Albrecht, D., Ehrenreich, A., Veith, B., Evers, S., Maurer, K. H., Hecker, M. and Schweder, T. (2007). The glucose and nitrogen starvation response of *Bacillus licheniformis*. *Proteomics* **7** (3): 413-423.
- Waeonukul, R., Kyu, K. L., Sakka, K. and Ratanakhanokchai, K. (2008). Effect of carbon sources on the induction of xylanolytic-cellulolytic multienzyme complexes in *Paenibacillus curdlanolyticus* strain B-6. *Bioscience Biotechnology and Biochemistry* **72** (2): 321-328.
- Waldo, D. R., Smith, L. W. and Cox, E. L. (1972). Model of cellulose disappearance from rumen. *Journal of Dairy Science* **55** (1): 125-129.
- Wang, J. Q., Xue, Y. F., Feng, X. L., Li, X. L., Wang, H., Li, W., Zhao, C. F., Cheng, X. J., Ma, Y. H., Zhou, P. J., Yin, J. N., Bhatnagar, A., Wang, R. and Liu, S. Q.

- (2004). An analysis of the proteomic profile for *Thermoanaerobacter tengcongensis* under optimal culture conditions. *Proteomics* **4** (1): 136-150.
- Weatherly, D. B., Atwood, J. A., Minning, T. A., Cavola, C., Tarleton, R. L. and Orlando, R. (2005). A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol Cell Proteomics* **4** (6): 762-772.
- Weimer, P. J., Lopezguisa, J. M. and French, A. D. (1990). Effect of cellulose fine-structure on kinetics of its digestion by mixed ruminal microorganisms *in vitro*. *Applied and Environmental Microbiology* **56** (8): 2421-2429.
- Weimer, P. J. (1998). Manipulating ruminal fermentation: A microbial ecological perspective. *Journal of Animal Science* **76** (12): 3114-3122.
- Whistler, R. L. and Richards, E. L. (1970). Hemicelluloses. *The Carbohydrates*. Eds. Pigman, W. and Horton, D. New York., Academic Press.: 447-469.
- Whitehead, T. R. (1993). Analyses of the gene and amino acid sequence of the *Prevotella (Bacteroides) Ruminicola* 23 xylanase reveals unexpected homology with endoglucanases from other genera of bacteria. *Current Microbiology* **27** (1): 27-33.
- Whitford, M. F., Forster, R. J., Beard, C. E., Gong, J. H. and Teather, R. M. (1998). Phylogenetic analysis of rumen bacteria by comparative sequence analysis of cloned 16S rRNA genes. *Anaerobe* **4** (3): 153-163.
- Wildgruber, R., Harder, A., Obermaier, C., Boguth, G., Weiss, W., Fey, S. J., Larsen, P. M. and Gorg, A. (2000). Towards higher resolution: Two-dimensional Electrophoresis of *Saccharomyces cerevisiae* proteins using overlapping narrow immobilized pH gradients. *Electrophoresis* **21** (13): 2610-2616.
- Williams, A. G. and Coleman, G. S. (1997). The rumen protozoa. *The Rumen Microbial Ecosystem*. Eds. Hobson, P. N. and Stewart, C. S. London, Blackie Academic and Professional Publishers.: 73-139.
- Wolff, S., Otto, A., Albrecht, D., Zeng, J. S., Buttner, K., Gluckmann, M., Hecker, M. and Becher, R. (2006). Gel-free and gel-based proteomics in *Bacillus subtilis* - A comparative study. *Molecular & Cellular Proteomics* **5** (7): 1183-1192.
- Wolff, S., Antelmann, H., Albrecht, D., Becher, D., Bernhardt, J., Bron, S., Buttner, K., van Dijk, J. M., Eymann, C., Otto, A., Tam, L. T. and Hecker, M. (2007). Towards the entire proteome of the model bacterium *Bacillus subtilis* by gel-based and gel-free approaches. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* **849** (1-2): 129-140.
- Wolff, S., Hahne, H., Hecker, M. and Becher, D. (2008). Complementary analysis of the vegetative membrane proteome of the human pathogen *Staphylococcus aureus*. *Molecular & Cellular Proteomics* **7** (8): 1460-1468.
- Wong, K. K. Y., Tan, L. U. L. and Saddler, J. N. (1988). Multiplicity of beta-1,4-xylanase in microorganisms - functions and applications. *Microbiological Reviews* **52** (3): 305-317.
- Woodward, J. R. and Fincher, G. B. (1983). Water-soluble (1,3),(1,4)-beta-D-glucans from barley (*Hordeum vulgare*) endosperm. II. Fine structure. *Carbohydrate Polymers* **3**: 207-225.
- Wright, A., Wait, R., Begum, S., Crossett, B., Nagy, J., Brown, K. and Fairweather, N. (2005). Proteomic analysis of cell surface proteins from *Clostridium difficile*. *Proteomics* **5** (9): 2443-2452.
- Wright, A. D. G., Auckland, C. H. and Lynn, D. H. (2007). Molecular diversity of methanogens in feedlot cattle from Ontario and Prince Edward Island, Canada. *Applied and Environmental Microbiology* **73** (13): 4206-4210.

- Wu, W., Lange, L., Skovlund, D. A. and Liu, Y. (2003). Cloning and sequences of cellobiohydrolase II from thermophilic fungi and their use for ethanol production and in detergents. Novozymes, USA. 2004056981.
- Yamazaki, S. and Tove, S. B. (1977). Purification and properties of an oxygen consuming substance from *Butyrivibrio fibrisolvens*. *Federation Proceedings* **36** (3): 789-789.
- Yang, H. J. and Xie, C. Y. (2010). Assessment of fibrolytic activities of 18 commercial enzyme products and their abilities to degrade the cell wall fraction of corn stalks in *in vitro* enzymatic and ruminal batch cultures. *Animal Feed Science and Technology* **159** (3-4): 110-121.
- Yang, J., Hou, X. and Gao, A. (2009). Quantity of microorganisms associated with solid and liquid phases of rumen contents using RT-PCR. *Scientia Agricultura Sinica* **42** (6): 2126-2132.
- Yang, W. Z., Beauchemin, K. A. and Rode, L. M. (2000). A comparison of methods of adding fibrolytic enzymes to lactating cow diets. *Journal of Dairy Science* **83** (11): 2512-2520.
- Yates, J. R., Speicher, S., Griffin, P. R. and Hunkapiller, T. (1993). Peptide mass maps - a highly informative approach to protein identification. *Analytical Biochemistry* **214** (2): 397-408.
- Ye, Z. H. and Varner, J. E. (1991). Tissue-specific expression of cell-wall proteins in developing soybean tissues. *Plant Cell* **3** (1): 23-37.
- Yoda, K., Toyoda, A., Mukoyama, Y., Nakamura, Y. and Minato, H. (2005). Cloning, sequencing, and expression of a *Eubacterium cellulosolvens* 5 gene encoding an endoglucanase (Cel5A) with novel carbohydrate-binding modules, and properties of Cel5A. *Applied and Environmental Microbiology* **71** (10): 5787-5793.
- Yoshimatsu, M., Toyoda, A., Onizawa, N., Nakamura, Y. and Minato, H. (2007). Biochemical characterization of cellulose-binding proteins (CBPA and CBPB) from the rumen cellulolytic bacterium *Eubacterium cellulosolvens* 5. *Bioscience Biotechnology and Biochemistry* **71**: 2577-2580.
- Yother, J. and White, J. M. (1994). Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein Pspa. *Journal of Bacteriology* **176** (10): 2976-2985.
- Yu, P. Q., McKinnon, J. J., Maenz, D. D., Olkowski, A. A., Raczy, V. J. and Christensen, D. A. (2003). Enzymic release of reducing sugars from oat hulls by cellulase, as influenced by *Aspergillus* ferulic acid esterase and *Trichoderma* xylanase. *Journal of Agricultural and Food Chemistry* **51** (1): 218-223.
- Yuan, J., Zhu, L., Liu, X. K., Zhang, Y., Ying, T. Y., Wang, B., Wang, J. J., Dong, H., Feng, E. L., Li, Q., Wang, J., Wang, H. X., Wei, K. H., Zhang, X. M., Huang, C. F., Huang, P. T., Huang, L. Y., Zeng, M. and Wang, H. L. (2006). A proteome reference map and proteomic analysis of *Bifidobacterium longum* NCC2705. *Molecular & Cellular Proteomics* **5** (6): 1105-1118.
- Zabackis, E., Huang, J., Muller, B., Darvill, A. G. and Albersheim, P. (1995). Characterization of the cell-wall polysaccharides of *Arabidopsis thaliana* leaves. *Plant Physiology* **107** (4): 1129-1138.
- Zhang, B., VerBerkmoes, N. C., Langston, M. A., Uberbacher, E., Hettich, R. L. and Samatova, N. F. (2006). Detecting differential and correlated protein expression in label-free shotgun proteomics. *Journal of Proteome Research* **5**: 2909-2918.
- Zhang, J. X. and Flint, H. J. (1992). A bifunctional xylanase encoded by the *xynA* gene of the rumen cellulolytic bacterium *Ruminococcus flavefaciens* 17 comprises 2 dissimilar domains linked by an asparagine glutamine-rich sequence. *Molecular Microbiology* **6** (8): 1013-1023.

- Zhang, W. Z. and Chait, B. T. (2000). Profound: An expert system for protein identification using mass spectrometric peptide mapping information. *Analytical Chemistry* **72** (11): 2482-2489.
- ZoBell, D. R., Wiedmeier, R. D., Olson, K. C. and Treacher, R. (2000). The effect of an exogenous enzyme treatment on production and carcass characteristics of growing and finishing steers. *Animal Feed Science and Technology* **87** (3-4): 279-285.
- Zverlov, V. V., Liebl, W., Bachleitner, M. and Schwarz, W. H. (1998). Nucleotide sequence of *arfB* of *Clostridium stercorarium*, and prediction of catalytic residues of alpha L-arabinofuranosidases based on local similarity with several families of glycosyl hydrolases. *FEMS Microbiology Letters* **164** (2): 337-343.
- Zykwinska, A., Thibault, J. F. and Ralet, M. C. (2007). Organization of pectic arabinan and galactan side chains in association with cellulose microfibrils in primary cell walls and related models envisaged. *Journal of Experimental Botany* **58** (7): 1795-1802.

Appendix A

Table A.1. Summary of the functional groups and sub-groups used to classify *B. proteoclasticus* protein functions.

Role	Designation	Function	Protein frequency	Proportion of proteome
Amino acid biosynthesis	A	Arginine, aromatic amino acids, aspartate/asparagine, cysteine/methionine, glutamate/glutamine, glycine/serine/threonine, histidine, lysine, proline, valine/leucine/isoleucine.	90	2.4%
Carbohydrate metabolism	C	Aminosugars, fucose, galactose, glucuronate, glycogen biosynthesis, glycolysis/gluconeogenesis, mannose, other, pentose phosphate pathway, polysaccharide degradation, rhamnose.	220	5.8%
Cell cycle	D	Cell division, chromosome replication, genome segregation.	50	1.3%
Cell envelope	I	Cell surface proteins, cell wall hydrolysis, exopolysaccharides, other, peptidoglycan biosynthesis.	329	8.6%
Cellular processes	B	Chemotaxis, motility, other, oxidative stress response.	96	2.5%
Central metabolism	G	One-carbon metabolism, other, polyamine biosynthesis.	13	0.3%
Energy metabolism	E	Acetate, ATP synthase, butyrate, electron transport, formate, other, TCA.	100	2.6%
Hypothetical	H	Conserved, hypothetical.	1342	35.2%
Lipid metabolism	L	Fatty acid biosynthesis, isoprene biosynthesis, other, phospholipid biosynthesis.	41	1.1%
Mobile elements	M	Transposases	36	0.9%
Nucleic acid metabolism	N	DNA-binding proteins, helicases, other, pyrimidines, recombination and repair, restriction and modification, site-specific recombination.	103	2.7%
Nucleotide metabolism	O	Purines, purines and pyrimidines, pyrimidines,	60	1.6%
Plasmid functions	F	Conjugation, other, plasmid replication.	39	1.0%
Protein fate	P	Other, protein degradation, protein folding, protein modification, protein secretion.	110	2.9%

Table continues

Role	Designation	Function	Protein frequency	Proportion of proteome
Protein synthesis	Q	Other, ribosomal proteins, RNA processing, translation factors, tRNA aminoacylation.	151	4.0%
Regulation	R	Other, protein interactions, RNA interactions, transcriptional regulators.	141	3.7%
Signal transduction	S	Other, two-component systems.	155	4.1%
Transcription	W	Other, RNA polymerase, transcription factors.	22	0.6%
Transporters	T	Amino acids, anions, carbohydrates, cations, other.	282	7.4%
Unknown function	U	Enzyme, general.	396	10.4%
Vitamins and cofactors	V	Biotin, cobalamin and heme, coenzyme A, folic acid, nicotinate, other, riboflavin, thiamine.	40	1.0%

Appendix B

Table B.1. Predicted cytosolic proteins identified in the mid-log and stationary phase harvested *B. proteoclasticus* culture medium.

Protein	Locus	PHX	Score	pI	kDa	Pep.	Cov.
<u>Amino acid biosynthesis</u>							
Cysteine synthase, CysK	Bpr_I1089	N	2.4e-13	5.0	32.3	14	65%
Diaminopimelate dehydrogenase	Bpr_I0298	Y	9.6e-11	5.6	35.8	15	52%
Glu/Leu/Phe/Val dehydrogenase	Bpr_I2129	Y	4.8e-18	5.4	48.6	21	55%
<u>Carbohydrate metabolism</u>							
2-deoxy-D-gluconate 3-dehydrogenase, KduD	Bpr_I0930	Y	3.0e-08	4.6	30.2	10	33%
Aldose-1-epimerase	Bpr_I0228	N	7.6e-12	4.8	38.3	15	56%
Fructose-1,6-bisphosphate aldolase, FbaA	Bpr_I2903	Y	1.2e-13	5.1	30.5	18	65%
Glucose-6-phosphate isomerase, Gpi	Bpr_I0035	Y	2.4e-09	5.0	57.1	14	32%
Glyceraldehyde-3-phosphate dehydrogenase, Gap	Bpr_I2050	Y	2.4e-07	5.7	36.9	15	34%
Phosphoenolpyruvate carboxykinase, PckA	Bpr_I0091	Y	1.5e-14	4.9	59.5	16	29%
Transaldolase	Bpr_I1511	Y	6.1e-11	4.8	23.4	11	57%
Triosephosphate isomerase, TpiA	Bpr_I2048	Y	1.5e-14	4.6	26.8	15	62%
Xylosidase/arabinofuranosidase, Xsa43E	Bpr_I2319	N	2.1e-06	4.3	35.3	9	39%
<u>Cell cycle</u>							
Single-stranded DNA binding protein, Ssb1	Bpr_I0240	N	5.0e-05	4.7	16.6	7	48%
<u>Cell envelope</u>							
NAD-dependent epimerase/dehydratase	Bpr_I2537	Y	9.6e-12	4.7	36.0	12	43%

Table continues

Protein	Locus	PHX	Score	pI	kDa	Pep.	Cov.
Nucleotide sugar dehydrogenase	Bpr_I2538	Y	1.4e-05	4.9	48.6	9	24%
<i>Energy metabolism</i>							
Butyrate kinase, Buk	Bpr_I2323	Y	1.5e-13	5.1	38.7	19	63%
Butyryl-CoA dehydrogenase, Bcd	Bpr_I2485	Y	1.2e-12	6.1	42.5	21	51%
Crotonase, Crt	Bpr_I2487	Y	3.0e-13	4.7	28.0	16	62%
Electron transfer flavoprotein α -subunit, EtfA	Bpr_I2483	Y	1.2e-05	4.8	37.4	8	30%
Electron transfer flavoprotein β -subunit, EtfB	Bpr_I2484	Y	1.9e-09	5.2	28.4	14	47%
Methylmalonyl-CoA decarboxylase α -subunit, MmdA	Bpr_I1226	Y	3.0e-21	4.5	50.9	27	55%
Phosphate butyryltransferase, Ptb	Bpr_I2324	Y	9.6e-08	5.1	33.3	9	34%
Pyruvate phosphate dikinase, PpdK	Bpr_I1154	Y	7.6e-17	4.8	104.3	25	24%
Thiolase, ThlA1	Bpr_I2488	Y	3.0e-16	5.0	42.0	28	65%
<i>Hypothetical and unknown function</i>							
Hypothetical protein	Bpr_I2329	Y	1.9e-09	4.7	23.5	9	41%
Hypothetical protein	Bpr_I2455	N	1.2e-05	4.7	47.7	8	22%
Hypothetical protein	Bpr_I2744	N	3.8e-08	4.8	28.5	10	47%
SAM-dependent methyltransferase	Bpr_I0663	N	3.8e-08	5.0	32.6	15	54%
<i>Nucleotide and nucleic acid metabolism</i>							
Adenine phosphoribosyl transferase, Apt	Bpr_I1345	Y	4.5e-07	5.1	19.3	8	55%
Exodeoxyribonuclease III, Xth	Bpr_III162	N	1.4e-05	5.6	24.8	9	37%
IMP cyclohydrolase, PurO	Bpr_I0731	N	6.1e-10	4.7	32.2	15	51%

Table continues

Table B.1 continued

Protein	Locus	PHX	Score	pI	kDa	Pep.	Cov.
Purine nucleoside phosphorylase, DeoD2	Bpr_I2948	Y	9.6e-14	5.1	26.1	16	50%
Uridylate kinase, PyrH	Bpr_I0953	N	7.6e-09	5.4	25.5	11	58%
<i>Protein synthesis and fate</i>							
Chaperone protein, DnaK	Bpr_I2625	Y	1.2e-12	4.5	66.8	18	34%
GTP-binding protein, YchF	Bpr_I1872	Y	7.6e-09	4.5	39.9	13	32%
Peptidase, U62 family	Bpr_I2456	N	2.4e-05	4.8	52.3	10	23%
Peptidyl-prolyl cis-trans isomerase cyclophilin-type	Bpr_I0739	Y	5.9e-05	4.5	19.9	10	56%
Ribosomal protein L10, RplJ	Bpr_I2371	Y	5.8e-07	4.6	20.4	11	60%
Ribosomal protein L4, RplD	Bpr_I0595	Y	8.4e-06	10.2	22.9	9	33%
Ribosomal protein L7/L12, RplL	Bpr_I2370	Y	7.6e-08	4.5	13.0	8	55%
Translation elongation factor EF-G-like protein	Bpr_I1098	N	3.8e-17	4.9	77.3	38	51%
Translation elongation factor G, Fusa	Bpr_I2365	Y	9.6e-11	4.6	78.0	13	25%
Translation elongation factor Tu, TufA	Bpr_I2364	Y	1.9e-15	4.8	43.6	19	57%
<i>Vitamins and cofactors</i>							
DNA-directed RNA polymerase α -subunit, RpoA	Bpr_I0623	Y	2.4e-11	4.4	35.1	15	44%

Table B.2. Summary of the non-polysaccharide degrading proteins identified in the *B. proteoclasticus* cytosol by 2-DE.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Amino acid biosynthesis</u>								
Acetylornithine aminotransferase, ArgD	Bpr_I1809	C	N	1.7e-04	5.1	43.9	11	34%
Aspartate/tyrosine/aromatic aminotransferase	Bpr_I2631	C	N	3.0e-14	4.7	43.8	15	46%
Aspartate-semialdehyde dehydrogenase, Asd	Bpr_I1664	C	Y	7.6e-18	5.5	40.1	17	50%
Branched-chain amino acid aminotransferase, IlvE	Bpr_I1650	C	N	2.4e-12	5.2	39.2	13	32%
Cysteine synthase, CysK	Bpr_I1089	C	N	1.9e-13	5.0	32.3	18	72%
Diaminopimelate dehydrogenase	Bpr_I0298	C	Y	9.6e-16	5.6	35.8	16	49%
Dihydrodipicolinate reductase, DapB	Bpr_I2453	C	N	2.7e-06	4.9	27.0	9	46%
Glu/Leu/Phe/Val dehydrogenase	Bpr_I2129	C	Y	1.2e-30	5.4	48.6	31	64%
Imidazole glycerol phosphate synthase glutamine amidotransferase subunit	Bpr_I1240	C	N	8.0e-03	4.7	22.5	8	44%
Ketol-acid reductoisomerase, IlvC	Bpr_I1657	C	Y	3.8e-16	5.1	37.3	20	64%
NADPH-dependent glutamate synthase, GltA3	Bpr_I1306	C	Y	1.9e-04	5.0	49.3	8	20%
OAH/OAS sulfhydrylase	Bpr_I2467	C	Y	9.6e-19	5.6	45.9	16	56%
Phosphoserine aminotransferase, SerC	Bpr_I1341	C	N	2.4e-15	5.0	40.1	20	59%
Pyrroline-5-carboxylate reductase, ProC	Bpr_I2765	C	N	4.6e-04	5.2	28.1	7	29%
Threonine synthase, ThrC	Bpr_I1058	C	N	1.5e-10	4.8	54.4	17	48%
<u>Carbohydrate metabolism</u>								
1-phosphofructokinase, PfkB	Bpr_I2103	C	N	4.8e-13	5.0	32.4	13	56%
2,3-bisphosphoglycerate-independent phosphoglycerate mutase, GpmA	Bpr_I1294	C	Y	6.1e-18	4.7	56.5	18	48%

Table continues

Table B.2 continued

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
2-dehydro-3-deoxygluconokinase, KdgK	Bpr_I0932	C	Y	3.8e-14	5.5	37.2	18	34%
2-keto-3-deoxygluconate 6-phosphate aldolase/2-keto-4-hydroxyglutarate aldolase	Bpr_I0931	C	Y	3.0e-14	5.7	34.3	17	62%
5-keto 4-deoxyuronate isomerase, KduI	Bpr_I0929	C	Y	8.5e-05	4.8	32.7	13	45%
6-phosphofructokinase, PfkA1	Bpr_I0224	C	N	1.5e-08	7.1	39.0	14	37%
6-phosphofructokinase, PfkA4	Bpr_I2767	C	Y	4.0e-07	8.1	34.8	14	41%
6-phosphogluconolactonase	Bpr_I1331	C	Y	6.3e-05	4.8	39.3	8	20%
Aldose 1-epimerase	Bpr_I0228	C	N	3.0e-15	4.8	38.3	16	56%
Aldose 1-epimerase family protein	Bpr_I1782	C	N	7.6e-13	5.3	33.9	14	46%
Altronate oxidoreductase, UxaB	Bpr_I1275	C	Y	2.4e-10	4.6	56.0	16	36%
Deoxyribose-phosphate aldolase, DeoC	Bpr_I1062	C	Y	7.6e-16	5.2	24.0	15	77%
Fructose-1,6-bisphosphate aldolase, FbaA	Bpr_I2903	C	Y	3.0e-13	5.1	30.5	14	49%
Galactokinase, GalK	Bpr_I2843	C	Y	1.4e-03	4.6	43.4	7	16%
Glucose-6-phosphate isomerase, Gpi	Bpr_I0035	C	Y	1.2e-15	5.0	57.1	16	41%
Glyceraldehyde-3-phosphate dehydrogenase, Gap	Bpr_I2050	C	Y	7.6e-10	5.7	36.9	11	34%
Lactaldehyde reductase, FucO	Bpr_I2065	C	Y	1.8e-03	4.7	41.2	9	17%
L-fucose isomerase related protein	Bpr_I0185	C	Y	1.5e-11	4.9	55.4	17	39%
Phosphoenolpyruvate carboxykinase, PckA	Bpr_I0091	C	Y	1.5e-14	4.9	59.5	18	32%
Phosphoglycerate kinase, Pgc	Bpr_I2049	C	Y	6.1e-19	5.4	43.7	22	59%
Pyruvate kinase, PykA	Bpr_I0083	C	Y	1.3e-03	4.8	51.7	9	20%
Transaldolase	Bpr_I1511	C	Y	5.9e-07	4.8	23.4	11	71%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Transketolase subunit A, TktA3	Bpr_I2813	C	Y	3.4e-02	5.4	30.0	6	19%
Transketolase subunit B, TktB3	Bpr_I2812	C	Y	2.4e-07	5.5	33.0	10	37%
Triosephosphate isomerase, TpiA	Bpr_I2048	C	Y	1.2e-18	4.6	26.8	18	64%
UDP-galactose 4-epimerase, GalE	Bpr_I0192	C	Y	4.8e-14	5.2	37.2	18	47%
UTP-glucose-1-phosphate uridylyltransferase, GalU	Bpr_I0191	C	Y	1.9e-20	4.6	45.8	23	56%
Xylulokinase, XylB	Bpr_I0173	C	Y	1.5e-10	4.9	53.7	14	38%
<u>Cell cycle</u>								
DNA polymerase III β -subunit, DnaN	Bpr_I0002	C	N	1.2e-13	4.7	41.4	17	50%
Single-stranded DNA binding protein, Ssb1	Bpr_I0240	C	N	7.6e-10	4.7	16.6	12	63%
<u>Cell envelope</u>								
Aminotransferase DegT/DnrJ/EryC1/StrS family	Bpr_I0345	C	Y	4.8e-11	5.8	44.9	14	33%
Aminotransferase DegT/DnrJ/EryC1/StrS family	Bpr_I2311	C	N	3.8e-10	5.0	51.0	18	35%
Aminotransferase DegT/DnrJ/EryC1/StrS family	Bpr_I2543	C	N	1.2e-08	5.1	47.3	14	37%
Cell wall binding domain-containing protein	Bpr_I0264	S	Y	3.0e-16	4.3	195.3	31	19%
dTDP-4-dehydrorhamnose 3,5-epimerase, RfbC2	Bpr_I0548	C	Y	9.6e-11	4.7	20.6	11	49%
dTDP-4-dehydrorhamnose reductase, RfbD1	Bpr_I2582	C	N	3.0e-07	4.8	32.1	9	41%
Glycosyl transferase GT2 family	Bpr_I2565	C	N	2.3e-02	5.2	38.1	8	29%
NAD dependent epimerase/dehydratase	Bpr_I2647	C	N	1.5e-14	5.5	38.8	17	48%
NAD dependent epimerase/dehydratase	Bpr_I2930	C	N	4.8e-11	8.8	31.5	11	44%
NAD-dependent epimerase/dehydratase	Bpr_I0517	C	N	1.9e-21	4.9	42.0	18	51%

Table continues

Table B.2 continued

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
NAD-dependent epimerase/dehydratase	Bpr_I2310	C	N	1.5e-02	5.0	34.2	10	33%
NAD-dependent epimerase/dehydratase	Bpr_I2537	C	Y	1.5e-13	4.7	36.0	14	52%
Nucleotide sugar dehydrogenase	Bpr_I0828	C	N	4.7e-02	5.0	46.1	8	18%
Phosphoglucomutase/phosphomannomutase family protein	Bpr_I0554	C	Y	7.6e-09	4.7	66.0	15	20%
S-adenosyl-methyltransferase, MrkW	Bpr_I1869	C	N	1.9e-09	7.0	35.3	13	41%
<u>Cellular processes</u>								
Chemotaxis protein methyltransferase, CheR	Bpr_I2033	C	N	1.5e-09	8.4	30.5	14	42%
Chemotaxis-specific methylesterase, CheB	Bpr_I1384	C	N	4.6e-06	8.1	38.6	10	32%
Flagellar hook protein, FlgE1	Bpr_I1367	C	Y	1.5e-04	4.2	112.3	12	14%
Flagellar motor switch protein, FliG	Bpr_I1359	C	N	9.6e-09	4.3	38.6	9	33%
Superoxide dismutase, SodA	Bpr_I0467	C	N	1.9e-13	5.3	24.1	11	61%
<u>Central metabolism</u>								
Agmatine deiminase, AguA	Bpr_I1200	C	N	1.2e-08	4.2	49.5	12	30%
Carboxynorspermidine dehydrogenase	Bpr_I1198	C	Y	6.9e-02	4.6	47.4	9	20%
N-carbamoylputrescine amidohydrolase, AguB	Bpr_I1201	C	N	4.8e-15	5.0	33.8	16	56%
S-adenosylmethionine synthetase, MetK	Bpr_I2769	C	Y	4.8e-16	4.9	43.1	16	38%
<u>Energy metabolism</u>								
2-enoate reductase	Bpr_I1763	C	N	1.3e-04	7.3	76.0	9	17%
2-enoate reductase	Bpr_I1977	C	N	4.8e-02	6.2	77.4	4	12%
3-hydroxybutyryl-CoA dehydrogenase, Hbd	Bpr_I2486	S	Y	6.1e-14	5.2	31.2	13	47%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
ATP synthase F1 α -subunit, AtpA1	Bpr_I0166	C	N	1.9e-10	5.3	55.0	15	31%
ATP synthase F1 β -subunit, AtpD1	Bpr_I0168	C	Y	4.8e-16	4.6	50.3	15	41%
Butyrate kinase, Buk	Bpr_I2323	C	Y	6.1e-22	5.1	38.7	18	66%
Butyryl-CoA dehydrogenase, Bcd	Bpr_I2485	C	Y	3.8e-17	6.1	42.5	21	60%
Crotonase, Crt	Bpr_I2487	C	Y	1.2e-14	4.7	28.0	19	65%
Electron transfer flavoprotein α -subunit, EtfA	Bpr_I2483	C	Y	1.2e-08	4.8	37.4	12	36%
Electron transfer flavoprotein β -subunit, EtfB	Bpr_I2484	C	Y	1.5e-11	5.2	28.4	14	53%
Isocitrate dehydrogenase, Icd	Bpr_I1102	C	N	4.8e-13	5.1	45.1	18	42%
Methylmalonyl-CoA decarboxylase α -subunit, MmdA	Bpr_I1226	C	Y	3.9e-07	4.5	50.9	12	28%
Phosphate acetyltransferase, Pta	Bpr_III011	C	Y	1.3e-02	5.2	14.4	7	54%
Phosphate butyryltransferase, Ptb	Bpr_I2324	C	Y	4.8e-16	5.1	33.3	19	72%
Pyruvate formate lyase, PflB	Bpr_I0112	C	Y	3.0e-24	5.7	84.9	30	50%
Pyruvate phosphate dikinase, PpdK	Bpr_I1154	C	Y	7.6e-15	4.8	104.3	28	28%
Pyruvate:ferredoxin oxidoreductase	Bpr_I0269	C	Y	2.4e-14	5.2	127.3	25	21%
Thiolase, ThlA1	Bpr_I2488	C	Y	2.4e-14	5.0	42.0	15	51%
Thiolase, ThlA2	Bpr_I2475	S	Y	3.8e-08	6.0	41.0	12	24%
Thioredoxin-disulfide reductase, TrxB	Bpr_I2491	C	Y	9.6e-08	4.5	33.7	13	43%
<i>Hypothetical</i>								
Hypothetical protein	Bpr_I0280	C	N	3.8e-15	8.4	23.2	16	76%
Hypothetical protein	Bpr_I0417	C	N	7.1e-03	4.9	39.8	11	29%

Table continues

Table B.2 continued

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Hypothetical protein	Bpr_I0592	C	N	6.1e-10	7.3	32.3	11	38%
Hypothetical protein	Bpr_I0874	C	N	1.0e-05	4.5	14.8	5	36%
Hypothetical protein	Bpr_I0962	C	N	3.8e-09	4.3	17.3	9	49%
Hypothetical protein	Bpr_I1177	C	N	2.4e-11	4.3	24.7	12	65%
Hypothetical protein	Bpr_I1179	C	Y	1.2e-07	4.7	24.9	11	45%
Hypothetical protein	Bpr_I1180	C	Y	9.6e-18	7.4	36.4	18	53%
Hypothetical protein	Bpr_I1213	C	Y	1.4e-05	4.3	26.5	8	39%
Hypothetical protein	Bpr_I1281	C	N	1.5e-13	4.6	34.1	18	52%
Hypothetical protein	Bpr_I1477	C	N	2.9e-04	4.2	19.1	5	49%
Hypothetical protein	Bpr_I1608	C	N	3.1e-06	4.6	47.5	10	26%
Hypothetical protein	Bpr_I2291	S	Y	7.1e-07	4.2	42.6	8	25%
Hypothetical protein	Bpr_I2455	C	N	6.1e-26	4.7	47.7	26	58%
Hypothetical protein	Bpr_I2494	C	N	6.1e-08	4.3	33.0	11	41%
Hypothetical protein	Bpr_I2583	S	Y	8.7e-06	4.3	44.3	9	24%
Hypothetical protein	Bpr_I2606	C	N	4.2e-05	4.8	41.1	13	31%
Hypothetical protein	Bpr_I2608	C	N	6.1e-16	7.3	32.8	17	44%
Hypothetical protein	Bpr_I2619	C	Y	3.8e-08	4.5	8.6	6	46%
Hypothetical protein	Bpr_I2634	C	N	4.8e-07	9.8	16.4	8	50%
Hypothetical protein	Bpr_I2744	C	N	2.4e-08	4.8	28.5	10	51%
Hypothetical protein	Bpr_III133	C	N	1.6e-03	5.6	57.9	10	16%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Lipid metabolism</u>								
3-oxoacyl-(acyl-carrier-protein) synthase, FabF	Bpr_I1268	C	N	2.0e-06	5.3	44.5	11	28%
Glycerol kinase, GlpK	Bpr_I1744	C	N	3.6e-05	4.9	55.6	9	18%
<u>Nucleic acid metabolism</u>								
Excinuclease ABC A subunit, UvrA	Bpr_I2597	C	N	3.8e-12	7.4	106.8	20	26%
Type I restriction modification system S subunit	Bpr_IV100	C	N	1.5e-02	6.1	46.5	7	18%
<u>Nucleotide metabolism</u>								
5-aminoimidazole-4-carboxamide ribonucleotide transformylase	Bpr_I0732	C	Y	1.5e-12	5.1	44.4	12	38%
Adenylate kinase, Adk	Bpr_I0615	C	N	1.5e-10	5.1	23.8	12	72%
Adenylosuccinate lyase, PurB	Bpr_I2212	C	Y	1.2e-10	5.4	53.7	16	28%
Dihydroorotate dehydrogenase, PyrD	Bpr_I2788	C	N	7.6e-10	5.1	32.3	13	47%
IMP cyclohydrolase, PurO	Bpr_I0731	C	N	4.0e-06	4.7	32.2	7	21%
Inosine-uridine preferring nucleoside hydrolase	Bpr_I0724	C	N	7.4e-03	4.5	32.9	9	25%
Phosphoribosylamine-glycine ligase, PurD	Bpr_I0870	C	N	1.8e-06	4.6	46.3	12	29%
Phosphoribosylaminoimidazole-succinocarboxamide synthase, PurC	Bpr_I1144	C	Y	3.8e-18	4.7	33.3	21	75%
Purine nucleoside phosphorylase, DeoD2	Bpr_I2948	C	Y	4.8e-13	5.1	26.1	16	55%
Purine-nucleoside phosphorylase, DeoD1	Bpr_I1527	C	Y	3.0e-12	4.6	29.5	13	67%
Uridine phosphorylase, Udp	Bpr_I1561	C	N	9.6e-13	5.4	28.1	14	69%
Uridylate kinase, PyrH	Bpr_I0953	C	N	1.9e-13	5.4	25.5	15	55%

Table continues

Table B.2 continued

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Plasmid functions</u>								
Single stranded DNA binding protein, Ssb3	Bpr_II423	C	N	1.9e-10	4.7	16.3	12	60%
<u>Protein fate</u>								
Chaperone protein, DnaK	Bpr_I2625	C	Y	9.6e-10	4.5	66.8	16	28%
Chaperonin, GroEL	Bpr_I1193	C	Y	1.1e-06	5.0	60.0	19	32%
Oligoendopeptidase, PepF1	Bpr_I0779	C	Y	1.9e-13	4.7	68.9	22	32%
Peptidase M16 family	Bpr_I1215	C	N	3.0e-39	4.6	111.5	44	46%
Peptidase U62 family	Bpr_I2456	C	N	6.1e-09	4.8	52.3	15	33%
Peptidyl-prolyl cis-trans isomerase FKBP-type	Bpr_I1031	C	Y	2.5e-06	4.1	33.4	8	25%
Serine protease subtilisin family	Bpr_I2629	S	Y	2.8e-05	3.8	153.3	15	12%
<u>Protein synthesis</u>								
Aspartyl-tRNA synthetase, AspS	Bpr_I0140	C	Y	3.8e-14	4.7	69.2	23	35%
Endoribonuclease L-PSP	Bpr_I2149	C	N	2.8e-05	4.4	13.5	4	56%
Methionyl-tRNA synthetase, MetG	Bpr_I2807	C	Y	9.6e-09	4.9	76.4	21	32%
Polyribonucleotide nucleotidyltransferase	Bpr_I0975	C	Y	7.6e-15	4.9	76.5	21	37%
Ribosomal protein L1, RplA	Bpr_I2299	C	Y	1.2e-12	9.8	24.6	16	58%
Ribosomal protein L10, RplJ	Bpr_I2371	C	Y	1.8e-04	4.6	20.4	8	32%
Ribosomal protein L13, RplM	Bpr_I0630	C	Y	6.1e-08	9.9	16.2	11	77%
Ribosomal protein L18, RplR	Bpr_I0610	C	Y	3.0e-06	10.5	13.5	9	48%
Ribosomal protein L21, RplU	Bpr_I1443	C	Y	7.6e-03	10.3	11.3	5	41%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Ribosomal protein L22, RplV	Bpr_I0599	C	Y	4.4e-06	10.7	15.1	8	56%
Ribosomal protein L3, RplC	Bpr_I0594	C	Y	1.9e-24	10.6	24.3	25	84%
Ribosomal protein L4, RplD	Bpr_I0595	C	Y	4.0e-07	10.2	22.9	9	47%
Ribosomal protein L5, RplE	Bpr_I0606	C	Y	3.8e-11	9.8	20.4	15	63%
Ribosomal protein L6, RplF	Bpr_I0609	C	Y	2.4e-15	10.3	19.4	14	60%
Ribosomal protein S1, RpsA	Bpr_I2035	C	Y	1.5e-19	4.6	41.9	21	54%
Ribosomal protein S2, RpsB	Bpr_I1395	C	Y	9.6e-09	8.4	27.6	11	39%
Ribosomal protein S3, RpsC	Bpr_I0600	C	Y	9.6e-08	9.6	24.6	12	53%
Ribosomal protein S4, RpsD	Bpr_I0622	C	Y	2.4e-11	10.6	24.1	13	43%
Ribosomal protein S5, RpsE	Bpr_I0611	C	Y	3.0e-17	10.5	17.9	15	79%
Ribosomal protein S7, RpsG	Bpr_I2366	C	Y	1.2e-11	10.5	17.5	13	70%
Ribosomal protein S9, RpsI	Bpr_I0631	C	Y	3.8e-13	11.1	15.4	11	67%
Single stranded nucleic acid binding protein	Bpr_I2957	C	Y	3.0e-10	6.3	28.3	14	52%
Threonyl-tRNA synthetase, ThrZ	Bpr_I2374	C	Y	1.2e-11	5.0	73.7	22	26%
Translation elongation factor EF-G-like protein	Bpr_I1098	C	N	3.0e-26	4.9	77.3	31	46%
Translation elongation factor G, Fusa	Bpr_I2365	C	Y	9.6e-15	4.6	78.0	21	27%
Translation elongation factor Ts, Tsf	Bpr_I1396	C	Y	3.0e-03	4.8	34.9	10	33%
Translation elongation factor Tu, TufA	Bpr_I2364	C	Y	3.0e-21	4.8	43.6	19	62%
tRNA (guanine-N7)-methyltransferase, TrmB	Bpr_I2923	C	N	4.4e-07	7.6	25.1	10	51%

Table continues

Table B.2 continued

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Regulation</u>								
Anti-sigma factor antagonist	Bpr_I0800	C	Y	4.9e-07	4.2	10.8	7	89%
Anti-sigma factor antagonist/phosphotransferase domain-containing protein	Bpr_I0249	C	N	9.6e-14	4.6	49.8	20	48%
Cold shock domain protein, CspD1	Bpr_I2451	C	Y	2.3e-06	4.6	7.3	5	89%
<u>Signal transduction</u>								
GGDEF domain-containing protein	Bpr_I1183	C	N	1.8e-05	4.6	53.0	10	24%
Hpr kinase/phosphatase, HprK	Bpr_I0099	C	N	1.3e-04	5.1	35.9	11	40%
PhoH family protein	Bpr_I1309	C	N	1.2e-07	8.6	38.6	10	34%
S-ribosylhomocysteinase, LuxS	Bpr_I0033	C	Y	7.6e-13	5.0	18.2	11	54%
<u>Transcription</u>								
DNA directed RNA polymerase β -subunit, RpoB	Bpr_I2369	C	Y	9.6e-13	4.6	145.5	22	20%
DNA directed RNA polymerase β' -subunit, RpoC	Bpr_I2368	C	Y	1.5e-23	6.6	139.3	36	30%
DNA-directed RNA polymerase α -subunit, RpoA	Bpr_I0623	C	Y	3.8e-18	4.4	35.1	17	58%
Transcription antitermination factor, NusB	Bpr_I0979	C	N	7.3e-07	4.3	15.3	8	57%
<u>Transporters</u>								
ABC transporter ATP-binding protein	Bpr_I2802	C	Y	6.1e-11	6.7	41.1	17	43%
Amino acid ABC transporter substrate-binding protein	Bpr_I1826	S	Y	5.8e-04	4.0	31.6	7	21%
Oligopeptide ABC transporter substrate-binding protein, OppA1	Bpr_I1276	S	Y	1.2e-20	4.0	83.1	25	35%
Peptide/nickel ABC transporter periplasmic protein	Bpr_I2750	S	Y	8.0e-07	3.9	57.9	9	23%
Peptide/nickel ABC transporter substrate-binding protein	Bpr_I1860	S	Y	7.6e-11	4.1	58.4	16	41%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Sugar ABC transporter substrate-binding protein	Bpr_I0182	S	Y	7.6e-07	4.1	63.3	13	27%
Sugar ABC transporter substrate-binding protein	Bpr_I0237	S	Y	8.5e-03	3.9	64.8	6	14%
Sugar ABC transporter substrate-binding protein	Bpr_I0313	S	Y	1.1e-03	3.8	55.2	7	11%
Sugar ABC transporter substrate-binding protein	Bpr_I0937	S	Y	9.6e-14	4.1	61.3	15	35%
Sugar ABC transporter substrate-binding protein	Bpr_I1589	S	Y	7.3e-06	3.9	61.1	13	34%
Sugar ABC transporter substrate-binding protein	Bpr_I1667	S	Y	7.6e-11	3.9	47.8	11	42%
Sugar ABC transporter substrate-binding protein	Bpr_I1720	S	Y	7.4e-04	4.0	49.3	9	18%
Sugar ABC transporter substrate-binding protein	Bpr_I2010	S	Y	1.9e-06	3.7	47.3	9	35%
Sugar ABC transporter substrate binding protein	Bpr_I2344	S	Y	4.4e-04	3.9	52.3	10	27%
Sugar ABC transporter substrate-binding protein	Bpr_I2443	S	Y	1.9e-11	4.0	57.0	16	31%
Xylose ABC transporter substrate-binding protein	Bpr_I1173	S	Y	9.6e-12	4.2	38.4	14	57%
<u>Unknown function</u>								
Acetyltransferase	Bpr_I1661	C	N	4.8e-08	6.7	24.9	10	49%
Acetyltransferase GNAT family	Bpr_I0656	C	N	3.0e-07	6.7	21.0	9	51%
Aminotransferase domain-containing protein	Bpr_I1980	C	N	1.5e-11	5.5	48.3	12	24%
FAD dependent oxidoreductase	Bpr_I2038	C	N	2.4e-06	6.7	45.2	10	26%
Hydrolase α/β -fold family	Bpr_I2806	C	N	6.1e-08	8.6	36.5	14	46%
NUDIX domain-containing protein	Bpr_I1938	C	N	7.6e-12	4.2	18.8	12	70%
Phosphoribulokinase/uridine kinase family protein	Bpr_I0923	C	N	1.9e-16	6.9	63.1	18	35%
Pyridoxamine 5'-phosphate oxidase family protein	Bpr_I1023	C	N	1.2e-05	5.3	15.0	6	61%

Table continues

Table B.2 continued

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
UBA/TS-N domain-containing protein	Bpr_III040	C	N	4.5e-07	4.5	13.6	8	42%
<i>Vitamins and cofactors</i>								
Cobalamin biosynthesis protein. CobW1	Bpr_I2021	C	Y	3.8e-08	4.4	36.9	9	32%
FeS assembly protein. SufD	Bpr_I0053	C	N	7.6e-12	4.5	42.7	11	37%
Methylenetetrahydrofolate dehydrogenase/cyclohydrolase. FolD	Bpr_I1303	C	N	1.5e-05	5.0	30.6	9	27%

Table B.3. Summary of the non-polysaccharidase degrading proteins identified in the *B. proteoclasticus* cytosol by 1-D LC-MS/MS.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<i>Amino acid biosynthesis</i>								
Anthranilate phosphoribosyltransferase, TrpD	Bpr_I0012	C	N	1.5e-12	5.1	36.6		26%
Aspartate/tyrosine/aromatic aminotransferase	Bpr_I2631	C	N	2.3e-13	4.7	43.8		18%
Aspartate-semialdehyde dehydrogenase, Asd	Bpr_I1664	C	Y	8.2e-14	5.5	40.1		54%
ATP phosphoribosyltransferase regulatory subunit, HisZ	Bpr_I1451	C	N	2.0e-06	4.5	41.3		11%
Branched-chain amino acid aminotransferase, IlvE	Bpr_I1650	C	N	3.4e-08	5.2	39.2		21%
Chorismate mutase/prephenate dehydratase, PheA	Bpr_I1730	C	N	1.4e-09	5.0	42.4		13%
Cysteine synthase, CysK	Bpr_I1089	C	N	8.9e-09	5.0	32.3		14%
D-3-phosphoglycerate dehydrogenase, SerA	Bpr_I1342	C	N	4.4e-13	5.3	42.0		26%
Diaminopimelate decarboxylase, LysA	Bpr_I1243	C	N	4.8e-12	4.8	48.0		9%
Diaminopimelate dehydrogenase	Bpr_I0298	C	Y	9.9e-15	5.6	35.8		41%
Diaminopimelate epimerase, DapF	Bpr_I2388	C	N	2.2e-05	4.4	33.5		6%
Dihydroxy-acid dehydratase, IlvD	Bpr_I1338	C	N	1.0e-30	6.0	58.9		30%
Glu/Leu/Phe/Val dehydrogenase	Bpr_I2129	C	Y	1.0e-30	5.4	48.6		57%
Histidinol dehydrogenase, HisD	Bpr_I1449	C	N	7.8e-09	4.8	47.0		12%
Ketol-acid reductoisomerase, IlvC	Bpr_I1657	C	Y	2.2e-15	5.1	37.3		64%
NADPH-dependent glutamate synthase, GltA3	Bpr_I1306	C	Y	1.0e-30	5.0	49.3		17%
OAH/OAS sulfhydrylase	Bpr_I0317	C	N	6.4e-11	5.4	45.6		12%
OAH/OAS sulfhydrylase	Bpr_I2467	C	Y	2.1e-14	5.6	45.9		25%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Phosphoserine aminotransferase, SerC	Bpr_I1341	C	N	3.8e-14	5.0	40.1		51%
Phosphoserine phosphatase/homoserine phosphotransferase bifunctional protein, ThrH	Bpr_I1141	C	N	1.4e-11	4.5	22.4		10%
Serine hydroxymethyltransferase, GlyA	Bpr_I1796	C	Y	8.6e-13	6.2	44.9		21%
Threonine dehydratase, IlvA	Bpr_I1571	C	N	5.4e-10	6.0	44.4		8%
Threonine synthase, ThrC	Bpr_I1058	C	N	1.1e-16	4.8	54.4		35%
Tryptophan synthase α -subunit, TrpA	Bpr_I0008	C	N	2.0e-07	4.6	27.7		10%
Tryptophan synthase β -subunit, TrpB2	Bpr							

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Glucokinase, Glk	Bpr_I0100	C	N	2.5e-10	4.9	32.9		19%
Glucose-1-phosphate adenylyltransferase, GlgC1	Bpr_I0290	C	N	8.1e-10	4.7	46.9		18%
Glucose-6-phosphate isomerase, Gpi	Bpr_I0035	C	Y	1.0e-30	5.0	57.1		53%
Glucuronate isomerase, UxaC	Bpr_I1591	C	Y	1.9e-12	5.0	54.4		13%
Glyceraldehyde-3-phosphate dehydrogenase, Gap	Bpr_I2050	C	Y	1.0e-30	5.7	36.9		41%
Glycogen synthase ADP-glucose type, GlgA	Bpr_I1257	C	N	1.3e-07	5.6	56.0		9%
Lactaldehyde reductase, FucO	Bpr_I2065	C	Y	1.1e-15	4.7	41.2		35%
L-fucose isomerase related protein	Bpr_I0185	C	Y	1.0e-30	4.9	55.4		62%
L-ribulose-5-phosphate 4-epimerase, AraD	Bpr_I2814	C	Y	1.0e-10	6.0	26.0		11%
Phosphoenolpyruvate carboxykinase, PckA	Bpr_I0091	C	Y	1.0e-30	4.9	59.5		56%
Phosphoglycerate kinase, Pgk	Bpr_I2049	C	Y	1.0e-30	5.4	43.7		69%
Pyruvate kinase, PykA	Bpr_I0083	C	Y	1.6e-08	4.8	51.7		9%
Ribose-phosphate pyrophosphokinase, PrsA1	Bpr_I0286	C	Y	2.7e-13	5.7	45.7		22%
Transaldolase	Bpr_I1511	C	Y	4.7e-14	4.8	23.4		70%
Transketolase subunit A, TktA3	Bpr_I2813	C	Y	1.1e-16	5.4	30.0		29%
Transketolase subunit B, TktB3	Bpr_I2812	C	Y	1.0e-30	5.5	33.0		37%
Triosephosphate isomerase, TpiA	Bpr_I2048	C	Y	1.0e-30	4.6	26.8		84%
Xylulokinase, XylB	Bpr_I0173	C	Y	1.8e-14	4.9	53.7		27%
<i>Cell cycle</i>								
Cell division protein, FtsZ	Bpr_I1862	C	Y	3.7e-14	5.0	42.9		27%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
DNA polymerase III β -subunit, DnaN	Bpr_I0002	C	N	5.1e-12	4.7	41.4		25%
FtsK/SpoIIIE family protein	Bpr_I1301	M	N	6.7e-05	5.0	105.3		5%
Single-stranded DNA binding protein, Ssb1	Bpr_I0240	C	N	2.2e-14	4.7	16.6		27%
<i>Cell envelope</i>								
Aminotransferase DegT/DnrJ/EryC1/StrS family	Bpr_I0345	C	Y	1.1e-16	5.8	44.9		31%
Aminotransferase DegT/DnrJ/EryC1/StrS family	Bpr_I2543	C	N	1.9e-11	5.1	47.3		11%
dTDP-4-dehydrorhamnose 3,5-epimerase, RfbC2	Bpr_I0548	C	Y	2.7e-08	4.7	20.6		14%
dTDP-4-dehydrorhamnose reductase, RfbD1	Bpr_I2582	C	N	4.5e-10	4.8	32.1		15%
dTDP-glucose 4,6-dehydratase, RfbB	Bpr_I0545	C	Y	2.2e-11	5.5	44.1		20%
Glycosyl transferase GT28 family	Bpr_I2555	C	N	9.5e-07	4.9	40.1		9%
NAD dependent epimerase/dehydratase	Bpr_I0348	C	Y	3.2e-12	5.9	38.9		17%
NAD dependent epimerase/dehydratase	Bpr_I2647	C	N	2.1e-10	5.5	38.8		8%
NAD-dependent epimerase/dehydratase	Bpr_I0517	C	N	5.4e-14	4.9	42.0		9%
NAD-dependent epimerase/dehydratase	Bpr_I2537	C	Y	1.1e-15	4.7	36.0		23%
Nucleotide sugar dehydrogenase	Bpr_I0828	C	N	2.6e-11	5.0	46.1		11%
Nucleotide sugar dehydrogenase	Bpr_I2538	C	Y	1.8e-09	4.9	48.6		6%
Oxidoreductase GFO/IDH/MOCA family	Bpr_I0406	C	N	2.2e-12	5.6	40.5		9%
Phosphoglucosyltransferase/phosphomannosyltransferase family protein	Bpr_I0554	C	Y	1.0e-30	4.7	66.0		38%
Polysaccharide biosynthesis protein	Bpr_I2562	C	Y	5.3e-14	5.1	43.6		23%
Rod shape-determining protein, MreB1	Bpr_I1486	C	Y	1.1e-15	5.5	36.6		20%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
UDP-galactopyranose mutase, Glf	Bpr_I2305	C	N	3.6e-11	4.8	46.2		9%
<i>Cellular processes</i>								
Flagellin, FliC1	Bpr_I0488	C	Y	1.7e-10	5.2	31.4		27%
Flagellin, FliC2	Bpr_I0489	C	Y	1.0e-30	5.2	30.8		28%
Flavin reductase domain-containing protein	Bpr_I0363	C	N	7.5e-11	5.1	23.4		18%
Rubrerythrin, Rbr1	Bpr_I0031	C	Y	3.4e-09	5.1	19.7		34%
Rubrerythrin, Rbr2	Bpr_I0362	C	Y	1.6e-12	4.9	21.3		34%
Superoxide dismutase, SodA	Bpr_I0467	C	N	1.0e-30	5.3	24.1		67%
<i>Central metabolism</i>								
Carboxynorspermidine dehydrogenase	Bpr_I1198	C	Y	2.9e-11	4.6	47.4		9%
Formate-tetrahydrofolate ligase, Fhs	Bpr_I1302	C	N	1.0e-30	6.0	60.0		24%
S-adenosylmethionine synthetase, MetK	Bpr_I2769	C	Y	1.9e-13	4.9	43.1		57%
<i>Energy metabolism</i>								
3-hydroxybutyryl-CoA dehydrogenase, Hbd	Bpr_I2486	S	Y	1.0e-30	5.2	31.2		71%
Acetate kinase, AckA	Bpr_I1580	C	Y	7.0e-13	5.9	42.7		29%
Acetyl-CoA synthetase, AscA	Bpr_I1274	C	N	2.3e-13	6.0	64.9		9%
Aconitate hydratase, AcnA	Bpr_I2797	C	N	5.9e-10	5.3	82.7		8%
ATP synthase F1 α -subunit, AtpA1	Bpr_I0166	C	N	1.9e-12	5.3	55.0		27%
ATP synthase F1 β -subunit, AtpD1	Bpr_I0168	C	Y	6.9e-13	4.6	50.3		45%
ATP synthase F1 β -subunit, AtpD2	Bpr_I1162	C	N	3.0e-08	4.5	50.9		16%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Butyrate kinase, Buk	Bpr_I2323	C	Y	9.9e-15	5.1	38.7		74%
Butyryl-CoA dehydrogenase, Bcd	Bpr_I2485	C	Y	1.0e-30	6.1	42.5		73%
Crotonase, Crt	Bpr_I2487	C	Y	1.0e-30	4.7	28.0		55%
Electron transfer flavoprotein α -subunit, EtfA	Bpr_I2483	C	Y	1.0e-30	4.8	37.4		46%
Electron transfer flavoprotein β -subunit, EtfB	Bpr_I2484	C	Y	2.2e-15	5.2	28.4		44%
Iron-containing alcohol dehydrogenase	Bpr_I1425	C	Y	1.0e-30	5.1	44.4		52%
Isocitrate dehydrogenase, Icd	Bpr_I1102	C	N	1.1e-15	5.1	45.1		16%
Malate dehydrogenase, Mdh	Bpr_I2277	C	N	2.4e-09	4.8	41.4		9%
Methylmalonyl-CoA decarboxylase α -subunit, MmdA	Bpr_I1226	C	Y	5.3e-14	4.5	50.9		24%
Oxaloacetate decarboxylase α -subunit, OadA	Bpr_I1230	C	Y	1.1e-15	5.2	52.4		18%
Phosphate acetyltransferase, Pta	Bpr_III010	C	N	2.2e-16	4.7	22.3		32%
Phosphate acetyltransferase, Pta	Bpr_III011	C	Y	7.9e-13	5.2	14.4		59%
Phosphate butyryltransferase, Ptb	Bpr_I2324	C	Y	5.6e-16	5.1	33.3		57%
Pyruvate carboxyltransferase domain-containing protein	Bpr_I1100	C	N	5.4e-05	5.4	57.2		7%
Pyruvate formate lyase, PflB	Bpr_I0112	C	Y	1.0e-30	5.7	84.9		65%
Pyruvate phosphate dikinase, PpdK	Bpr_I1154	C	Y	1.0e-30	4.8	104.3		52%
Pyruvate:ferredoxin oxidoreductase	Bpr_I0269	C	Y	1.0e-30	5.2	127.3		48%
Thiolase, ThlA1	Bpr_I2488	C	Y	1.0e-30	5.0	42.0		68%
Thiolase, ThlA2	Bpr_I2475	S	Y	1.0e-30	6.0	41.0		66%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<i>Hypothetical</i>								
Hypothetical protein	Bpr_I0212	C	Y	1.1e-16	5.2	41.6		28%
Hypothetical protein	Bpr_I0219	S	Y	8.9e-11	3.9	30.3		5%
Hypothetical protein	Bpr_I0280	C	N	2.4e-11	8.4	23.2		15%
Hypothetical protein	Bpr_I0351	C	N	3.6e-14	5.0	34.4		6%
Hypothetical protein	Bpr_I0566	M	N	5.4e-09	9.9	28.0		5%
Hypothetical protein	Bpr_I0799	C	N	1.9e-08	4.1	13.5		14%
Hypothetical protein	Bpr_I0801	C	N	1.1e-11	4.3	23.4		21%
Hypothetical protein	Bpr_I0874	C	N	2.1e-10	4.5	14.8		13%
Hypothetical protein	Bpr_I0876	S	N	2.0e-04	4.2	63.1		3%
Hypothetical protein	Bpr_I0952	C	Y	2.2e-15	3.8	37.2		26%
Hypothetical protein	Bpr_I1041	C	N	8.1e-12	4.1	9.3		39%
Hypothetical protein	Bpr_I1177	C	N	3.6e-13	4.3	24.7		9%
Hypothetical protein	Bpr_I1179	C	Y	1.0e-30	4.7	24.9		51%
Hypothetical protein	Bpr_I1180	C	Y	1.0e-30	7.4	36.4		66%
Hypothetical protein	Bpr_I1300	C	Y	3.6e-12	4.7	21.8		50%
Hypothetical protein	Bpr_I1457	C	N	3.8e-09	4.2	25.9		18%
Hypothetical protein	Bpr_I1473	C	N	1.8e-10	5.4	24.1		18%
Hypothetical protein	Bpr_I1477	C	N	3.5e-12	4.2	19.1		29%
Hypothetical protein	Bpr_I1479	C	N	5.8e-07	5.3	15.7		22%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Hypothetical protein	Bpr_I2144	C	N	1.1e-09	4.8	84.8		5%
Hypothetical protein	Bpr_I2200	S	Y	2.1e-09	4.2	37.5		8%
Hypothetical protein	Bpr_I2329	C	Y	3.5e-12	4.7	23.5		44%
Hypothetical protein	Bpr_I2455	C	N	1.0e-30	4.7	47.7		26%
Hypothetical protein	Bpr_I2606	C	N	3.3e-10	4.8	41.1		16%
Hypothetical protein	Bpr_I2619	C	Y	1.6e-13	4.5	8.6		43%
Hypothetical protein	Bpr_I2633	C	N	1.7e-07	5.9	12.0		34%
Hypothetical protein	Bpr_I2744	C	N	2.7e-14	4.8	28.5		30%
Hypothetical protein	Bpr_II050	C	N	1.7e-05	8.0	12.6		20%
Hypothetical protein	Bpr_II102	C	N	5.3e-11	3.7	5.9		56%
Hypothetical protein	Bpr_III164	C	N	6.1e-14	3.9	60.2		18%
<i>Lipid metabolism</i>								
1-deoxy-D-xylulose-5-phosphate reductoisomerase, Dxr	Bpr_I0957	C	N	8.3e-09	5.5	41.8		7%
1-hydroxy-2-methyl-2-(E)-butenyl 4-diphosphate synthase, IspG	Bpr_I0959	C	N	1.6e-14	6.5	37.6		22%
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase, IspF	Bpr_I1831	C	N	1.7e-07	5.0	16.9		11%
3-oxoacyl-(acyl-carrier-protein) synthase, FabF	Bpr_I1268	C	N	1.0e-30	5.3	44.5		48%
Acetyl-CoA carboxylase biotin carboxyl carrier protein, AccB	Bpr_I1269	C	N	2.2e-15	4.5	15.1		45%
Fatty acid/phospholipid synthesis protein, PlsX	Bpr_I1576	C	N	1.6e-11	6.2	35.8		9%
Glycerol kinase, GlpK	Bpr_I1744	C	N	1.0e-30	4.9	55.6		25%
Glycerophosphoryl diester phosphodiesterase	Bpr_I2042	C	N	1.3e-05	4.3	27.3		21%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Nucleic acid metabolism</u>								
DNA-binding protein	Bpr_I1070	C	Y	1.0e-12	11.0	9.7		50%
DNA-binding protein	Bpr_II236	C	Y	4.0e-11	11.1	9.6		55%
DNA-binding protein	Bpr_III252	C	Y	2.0e-11	11.0	9.9		46%
DNA-binding protein	Bpr_IV068	C	Y	1.0e-07	11.1	9.9		13%
DNA-repair protein, RecN	Bpr_I0987	C	N	1.3e-05	4.5	62.4		0%
<u>Nucleotide metabolism</u>								
5-aminoimidazole-4-carboxamide ribonucleotide transformylase	Bpr_I0732	C	Y	2.0e-10	5.1	44.4		12%
Adenine phosphoribosyl transferase, Apt	Bpr_I1345	C	Y	3.9e-11	5.1	19.3		26%
Adenylate kinase, Adk	Bpr_I0615	C	N	2.8e-10	5.1	23.8		24%
Adenylosuccinate lyase, PurB	Bpr_I2212	C	Y	7.3e-13	5.4	53.7		12%
GMP synthase, GuaA	Bpr_I2780	C	N	1.6e-08	5.0	57.3		8%
Hypoxanthine phosphoribosyltransferase, Hpt	Bpr_I2906	C	Y	8.3e-14	4.8	19.6		25%
IMP dehydrogenase, GuaB	Bpr_I2437	C	Y	3.3e-11	6.0	52.2		11%
Orotate phosphoribosyltransferase, PyrE1	Bpr_I0863	C	Y	1.1e-15	4.7	25.4		32%
Phosphoribosylamine-glycine ligase, PurD	Bpr_I0870	C	N	1.3e-09	4.6	46.3		12%
Phosphoribosylaminoimidazole-succinocarboxamide synthase, PurC	Bpr_I1144	C	Y	6.1e-12	4.7	33.3		20%
Purine nucleoside phosphorylase, DeoD2	Bpr_I2948	C	Y	1.2e-10	5.1	26.1		37%
Purine-nucleoside phosphorylase, DeoD1	Bpr_I1527	C	Y	1.3e-14	4.6	29.5		59%
Thymidylate synthase, ThyA	Bpr_I2288	C	N	6.5e-09	5.8	32.7		6%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Uracil phosphoribosyltransferase, Upp	Bpr_I0274	C	Y	4.0e-11	6.9	24.1		15%
Uridine phosphorylase, Udp	Bpr_I1561	C	N	4.9e-11	5.4	28.1		9%
Uridylate kinase, PyrH	Bpr_I0953	C	N	3.3e-13	5.4	25.5		41%
<u>Plasmid functions</u>								
Single stranded DNA binding protein, Ssb3	Bpr_II423	C	N	1.4e-11	4.7	16.3		12%
<u>Protein fate</u>								
Aminoacyl-histidine dipeptidase, PepD	Bpr_I0786	C	Y	7.9e-10	4.5	51.7		10%
Chaperone protein, DnaK	Bpr_I2625	C	Y	6.8e-14	4.5	66.8		23%
Chaperone protein, GrpE	Bpr_I2626	C	N	1.3e-08	4.4	26.2		16%
Chaperonin, GroEL	Bpr_I1193	C	Y	1.0e-30	5.0	60.0		46%
Metallopeptidase M24 family	Bpr_I1063	C	Y	1.1e-15	4.9	68.6		7%
Oligoendopeptidase, PepF1	Bpr_I0779	C	Y	2.5e-12	4.7	68.9		20%
Peptidase M18 family	Bpr_I1080	C	Y	1.9e-09	5.0	54.9		14%
Peptidase M29 family	Bpr_I0866	C	N	1.9e-09	4.6	81.1		6%
Peptidase U62 family	Bpr_I2456	C	N	3.2e-11	4.8	52.3		13%
Peptidyl-prolyl cis-trans isomerase cyclophilin-type	Bpr_I0739	C	Y	1.4e-09	4.5	19.9		51%
Peptidyl-prolyl cis-trans isomerase FKBP-type	Bpr_I1031	C	Y	3.3e-12	4.1	33.4		9%
Trigger factor, Tig	Bpr_I2409	C	Y	1.1e-15	4.6	54.2		25%
Tripeptide aminopeptidase, PepT	Bpr_I1493	C	N	2.4e-13	4.7	45.8		12%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<i>Protein synthesis</i>								
Alanyl-tRNA synthetase, AlaS	Bpr_I1221	C	Y	3.8e-11	5.0	96.5		6%
Asparaginyl-tRNA synthetase, AsnS	Bpr_I2782	C	Y	5.4e-11	4.9	53.0		13%
Aspartyl-tRNA synthetase, AspS	Bpr_I0140	C	Y	1.7e-10	4.7	69.2		21%
Endoribonuclease L-PSP	Bpr_I2149	C	N	7.3e-10	4.4	13.5		41%
Glutamyl-tRNA synthetase, GltX	Bpr_I1497	C	Y	4.9e-10	4.8	56.3		8%
GTP-binding protein, YchF	Bpr_I1872	C	Y	1.9e-11	4.5	39.9		11%
Leucyl-tRNA synthetase, LeuS	Bpr_I2143	C	Y	2.6e-14	5.0	92.4		9%
Lysyl-tRNA synthetase, LysS	Bpr_I2757	C	Y	1.5e-13	4.8	59.9		22%
Methionyl-tRNA synthetase, MetG	Bpr_I2807	C	Y	5.7e-12	4.9	76.4		16%
Peptide chain release factor 2, PrfB	Bpr_I2321	C	Y	3.0e-14	4.4	42.6		9%
Polyribonucleotide nucleotidyltransferase	Bpr_I0975	C	Y	1.1e-15	4.9	76.5		37%
Ribosomal protein L1, RplA	Bpr_I2299	C	Y	5.9e-14	9.8	24.6		34%
Ribosomal protein L2, RplB	Bpr_I0597	C	Y	3.5e-12	11.0	30.4		30%
Ribosomal protein L3, RplC	Bpr_I0594	C	Y	3.9e-13	10.6	24.3		38%
Ribosomal protein L4, RplD	Bpr_I0595	C	Y	1.6e-07	10.2	22.9		23%
Ribosomal protein L5, RplE	Bpr_I0606	C	Y	6.8e-11	9.8	20.4		53%
Ribosomal protein L6, RplF	Bpr_I0609	C	Y	2.0e-14	10.3	19.4		43%
Ribosomal protein L7/L12, RplL	Bpr_I2370	C	Y	1.1e-09	4.5	13.0		30%
Ribosomal protein L11, RplK	Bpr_I2300	C	Y	1.0e-30	10.3	14.9		21%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Ribosomal protein L13, RplM	Bpr_I0630	C	Y	7.7e-12	9.9	16.2		47%
Ribosomal protein L14, RplN	Bpr_I0604	C	N	3.8e-06	10.8	13.4		44%
Ribosomal protein L15, RplO	Bpr_I0613	C	Y	2.6e-09	11.0	15.6		18%
Ribosomal protein L17, RplQ	Bpr_I0624	C	Y	8.3e-06	10.7	20.3		7%
Ribosomal protein L18, RplR	Bpr_I0610	C	Y	1.2e-12	10.5	13.5		13%
Ribosomal protein L20, RplT	Bpr_I0785	C	Y	8.1e-10	11.9	13.2		24%
Ribosomal protein L21, RplU	Bpr_I1443	C	Y	3.3e-08	10.3	11.3		44%
Ribosomal protein L22, RplV	Bpr_I0599	C	Y	1.4e-09	10.7	15.1		30%
Ribosomal protein L23, RplW	Bpr_I0596	C	Y	3.5e-09	10.4	11.1		30%
Ribosomal protein L29, RpmC	Bpr_I0602	C	Y	2.8e-09	10.4	7.9		38%
Ribosomal protein L31, RpmE	Bpr_I1823	C	Y	2.3e-10	9.0	7.7		53%
Ribosomal protein S1, RpsA	Bpr_I2035	C	Y	1.0e-30	4.6	41.9		50%
Ribosomal protein S2, RpsB	Bpr_I1395	C	Y	8.6e-07	8.4	27.6		20%
Ribosomal protein S3, RpsC	Bpr_I0600	C	Y	4.2e-11	9.6	24.6		22%
Ribosomal protein S4, RpsD	Bpr_I0622	C	Y	1.6e-12	10.6	24.1		32%
Ribosomal protein S5, RpsE	Bpr_I0611	C	Y	1.7e-09	10.5	17.9		31%
Ribosomal protein S7, RpsG	Bpr_I2366	C	Y	6.3e-14	10.5	17.5		60%
Ribosomal protein S8, RpsH	Bpr_I0608	C	Y	1.0e-30	10.0	14.5		50%
Ribosomal protein S9, RpsI	Bpr_I0631	C	Y	1.9e-10	11.1	15.4		30%
Ribosomal protein S11, RpsK	Bpr_I0621	C	Y	3.8e-12	11.6	13.9		31%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Ribosomal protein S13, RpsM	Bpr_I0620	C	N	4.1e-11	11.3	17.6		24%
Ribosomal protein S15, RpsO	Bpr_I0974	C	Y	7.5e-12	10.8	10.1		21%
Ribosomal protein S16, RpsP	Bpr_I1565	C	Y	5.6e-15	10.7	8.9		40%
Ribosomal protein S17, RpsQ	Bpr_I0603	C	Y	2.7e-05	10.8	10.0		13%
Ribosomal protein S19, RpsS	Bpr_I0598	C	Y	7.1e-09	11.1	10.6		30%
Ribosomal protein S20, RpsT	Bpr_I2019	C	Y	3.7e-11	11.4	9.4		17%
Ribosome recycling factor, Frr	Bpr_I0954	C	Y	4.6e-13	6.1	20.9		30%
RNA binding protein	Bpr_I1564	C	N	2.4e-07	7.5	8.1		44%
Seryl-tRNA synthetase, SerS	Bpr_I2950	C	Y	3.2e-13	5.2	49.3		14%
Single stranded nucleic acid binding protein	Bpr_I2957	C	Y	3.8e-13	6.3	28.3		19%
Threonyl-tRNA synthetase, ThrZ	Bpr_I2374	C	Y	4.9e-13	5.0	73.7		20%
Translation elongation factor EF-G-like protein	Bpr_I1098	C	N	1.1e-16	4.9	77.3		51%
Translation elongation factor G, Fusa	Bpr_I2365	C	Y	1.4e-15	4.6	78.0		37%
Translation elongation factor Ts, Tsf	Bpr_I1396	C	Y	1.1e-15	4.8	34.9		49%
Translation elongation factor Tu, TufA	Bpr_I2364	C	Y	1.0e-30	4.8	43.6		73%
Valyl-tRNA synthetase, ValS	Bpr_I2156	C	Y	2.4e-14	4.9	102.8		19%
<u>Regulation</u>								
Anti-sigma factor antagonist	Bpr_I0800	C	Y	8.4e-07	4.2	10.8		29%
Arginine repressor, ArgR	Bpr_I0986	C	N	1.5e-10	4.6	16.7		20%
Cold shock domain protein, CspD1	Bpr_I2451	C	Y	1.0e-30	4.6	7.3		80%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
HTH domain-containing protein	Bpr_III238	C	N	1.2e-08	4.7	19.7		21%
<u>Signal transduction</u>								
Hpr kinase/phosphatase, HprK	Bpr_I0099	C	N	7.1e-12	5.1	35.9		16%
Phosphocarrier HPr family	Bpr_I0104	C	Y	2.2e-09	4.8	9.3		27%
Response regulator domain-containing protein	Bpr_I1147	C	N	1.9e-09	4.7	29.6		6%
S-ribosylhomocysteinase, LuxS	Bpr_I0033	C	Y	9.7e-05	5.0	18.2		23%
Two-component system response regulator	Bpr_I2833	C	N	7.4e-07	4.7	27.2		6%
<u>Transcription</u>								
DNA directed RNA polymerase α -subunit, RpoA	Bpr_I0623	C	Y	3.3e-15	4.4	35.1		48%
DNA directed RNA polymerase β -subunit, RpoB	Bpr_I2369	C	Y	5.9e-14	4.6	145.5		14%
DNA directed RNA polymerase β' - subunit, RpoC	Bpr_I2368	C	Y	2.9e-10	6.6	139.3		6%
RNA polymerase sigma factor sigma-70 family	Bpr_I1467	C	N	4.3e-08	4.0	29.3		10%
S1 RNA binding domain-containing protein	Bpr_I0204	C	Y	4.7e-07	5.8	97.4		5%
Transcription antitermination factor, NusB	Bpr_I0979	C	N	3.1e-09	4.3	15.3		27%
Transcription elongation factor GreA/GreB1	Bpr_I2758	C	Y	2.4e-08	4.7	18.1		19%
Transcription termination/antitermination factor, NusG	Bpr_I2301	C	Y	9.7e-07	4.4	23.6		11%
<u>Transporters</u>								
ABC transporter ATP-binding protein	Bpr_I1145	C	Y	5.3e-04	4.5	61.7		5%
ABC transporter ATP-binding protein	Bpr_I2802	C	Y	2.0e-14	6.7	41.1		20%
ABC transporter permease protein	Bpr_I0776	M	N	1.5e-11	4.6	131.8		0%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Branched-chain amino acid ABC transporter substrate-binding protein	Bpr_I0334	C	Y	5.9e-13	4.5	48.5		19%
Mechanosensitive ion channel protein MscL family	Bpr_I1598	M	N	1.4e-13	6.5	17.2		10%
Oligopeptide ABC transporter substrate-binding protein, OppA1	Bpr_I1276	S	Y	1.0e-30	4.0	83.1		38%
Peptide/nickel ABC transporter periplasmic protein	Bpr_I2750	S	Y	1.1e-13	3.9	57.9		23%
Peptide/nickel ABC transporter substrate-binding protein	Bpr_I1860	S	Y	3.5e-12	4.1	58.4		15%
PTS system HPr phosphocarrier	Bpr_I2105	C	Y	3.6e-14	4.3	9.3		69%
PTS system I PEP-phosphotransferase	Bpr_I2106	C	N	7.9e-13	4.6	62.2		19%
PTS system IIABC fructose-specific family	Bpr_I2104	M	Y	4.4e-15	6.8	67.1		9%
Sugar ABC transporter permease protein	Bpr_I0180	M	N	1.0e-06	10.1	35.4		5%
Sugar ABC transporter permease protein	Bpr_I2444	M	N	9.7e-07	9.8	34.6		5%
Sugar ABC transporter substrate-binding protein	Bpr_I2344	S	Y	9.7e-12	3.9	52.3		26%
Sugar ABC transporter substrate-binding protein	Bpr_I0117	S	Y	1.0e-30	4.2	68.4		43%
Sugar ABC transporter substrate-binding protein	Bpr_I0182	S	Y	1.0e-30	4.1	63.3		66%
Sugar ABC transporter substrate-binding protein	Bpr_I0237	S	Y	7.6e-09	3.9	64.8		5%
Sugar ABC transporter substrate-binding protein	Bpr_I0313	S	Y	1.0e-30	3.8	55.2		26%
Sugar ABC transporter substrate-binding protein	Bpr_I0937	S	Y	4.4e-16	4.1	61.3		54%
Sugar ABC transporter substrate-binding protein	Bpr_I1589	S	Y	1.0e-30	3.9	61.1		40%
Sugar ABC transporter substrate-binding protein	Bpr_I1667	S	Y	1.0e-30	3.9	47.8		63%
Sugar ABC transporter substrate-binding protein	Bpr_I1720	S	Y	1.1e-15	4.0	49.3		48%
Sugar ABC transporter substrate-binding protein	Bpr_I2010	S	Y	1.0e-30	3.7	47.3		37%

Table continues

Table B.3 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Sugar ABC transporter substrate-binding protein	Bpr_I2264	S	Y	1.1e-16	3.9	47.5		29%
Sugar ABC transporter substrate-binding protein	Bpr_I2443	S	Y	1.0e-30	4.0	57.0		40%
Sugar ABC transporter substrate-binding protein	Bpr_III244	S	Y	1.0e-30	4.1	35.2		60%
Xylose ABC transporter ATP-binding protein	Bpr_I1174	C	N	2.7e-13	6.2	57.5		20%
Xylose ABC transporter substrate-binding protein	Bpr_I1173	S	Y	1.0e-30	4.2	38.4		56%
<u>Unknown function</u>								
Acetyltransferase GNAT family	Bpr_I0656	C	N	1.4e-10	6.7	21.0		31%
Acetyltransferase GNAT family	Bpr_III121	C	N	6.6e-11	5.0	30.3		15%
Acetyltransferase GNAT family	Bpr_III122	C	N	9.4e-13	6.2	20.1		15%
Acetyltransferase GNAT family	Bpr_III192	C	N	2.5e-07	5.4	22.8		14%
ACT domain-containing protein	Bpr_I1609	C	Y	3.5e-10	4.5	9.9		16%
Aminotransferase domain-containing protein	Bpr_I1980	C	N	7.9e-12	5.5	48.3		11%
Archaeal ATPase family protein	Bpr_I1663	C	N	1.2e-06	6.0	53.2		4%
Bmp family protein	Bpr_I1560	S	Y	1.5e-11	3.7	40.7		27%
CoA-binding domain-containing protein	Bpr_I1206	C	N	1.9e-07	6.7	23.9		11%
Exopolyphosphatase-related protein	Bpr_I1009	C	N	7.8e-10	5.0	35.7		23%
FAD-dependent pyridine nucleotide-disulphide oxidoreductase	Bpr_I1742	C	N	8.5e-09	5.4	46.2		11%
HIT domain-containing protein	Bpr_I0077	C	N	5.2e-13	6.2	14.8		27%
Hydrolase α / β -fold family	Bpr_I2806	C	N	2.2e-08	8.6	36.5		16%
Nonspecific acid phosphatase	Bpr_I2643	S	N	1.0e-11	4.4	41.3		29%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Oxidoreductase aldo/keto reductase family	Bpr_I0618	C	N	1.2e-10	6.0	38.0		11%
Oxidoreductase aldo/keto reductase family	Bpr_I2938	C	Y	7.6e-12	5.2	43.3		18%
Oxidoreductase NAD-binding domain-containing protein	Bpr_I1305	C	N	1.4e-10	4.9	31.8		21%
PilZ domain-containing protein	Bpr_I1091	C	N	1.7e-11	4.7	13.0		59%
Pyridoxamine 5'-phosphate oxidase family protein	Bpr_I1023	C	N	3.6e-09	5.3	15.0		47%
SAM-dependent methyltransferase	Bpr_I0663	C	N	1.2e-11	5.0	32.6		22%
SAM-dependent methyltransferase/acetyltransferase GNAT family	Bpr_I2871	C	N	3.2e-12	5.2	46.7		7%
SEC-C domain-containing protein	Bpr_I2375	C	Y	1.9e-12	4.8	19.4		17%
SPFH domain/band 7 family protein	Bpr_I1185	C	N	1.2e-12	5.2	33.8		12%
SpoVG family protein	Bpr_I0292	C	N	1.7e-09	4.9	10.9		12%
<u>Vitamins and cofactors</u>								
Cobalamin biosynthesis protein, CobW1	Bpr_I2021	C	Y	2.3e-09	4.4	36.9		6%
Nicotinate (nicotinamide) nucleotide adenylyltransferase, NadD	Bpr_I1438	C	N	4.1e-12	6.1	23.0		33%

Table B.4. Summary of the non-polysaccharide degrading proteins identified in the *B. proteoclasticus* cell-associated fraction.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Amino acid biosynthesis</u>								
Branched-chain amino acid aminotransferase, IlvE	Bpr_I1650	C	N	1.2E-07	5.2	39.2		24%
Cysteine synthase, CysK	Bpr_I1089	C	N	1.5E-19	5.0	32.3		63%
Diaminopimelate dehydrogenase	Bpr_I0298	C	Y	2288	5.6	35.8		35%
Glu/Leu/Phe/Val dehydrogenase	Bpr_I2129	C	Y	9.6E-12	5.4	48.6		43%
Phosphoserine aminotransferase, SerC	Bpr_I1341	C	N	4.8E-11	5.0	40.1		52%
<u>Carbohydrate metabolism</u>								
2,3-bisphosphoglycerate-independent phosphoglycerate mutase, GpmA	Bpr_I1294	C	Y	1131	4.7	56.5		21%
Aldose 1-epimerase	Bpr_I0228	C	N	3.8E-09	4.8	38.3		34%
Endo-1,4- β -xylanase, Xyn10B	Bpr_I0026	S	Y	4.8e-13	4.3	136.9		31%
Fructose-1,6-bisphosphate aldolase, FbaA	Bpr_I2903	C	Y	4172	5.1	30.5		60%
Glucose-6-phosphate isomerase, Gpi	Bpr_I0035	C	Y	7.6E-12	5.0	57.1		41%
Glyceraldehyde-3-phosphate dehydrogenase, Gap	Bpr_I2050	C	Y	1159	5.7	36.9		22%
Lactaldehyde reductase, FucO	Bpr_I2065	C	Y	9.6E-10	4.7	41.2		41%
Phosphoenolpyruvate carboxykinase, PckA	Bpr_I0091	C	Y	1479	4.9	59.5		25%
Transketolase subunit A, TktA3	Bpr_I2813	C	Y	1.6E-05	5.4	30.0		26%
Triosephosphate isomerase, TpiA	Bpr_I2048	C	Y	7.6E-15	4.6	26.8		47%
UDP-galactose 4-epimerase, GalE	Bpr_I0192	C	Y	1.5E-08	5.2	37.2		38%
Xylosidase/arabinofuranosidase, Xsa43J	Bpr_I2935	S	Y	1.9e-07	4.2	251.9		9%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Xylulokinase, XylB	Bpr_I0173	C	Y	1.0E-05	4.9	53.7		17%
<u>Cell envelope</u>								
NAD-dependent epimerase/dehydratase	Bpr_I2537	C	Y	340	4.7	36.0		16%
<u>Central metabolism</u>								
Agmatine deiminase, AguA	Bpr_I1200	C	N	1018	4.2	49.5		15%
<u>Energy metabolism</u>								
Aconitate hydratase, AcnA	Bpr_I2797	C	N	6.1E-14	5.3	82.7		36%
Butyrate kinase, Buk	Bpr_I2323	C	Y	2.4E-15	5.1	38.7		55%
Crotonase, Crt	Bpr_I2487	C	Y	6726	4.7	28.0		51%
Electron transfer flavoprotein α -subunit, EtfA	Bpr_I2483	C	Y	653	4.8	37.4		17%
Phosphate acetyltransferase, Pta	Bpr_III011	C	Y	581	5.2	14.4		32%
Phosphate butyryltransferase, Ptb	Bpr_I2324	C	Y	9.4E-06	5.1	33.3		38%
Pyruvate phosphate dikinase, PpdK	Bpr_I1154	C	Y	1.2E-15	4.8	104.3		20%
Thiolase, ThlA1	Bpr_I2488	C	Y	3815	5.0	42.0		32%
<u>Hypothetical</u>								
Hypothetical protein	Bpr_I2329	C	Y	1910	4.7	23.5		43%
Hypothetical protein	Bpr_I2455	C	N	2.4E-11	4.7	47.7		34%
Hypothetical protein	Bpr_I2583	S	Y	9.6E-15	4.3	44.3		46%
Hypothetical protein	Bpr_I2744	C	N	748	4.8	28.5		26%

Table continues

Table B.4 continued.

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
<u>Lipid metabolism</u>								
3-oxoacyl-(acyl-carrier-protein) synthase, FabF	Bpr_I1268	C	N	1187	5.3	44.5		16%
<u>Nucleotide metabolism</u>								
Phosphoribosylamine-glycine ligase, PurD	Bpr_I0870	C	N	2.4E-07	4.6	46.3		29%
Purine-nucleoside phosphorylase, DeoD1	Bpr_I1527	C	Y	998	4.6	29.5		25%
<u>Protein fate</u>								
Chaperone protein, DnaK	Bpr_I2625	C	Y	875	4.5	66.8		13%
<u>Protein synthesis</u>								
Ribosomal protein S1, RpsA	Bpr_I2035	C	Y	6.9E-07	4.6	41.9		30%
Translation elongation factor EF-G-like protein	Bpr_I1098	C	N	7.6E-12	4.9	77.3		27%
Translation elongation factor G, FusA	Bpr_I2365	C	Y	3.8E-21	4.6	78.0		39%
Translation elongation factor Ts, Tsf	Bpr_I1396	C	Y	1078	4.8	34.9		26%
Translation elongation factor Tu, TufA	Bpr_I2364	C	Y	9.6E-12	4.8	43.6		54%
<u>Transporters</u>								
ABC transporter ATP-binding protein	Bpr_I2802	C	Y	1698	6.7	41.1		17%
Oligopeptide ABC transporter substrate-binding protein, OppA1	Bpr_I1276	S	Y	6.1E-21	4.0	83.1		40%
Peptide/nickel ABC transporter periplasmic protein	Bpr_I2750	S	Y	1883	3.9	57.9		19%
Sugar ABC transporter substrate-binding protein	Bpr_I2344	S	Y	2.4E-09	3.9	52.3		19%
Sugar ABC transporter substrate-binding protein	Bpr_I0117	S	Y	2478	4.2	68.4		20%
Sugar ABC transporter substrate-binding protein	Bpr_I0182	S	Y	8045	4.1	63.3		46%

Table continues

Protein	Locus	Location	PHX	Score	pI	kDa	Pep.	Cov.
Sugar ABC transporter substrate-binding protein	Bpr_I0937	S	Y	3.0E-10	4.1	61.3		33%
Sugar ABC transporter substrate-binding protein	Bpr_I1589	S	Y	3103	3.9	61.1		33%
Sugar ABC transporter substrate-binding protein	Bpr_I1667	S	Y	1504	3.9	47.8		20%
Sugar ABC transporter substrate-binding protein	Bpr_I1720	S	Y	6565	4.0	49.3		31%
Sugar ABC transporter substrate-binding protein	Bpr_I2010	S	Y	3478	3.7	47.3		35%
Sugar ABC transporter substrate-binding protein	Bpr_I2443	S	Y	1.2E-06	4.0	57.0		29%
Sugar ABC transporter substrate-binding protein	Bpr_III244	S	Y	8046	4.1	35.2		50%

Table B.5. BLASTp analysis of the differentially abundant proteins detected in the mid-log phase *B. proteoclasticus* cytosol by 2-DE.

Spot	<i>B. proteoclasticus</i> protein	Locus	Homologue ^a	Organism ^b	gi	Expect ^c	Identity	Region ^d
C853	Adenylosuccinate lyase, PurB	Bpr_I2212	Adenylosuccinate lyase	<i>Ruminococcus</i> sp. SR1/5	291548168	1e-180	83%	1-477
C1135	Amino acid ABC transporter substrate-binding protein	Bpr_I1826	Amino acid ABC transporter, periplasmic amino acid-binding protein	<i>Faecalibacterium prausnitzii</i> A2-165	257438674	1e-82	62%	62-295
C1010	Anti-sigma factor antagonist/phosphotransferase domain-containing protein	Bpr_I0249	STAS domain-containing protein	<i>Prevotella ruminicola</i> 23	294673555	1e-41	27%	20-430
	Ribosomal protein S1, RpsA	Bpr_I2035	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	<i>Roseburia intestinalis</i> L1-82	240146046	1e-160	76%	1-356
C654	DNA-directed RNA polymerase α subunit, RpoA	Bpr_I0623	DNA-directed RNA polymerase, alpha subunit	<i>Clostridium hathewayi</i> DSM 13479	288870169	1e-155	87%	1-313
C837	Fructose-1,6-bisphosphate aldolase, FbaA	Bpr_I2903	Fructose-1,6-bisphosphate aldolase, class II, various bacterial and amitochondriate protist	<i>Coprococcus</i> sp. ART55/1	295094301	4e-149	89%	1-287
C733	IMP cyclohydrolase, PurO	Bpr_I0731	IMP cyclohydrolase-like protein.	<i>Roseburia intestinalis</i> XB6B4 Length=242	291539658	2e-107	77%	48-286
C733	Translation elongation factor Tu, TufA	Bpr_I2364	Translation elongation factor 1A (EF-1A/EF-Tu)	Butyrate-producing bacterium SS3/4	291562252	1e-180	82%	1-396
C792	NADPH-dependent glutamate synthase, GltA3	Bpr_I1306	Glutamate synthase (NADPH), homotetrameric	<i>Ruminococcus torques</i> L2-14	291550425	1e-180	80%	2-461
C637	Pullulanase, Pul13A	Bpr_III161	Pullulanase	<i>Butyrivibrio crossotus</i> DSM 2876	260438049	1e-180	46%	1-869

Table continues

Spot	<i>B. proteoclasticus</i> protein	Locus	Homologue ^a	Organism ^b	gi	Expect ^c	Identity	Region ^d
C709	Ribosomal protein S1, RpsA	Bpr_I2035	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	<i>Roseburia intestinalis</i> L1-82	240146046	1e-160	76%	1-356
C1072	Serine protease subtilisin family	Bpr_I2629	Bacterial Ig-like domain (group 2)./Subtilase family.	<i>Butyrivibrio fibrisolvens</i> 16/4	291519087	1e-180	62%	1-1347
C1027	Xylosidase/arabinofuran-oxidase, Xsa43E	Bpr_I2319	Carbohydrate-binding family 6 protein	<i>Clostridium thermocellum</i> ATCC 27405	125974681	3e-107	65%	11-303

^a BLASTp top ranked homologous protein.

^b Organism that expresses the BLASTp top ranked homologous protein.

^c Expectation score is the statistical probability of the top ranked protein match being a false positive identification, $p < 0.05$.

^d Values denote the first and last residues of the homologous region of the identified in *B. proteoclasticus* protein.

Table B.6. BLASTp analysis of the differentially abundant proteins detected in the stationary phase *B. proteoclasticus* cytosol by 2-DE.

Spot	<i>B. proteoclasticus</i> protein	Locus	Homologue ^a	Organism ^b	gi	Expect	Identity	Region
C1135	Amino acid ABC transporter SBP	Bpr_I1826	Amino acid ABC transporter, periplasmic amino acid-binding protein	<i>Faecalibacterium prausnitzii</i> A2-165	257438674	1e-82	62%	62-295
C1120	β-Galactosidase, Bga35A	Bpr_I2006	Galactosidase	<i>Cellulosilyticum ruminicola</i>	280977827	1e-180	52%	12-733
	β-xylosidase, Xyl3A	Bpr_I0184	Glycoside hydrolase, family 3 domain-containing protein	<i>Ruminococcus</i> sp. 5_1_39B_FAA	253579611	1e-180	63%	3-707
C704	Peptidase	Bpr_I2455	Peptidase U62 modulator of DNA gyrase	<i>Clostridium carboxidivorans</i> P7	255526201	4e-52	33%	1-422
	Phosphoribosylamine-glycine ligase, PurD	Bpr_I0870	Phosphoribosylamine-glycine ligase	<i>Butyrivibrio crossotus</i> DSM 2876	260437894	1e-180	81%	1-423
C784	Peptidase	Bpr_I2456	Zn-dependent protease, TldD	<i>Clostridium acetobutylicum</i> ATCC 824	15893521	2e-122	51%	1-477
C601	Oligopeptide ABC transporter SBP, OppA1	Bpr_I1276	Hypothetical protein	<i>Roseburia inulinivorans</i> DSM 16841	225375060	1e-180	54%	1-757
C1057	Sugar (D-xylose) ABC transporter SBP	Bpr_I1173	Putative solute-binding component of ABC transporter	<i>Clostridium phytofermentans</i> ISDg	160879730	4e-161	68%	19-359
C785	Xylulokinase, XylB1	Bpr_I0173	Xylulokinase	<i>Roseburia intestinalis</i> L1-82	240146123	1e-180	76%	1-489
	Aminotransferase, DegT/DnrJ/EryC1/StrS family	Bpr_I2311	DegT/DnrJ/EryC1/StrS aminotransferase	<i>Clostridium saccharolyticum</i> WM1	302384934	1e-179	64%	8-453

^a BLASTp top ranked homologous protein.

^b Organism that expresses the BLASTp top ranked homologous protein.

Table B.7. BLASTp analysis of the differentially abundant proteins detected in the mid-log phase *B. proteoclasticus* cytosol by 1-D LC-MS/MS.

<i>B. proteoclasticus</i> protein	Locus	Homologue	Organism	gi	Expect	Identity	Region
ATP synthase F1 β subunit, ATP D2	Bpr_I1162	ATP synthase F1, β -subunit	<i>Clostridium</i> sp. M62/1	283795522	1e-180	82%	1-469
β -Xylosidase Xyl3A	Bpr_I0184	Glycoside hydrolase, family 3 domain-containing protein	<i>Ruminococcus</i> sp. 5_1_39B_FAA	253579611	1e-180	63%	3-707
Bmp family protein	Bpr_I1560	Nucleoside-binding protein	<i>Butyrivibrio fibrisolvens</i> 16/4	291519926	3e-154	79%	50-393
DNA-binding protein	Bpr_II236	DNA-binding protein	<i>Butyrivibrio proteoclasticus</i> B316	302670433	6e-35	86%	1-92
DNA-directed RNA polymerase subunit, RpoA	Bpr_I0623	DNA-directed RNA polymerase, α -subunit	<i>Clostridium hathewayi</i> DSM 13479	288870169	1e-155	87%	1-313
Fatty acid/phospholipid synthesis protein, PlsX	Bpr_I1576	Phosphate:acyl-(acyl carrier protein) acyltransferase	<i>Ruminococcus torques</i> L2-14	291550075	1e-135	70%	1-334
Orotate phosphoribosyl-transferase, PyrE1	Bpr_I0863	Orotate phosphoribosyltransferase	<i>Clostridium</i> cf. <i>saccharolyticum</i> K10	295091343	2e-95	76%	1-225
PTS system HPr phosphocarrier	Bpr_I2105	Phosphotransferase System HPr (HPr) Family	<i>Butyrivibrio fibrisolvens</i> 16/4	291518288	4e-30	73%	1-87
Sugar ABC transporter SBP	Bpr_I1720	Extracellular solute-binding protein family 1	<i>Acetivibrio cellulolyticus</i> CD2	302592083	5e-69	34%	1-455
Sugar ABC transporter SBP	Bpr_I2443	Extracellular solute-binding protein family 1	<i>Clostridium lentocellum</i> DSM 5427	296441832	1e-40	30%	1-522
Sugar ABC transporter SBP	Bpr_I2264	Extracellular solute-binding protein, family 1	<i>Ruminococcus</i> sp. 5_1_39B_FAA	253579691	6e-120	49%	1-441

Table continues

Table B.7 continued.

<i>B. proteoclasticus</i> protein	Locus	Homologue	Organism	gi	Expect	Identity	Region
Sugar ABC transporter SBP	Bpr_I0117	Putative bacterial extracellular solute-binding protein	<i>Roseburia intestinalis</i> L1-82	240144557	1e-166	48%	25-617
Sugar ABC transporter SBP	Bpr_I1667	Bacterial extracellular solute-binding protein	<i>Turicibacter</i> sp. PC909	293376212	6e-58	34%	3-419
Xylose ABC transporter ATP-binding protein	Bpr_I1174	D-xylose ABC transporter, ATP-binding protein	<i>Bryantella formatexigens</i> DSM 14469	255505494	1e-180	80%	1-509

Appendix C



Figure C.1. Cel5C sequence coverage of tryptic peptides extracted from spot 1 (A) and spot 2 (B). Note the absence of peptides matching to the N- or C-terminus of the full length protein sequence in both identifications.

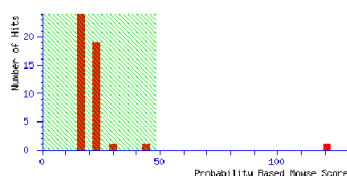
A

Search title : B316 260606 Stat Ext Xylose pI 3-5.6 180906 Spot 302 accum
 Database : B316 July2008 (3818 sequences; 1305723 residues)
 Timestamp : 12 Jan 2010 at 01:43:36 GMT
 Top Score : 121 for **Bpr_I2473**, Bpr_I2473 pectin methyltransferase Pme8B 3017405:3025603 reverse MW:294078

	B316	Decoy
Protein hits above identity threshold	1	0
Highest scoring protein hit	121	39

Probability Based Mowse Score

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Protein scores greater than 48 are significant ($p < 0.05$)



Concise Protein Summary Report

Format As:
 Significance threshold $p < 0.05$ Max. number of hits

- Bpr_I2473** Mass: 294580 Score: **121** Expect: 2e-009 Queries matched: 23
 Bpr_I2473 pectin methyltransferase Pme8B 3017405:3025603 reverse MW:294078

Bpr_I1355 Mass: 9272 Score: 20 Expect: 36 Queries matched: 2
 Bpr_I1355 hypothetical protein 293691:293936 reverse MW:9277

Bpr_I1975 Mass: 35172 Score: 19 Expect: 49 Queries matched: 3
 Bpr_I1975 methyl-accepting protein 2406829:2407788 reverse MW:35137

Bpr_I1419 Mass: 19571 Score: 18 Expect: 65 Queries matched: 2
 Bpr_I1419 hypothetical protein 1751371:1751880 forward MW:19526

B

Nominal mass (M_0): 294580; Calculated pI value: 4.41
 NCBI BLAST search of **Bpr_I2473** against nr
 Unformatted [sequence string](#) for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
 Number of mass values searched: 86
 Number of mass values matched: 23
 Sequence Coverage: 13%

Matched peptides shown in **Bold Red**

```

1 MRPRTWHEKA GAIMLSAAMS LSTFSAALPD GVYVAINGOV LQADTSASAE
51 TSTQDSSEAGS DEASVSATIE TSATDSQET DAALESTOE SSWSEETDT
101 DSASKQDAS ESEEGAILL LQENKTSORN FDSKRVWMD FGAENLGDAY
151 HNRKLDGTHN SPISVTAET SPVTAASFTD DGVYVWDSG IYTHRLDET
201 NTSILTRVDAM SLGSDQDVTY GULYVSVDSG TQMYVVALEK KADQVITAFV
251 ASNGTDSIVH FQNVSDASDD ASHVTLLGSS TVTKMTFFPA KDAQVFLFA
301 TEKVLAKRVY REBAQVGLT GSVEGSEGS SEDIVTINQV HGNVVRATVA
351 DQVYTAQAS GFYVNSLAG ADQVATTSOW NVVSSQETM CIVTEVDGV
401 EVSQRKTGIA DKDMKPYSA AEFFAAQDE KSVYVPEAL TLDSALIEY
451 LKLQKGVTS VSVADKETA TERFAGVEDY ELVDELITLD KSSMDVIEF
501 EAPFLKRVYI VPKAGTLSD KDAVFFPRL DAESEFAAGC TVIEFPTGSD
551 IALHQGVQAV EVSNTALQV KLSLLING ADVTYVTS SVTHNRFSD
601 PFTSKKACV TESTINGLSM TNGSENNVY LYSAGKIEY FVSGDQIQV
651 TANYQSYFF ENDTSPNVV KTSSTINET FVYVKGAG TIDVTVLOSS
701 LTKLATTYQ TEKREILVQ SEDAGVIAE HMLANVRHM DRDQGVYI
751 SEEPANVEIN LVIVDAVTL ANASNPHTI LKNSVDIDD NAVRITVYG
801 HGVTVYMSG DYKVDADVLA ANKNGVYAV INPSGOTGY WNAVVVDAA
851 GFPAKDIIEI HSNQGVSKP AAEVTVVQS SAKESGIAGA SNASGTTVQ
901 DQVYTERAA LALYGAATEA YENKILIGR QDTLYNHTT IAAFYKCAV
951 GQTDYIFGM TAVFACDLV LMTSDNVWV AVTAPQGS ASERGVLMN
1001 CTVKSTIPGV DTSKTVYVK GQFQDQWAM TSEAVFDTV VEATQWANGA
1051 YDENARTSLI QPVGNFSLG QGSAHVFEG TFFVSGDMS KESASVQGP
1101 KENVLADDTA IYVAFLEWV QPAGNNHNM LTFPSGKSE VAPPSSET
1151 PVSENNFTF ESSTLEFPA GAKADGEMK VQDDVFII VSVKSVQDS
1201 SPTWAEKYS QVRLHLOGAA ATERNSVFT TSAATVWV WACDAGQGH
1251 TLMHSGRTV ATTAETAYN SPVSTIQLD EAGTILGET ANNYITFTV
1301 TIDKARESVV STLESADLA FAAGAKADCE SEKAGTDQVY TLVYSAUSI
1351 DTSSKTVDDG YTSQRIHFQ GAATERNAI KFTTSADAT VVWMAACGDA
1401 QKQAILDST GNAVATTEA AVNPSPIST LKTAAGTY LGATANNVY
1451 FRTVTDQAP VEYVAEYS VYAVTSISV LNKDQWID VSDATGLM
1501 QGSLVVEHL DADNVVELL KSLADINEFT FAPTPNSGD VYFRALSRD
1551 DEEDVKSIVR TDAFSPVLP KAPSPKNAVN KGATVVVF YVPEAEYS
1601 LTATDKTDA SLKAVITAM TPOVDESEY STFFVGLNV KHLALVAN
1651 RQQVQDSVY MDEVTESG KEVTSBAPQ STVSGDMS TNGDQSVY
1701 WNIHQWGLV PASTDGLSF YVAVANWHF TLATATIDT WFTNQEGF
1751 GLMAADRVVY HNAAVFNN STGSGSSVE YYVDTNQEV TIDTSAKIT
1801 HRLQAGXK VVTVENLAR LEANDTEYN NEPSNNPL STSGQSGVY
1851 TMLFSGES QVTVLIANG LAQVLSIQK NRPQVTSR DQNVVITK
1901 KFYDTALNK LQENVTYGF FAATFKATF SDIELTIDP SEDAPAEER
1951 VTVVVDYKI ISATVNTAK VOLQVAGAD GVLITIDAG NAVVNRVAV
2001 AGVYSTVY LQWGANVY TTPVADNP GQSEYGLSS GVYVYVYV
2051 KYATVNDCEI TVRADGHEF ATGTANDVD IYAVKYVQ GQITVAAET
2101 YELISTVTIA RIGDGTPEF IMNTGGRRA IFDNNKAG FVLAGNVYI
2151 NGQIDCHSN SLKQVQBS HTLEDIRY ENQNTQVS SVLTIDGRM
2201 WRYLLIANG TVYKADRV EDAGFAAPK TQDQVDS CTFVNDAG
2251 WDLFAVESG SIQVTLMS VAFANGVVD GTHGNGNF RIGGSEMSG
2301 HGLINSVAV HKAGIDENS GDIQVYEM SPVNGNNVA LYTNDANTN
2351 VYVDVLYR TIGTCHNI KAGTQDTH IYGRHFFH DAGASRSDG
2401 VANDVYTS DAVVANDP VAVRALEBA DQTDGHT KLESTJAL
2451 AAGLDABEY VAGVDAQEA VDEREISGS SDQAKVVE DPQPEEREI
2501 KGFVTKYR TVFTGDEK LIGLKKIGE LVYFRITGV LNSVITIDG
2551 HRYFDGSH AVNDRSHG ATFFDETH EYDAVVDV SHVYVLDG
2601 KTFQKFAIT GSHVYINDO QVYTSATF QVSTVYVA EGVYVNYIA
2651 NIEGKYYV DQGRVKGK TVVDNKYF GQDAMLET FVTQGSTY
2701 LQSDGMLKH TVTVVDVY SANSAGNVY KK
  
```

Figure C.2. MASCOT summary of MALDI-TOF Pme8B identification (A) and peptide sequence coverage (B) of protein extracted from spot 7.

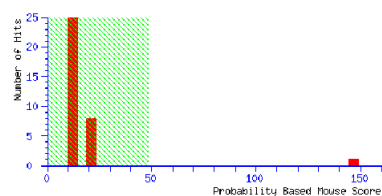
A

Search title : B316 260606 Stat Ext Xylose pI 3-5.6 180906 Spot 304 accum
 Database : B316 July2008 (3818 sequences; 1305723 residues)
 Timestamp : 6 Jan 2010 at 06:36:15 GMT
 Top Score : 147 for **Bpr_I2372**, Bpr_I2372 pectate lyase Pel1A 2900147:2903335 reverse MW:115219

B316 Decoy		
Protein hits above identity threshold	1	0
Highest scoring protein hit	147	29

Probability Based Mowse Score

Protein score is $-10 \cdot \log(P)$, where P is the probability that the observed match is a random event.
 Protein scores greater than 48 are significant ($p < 0.05$).



Concise Protein Summary Report

Format As: Concise Protein Summary [Help](#)

Significance threshold $p < 0.05$ Max. number of hits: AUTO

Re-Search All Search Unmatched

1. [Bpr_I2372](#) Mass: 115490 Score: **147** Expect: 7.6e-012 Queries matched: 20
 Bpr_I2372 pectate lyase Pel1A 2900147:2903335 reverse MW:115219
[Bpr_I1613](#) Mass: 83759 Score: 16 Expect: 1.1e+002 Queries matched: 4
 Bpr_I1613 DNA polymerase I PolA 1976359:1978581 reverse MW:83299
[Bpr_I2244](#) Mass: 27294 Score: 15 Expect: 1.3e+002 Queries matched: 2
 Bpr_I2244 bis(5-nucleosyl)-tetraphosphatase ApaH 2738051:2738761 reverse MW:27083
[Bpr_I2384](#) Mass: 10073 Score: 14 Expect: 1.4e+002 Queries matched: 2
 Bpr_I2384 hypothetical protein 2916196:2916459 reverse MW:10022

B Match to: **Bpr_I2372** Score: **147** Expect: **7.6e-012**
Bpr_I2372 pectate lyase Pel1A 2900147:2903335 reverse MW:115219

Nominal mass (M_r): **115490**; Calculated pI value: **4.35**

NCBI BLAST search of [Bpr_I2372](#) against nr
 Unformatted [sequence string](#) for pasting into other applications

Fixed modifications: Carbamidomethyl (C)
 Variable modifications: Oxidation (M)
 Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
 Number of mass values searched: **49**
 Number of mass values matched: **20**
 Sequence Coverage: **25%**

Matched peptides shown in **Bold Red**

```

1 MLRRGYQHLG KTNNFKGGYM LKRLNTRLG LAMAGVLIAT SAFAGSKVDA
51 AAAVEAPVVM EDSGWLESAY VTWRAVDGAS GYNVYVNDGG GQVQLDDELI
101 RLYSSDSGEK YRADALGLS AGNYSFTVVP LEDGDEDTSA AVSTDVSVYQ
151 NHDRSGFAFV NGDSNGAYNY DGTLRNDAVV LVITNETKDS VAMDVVTNAK
201 GATTPATGLQ NILNLYKKGY DLRLDVRV FV GQITDFATME GGDIVISGAG
251 DNKRKSCGIT FEGVGEDATA DGWGLRIKNA SNVEVRNIGF MNCDSSGDD
301 LGLQSQDDHV WVNHCDFEYF YAGGDSQDAK GDGALDTKKS TFVTHSYNHF
351 YDTGKSNLQG MKSESTENYI TYHHNWDHS DSRHPRITC SVHIYNNYVD
401 GNAKYGVGVVT MGASAFVEAN YYRNCKFPML ISEQSSDVMA DWETLTRNED
451 YGTFSSENGG IKSFNMYE GQQSYVTYQE NSTEFDAYEV SAADEQVPE
501 VVSYKGGTSY NNFDTSIMY SYSADSPQEA KEKVVSFAGR ENGDDFNWVF
551 NNEVDDADYG VNRALKALN SYQTNLVSFG GGSTAGEAGN SGEESSENS
601 GAGEASSSDS GAGEASSDNS GEASSDSSSE QHEETPVNNG SYVHNFTASG
651 KDSDFFTIEG NLSTGKGTVT YNGKLLTQCL KIESKTSIKF AAPSDGQLIL
701 VFGDRDSSIK LDGTAVKEAT NVNMLDVRKG DHELTKKDTE NLFYMEYVSN
751 DQPSGDDPGN DDPASEASSE AASEASSEAA GEASSEAAASE SSSEAAAGEAS
801 SEAAESSSE AAGEASSEAA SEASSDEAGE ASSDEPTGS SEEIAEPEPD
851 EPSSSNGEWL TYGKTYVVL TDGTVKVTGIY TIDGTYTFD ANGVRKTNVF
901 VTDGDKTHFF GADGKMTVGW LTRYFKTYFF DENGNMVTGF AYVDGIYHYH
951 DEDGVMFKNG NKTINEKKYF FNSDGMVGTG WLERYGSTYY YDESGAMVYG
1001 LQIIDGNMHF FRDNGKLAKD MMVTVDKTY YFDNDGNMYV GELTRWFHKY
1051 IFDENGVLIV KK

```

Figure C.3. MASCOT summary of MALDI-TOF Pel1A identification (A) and peptide sequence coverage (B) of protein extracted from spot 6.