

**Nutrient Acquisition in a Human Gut Symbiont:
Molecular Analysis of the Carbohydrate Utilisation Apparatus of
*Bacteroides thetaiotaomicron***

A thesis submitted for the degree of Doctor of Philosophy

Newcastle University

2006-2009

Hongjun Zheng

Institute for Cell and Molecular Biosciences

Newcastle University

路漫漫其修远兮, 吾将上下而求索。

-----屈原 《离骚》

The road of search is so long, but I will keep on exploration.

----- *The Sorrow of Separation by Qu Yuan (340-278 BC)*

Acknowledgements

First of all, I must thank Dave for all his kind supervision in the three years. It is highly appreciated. Secondly, I want to thank all mates of Lab M2035, past and present, for their friendship, advice. I also would like to say “Thanks a lot” to Dr. Susan Firbank and Mr. Carl Morland for their wonderful association.

I thank Dorothy Hodgkin Postgraduate Awards scheme from the bottom of my heart for offering the education fee and living cost during the three years. The UK government, the Biotechnology and Biological Sciences Research Council and the company GlaxoSmithKline as my sponsors lighted up my life totally.

Also, thank you, dear mum and grandma for your support and endless love.

Abstract

The gut microbiota play a significant role in human health and nutrition, although the mechanisms these organisms use to survive in this densely populated environment are not well understood. *Bacteroides thetaiotaomicron* is a dominant member of the gut bacterial community whose genome sequence reveals large expansions in protein families involved in the sensing, acquisition and utilisation of complex carbohydrates, pointing to the ability to access a wide range of glycans as playing a significant role in becoming a successful resident of the human gut. Here we have characterised components of the *B. thetaiotaomicron* polysaccharide utilisation apparatus at molecular level, focusing mainly on fructan sensing, binding and degradation systems.

Microarray data from our collaborators revealed that growth of *B. thetaiotaomicron* on inulin (β -2,1-linked fructan) specifically upregulated a locus of nine genes BT1757-BT1765 and an orphan gene BT3082, encoding a glycoside hydrolase from family 32 (GH32). The locus contains three other GH32s, two predicted polysaccharide binding outer membrane proteins (SusC and SusD homologues), a fructokinase, and an inner membrane monosaccharide transporter. Together these components form a polysaccharide utilisation locus (PUL). The nearest regulatory gene to the PUL is BT1754, a hybrid two component system. Here we show that the periplasmic sensor domain of BT1754 (BT1754peri) binds specifically to fructose, with a K_d of $\sim 2 \mu\text{M}$ and a stoichiometry of 1:1, but not fructooligosaccharides or other monosaccharides. The crystal structure of BT1754peri revealed a two domain periplasmic binding protein (PBP)-fold with the ligand fructose sandwiched between the two domains. BT1754 is the first periplasmic sensor histidine kinase domain to display a non-PAS fold. The structure in combination with biophysical and site-directed mutagenesis studies also shows how the protein displays such specificity in ligand recognition and provides insights into the mechanism of signal transduction across the inner membrane.

The four glycoside hydrolase family 32 members regulated by BT1754 were also biochemically and structurally characterised in this thesis. Three of four GH32 enzymes BT1759, BT1765 and BT3082, digest both β -2,1-linked (inulin) and β -2,6-linked (levan) fructans, indicating that levan is also utilised by the same locus. BT1759 and BT3082 are exo-acting enzymes releasing fructose from both long-chain and short-chain inulins and levans, while BT1765 is also exo-acting and produces fructose, but preferred short-chain sugars. BT1760 is unusual for GH32 as it is specific for levan and has an endo-like activity, releasing a range of different size oligosaccharides from the polysaccharide. The crystal structures of wild type BT3082 and a nucleophile mutant in complex with substrate (kestose) were solved to 2.2 Å and revealed a typical GH32 β -propeller fold as the catalytic domain. Like all other GH32s solved to date, the enzyme has a C-terminal β -sheet domain of unknown function that was shown to be necessary for correct folding of the enzyme. BT3082 also has a unique N-terminal β -sheet domain that was shown to be essential for enzyme activity but not correct folding. Extensive site-directed mutagenesis was carried out to provide insight into the relative importance of different residues in substrate binding and catalysis in BT3082. The crystal structure of wild type BT1760 was solved to 2.6 Å, revealing a surprisingly similar structure to that of the exo-acting enzymes. A rationale for the endo-like activity of this enzyme and the role each of the four GH32s in fructan utilisation by *B. thetaiotaomicron* is discussed.

The outer membrane SusD homologue BT1762 from the fructan locus was shown to bind preferentially to long chain levans, with no recognition of inulin. The crystal structure of BT1762 was solved to 1.9 Å and was shown to share the same novel α -helical fold as SusD. Site-directed mutagenesis of a number of residues in the same region as the SusD binding site showed that while the location of the ligand binding sites are conserved between these two proteins, the identity of the residues involved in polysaccharide recognition are not the same. A model is proposed for levan recognition in BT1762 and also its role in polysaccharide utilisation.

B. thetaiotaomicron also has at least 12 ECF sigma/anti-sigma factor gene pairs likely involved in polysaccharide utilisation. Here we show that these systems form a trans-envelope signalling apparatus with their cognate SusC-transducer homologue, but that there is no cross talk between different systems. The significance of this finding in relation to survival of this important gut bacterium is discussed.

Contents

Abstract.....	4
Contents.....	5
List of Figures.....	9
List of Tables.....	12
List of genes in this study.....	13
Abbreviations.....	14
Chapter 1 Introduction.....	16
1.1 Carbohydrates and Nutrition.....	16
1.1.1 Monosaccharides.....	16
1.1.2 Disaccharides, oligosaccharides, and polysaccharides.....	17
1.1.3 Fructans.....	18
1.2 Niche of the human gut.....	20
1.2.1 The roles of gut bacteria.....	21
1.2.1.1 In normal human health.....	21
1.2.1.2 In nutrition.....	22
1.2.1.3 Possible harmful effects of gut bacteria.....	23
1.2.2 <i>Bacteroides thetaiotaomicron</i>	23
1.2.3 Bacterial extracellular sensing and signalling systems.....	25
1.2.3.1 ECF sigma/anti-sigma factors.....	26
1.2.3.2 Two component systems.....	28
1.2.3.2.1 Periplasmic-sensing HK sensor domains.....	33
1.2.3.2.2 Role of transmembrane region in signal transduction.....	34
1.2.3.2.3 Cytoplasmic domains of sensor HKs.....	35
1.2.3.2.4 Response Regulator (RR) domains.....	37
1.2.3.3 Hybrid two component systems.....	38
1.2.4 Polysaccharide utilisation in <i>Bacteroides thetaiotaomicron</i>	39
1.2.4.1 SusC.....	40
1.2.4.2 SusD.....	42
1.2.5 Glycoside hydrolases.....	43
1.2.6 Other systems employed by bacteria for carbohydrate utilisation.....	47
1.3 Objectives of this study.....	47
Chapter 2 Materials and Methods.....	48
2.1 Bacterial strains and plasmids.....	48
2.2 Growth media.....	48
2.3 Selective media.....	49
2.4 Chemicals, enzymes and kits.....	49
2.5 Sterilisation.....	49
2.6 Storage of DNA and bacteria.....	49
2.7 Centrifugation.....	50
2.8 Chemical competent <i>E. coli</i>	50
2.9 Transformation of chemical competent <i>E. coli</i>	50
2.10 Purification of DNA.....	50
2.11 Primers.....	51
2.12 Polymerase chain reaction (PCR).....	51
2.13 Site directed mutagenesis.....	52
2.14 Agarose gel electrophoresis of DNA.....	53
2.15 Visualisation of DNA and photography of agarose gels.....	54
2.16 Determination of DNA fragment size.....	54
2.17 DNA extraction from agarose gels.....	55
2.18 DNA digestion, ligation.....	55
2.19 DNA Sequencing.....	55
2.20 Purification of DNA by phenol.....	56
2.21 Determination of DNA concentration.....	56
2.22 Protein over expression in <i>E. coli</i>	56
2.23 Preparation of periplasm protein.....	57
2.24 Preparation of cell free extracts (CFEs).....	57
2.24.1 His tagged protein.....	57

2.24.2 GST tagged protein.....	57
2.24.3 Untagged protein	57
2.25 Insoluble Protein.....	58
2.26 Protein Purification.....	58
2.26.1 Immobilised Metal Affinity Chromatography (IMAC).....	58
2.26.2 Ion Exchange Chromatography (IEC).....	58
2.26.3 Gel Filtration Chromatography (GFC).....	59
2.27 Determination of protein concentration by spectrophotometer.....	59
2.28 Protease Digestion.....	59
2.29 Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE).....	60
2.30 Affinity Gel Electrophoresis (AGE).....	62
2.31 Protein N-terminal Sequencing and molecular weight measurement.....	63
2.32 Isothermal Titration Calorimetry (ITC).....	64
2.33 Circular Dichroism (CD).....	65
2.34 Fluorescence Spectroscopy.....	65
2.35 Differential Scanning Calorimetry (DSC).....	66
2.36 Protein Crystallization.....	67
2.36.1 Sitting drop vapour diffusion.....	67
2.36.2 Hanging drop vapour diffusion.....	67
2.36.3 Collection of crystal diffraction data and structure refinement.....	68
2.37 Protein cross linking.....	69
2.38 Carbohydrate Digestion.....	69
2.39 Thin Layer Chromatography (TLC).....	69
2.40 DNSA assay.....	70
2.41 High performance liquid chromatography (HPLC).....	71
2.42 Bioinformatics Softwares.....	72
2.42.1 NTI vector.....	72
2.42.2 PyMol.....	72
2.42.3 Tcoffee.....	73
2.42.4 SignalP.....	73
2.42.5 TMHMM and TMPred.....	73
2.42.6 SMART.....	73
2.42.7 Swiss Model.....	73
2.42.8 DNASTar.....	73
2.42.9 ClustalW2.....	73
2.42.10 BLAST.....	74
2.42.11 DALI.....	74
2.42.12 SWEET2.....	74
2.42.13 LipoP.....	74
2.42.14 Multalin.....	74

Chapter 3 Characterisation of the periplasmic sensor domain of the BT1754 hybrid two-component system

.....	75
3.1 Introduction.....	75
3.2 Results.....	77
3.2.1 Cloning of the predicted periplasmic domain of BT1754.....	77
3.2.1.1 Expression and purification of BT1754peri.....	79
3.2.1.2 Carbohydrate binding studies with BT1754peri.....	79
3.2.1.3 Biophysical studies on BT1754peri.....	83
3.2.1.3.1 BT1754peri oligomeric state in the presence and absence of fructose.....	83
3.2.1.3.2 BT1754peri is thermally stabilized in the presence of fructose.....	85
3.2.1.3.3 BT1754peri is resistant to protease digestion in the presence of fructose.....	86
3.2.2 BT1754peri crystallization.....	87
3.2.3 BT1754peri fold and ligand binding site.....	88
3.2.4 The deletion analysis of BT1754peri.....	94
3.2.5 Mutation of residues involved in sugar binding in BT1754peri.....	94
3.2.5.1 Mutated proteins expression.....	96
3.2.5.2 ITC analysis of mutants.....	96
3.2.5.3 Analysis of ligand binding by fluorescence.....	99
3.2.5.4 CD analysis.....	99
3.2.6 Structure comparisons.....	100

3.3 Discussion.....	105
3.3.1 BT1754peri binds specifically to fructose.....	105
3.3.2 Signal transduction in BT1754.....	106
Chapter 4: Glycoside hydrolase family 32.....	111
4.1 Introduction.....	111
4.2 Results.....	113
4.2.1 Cloning and expression of the GH32s.....	115
4.2.1.1 Signal peptide prediction.....	115
4.2.1.2 Primers and vectors.....	116
4.2.1.3 Proteins expression.....	116
4.2.2 Biochemical characterisation of the GH32s.....	117
4.2.2.1 Enzyme activity against fructans and fructoligosaccharides.....	117
4.2.2.2 Product analysis by HPLC.....	119
4.2.2.3 TLC.....	121
4.2.3 Structural analysis of <i>B. thetaiotaomicron</i> GH32s.....	122
4.2.3.1 BT3082 fructosidase.....	122
4.2.3.1.1 Structure of BT3082.....	122
4.2.3.1.2 BT3082 active site.....	126
4.2.3.1.3 Deletion analysis of BT3082.....	132
4.2.3.2 Structure of BT1760 levanase.....	134
4.3 Discussion.....	137
4.3.1 GH32 enzyme activity.....	137
4.3.2 Role of the N- and C-terminal domains in enzyme activity.....	140
4.3.3 Active site of BT3082.....	141
4.3.4 BT1760.....	142
4.3.5 Applications of GH32s.....	145
Chapter 5 SusD homologue BT1762.....	146
5.1 Introduction.....	146
5.2 Results.....	147
5.2.1 Sequence analysis.....	147
5.2.2 Protein expression.....	149
5.2.3 Ligand binding studies.....	150
5.2.4 BT1762 crystallization.....	152
5.2.5 BT1762 structure.....	153
5.2.6 Mutational analysis of predicted binding site of BT1762.....	162
5.2.6.1 BT1762 mutant protein production.....	162
5.2.6.2 Ligand binding analysis of BT1762 mutants.....	163
5.3 Discussion.....	166
Chapter 6 Analysis of the trans-envelope signalling apparatus of <i>B. thetaiotaomicron</i>.....	171
6.1 Introduction.....	171
6.2 Results.....	180
6.2.1 Protein Expression.....	181
6.2.1.1 Expression of the N-terminal periplasmic domain of the SusC transducers.....	183
6.2.1.2 Expression of the periplasmic domain of the anti-sigma factors.....	184
6.2.1.3 Expression of the cytoplasmic domain of the anti-sigma factors.....	190
6.2.1.4 Expression of sigma factors.....	191
6.2.2 Protein Interaction Studies.....	192
6.2.2.1 The N-terminal periplasmic extension of SusC transducers binds specifically to the periplasmic domain of their cognate anti-sigma factor.....	192
6.2.2.2 The cytoplasmic domains of anti-sigma factors bind specifically to their cognate ECF-sigma factors.....	194
6.2.3 Protein Crystallization.....	196
6.3 Discussion.....	201
Chapter 7 Final Discussion.....	204
Future work.....	208
References.....	209
Appendices.....	231
Appendix A: Chemicals, Media, and Enzymes used in this study.....	231
A1 Chemicals.....	231
A2 Media.....	234

A3 Enzymes.....	234
A4 DNA.....	234
A5 Kits.....	235
Appendix B Crystallization Screens.....	236
B1 Newcastle Screen Composition Table.....	236
B2 The Classics Suite™ (Qiagen) Composition Table.....	239
B3 The JCSG+ Suite™ (Qiagen) Composition Table.....	242
B4 The PACT Suite™ (Qiagen) Composition Table.....	245
B5 The AmSO4 Suite™ (Qiagen) Composition Table.....	247
Appendix C: Vectors used in this Study.....	250
C1 pET-21a.....	250
C2 pET-22b.....	251
C3 pET-28b.....	252
C4 pET-32b.....	253
C5 pET-39b.....	254
C6 pCR®-Blunt.....	255
C7 minipRSET-A.....	256
C8 minipRSET-Trx.....	257
C9 pGEX-6p-1.....	258

List of Figures

Figure 1. 1 Isomers of fructose.....	17
Figure 1. 2 Mannose, glucose, and galactose.....	17
Figure 1. 3 Sucrose.....	18
Figure 1. 4 Fructans.....	19
Figure 1. 5 The major indigenous bacterial species found in the human GI tract.....	21
Figure 1. 6 The Epithelium barrier.....	22
Figure 1. 7 Electron microscope pictures of <i>Bacteroides thetaiotaomicron</i>	24
Figure 1. 8 Typical PUL organization in <i>B. thetaiotaomicron</i>	25
Figure 1. 9 The subregions of σ^{70}	26
Figure 1. 10 ECF Sigma/anti-sigma factor complex.....	27
Figure 1. 11 A classical TCS.....	30
Figure 1. 12 Three major groups of TCSs.....	31
Figure 1. 13 The diversity of TCS pathways.....	32
Figure 1. 14 PAS structures.....	33
Figure 1. 15 TM structure.....	35
Figure 1. 16 The structure of the cytoplasmic domains of a sensor HK.....	36
Figure 1. 17 HAMP structure.....	36
Figure 1. 18 TCS RR multidomains.....	37
Figure 1. 19 Hybrid two component system.....	38
Figure 1. 20 The relative locations of the <i>sus</i> genes of the <i>B. thetaiotaomicron</i>	40
Figure 1. 21 Some known structures of β -barrel membrane proteins.....	41
Figure 1. 22 Conserved domain alignment of SusC to other OMPs.....	41
Figure 1. 23 SusD structure.....	42
Figure 1. 24 Ribbon representation shows example structures of GH clans.....	44
Figure 1. 25 Mechanism of retaining and inverting.....	45
Figure 1. 26 Structural features of endo and exo-acting GHs.....	46
Figure 1. 27 <i>B. thetaiotaomicron</i> glycoside hydrolase family distribution.....	47
Figure 2. 1 DNA standards.....	54
Figure 2. 2 Protein markers used for SDS-PAGE.....	62
Figure 2. 3 MALDI-TOF mass spectrum standards.....	63
Figure 2. 4 ITC instrument used in this study.....	64
Figure 2. 5 Circular dichroism instrument used in this study.....	65
Figure 2. 6 DSC instrument used in this study.....	66
Figure 2. 7 Robot crystallization instrument and its sitting drop method.....	67
Figure 2. 8 Hanging drop crystallization method.....	68
Figure 2. 9 TLC shows the isolated fragments of digested levan.....	70
Figure 2. 10 DNSA assay example.....	71
Figure 2. 11 Fructose, glucose, and sucrose as standard of HPLC.....	72
Figure 3. 1 Polysaccharide utilisation loci of <i>B. thetaiotaomicron</i>	76
Figure 3. 2 The inulin utilisation locus regulated by BT1754.....	77
Figure 3. 3 The prediction of transmembrane regions and orientation of BT1754.....	78
Figure 3. 4 BT1754 amino acid sequence.....	78
Figure 3. 5 Purification of BT1754peri.....	79
Figure 3. 6 BT1754peri does not bind inulin or fructooligosaccharides with a terminal glucose.....	80
Figure 3. 7 BT1754peri binds fructose.....	81
Figure 3. 8 TLC showing purification of inulin oligosaccharides produced by partial acid hydrolysis.....	81
Figure 3. 9 Structures of sugars tested for binding to BT1754peri.....	82
Figure 3. 10 BT1754peri is a monomer in the absence and presence of ligand as assessed by gel filtration.....	84
Figure 3. 11 Crosslinking analysis with BT1754peri in the presence and absence of fructose.....	85
Figure 3. 12 DSC data showing the melt peaks of BT1754peri.....	85
Figure 3. 13 BT1754peri protease K digestion.....	86
Figure 3. 14 The MALDI-TOF map of digested fragment of BT1754.....	87
Figure 3. 15 BT1754peri for crystallization.....	87
Figure 3. 16 The ClustalW alignment of BT1754peri with <i>T. tengcongensis</i> ribose binding protein (RBP).....	88

Figure 3. 17 BT1754peri structure.....	89
Figure 3. 18 BT1754peri binding site.....	90
Figure 3. 19 DALI analysis of BT1754peri.....	92
Figure 3. 20 Tcoffee alignments of BT1754peri homologs from DALI.....	94
Figure 3. 21 The mutants of BT1754peri.....	96
Figure 3. 22 ITC data hint the binding ability of BT1754peri soluble mutants against fructose.....	97
Figure 3. 23 ITC data hint the binding ability of BT1754peri insoluble mutants against fructose.....	98
Figure 3. 24 Fluorescence spectroscopy to assess the binding of refolded WT BT1754peri and W45A to fructose.....	99
Figure 3. 25 CD data showing the secondary structures of soluble BT1754peri mutants.....	100
Figure 3. 26 CD data showing the secondary structures of insoluble BT1754peri mutants.....	100
Figure 3. 27 The crystal structure of ribose binding protein from <i>Thermoanaerobacter tengcongensis</i>	101
Figure 3. 28 The structures overlap of BT1754peri and RBP.....	102
Figure 3. 29 The binding sites overlay of BT1754peri and RBP.....	103
Figure 3. 30 Ribose and fructose overlay.....	104
Figure 3. 31 Hinge prediction of BT1754peri according to MBP structure.....	107
Figure 3. 32 Model for structure of BT1754 sensor HK.....	109
Figure 3. 33 Model of piston like motion of C-terminal helices in BT1754 on fructose binding.....	110
Figure 4. 1 GH32 structures solved to date.....	112
Figure 4. 2 Gene map of the unique locus upregulated in response to growth of <i>B. thetaiotaomicron</i> on inulin.....	113
Figure 4. 3 GH32 multiple sequence alignment.....	114
Figure 4. 4 Signal peptide prediction of GH32 members of <i>B. thetaiotaomicron</i>	116
Figure 4. 5 IMAC purified GH32s from <i>B. thetaiotaomicron</i> used in this study.....	117
Figure 4. 6 HPLC data of <i>Bacteroides</i> GH32s against polymeric fructans.....	119
Figure 4. 7 HPLC data show product profile of BT1759, BT1765, and BT3082 against kestopentaose.....	120
Figure 4. 8 BT1760 digest levan oligos.....	121
Figure 4. 9 TLC shows the process of levan digestion against BT1760 and BT3082 separately.....	122
Figure 4. 10 Structure of BT3082 nucleophile mutant (D131A) with kestose bound.....	123
Figure 4. 11 Structure of BT3082 fructosidase.....	124
Figure 4. 12 Overlay of BT3082 WT with D131A.....	124
Figure 4. 13 DALI results for BT3082 N-terminal domain.....	125
Figure 4. 14 DALI results for BT3082 β -sandwich.....	125
Figure 4. 15 DALI results for BT3082 catalytic domain.....	126
Figure 4. 16 BT3082 D131A binding site to kestose.....	127
Figure 4. 17 Enzyme activity of the BT3082 active site mutants against sucrose, inulin and levan.....	131
Figure 4. 18 CD spectroscopy data showing the BT3082 WT and 17 mutants.....	132
Figure 4. 19 BLAST result for BT3082.....	133
Figure 4. 20 Expression of Δ N BT3082.....	133
Figure 4. 21 Structure of BT1760 endo levanase.....	134
Figure 4. 22 DALI result of the top 40 structures similar to BT1760.....	135
Figure 4. 23 DALI result of β -propeller catalytic module of BT1760.....	136
Figure 4. 24 Model for location and role of GH32s in <i>B. thetaiotaomicron</i>	138
Figure 4. 25 Overlay shows kestose difference of BT3082 and 1-FEH from <i>Chicorium intybus</i>	139
Figure 4. 26 Three domains of BT3082.....	140
Figure 4. 27 BT1760 WT overlays BT3082 WT.....	142
Figure 4. 28 Surface representation shows the overlay GH32s β -propeller catalytic domains with kestose in the active site.....	143
Figure 4. 29 Modeled structures of fructooligosaccharides.....	144
Figure 4. 30 BT1760 catalytic pocket with kestose positioned from BT3082.....	145
Figure 5. 1 SusD location and structure.....	147
Figure 5. 2 Alignment of SusD, BT1762, and four other SusD homologues from <i>B. thetaiotaomicron</i>	149
Figure 5. 3 N-terminal His tagged BT1762 purification.....	150
Figure 5. 4 ITC data show the binding ability of BT1762 to different fructans.....	151
Figure 5. 5 Native BT1762 for crystallization.....	152
Figure 5. 6 Se-Met BT1762 for crystallization.....	153
Figure 5. 7 Structure of BT1762.....	154
Figure 5. 8 BT1762 disulphide ring.....	155
Figure 5. 9 DALI analysis for BT1762 structure.....	155

Figure 5. 10 Cartoon representations show the overlay of BT1762, BT1043, and SusD.....	156
Figure 5. 11 Alignment of BT1762 and SusD.	157
Figure 5. 12 Surface representation shows the possible binding site of BT1762.	158
Figure 5. 13 Key residues around ligands of SusD and BT1043.....	159
Figure 5. 14 Surface and ribbon representation show the aromatic residues of potential binding area.....	160
Figure 5. 15 Putative ligand binding site of BT1762.....	161
Figure 5. 16 SDS-PAGE showing purified BT1762 mutants.	163
Figure 5. 17 ITC data show the BT1762 mutants binding ability against levan.....	164
Figure 5. 18 CD spectroscopy BT1762 wild type and mutants.	165
Figure 5. 19 Structures of levan hexaose, inulin hexaose, and maltoheptaose.	167
Figure 5. 20 SusD and BT1762 ligand binding sites.	168
Figure 5. 21 Overlay the disulphide rings from BT1762 and quinohemoprotein alcohol dehydrogenase.	169
Figure 5. 22 Model for the function of BT1762.	170
Figure 6. 1 Schematic of the FecAIR transmembrane signaling apparatus from <i>E. coli</i>	172
Figure 6. 2 Domain structures of FecAIR.	173
Figure 6. 3 TonB-dependent system.	173
Figure 6. 4 Diversity of TonB-dependent proteins.	174
Figure 6. 5 Alignment of ECF sigma factors to FecI.....	175
Figure 6. 6 The alignment of the cytoplasmic domain (N-terminal) of anti-sigma factors to FecR.....	176
Figure 6. 7 The alignment of the periplasmic domain (C-terminal) of anti-sigma factors to FecR.....	178
Figure 6. 8 The alignment of the SusC transducers of <i>B. thetaiotaomicron</i> with FecA of <i>E. coli</i>	179
Figure 6. 9 ECF-PUL gene clusters for this study.	180
Figure 6. 10 Model of the predicted transenvelope signalling system in <i>B. thetaiotaomicron</i> based on the <i>E. coli</i> Fec system.	181
Figure 6. 11 Domains expressed in this study.....	182
Figure 6. 12 Amplification of BT4705 sigma factor from <i>B. thetaiotaomicron</i> genomic DNA.....	183
Figure 6. 13 A map shows the predicted signal peptide cleavage site of BT0754 SusC transducer.	183
Figure 6. 14 SDS-PAGE of purified BT4707N SusC transducer.	184
Figure 6. 15 Transmembrane prediction for BT0753 anti-sigma factor.	185
Figure 6. 16 SDS PAGE of purified BT1876C fused to thioredoxin.....	186
Figure 6. 17 BT3518C express two bands in pET32b.....	187
Figure 6. 18 BT0753C expressed as two bands when fused to three tags.	188
Figure 6. 19 MADL-TOF map shows the C-terminal fraction of BT0753 peri.	188
Figure 6. 20 Alignment of partial periplasmic domains of anti-sigma factors from <i>B. thetaiotaomicron</i> and <i>E. coli</i> FecR showing position of conserved cleavage site.	189
Figure 6. 21 SDS-PAGE shows purified BT0753CΔ.....	190
Figure 6. 22 SDS-PAGE of purified N-terminal cytoplasmic domain of BT0753 anti sigma factor (BT0753N).	191
Figure 6. 23 SDS-PAGE of purified N-terminal cytoplasmic domain of BT0752 sigma factor.....	192
Figure 6. 24 Native gel of SusC transducer N-extension and periplasmic domain of anti-sigma factor interaction from same PUL.....	192
Figure 6. 25 ITC shows the binding of SusC N-extension and periplasmic domain of anti-sigma factors from same PUL.	193
Figure 6. 26 ITC shows no cross binding of SusC N-extension and periplasmic domain of anti-sigma factors from different gene clusters.....	193
Figure 6. 27 Native gel shows binding of sigma to cytoplasmic domain of anti-sigma factor from the same PUL.	194
Figure 6. 28 ITC of sigma factors and anti-sigma factor cytoplasmic domains from the same PUL.....	195
Figure 6. 29 ITC shows no binding of BT0752 to BT3518N.....	195
Figure 6. 30 Sequence of both the BT0753CΔ and BT0754N genes in pET21a for co-expression.	197
Figure 6. 31 Agarose gel shows co-expression structure of BT0753CΔ and BT0754N in modified pET21a.....	198
Figure 6. 32 SDS-PAGE shows the co-expression of BT0753CΔ with BT0754N.....	198
Figure 6. 33 SDS-PAGE shows the co-expression of BT0753CΔ and BT0754N.....	199
Figure 6. 34 Gel filtration for the complex of BT0753CΔ and BT0754N.....	199
Figure 6. 35 SDS-PAGE shows untagged BT0752 was fished by tagged BT0753N.....	200
Figure 7. 1 Overall process of <i>B. thetaiotaomicron</i> sensing and utilisation of polysaccharides.....	207

List of Tables

Table 1. 1 Distribution of sigma/anti-sigma factor complex in organisms.	28
Table 1. 2 TCSs distribution in organisms.	29
Table 2. 1 <i>E. coli</i> strains in this study.	48
Table 2. 2 List of plasmids used in this study.	48
Table 2. 3 Growth media used in this study.	49
Table 2. 4 Antibiotics used in growth media.	49
Table 2. 5 General PCR working system.	52
Table 2. 6 General PCR programme.	52
Table 2. 7 Mutation PCR reaction system.	53
Table 2. 8 Mutation PCR programme.	53
Table 2. 9 DNA agarose gel running buffer and loading buffer.	53
Table 2. 10 Commercial primers for sequencing.	56
Table 2. 11 SDS-PAGE gel solutions, loading buffer and running buffer.	61
Table 2. 12 AGE gel solutions.	62
Table 2. 13 DNSA solution.	70
Table 2. 14 HPLC buffers and programme.	72
Table 3. 1 HTCSs of <i>B. thetaiotaomicron</i>	77
Table 3. 2 ITC data show the binding ability of BT1754peri against a group of ligands.	83
Table 3. 3 Hydrogen bonding and close contacts between BT1754peri residues and fructose.	91
Table 3. 4 Primers for BT1754peri deletion.	94
Table 3. 5 Primers for mutations of BT1754peri.	95
Table 3. 6 ITC data for binding of WT BT1754peri and P168A mutant to fructose.	98
Table 3. 7 The 13 paired residues of BT1754peri and RBP around ligands.	103
Table 4. 1 GH32 identities.	115
Table 4. 2 Primers for GH32 wild types.	116
Table 4. 3 Kinetic parameters for <i>B. thetaiotaomicron</i> GH32 enzymic activities against oligo and polysaccharides determined by reducing sugar assays.	118
Table 4. 4 Hydrogen bonding and close contacts between BT3082 active site residues and K3.	128
Table 4. 5 Primers of 16 more single mutations of BT3082.	129
Table 5. 1 ITC data show the binding ability of BT1762 to levan.	151
Table 5. 2 Primers for BT1762 mutations.	162
Table 5. 3 ITC parameter of the BT1762 mutants binding to Levan.	165
Table 6. 1 Gene clusters expressed in this chapter.	180
Table 6. 2 Proteins used in this chapter.	182
Table 6. 3 Primers for the N-terminal extension of SusC transducers studied.	184
Table 6. 4 The transmembrane regions and orientation of anti-sigma factors used in this study.	185
Table 6. 5 Primers for the full length of periplasmic domain of anti-sigma factors.	186
Table 6. 6 Primers for the second part of periplasmic domain of anti-sigma factors.	189
Table 6. 7 The primers for cytoplasmic domain of anti-sigma factors.	190
Table 6. 8 Primers for sigma factors.	191
Table 6. 9 ITC data of SusC N-extension and periplasmic domain of anti-sigma factors.	194
Table 6. 10 ITC data of the interaction between cognate sigma and anti-sigma factors.	196

List of genes in this study

Gene ID	Gene classification	Putative function
BT0752	ECF Sigma factor	Sensor
BT0753	Anti-sigma factor	Sensor
BT0754	SusC homologue	Outer membrane protein
BT1754	Hybrid two component system	Sensor
BT1759	Glycoside hydrolase family 32	levanase precursor
BT1760	Glycoside hydrolase family 32	levanase precursor
BT1762	SusD homologue	Outer membrane protein
BT1765	Glycoside hydrolase family 32	levanase precursor
BT1875	SusC homologue	Outer membrane protein
BT1876	Anti-sigma factor	Sensor
BT1877	ECF Sigma factor	Sensor
BT3082	Glycoside hydrolase family 32	levanase precursor
BT3517	ECF Sigma factor	Sensor
BT3518	Anti-sigma factor	Sensor
BT3519	SusC homologue	Outer membrane protein
BT4705	ECF Sigma factor	Sensor
BT4706	Anti-sigma factor	Sensor
BT4707	SusC homologue	Outer membrane protein

Abbreviations

AA	Amino Acid
ABC	ATP-binding cassette
ABP	arabinose-binding protein
AGE	Affinity Gel Electrophoresis
Amp	Ampicillin
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin, fraction V
BT	<i>Bacteroides thetaiotaomicron</i>
CAZY	Carbohydrate-Active enZymes
CD	Circular Dichroism
DNA	Deoxyribonucleic acid
DNSA	3,5-Dinitrophenol
DP	Degree of polymerisation
DSC	Differential Scanning Calorimetry
DTT	Dithiothrietol
ECF	Extracytoplasmic function
ECP	Glu-Cys-Pro motif
EDTA	Ethelene diamine tetra-aceticacid, disodium salt
FEH	Fructan exohydrolase
FPLC	Fast Protein Liquid Chromatography
Frc	D-Fructose
G7	Maltoheptaose
Gal	D-Galactose
GBP	Glucose binding protein
GH	glycoside hydrolase
GI	gastrointestinal
Glc	D-glucose
HBP	histidine-binding protein
HEPES	N-(2-Hydroxyethyl) piperazine-N'-(2-ethanesulfonic acid) sodium salt
HK	Histidine kinase
HPLC	High Performance Liquid Chromatography
HTCS	Hybrid two component system
HTH	Helix-turn-helix
ID	Identity
IM	Inner Membrane
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl- β -D-thiogalactopyranoside
ITC	Isothermal Calorimetry
IUBMB	International Union of Biochemistry and Molecular Biology
K2	Sucrose
K3	1-Kestose
K4	1,1-Kestotetraose
K5	1,1,1-Kestopentaose

Kan	Kanamycin
L2	Levan fraction, 2 fructose units
L3	Levan fraction, 3 fructose units
L4	Levan fraction, 4 fructose units
L5	Levan fraction, 5 fructose units
L6	Levan fraction, 6 fructose units
L7	Levan fraction, 7 fructose units
LB	Luria-Bertani Broth
MBP	Maltose binding protein
MOPS	3-(<i>N</i> -Morpholino) propane sulfonic acid
MW	Molecular Weight
NA	No activity
ND	Not determined
OM	Outer membrane
PAD	pulsed amperometric detection
PAGE	Polyacrylamide Gel Electrophoresis
PAS	Per-Arnt-Sim
PBP	Periplasmic binding protein
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PEG	Polyethylene Glycol
PEG550mme	Polyethylene Glycol 550 monomethyl ether
PINNACLE	Proteome investigations at Newcastle
PL	Polysaccharide Lyases
PULs	Polysaccharide utilisation loci
RBP	Ribose binding protein
RDP	Arg-Asp-Pro motif
rEK	recombinant EntroKinase
RNAP	RNA polymerase
RR	Response regulator
SDS	Sodium Dodecyl Sulphate
Se-Met	Seleno methionine
SMART	Simple Modular Architecture Research Tool
Sus	Starch utilisation system
TCEP	Tris (2-carboxyethyl) phosphine hydrochloride
TCS	Two component system
TEMED	N,N,N',N'-Tetramethylethylenediamine
TIGR	The Institute for Genomic Research
TLC	Thin layer chromatography
TM	Transmembrane
TPR	Tetratrico peptide repeat
WT	Wild type

Chapter 1 Introduction

1.1 Carbohydrates and Nutrition

Nutrition is the provision, to cells and organisms, of the materials necessary (in the form of food) to support life. Many common health problems can be prevented or alleviated with good nutrition. There are seven major classes of nutrients: carbohydrates, fats, fiber, minerals, proteins, vitamins, and water. Carbohydrates and proteins provide four kilocalories of energy per gram, while fats provide nine kilocalories per gram (Berg et al., 2002).

More than 100 years ago, carbohydrate referred to natural chemical structures that have a composition of the formula $(\text{CH}_2\text{O})_n$, where n is any number greater than two. However, many substances slightly different with this formula are also defined as carbohydrate, while many others agree with this formula are not carbohydrate. Nowadays, the definition of carbohydrate is a polyhydroxy aldehyde or ketone (Ajit et al., 1999).

Carbohydrates may be classified as monosaccharides, disaccharides, oligosaccharides, and polysaccharides by the number of sugar units they contain. Simple carbohydrates are metabolized quickly and thus raise blood sugar levels more quickly resulting in rapid increases in blood insulin levels compared to complex carbohydrates (Henry, 2001).

1.1.1 Monosaccharides

Monosaccharides are the simplest carbohydrates, such as glucose, fructose, mannose, galactose, xylose, and ribose. They consist of one sugar and are usually colorless, water-soluble, crystalline solids. Some monosaccharides have a sweet taste. In 1888, Emil Fischer discovered chain-like conformations of glucose, fructose, and mannose. After that, Walter Haworth found that sugars have a ring-like, rather than just a straight-line, arrangement of their carbon atoms. Pyranose ring is a six-membered ring of five carbons and one oxygen formed by the reaction of the OH- group with its aldehyde or ketone forming a hemiacetal or a hemiketal. A furanose is a five-membered furan-based ring structure formed by the reaction of the OH- group with its aldehyde or ketone forming a hemiacetal or a hemiketal. Both forms can exist in equilibrium (Figure 1. 1). The ratio of pyranose and furanose may shift with the change of solvent because of the variation of hydrogen bonds formed with solvent molecule, for example, water prefers pyranose, while dimethyl sulfoxide prefers furanose (Machie and Perlin.S.A, 1966). All monosaccharides are reducing sugars, which means its anomeric carbon is in the free form.

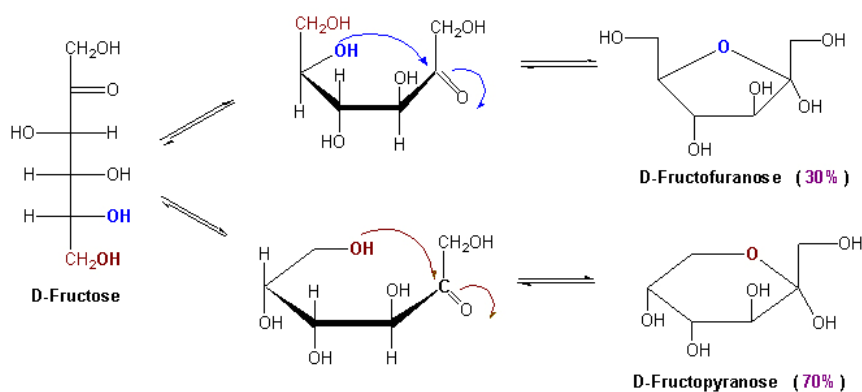


Figure 1. 1 Isomers of fructose.

Monosaccharide can change the conformation from chain-like to ring-like.

There are 4 different schemes to name a monosaccharide. According to the number of carbon atoms (3 to 10), monosaccharides are named into triose, tetrose, pentose, hexose, heptose, octose, nonose, and decose. According to the type of carbonyl group, monosaccharides are classified as aldose and ketose. Aldose has an aldehyde (CHO), while ketose has a ketone (C=O). According to the molecular configuration at the chiral carbon furthest from the aldehyde or ketone group, monosaccharides are classified as D and L. If the chirality at this carbon is equivalent to D-glyceraldehyde's C2, the sugar is D; if it is equivalent to L-glyceraldehyde's C2, the sugar is L. Usually, the D form is found in nature. Unlike glyceraldehyde, D/L designation on more complex sugars is not associated with their direction of light rotation. According to the anomeric carbon, in the chair configuration, the α -isomer has the OH- of the anomeric carbon in an axial position, whereas the β -anomer has the OH- of the anomeric carbon in the equatorial position. The α -anomer and β -anomer of one monosaccharide are always epimers and anomers. Epimers are diastereomers that differ in configuration of only one stereogenic center. Many sugars that are common in nature, such as galactose and mannose are epimers of glucose (Figure 1. 2). Mannose differs to glucose on OH- of C2, while galactose differs on OH- of C4.

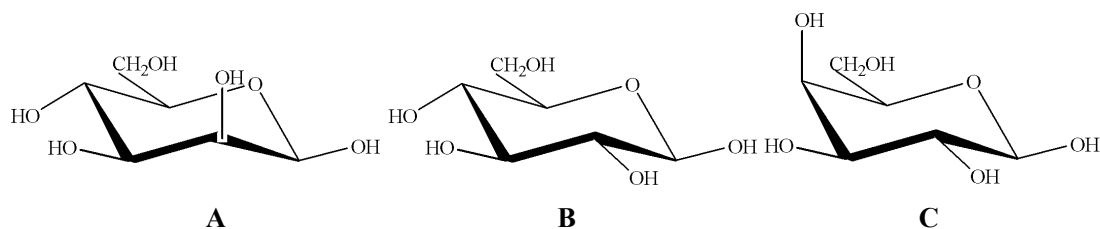


Figure 1. 2 Mannose, glucose, and galactose.

All these monosaccharides are major sugars in nature. A: β -D-mannose, B: β -D-glucose, C: β -D-galactose. A and C are epimers of B.

1.1.2 Disaccharides, oligosaccharides, and polysaccharides

A disaccharide is composed of two monosaccharides via a dehydration reaction (also called a condensation reaction or dehydration synthesis) that leads to the loss of a molecule of water and

formation of a glycosidic bond. The glycosidic bond can be formed between any hydroxyl groups on the component monosaccharide. So, even if both component sugars are the same (e.g., glucose), different bond combinations regiochemistry (D/L) and stereochemistry (α/β) result in disaccharides that are diastereoisomers with different chemical and physical properties.

There are two basic types of disaccharides: reducing disaccharides, in which the monosaccharide components are bonded by hydroxyl groups; and non-reducing disaccharides, in which the components bond through their anomeric centers (Matsuda, 1957). Most disaccharides are the reducing type, but a commonly occurring non-reducing disaccharide is sucrose, which is composed of a glucose and a fructose molecule linked by an α -1, 2 (from glucose) or β -2, 1 (from fructose) bond.

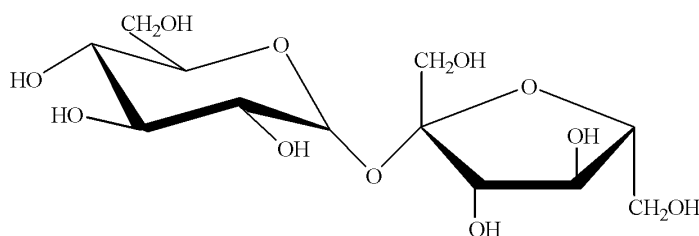


Figure 1.3 Sucrose.

Sucrose is non-reducing disaccharide formed by a glucose and a fructose.

An oligosaccharide is a saccharide polymer containing a small number (typically three to ten) of component sugars. Polysaccharides are relatively complex carbohydrates. They are usually large, often branched, macromolecules. They tend to be amorphous, insoluble in water, and have no sweet taste. Unlike proteins which are formed by 23 different amino acid residues, polysaccharides seldom contain more than 3 different kinds of monosaccharides or linkages, normally only 1 or 2 (Ajit et al., 1999). Examples include storage polysaccharides such as starch, fructan, and glycogen, and structural polysaccharides such as cellulose and chitin.

1.1.3 Fructans

Fructans are important storage energy sources, similar in function to starch, naturally occurring very widely in commercial plants such as wheat, chicory, artichokes, asparagus, green beans, leeks and onions, which are typically found in plants roots or rhizomes (Ritsema and Smeekens, 2003). Most plants which synthesize and store inulin do not store other materials such as starch (Roberfroid, 2005). Fructans have been shown to play a role in resistance to cold, and drought in plants (David and Cynthia, 1998; Pilon-Smits *et al.*, 1995; Spollen and Nelson, 1994). The major mechanism is the fast polymerization/depolymerization of fructans on varying conditions offering a flexible and efficient osmoregulatory region (Ritsema and Smeekens, 2003). Fructans are also synthesised by a

number of bacteria where they are involved in dental biofilm formation (Rozen *et al.*, 2001).

Fructans are homopolymers of β -D-fructose with a single glucose at the chain end, which lead to rather different polymer behavior as compared with the pyranose ring-based glucans. Fructans occur in three main types: inulin (2 \rightarrow 1 linkage), levan (2 \rightarrow 6 linkage) and graminan (mixed 2 \rightarrow 1 linkage and 2 \rightarrow 6 linkage (Ritsema and Smeekens, 2003). Inulins are only produced by plants whereas levans and graminins are produced mostly in monocots and bacteria (Ritsema and Smeekens, 2003). The difference between plant and bacterial levan appears mainly to be chain length; the average DP (degree of polymerisation) of bacterial levan is 10^5 to 10^6 fructose units, whereas plant levan is much shorter (only 6-42 fructose units) (Wack and Blaschek, 2006). The main reason is plants lack the key enzymes to produce high DP fructans, such as levansucrase (Hettwer *et al.*, 1995). The average DP of inulin is \sim 12 for the polymer isolated from chicory, but the chains range from \sim 2-60 sugar units.

Despite their presence in a range of food crops, fructans are not utilised by our bodies (although we can readily metabolise fructose) due to the lack of the β -fructosidase enzymes required for their depolymerisation (Magnus *et al.*, 2006). However our normal gut microbiota are able to utilise these polymers and thus fructans have been shown to act as prebiotic molecules, enabling our beneficial gut bacteria to outcompete pathogenic non-fructan metabolizing bacteria (Kolida and Gibson, 2007).

There is a growing demand for fructan oligo and polysaccharides as health promoting prebiotics (Kolida and Gibson, 2007). Fructose itself is the sweetest of native saccharides and is added to a lot of food and drink as a sweetener. Fructose is often recommended for diabetics because it does not trigger the production of insulin (Sato *et al.*, 1996). Inulin is also commonly used in low fat foods as a replacement for fat as it has a similar ‘mouth feel’, but very low calorific content.

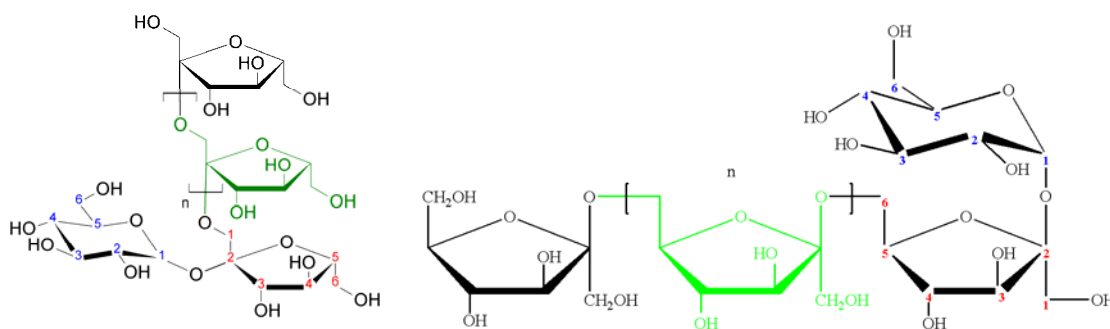


Figure 1. 4 Fructans.

Left: Inulin, β -D-fructose-[(2 \rightarrow 1)- β -D-fructose] $_n$ -(2 \rightarrow 1)- α -D-glucose, Right: Levan, β -D-fructose-[(2 \rightarrow 6)- β -D-fructose] $_n$ -(2 \rightarrow 1)- α -D-glucose. Graminan is branched mixture of (2 \rightarrow 1) and (2 \rightarrow 6) linkages. Note the terminal glucose at each chain end.

1.2 Niche of the human gut

The human body is home to a vast number of microbial cells that have coevolved with us and thus play a significant role in our normal biology. Of the various niches colonized by microbes, the gut is by far the most densely inhabited (Figure 1. 5). Indeed, 500-1000 species of gastrointestinal microbes (mostly bacterial) are present at densities of 10^{11} per gram of luminal contents, making it the most densely populated ecosystem on the planet (Berg, 1996). Most of these, up to 10^{14} microbial cells, are found in the large intestine, which is about 10 times of numbers of human cells (Backhed *et al.*, 2005). From the process of birth, a sterile new baby is present with a large mixture of microbes from the environment. Only some of these organisms will colonize in the pristine host leading to a complex ecosystem known as indigenous microbiota, while most of these bacteria will not be maintained in the human gastrointestinal tract (Savage, 1977). The composition of the microbiota can vary under some conditions, such as antibiotic treatment and dietary changes, but normally remains constant as defined by the cellular protein and fatty acid analysis as well as the 16S rRNA gene fingerprinting (Moore and Moore, 1995; Zoetendal *et al.*, 1998).

The *Bacteroidetes* and the *Firmicutes* are the major divisions present and constitute >99 % of all phylotypes present in the colon with an approximate 50:50 split in the numbers of each based on 16S rRNA (Backhed *et al.*, 2005; Eckburg *et al.*, 2005; Zoetendal *et al.*, 2006). Only one archaeal species *Methanobrevibacter smithii* was found, which produces methane. Four major fungal phyla: *Ascomycota*, *Basidiomycota*, *Chytridiomycota* and *Zygomycota* form ~ 2 % of the microflora, which are mainly present in the intestinal microbial community. Most fungi are nonpathogenic to human and animals, except *Cryptococcus neoformans*. Fungi produce immune-modulating compounds, which may alter immune system of host and take part in microbial immune homeostasis (Noverr *et al.*, 2004; Scupham *et al.*, 2006).

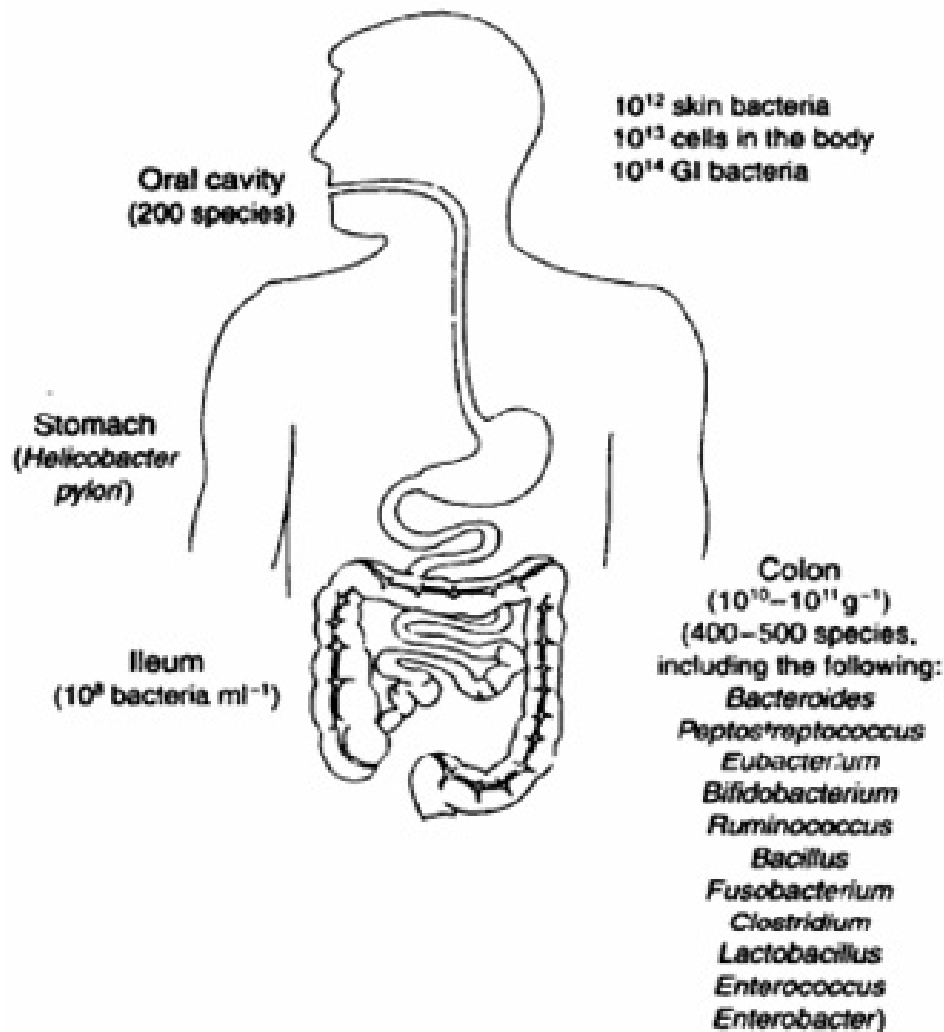


Figure 1. 5 The major indigenous bacterial species found in the human GI tract. The microflora found in the colon are listed in descending order of prominence. Cited from (Berg, 1996).

1.2.1 The roles of gut bacteria

Microbes in the gut play an important role in normal human health and nutrition. There are predicted 2 to 4 million genes are packed in the intestinal community, which will generate metabolic functions beyond those of our own genome (Hooper and Gordon, 2001; Hooper *et al.*, 2002).

1.2.1.1 In normal human health

Gut bacteria produce antimicrobial substances termed bacteriocins to inhibit their competitors to grow (Lievin *et al.*, 2000). Non-pathogenic bacteria adhere to the epithelial cells to prevent attachment of pathogenic enteroinvasive bacteria physically (Figure 1. 6) (Bernet *et al.*, 1994).

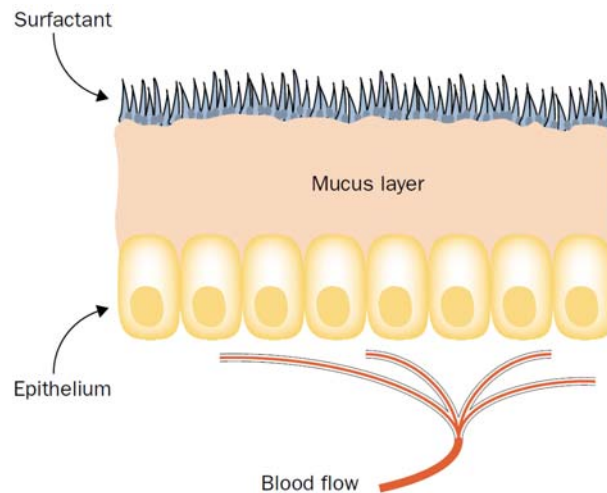


Figure 1. 6 The Epithelium barrier.

The epithelium has mucus layer as defensive barrier. But, GI bacteria are the important line outside of the surfactant to protect from exotic microbes from luminal environment. Cited from (Guarner and Malagelada, 2003).

Gut microflora have been shown to enhance the threshold of tolerance of the immune system to microbial antigens (Braun-Fahrlander *et al.*, 2002). GI bacteria can evade the immune response by changing their surface polysaccharide capsule, or precoating with IgA helps bacteria evade the human immune response (Krinis *et al.*, 2001; Smith and Macfarlane, 1996; van der Waaij *et al.*, 1996). Gut bacteria have been shown to play a role in defence by producing anti cancer molecules such as butyrate (Kyriakidis *et al.*, 2008; Mazmanian *et al.*, 2005).

Bacteria can improve human health as probiotics (Guarner and Schaafsma, 1998). Supplementation of *Lactobacillus rhamnosus* strain GG defends diarrhea in children (Oberhelman *et al.*, 1999). Oral probiotics elevate IgA level against rotavirus infection of children (Majamaa *et al.*, 1995). Actually, the newborn are predicted to need normal microbiota immediately after delivery to develop the systemic and mucosal immunity (Isolauri *et al.*, 2004).

1.2.1.2 In nutrition

Bacteria also play an important role in human nutrition, breaking down otherwise indigestible dietary polysaccharides, including resistant starches and pectins, to provide us with ~10-15 % of our daily calories from the fatty acid products of their fermentation (Backhed *et al.*, 2005). In a human adult colon, the daily substrate utilisation is about 20-60 g carbohydrates (Silvester *et al.*, 1995). Many members of the gut microflora (Finegold *et al.*, 1983) obtain energy by hydrolysis of complex carbohydrates of dietary, host polysaccharides (such as xylan, pectins, starch, and mucins) and glycoconjugates (glycoproteins and glycolipids) in the colon (Prins, 1977; Roberfroid *et al.*, 1995; Vercellotti *et al.*, 1977). They also can utilise simple sugars when present, but 98 % of simple sugars, such as glucose and sucrose, are absorbed in the small intestine (Hooper *et al.*, 2001;

Salyers and Leedle., 1983).

Colonic microorganisms have been shown to improve the absorption of metals such as iron, calcium, and magnesium and are known to synthesise several important vitamins (B12 and biotin) and help the absorption of others (e.g. vitamin K) (Conly *et al.*, 1994; Roberfroid *et al.*, 1995; Younes *et al.*, 2001). Bacteria used in yoghurt improve lactose digestion and consumption and calcium intakes (Labayen *et al.*, 2001). Furthermore, indigenous gut bacteria have been shown to stimulate the development of the epithelial blood vessels (Bons *et al.*, 1997).

1.2.1.3 Possible harmful effects of gut bacteria

Diets with high fat and meat, and low vegetables increase the risk of colon cancer possibly because such diets increase the amount of bacterial N-nitroso compounds in feces, which are well known promoters of colon cancer, inducing G→A transitions in codons 12 and 13 of K-ras (Bingham, 1999; Hill, 1991; Hughes *et al.*, 2001). Some intestinal bacteria can detoxify such substances whereas other bacteria, like *Clostridium* genera are thought to increase the growth of colonic tumours (Horie *et al.*, 1999; Wollowski *et al.*, 2001). Also, some research pointed out that overgrowth of gut bacteria may be harmful and can elicit gastritis, especially in the elderly (Husebye, 2005). Gut microbiota could be an extra contributor to the formation of obesity by small alteration in energy harvest from the diet (Flegal and Troiano, 2000; Turnbaugh *et al.*, 2006). Hopefully, the balance of benefit and harm will be achieved by controlling the diet (Rastall, 2004).

Although we have co-evolved with our indigenous microflora, it is possible to live without these microbes as evidenced by the existence of germfree animals in research labs (Luckey, 1972).

1.2.2 Bacteroides thetaiotaomicron

Bacteroides thetaiotaomicron is a Gram-negative anaerobic bacterium (Figure 1. 7) that is a prominent member of the normal human gut microflora, 13 % of *Bacteroidetes* and 6 % of total bacteria detected by 16S rRNA analysis (Eckburg *et al.*, 2005). Its 6.26 Mbp genome has been sequenced (ATCC 29148; originally isolated from the feces of a healthy adult human) (Xu *et al.*, 2003).

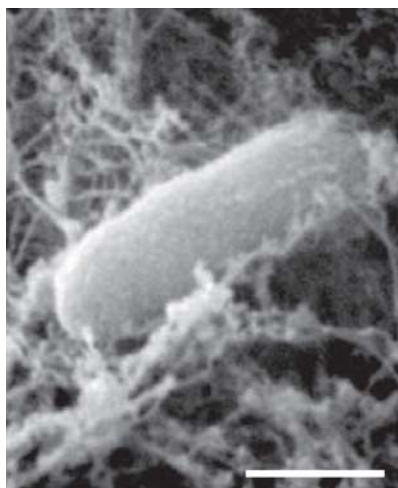


Figure 1. 7 Electron microscope pictures of *Bacteroides thetaiotaomicron*.

B. thetaiotaomicron embedded in the mucus layer overlying the epithelium in a monoassociated mouse gut. Scale bar is 0.5 μm . Cited from (Sonnenburg *et al.*, 2005).

The most striking feature is large expansions in groups of proteins involved in the sensing, acquisition and degradation of carbohydrates. These include 209 homologues of outer membrane polysaccharide-binding and transport proteins SusC (107) and SusD (102), 226 glycoside hydrolases (GH), as well as 15 polysaccharides lyases (PL), 20 sugar-specific transporters, and 21 permease subunits of ATP-binding cassette (ABC) transporters (Magnus *et al.*, 2006; Shipman *et al.*, 2000). The number of predicted GH genes in the proteome exceeds that in any other sequenced bacterium, including other human gut symbionts. In comparison, the human 2.85-Gb genome only contains 98 GH genes. The human genome is deficient in enzymes required to degrade of xylan-, pectin-, and arabinose-containing polysaccharides which are common components of dietary fiber while *B. thetaiotaomicron* has 64 corresponding enzymes and appears capable of cleaving most glycosidic bonds found in nature (Xu and Gordon, 2003). Coevolution of human glycan diversity and a large group of microbial GH ensures the community can rapidly respond to changes in the diet. Research on germ-free mice colonized by *B. thetaiotaomicron* showed that in the suckling period, gut *B. thetaiotaomicron* prefers host-derived polysaccharides and mono- and oligosaccharides from mother's milk. After weaning, *B. thetaiotaomicron* starts to utilise plant-derived dietary polysaccharides (Magnus *et al.*, 2006).

B. thetaiotaomicron also contains a large number of inner membrane spanning environmental sensors and regulators including 50 extracytoplasmic function sigma factors (ECF- σ), 26 anti-sigma (anti- σ) factors (typically found downstream of a co-transcribed sigma), 79 classic two-component systems (TCS) and 32 novel 'hybrid' two-component systems (HTCS), which comprise all components of a classical TCS in a single peptide (Magnus *et al.*, 2006; Xu *et al.*, 2003).

Many of these regulatory genes appear to be involved in controlling carbohydrate utilisation in *B. thetaiotaomicron*, as they are positioned adjacent to SusC and SusD homologues as well one or more catabolic genes such as GHs (Martens et al., 2008). These gene clusters are named polysaccharide utilisation loci (PULs) (Figure 1. 8). The arrangement in close proximity of environmental sensors-regulators with the outer membrane polysaccharide binding and transport proteins (SusC and SusD) and glycoside hydrolases suggests the co-regulation of these components to enable *B. thetaiotaomicron* to sense and respond to different types of carbohydrates depending on their availability in its local environment with high efficiency. Indeed recent data have shown that this is the case (Martens et al., 2008).

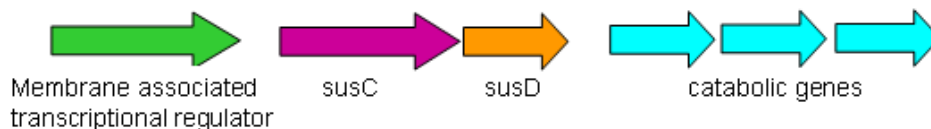


Figure 1. 8 Typical PUL organization in *B. thetaiotaomicron*.

Typical polysaccharide utilisation locus in *B. thetaiotaomicron* composed of a membrane associated transcriptional regulator (TCS, HTCS or ECF-sigma/anti sigma factor), a SusC homologue (β barrel porin), a SusD homologue (outer membrane lipoprotein) and several catabolic genes, commonly glycoside hydrolases or polysaccharide lyases. These genes are closely linked on the genome, but are not always present as operons.

The extensive number of environmental sensor regulators coupled with the availability of genetic systems for manipulating *B. thetaiotaomicron*, the relative ease of culturing this aero-tolerant anaerobe and a complete genome sequence makes *B. thetaiotaomicron* an outstanding model system for increasing our understanding of the mechanisms of extracellular signal perception and transduction in bacteria, furthermore, for a more thorough understanding of the relationship between humans and their intestinal microflora (Hooper *et al.*, 2003; Hooper *et al.*, 2001; Hooper *et al.*, 1999).

1.2.3 Bacterial extracellular sensing and signalling systems

Survival in the microbial world depends on the ability of the cell to sense and respond to changes in its environment. In bacteria this is commonly achieved using multidomain proteins that span the cytoplasmic membrane and contain an extracellular ‘input’ domain that senses the environmental condition, linked to a cytoplasmic ‘output’ domain that is able to mediate the appropriate cellular response, either at the transcriptional or post translational level (Hughes and Mathee, 1998). The main classes of extracellular sensor regulators are extracytoplasmic function (ECF) sigma/anti-sigma factors and two component systems (TCS).

1.2.3.1 ECF sigma/anti-sigma factors

Bacterial sigma factors can be divided into two large families: the σ^{70} and σ^{54} (Helmann and Chamberlin, 1988). The description of sigma factors reflects their molecular mass in kDa. Extracytoplasmic function (ECF) sigma factors are a subfamily of σ^{70} .

Different sigma factors regulate different sets of genes to adjust metabolism with cues from outside and inside cells (Brinkman *et al.*, 1999; Engels *et al.*, 2004; Gunesekere *et al.*, 2006). Sigma factors activate the transcription of genes by binding to specific promoters while associated with RNA polymerase (RNAP) (Carter *et al.*, 2002; Horsburgh and Moir, 1999; Newman *et al.*, 1999; Sun *et al.*, 2004; Wilson and Lamont, 2000; Yeoman *et al.*, 1999; Yeoman *et al.*, 2003). Sigma factors can not bind DNA in the absence of RNAP (Malhotra *et al.*, 1996). These sigma factors recognize promoters with similar elements around 10 and 35 bp upstream of the transcription initiation site.

σ^{70} could be classed as four groups according to gene structure and function (Helmann, 2002; Paget and Helmann, 2003). Group 1 belongs to the essential primary sigma factors, which are classed into 4 sequence-conserved regions (σ_1 , σ_2 , σ_3 , σ_4 , with subregions 1.1, 1.2, 2.1, 2.2, 2.3, 2.4, 3.0, 3.1, 3.2, 4.1 and 4.2) (Figure 1. 9) that fold into separate structural domains connected by linkers (Gross *et al.*, 1998; Malhotra *et al.*, 1996; Murakami *et al.*, 2002; Vassylyev *et al.*, 2002). The region 2 of sigma factors is the highest conserved region, in which subregions 2.3 and 2.4 are the promoter melting and recognition region (Malhotra *et al.*, 1996). Group 2 is close to the primary σ factors, but is dispensable to cell growth. Group 3 is secondary sigma factors, which are significantly smaller than group 1 and 2. In many cases, region 3 is absent.

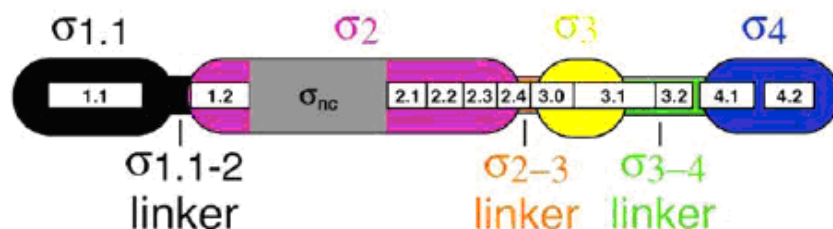


Figure 1. 9 The subregions of σ^{70} .
Cited from (Mooney *et al.*, 2005).

Group 4 contains ECF sigma factors which respond to signals from the extracytoplasmic environment including osmotic and oxidative stress, iron and haem uptake, temperature, pH, detergents, salt, and phage (Alvarez *et al.*, 2007; Arraiz *et al.*, 2001; Biville *et al.*, 2004; Burger *et*

al., 2000; Dona *et al.*, 2008; Horsburgh and Moir, 1999; Manganelli *et al.*, 2001; Maunsell *et al.*, 2006; Ryu *et al.*, 2006). ECF σ factors are more varied in sequence than other sigma factors. In many cases, they equal or exceed the number of other sigma factors found (Table 1. 1) (Helmann, 2002). Bacteria in poor environments and frequent variation of nutrients are especially rich in ECF sigma factors (Braun *et al.*, 2003b). Generally, the ECF sigma factors have several common characteristics (Helmann, 2002). They do not have a conserved subregion 3 which interacts with DNA -10 promoter element; instead they identify similar promoter elements with an 'AAC' motif in the -35 region. Finally, more than one ECF σ factor can target the same promoter for regulation, especially in bacteria containing many ECF σ factors (Huang and Helmann, 1998).

Normally, sigma factors are inhibited by anti-sigma factors (Anthony *et al.*, 2004; Brown and Hughes, 1995; Brutsche and Braun, 1997; Gorham *et al.*, 1996; Hughes and Mathee, 1998; Rowen and Deretic, 2000). The non-ECF anti-sigma factors are all cytoplasmic proteins, while the ECF anti-sigma factors are transmembrane proteins spanning the cytoplasmic membrane with a periplasmic sensor domain and a cytoplasmic domain that holds the sigma factor in an inactive form until the appropriate extracellular signal is received (Figure 1. 10) (Browning *et al.*, 2003). The activity of ECF anti-sigma factors is controlled by the interaction of the periplasmic domain of these proteins with specific molecules in the periplasm (Helmann, 2002). The ECF anti-sigma factors are sometimes also controlled by an anti-anti-sigma factor, which locates to membrane (Masuda *et al.*, 2004).

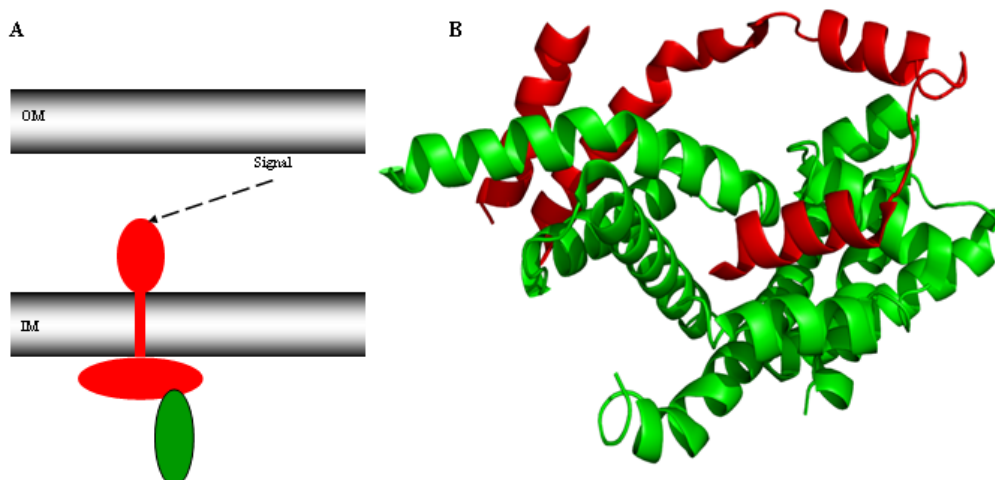


Figure 1. 10 ECF Sigma/anti-sigma factor complex.

Anti-sigma factor is colored in red, while sigma factor is colored in green, in both location diagram (A), and in cartoon representation (B). This sigma/anti-sigma factor is sigma 28/FilM complex in *E. coli* (PDB ID 1SC5) (Sorenson *et al.*, 2004). When signal is released to anti-sigma factor, the conformational change of FilM will liberate sigma factor to initiate gene transcription.

Organism	Total Sigma factors	ECF sigma factors
<i>Streptomyces coelicolor</i>	65	41
<i>Bacteroides thetaiotaomicron</i>	54	50
<i>Pseudomonas aeruginosa PAO1</i>	24	19
<i>Bacillus halodurans C-125</i>	19	11
<i>Caulobacter crescentus</i>	17	13
<i>Bacillus subtilis 168</i>	17	7
<i>Escherichia.coli K12 MG1655</i>	15	3
<i>Mycobacterium tuberculosis H37Rv</i>	13	10
<i>Clostridium perfringens</i>	12	3
<i>Vibrio cholerae El Tor N16961</i>	8	3
<i>Porphyromonas gingivalis</i>	7	5
<i>Enterococcus faecalis V583</i>	6	2
<i>Haemophilus influenza</i>	5	2
<i>Mycobacterium leprae TN</i>	4	2
<i>Aquifex aeolicus VF5</i>	4	0
<i>Bifidobacterium longum</i>	3	1
<i>Mycoplasma genitalium G-37</i>	1	0

Table 1. 1 Distribution of sigma/anti-sigma factor complex in organisms.

Based on the genome sequences at TIGR (www.tigr.org). See references (Helmann, 2002) and (Xu *et al.*, 2003).

ECF sigma/anti-sigma factors are often involved in trans-envelope signalling, a process that also involves a TonB-dependent outer membrane transducer protein. These outer membrane transducers are β -barrel porins that couple import of specific molecules with signalling via a periplasmic tail that contacts the cognate anti-sigma factor. In *B. thetaiotaomicron* many of the ECF-sigma/anti-sigma factor gene pairs are in PULS that contain SusC homologues with an N-terminal extension and are predicted to be TonB dependent transducers. More details of this trans envelope signalling system are present in chapter 6.

1.2.3.2 Two component systems

Two-component systems (TCS) are one of the most commonly encountered means of extracellular sensing and signalling in both Gram-positive and Gram-negative bacteria (Table 1. 2, and Figure 1. 11). TCS are also found in archaea, fungi, and some plants, but not in animals. The number of TCS in different bacteria varies considerably depending on the environmental niche of the organism (Chang and Stewart, 1998; Loomis *et al.*, 1997; Perraud *et al.*, 1999; Wurgler-Murphy and Saito, 1997). Most sequenced bacteria have more than one TCS, except *Mycoplasma* species which are parasites and very stable environments, so do not need many sensors. Organisms with constantly changing or poor environments will need more sensors. TCSs may share the same ancestor as heat-shock protein Hsp90, which indicated TCSs are conserved sensing proteins (Dutta *et al.*, 1999).

Organism	Histidine Kinase (HK)		Response Regulator (RR)
	typical	hybrid	
<i>Escherichia .coli</i>	24	5	32
<i>Streptococcus pneumoniae</i>	13	0	14
<i>Synechocystis sp.</i>	25	16	38
<i>Methanobacterium thermoautotrophicum</i>	14	1	9
<i>Dictyostelium discoideum</i>	0	11	1
<i>Saccharomyces cerevisiae</i>	0	1	2
<i>Arabidopsis thaliana</i>	2	9	16
<i>Bacteroides thetaiotaomicron</i>	48	41	31
<i>Pseudomonas aeruginosa</i>	50	16	58
<i>Streptomyces coelicolor</i>	64	1	70

Table 1. 2 TCSs distribution in organisms.

Based on (West and Stock, 2001) and (Zhang and Shi, 2005).

TCSs regulate a wide range of bacterial responses to environmental variations, including oxidative stress, pH, temperature, and the presence of nutrients and harmful compounds (Adams *et al.*, 1999; Chen *et al.*, 2008; Gottschalk *et al.*, 2008; Griffiths *et al.*, 2008; Krzeslak *et al.*, 2008; Lee *et al.*, 2008b; Li *et al.*, 2008; Marutani *et al.*, 2008; Nes and Eijsink, 1999; Putrins *et al.*, 2008; Sebert *et al.*, 2005; Shanks *et al.*, 2008; Williamson *et al.*, 2008; Zoetendal *et al.*, 2008).

Signal transduction in TCS is via a phosphorelay and the minimal TCS, as the name suggests, is composed of two components, a membrane associated sensor histidine kinase (HK) and a cytoplasmic response regulator (RR) (Figure 1. 11). The minimal HK is composed of three distinct domains; an extracellular or membrane embedded sensor or cytoplasmic domain that senses a specific environmental signal, a cytoplasmic phosphoacceptor domain and an ATP-dependent histidine kinase domain. The RR is a free cytoplasmic protein composed of a receiver domain and an output domain that once activated usually binds DNA to affect transcription and to elicit the appropriate cellular response (Lee *et al.*, 2008b; Stock *et al.*, 2000). Typically, the genes for the HK and the cognate RR are close to each other. But some bacteria show an orphan gene organization of TCS pairs, especially *Myxococcus xanthus* which contains 146 HKs and more than half are separated by two or more genes from RR genes (Mascher *et al.*, 2006).

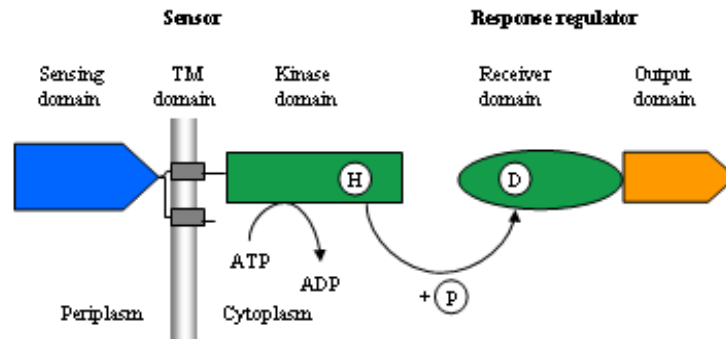


Figure 1. 11 A classical TCS.

TCS comprises a membrane associated sensor and a cytoplasmic response regulator. In the sensor there is an N-terminal extracellular sensing domain (blue arrow) and a C-terminal cytoplasmic kinase domain (green rectangles). In the response regulator, there is an N-terminal receiver domain (green circles) and a C-terminal output domain (orange arrow). The H in the kinase domain represents the conserved His residue that is phosphorylated, while the D in receiver domain of RR presents the Asp residue that receives the phosphate from the HK. The output domain can regulate at the level of transcription or post translation. The pathway is ATP dependent.

Based on the biological function of signal transduction, >18,000 sensor kinases in the SMART database (<http://smart.embl-heidelberg.de/>) are classified so far into three major groups: periplasmic-sensing HKs, transmembrane-sensing HKs (TMHK), and cytoplasmic-sensing HKs (Figure 1. 12) (Mascher *et al.*, 2006). The biggest group is periplasmic-sensing HKs including gram-positive bacteria which have no periplasm (extracellular HKs here are classified as periplasmic for simplicity). TMHK is the smallest group, which has high diversity, possessing 2 to 20 transmembrane regions connected with very short intra- or extracellular loops (20-50 residues). TMHK lacks the obvious extracellular sensing domain of the first group (50-300 residues). Their stimuli are either membrane associated or with in the membrane, such as turgor or solute gradients. Most sensors of gram-positive bacteria belong to this class. Cytoplasmic-sensing group are either free or with membrane-anchor proteins in cytoplasm. They detect the stimuli inside the cell, such as cell cycle, and cell development. Some HKs are mixed periplasmic/cytoplasmic-sensing HK to detect additional stimuli.

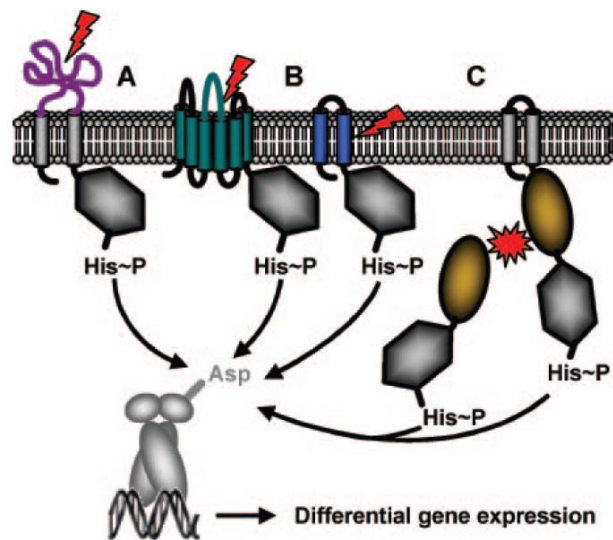


Figure 1. 12 Three major groups of TCSs.

A: Periplasmic-sensing HKs. **B:** Transmembrane-sensing HKs, with 2-20 transmembrane regions, either with short extracellular loops. **C:** Cytoplasmic-sensing HKs, either free or with membrane-anchor. The red arrow or red star presents the stimulus. The gray protein with Asp presents the RR. Cited from (Mascher *et al.*, 2006).

Periplasmic-sensing HKs comprise two regions: an N-terminal periplasmic sensing domain linked by TM helices on both sides (TM1 and TM2) and a C-terminal cytoplasmic transmitter domain (Figure 1. 11 and Figure 1. 12A). Usually, the binding of the sensor domain to a small extracellular signalling molecule or interaction with a physical stimulus causes the activation of the HK, which carries out an ATP-dependent autophosphorylation of a special conserved histidine residue on the phosphoacceptor (Belcheva and Golemi-Kotra, 2008; Filippou *et al.*, 2008). The phosphate is then transferred from the histidine to a conserved aspartate of the receiver domain of the RR. This phosphorylation event activates the output domain of the RR, usually via a conformational change (Whitworth *et al.*, 2008). This basic modular scheme allows adaption to a huge range of environmental signals by simply varying the sensory input domain, while the rest of the phosphorelay system remains highly conserved.

This kind of classical organization can vary in the nature of the cytoplasmic linker region between TM2 and the transmitter domain, or presence of an additional phosphorylation domain after the transmitter domain. The latter complex systems involve multiple His-donors and Asp-receivers and form complex multistep phosphorylation pathways (Figure 1. 13). Both prokaryotes and eukaryotes have these complex TCS, although as expected they are more common in the latter (Appleby *et al.*, 1996). A phylogenetic study revealed that there was no common ancestor for complex TCS and that the additional domains were likely a result of gene duplication from the original TCS (Zhang and Shi, 2005). Normally, in gram-negative bacteria, components of complex HKs are in one peptide, while in gram-positive bacteria, such as *B. subtilis*, these components are individual (Figure 1. 13b).

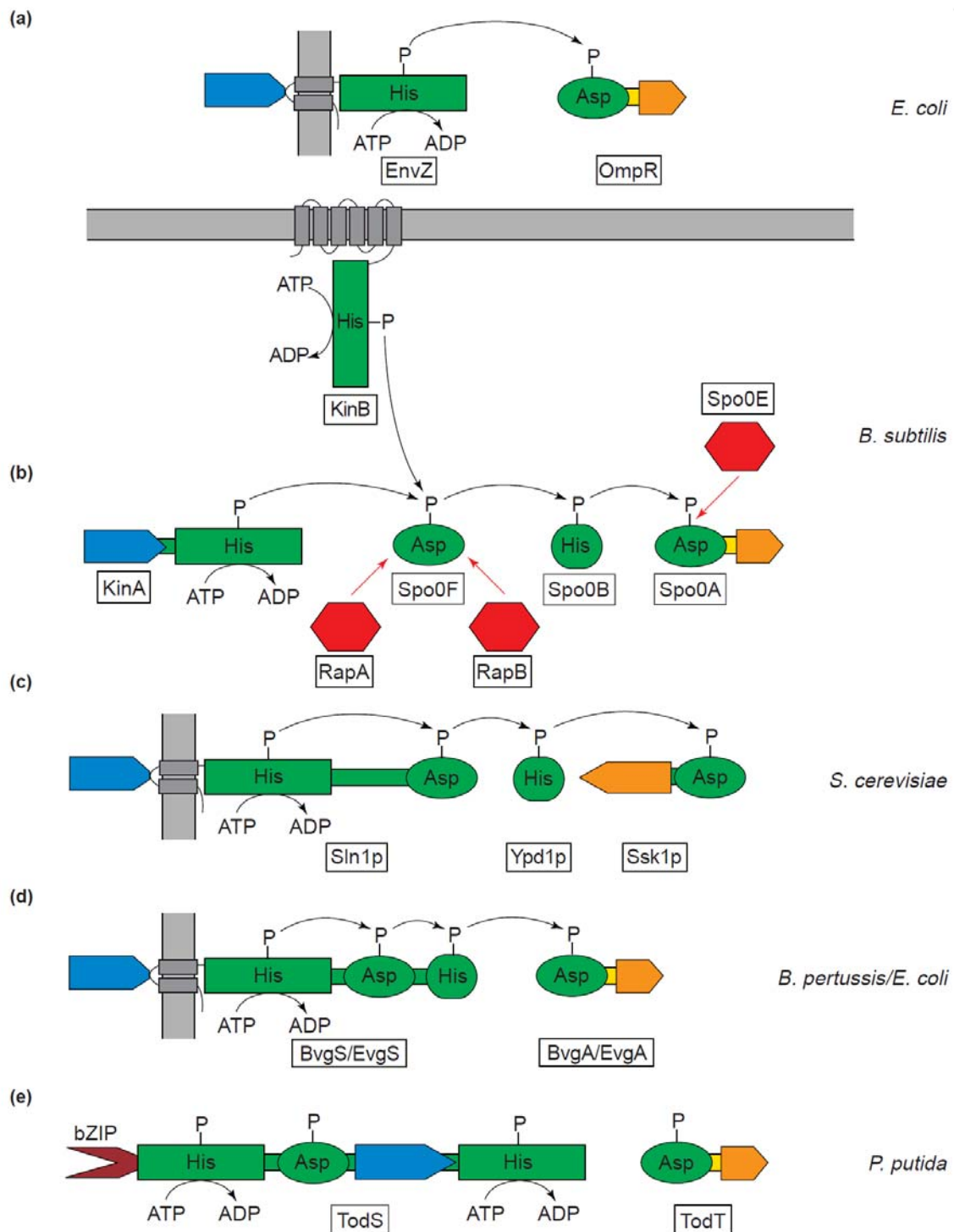


Figure 1. 13 The diversity of TCS pathways.

The HK and RR receiver are shown as green rectangles and green circles marked with His and Asp respectively, while the sensing input domains are blue arrows. Black arrows show the phosphate flow. Membranes are indicated as grey bars. Orange arrows represent the output domain of the RR. **(a)** The osmoregulatory EnvZ/OmpR system of *E. coli* is an example of ‘classical’ TCS. In such systems, phosphotransfer to the RR occurs in a single step (Siryaporn and Goulian, 2008). **(b)** The sporulation control system of *Bacillus subtilis* has two sensor HKs, one of which is free in the cytoplasm, a cytoplasmic Asp containing phosphotransfer protein (APT) and a cytoplasmic cytoplasmic His containing phosphotransfer protein (HPT), leading to a His–Asp–His–Asp relay. In this case, all domains form independent proteins (Weinrauch *et al.*, 1990). **(c)** The osmoregulatory SLN1-YPD1-SSK1 system from yeast *Saccharomyces cerevisiae* has a three-step phosphorylation via a cytoplasmic HPT (Maeda *et al.*, 1994). **(d)** The BvgAS/EvgAs system of *Bordetella pertussis/E. coli* has three-step phosphorylation as well, but the HPT is part of the same polypeptide as the membrane-associated HK (Beier *et al.*, 1995). **(e)** The TodST system of *Pseudomonas putida* has two HK domains and two Asp receiver domains, with the sensing domain after the first receiver domain. Cited from (Perraud *et al.*, 1999).

1.2.3.2.1 Periplasmic-sensing HK sensor domains

In the simplest cases, the signal detection mechanism of periplasmic-sensing HKs is direct interaction between the sensor domain and a small chemical molecule (Janausch *et al.*, 2002). In other cases, chemical stimuli interact with a periplasmic binding protein, which release or associate with the sensor of HKs to activate the systems such as VirA and PhoQ (Dathe and Wieprecht, 1999; Shimoda *et al.*, 1993). When binding the stimulus, the conformational changes of sensor domain may be propagated into the TM helices, transducing the signal across the membrane.

The sensor domains of periplasmic-sensing HKs have low sequence similarity among different systems, as expected due to the diversity of signals they respond to, but functionally, they can be grouped as PAS, GAF, PBPb, CAHCHE and Reg_prop domains according to sequence alignment (Mascher *et al.*, 2006).

PAS domains were initially identified in the *PER*, *ARNT*, and *STM* proteins, which are very common sensory folds used in a lot of proteins, such as CitA, DcuS, and PhoQ (Cheung and Hendrickson, 2008; Sevvana *et al.*, 2008; Tu *et al.*, 2006). PASs sense light, redox potential, energy, oxygen, voltage, and small ligands and involve in protein-protein interactions (Ayers and Moffat, 2008; Taylor and Zhulin, 1999; Zhulin *et al.*, 1997). PAS domains are known to bind metal ions, heme, flavin, C4 dicarboxylates, and adenine, which are often duplicated. Generally, PAS domain has a β -scaffold and several α -helices and loops (Figure 1. 14). PAS may appear in the linker region between TM2 and HK as well.

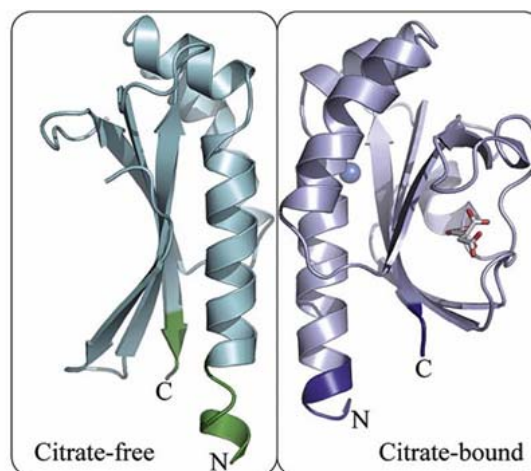


Figure 1. 14 PAS structures

The periplasmic domain PAS sensor histidine kinase CitA with the absence (left) and presence (right) of citrate shown as a monomer. This PAS sensor has a β -scaffold and several α -helices and loops around it. While ligand citrate bound, the scaffold and loops fold up to hold the citrate. Cited from (Sevvana *et al.*, 2008).

GAF is a large family of small molecule binding domain, predicted to bind cyclic nucleotides, such as cGMP/cAMP. GAF has PAS-like fold (Aravind and Ponting, 1997; Ho *et al.*, 2000). GAF could be looked as a subgroup of PAS.

PBPb is periplasmic binding protein (about 250 residues), which is periplasmic binding protein homolog, such as BvgS and Evgs in *Bordetella spp.* and *E. coli* respectively (Bantscheff *et al.*, 2000). PBPb are specific for polar amino acids and opines, which are normally found up to three repeats side by side in a single sensor domain of HKs (Tam and Saier, 1993).

CACHE domains (Ca^{2+} channels, chemotaxis receptors; 150-300 residues) have been shown to be involved in small-molecule binding, such as amino acids in methylation, carbohydrates in phosphotransferase system, and dicarboxylates in nitrogen fixation (Garrity *et al.*, 1998; Hanlon and Ordal, 1994; Reid and Poole, 1998).

Reg_prop is a group of novel sensors found in *B. thetaiotaomicron*, which has more than 1000 residues to form 14 tandem Reg_prop repeats which are predicted to be two seven-bladed β -propellers (Pons *et al.*, 2003). The function of these domains is unknown to date.

So far all structurally resolved sensor domains belong to the PAS superfamily (Miyatake *et al.*, 2000; Pappalardo *et al.*, 2003; Rajagopal and Moffat, 2003; Reinelt *et al.*, 2003). The functional specificity of a PAS is associated with the cofactor, such as 4'-hydroxycinnamic acid in the bacterial photoactive yellow protein, a heme in the oxygen sensor, flavin mononucleotide, and flavin adenine dinucleotide in redox potential sensors, even though some PASs can function independently (Borgstahl *et al.*, 1995; Christie *et al.*, 1999; Gong *et al.*, 1998; Soderback *et al.*, 1998; Watts *et al.*, 2006). PAS could sense light, oxygen, and redox potential, and regulate many effectors: kinases, phosphodiesterases, and transcription factors (Crosson *et al.*, 2003).

1.2.3.2.2 Role of transmembrane region in signal transduction

Little is known about the signal transduction mechanism via transmembrane helices mainly because of the constraints of a lipid bilayer (Matthews *et al.*, 2006). α -helices are hard to compress and extend, so they could make a small change in the conformation of a long distance. Generally, there are five postulated models for transmembrane signalling: association, piston, rotation, scissor model, and see saw model (Ottemann *et al.*, 1999) (Figure 1. 15A). A recent research revealed that the 2 transmembrane helices of a histidine kinase YycG from *Bacillus subtilis* form into piston model to complete the signalling process through a subtle alteration (Figure 1. 15B) (Szurmant *et al.*, 2008). The piston model is predicted to happen in histidine kinase CitA as well (Figure 1. 14B).

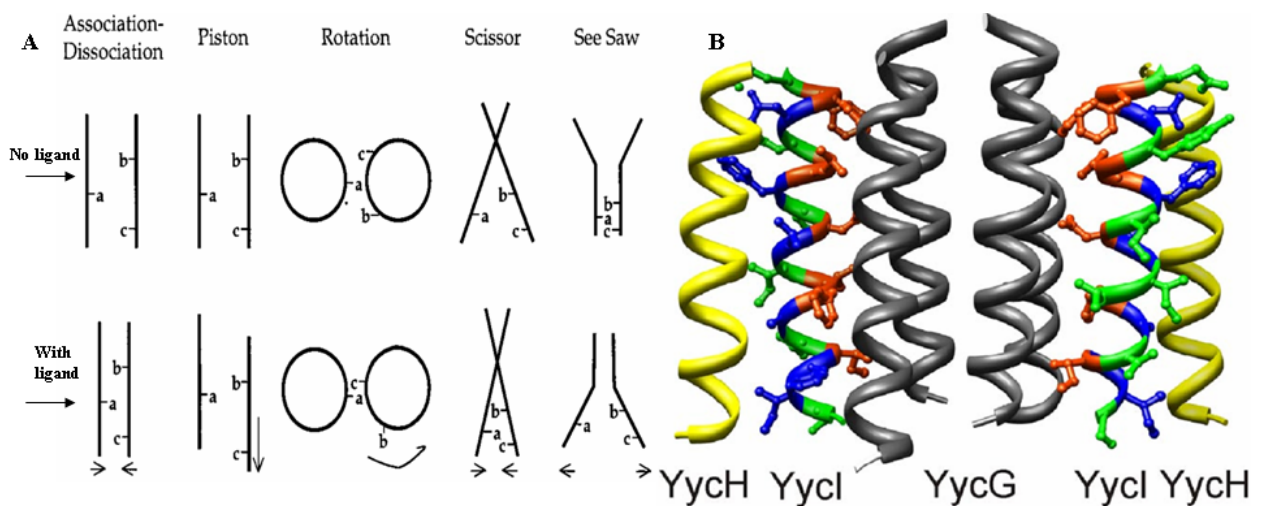


Figure 1. 15 TM structure.

A: The five postulated models for transmembrane signalling. The black lines are present α -helices in transmembrane domain from the side angle, while the cycles are present single helices from the angle of the top. The letters a, b, c, represent amino acids. Cited from (Ottemann *et al.*, 1999). **B:** The TM structure complex YycG/H/I of *Bacillus subtilis*. The 4 transmembrane helices of two homodimer histidine kinase YycG (gray) were associated with two other transmembrane helices of two homodimer histidine kinase YycH (yellow) and YycI (red-green-blue) to transducer the signal. The residues in YycI contacting with YycG are in orange sticks, while these contacting with YycH are in blue, and these free residues are in green. The signal stimuli from the extracellular domain to the TM complex YycG/H/I will be transmitted through by the subtle alteration of these proteins (Szurmant *et al.*, 2008).

1.2.3.2.3 Cytoplasmic domains of sensor HKs

Sensor HKs belong to an ancient enzyme family which uses ATP as the phosphate donor (Saier, 1993). The closest relatives are serine/threonine/tyrosine protein kinases, such as MAP kinase (Bourret *et al.*, 1991). In a small group of HKs, the phosphorylated residue is Tyr or Ser instead of His (Elich and Chory, 1997; Wu *et al.*, 1999; Yeh *et al.*, 1997). The cytoplasmic transmitter domain of the HK is highly conserved and has two sub domains: the His-containing phosphoacceptor domain, which usually forms a homodimer, and catalytic (kinase) domain which carries out trans-auto phosphorylation of the opposing monomer (Figure 1. 16).

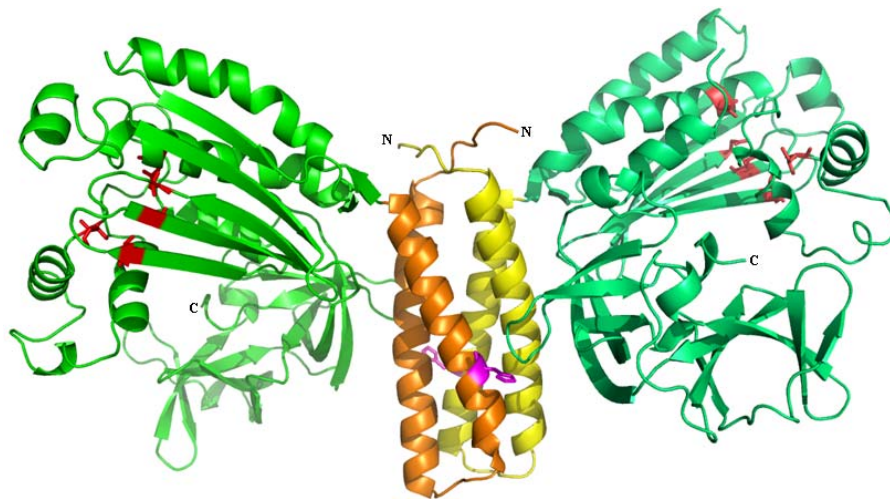


Figure 1. 16 The structure of the cytoplasmic domains of a sensor HK.

The cytoplasmic region of CheA sensor HK from *Thermotoga maritima*. PDB ID is 1B3Q (Bilwes *et al.*, 1999). The monomer A rotates 90 degrees around monomer B to form a homodimer by a central four-helix bundle (yellow from monomer A and orange from B), which has a conserved histidine (purple sticks) separately. This is the His-containing phosphoacceptor domain. The ATP-dependent catalytic (kinase) domains are present in limegreen for monomer A and green for monomer B with ATP binding sites colored hotpink for A, and red for B.

Some HK are bifunctional and play a role as a phosphatase as well. This kind of HK controls the phosphorylation level of its cognate RR which means it controls the signalling flow. The linker region between the TM2 and transmitter domain varies a lot in size, extent and function. HAMP or PAS domains are sometimes present as linkers. HAMP is helical domain of around 50 residues, originally found in *histidine kinases*, *adenylyl cyclases*, *methyl-accepting chemotaxis proteins*, and *phosphatases* (Aravind and Ponting, 1999; Fulop and Jones, 1999) (Figure 1. 17). The HAMP domain may be involved in signal transduction, possibly via rotation of its four α -helices (Hulko *et al.*, 2006; Inouye, 2006).

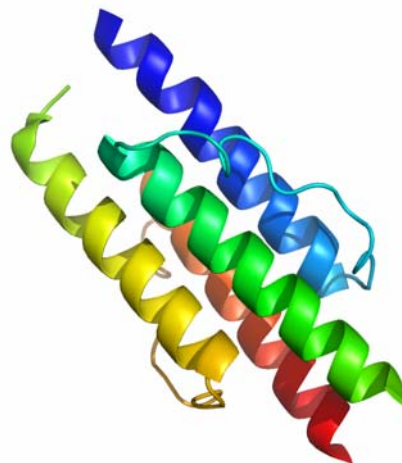


Figure 1. 17 HAMP structure.

HAMP is usually about 50 residues, formed into four helices and acts a linker domain between the membrane and the cytoplasmic domains of the HK. PDB ID is 2ASW (Inouye, 2006).

1.2.3.2.4 Response Regulator (RR) domains

Response regulators (RR) are usually free cytoplasmic proteins with two distinct domains; the N-terminal receiver domain and the C-terminal output domain. The receiver domain accepts the phosphate from the HK on a conserved Asp. This event activates the associated output domain, usually by a conformational change, to elicit the appropriate cellular response, either at the transcriptional or post-translational level. RRs are analogous to Ras family members of eukaryotic small GTPases (Artymiuk *et al.*, 1990). The structure of the N-terminal receiver domain is conserved due to its conserved function (Figure 1. 18) (Robinson *et al.*, 2000). Phosphorylation changes the electrostatic effect of the conserved Asp less than other residues, and phospho-Asp is the most stable phosphorylated residue, which has a half life of hours. This high energy bond between Asp and phospho group changes the conformation to the active form of the RR. The structure of C-terminal regulatory output domain varies according to the function (Galperin, 2006). Regulatory domains can bind DNA, RNA, protein, and even have enzyme activity. If the regulatory domain is a DNA binder, it normally has a helix-turn-helix (HTH) structure (Figure 1. 18b). In some cases, RRs have no attached regulatory domain, their mode of action may be different and involve more complex phosphorelay events (Stock *et al.*, 1989).

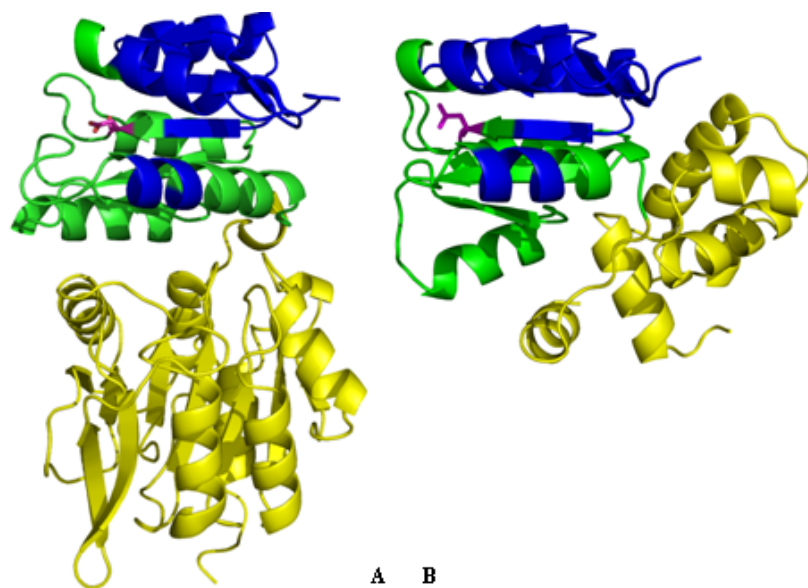


Figure 1. 18 TCS RR multidomains.

The conformational changed area of receiver domain are colored blue, the unaltered regions are green. An Asp residue showing the phosphorylation site is colored purple. The Output domain is yellow. **A:** The structure of methylesterase CheB (PDB ID 1A2O) (Djordjevic *et al.*, 1998). **B:** The structure of nitrate response NarL (PDB ID 1A04) (Baikalov *et al.*, 1998). The output domain is to bind to DNA, which has a helix-turn-helix structure.

TCSs sometimes act in concert with sigma factors to co-regulate gene expression as a positive and negative switch (Yan *et al.*, 2008). The processes mediated by TCS make them a novel intervention in antimicrobial therapy (Filippou *et al.*, 2008). TCSs can be used as drug targets because they are

not found in animals and often regulate important processes including virulence (Beier and Gross, 2006).

1.2.3.3 Hybrid two component systems

As mentioned earlier *B. thetaiotaomicron* has large number of novel hybrid two component systems (HTCS) that comprise all of the domains of a classical two component system in a single peptide (Figure 1. 19) (Sonnenburg *et al.*, 2006). In all known HTCS the output domain is an AraC-type helix–turn–helix DNA-binding domain (HTH_AraC) (Figure 1. 19). HTH_AraC domains are usually involved in the activation of transcription of the genes under their control. The independent C-terminal domain of an HTCS could bind DNA, which suggests the HTCS may release the HTH_AraC to regulate the gene expression (Miyazaki *et al.*, 2003). But the mechanism of how to terminate this regulation is unknown.

B. thetaiotaomicron has 32 HTCS, which is more than any other sequenced organism. In *B. thetaiotaomicron*, the genes for many of these HTCS are linked to PULs containing SusC/D homologues as well as glycoside hydrolases and lyases. Therefore it is likely that the HTCS are involved in regulation of polysaccharide utilisation in this organism.

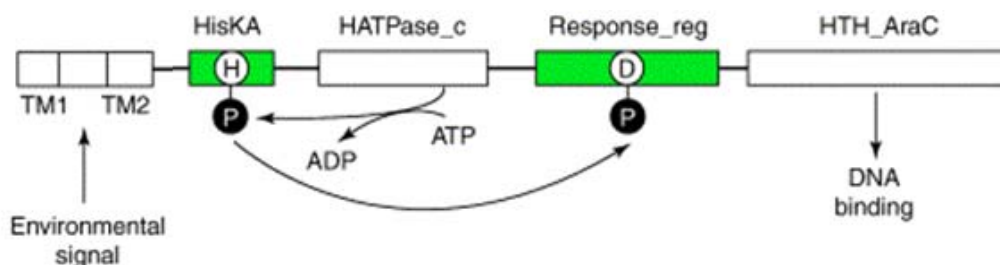


Figure 1. 19 Hybrid two component system.

Cartoon representation shows the sub-domains of HTCS. The N-terminal periplasmic sensor domain displays the most sequence variation and is anchored in the inner membrane via two transmembrane regions (TM) that flank its N and C-terminus (Sonnenburg *et al.*, 2006). The other four domains, HisKA, HATPase_C, Response_reg, and HTH_AraC are conserved, and located in the cytoplasm. Cited from (Xu *et al.*, 2004).

Genome sequencing has revealed that HTCS occurs in a range of different species including *Bacteroidetes*, *Proteobacteria*, and *Chloroflexi* (Miyazaki *et al.*, 2003; Sonnenburg *et al.*, 2006). *Bacteroidetes* as the major organisms of HTCS are located in gut habitats and niches, which usually contain a lot of sensor regulators so they seem to be an adaptation to minimise cross interaction.

1.2.4 Polysaccharide utilisation in *Bacteroides thetaiotaomicron*

The only previously studied PUL is the starch utilisation system (*sus*) of *B. thetaiotaomicron*. The *sus* is composed of seven genes *susABCDEFG* controlled by a membrane associated transcriptional regulator known as SusR (Figure 1. 20) (D'Elia and Salyers, 1996; Hooper *et al.*, 1999; Reeves *et al.*, 1996; Reeves *et al.*, 1997). SusR is a maltose dependent regulator which has an N-terminal periplasmic input domain and a C-terminal helix-turn-helix DNA binding output domain, similar to many TCS-RR output domains. The *susR* gene is not regulated itself, but its product activates the expression of downstream *sus* genes in the presence of maltose or any higher oligomers. The *susA* gene belongs to a single transcriptional unit, while other members belong to a downstream operon, which suggested that the regulation of *susA* is different to other proteins. Indeed, the expression of *susA* is less than *susB/C*, and sensitive to the gene amount of *susR* (D'Elia and Salyers, 1996). SusA, SusB, and SusG display neopullulanase, α -glycosidase and α -amylase activities, respectively (Reeves *et al.*, 1996). SusG is an extracellular lipoprotein that associates with the outer membrane, whereas SusA and SusB are periplasmic enzymes (Shipman *et al.*, 1999). SusDEFG are thought to form an OMP complex to bind and degrade long-chain starch molecules, which are then transported via SusC. The roles of SusE and SusF are unclear, but they may also be involved in polysaccharide binding in the complex as they are both found associated with the outer membrane (Shipman *et al.*, 2000). Interestingly, although SusC and SusD homologues are very common in the *B. thetaiotaomicron* genome, there are no SusE or SusF homologues, suggesting the role of these proteins is specific to starch utilisation.

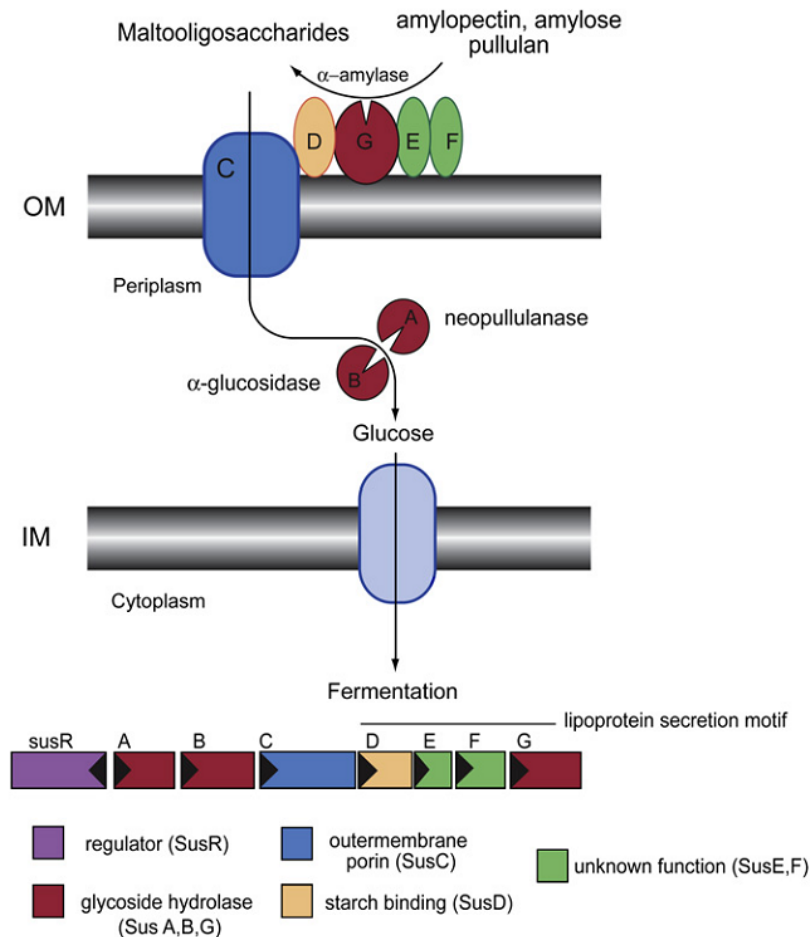


Figure 1. 20 The relative locations of the *sus* genes of the *B. thetaiotaomicron*.

SusR is the regulator of *sus* genes of *B. thetaiotaomicron*. SusA, SusB, and SusG are glycoside hydrolases that digest starch. SusC/D form an outer membrane complex that binds and actively transports starch binding into the cell. The functions of SusE/F are unclear, but appear to be part of the SusC/D complex. Cited from (Koropatkin *et al.*, 2008).

1.2.4.1 SusC

The outer membrane provides a semi permeable layer that protects Gram negative bacteria from many environmental insults. Porins are β -barrel proteins that span the outer membrane and enable molecules to pass back and forth across this barrier (Figure 1. 21). While small molecules ($<\sim 600$ Da) can diffuse across passively, large molecules have to be actively transported, usually by utilising the energy provided by the inner membrane TonB complex (Braun *et al.*, 2003a; Koebnik *et al.*, 2000; Walzer *et al.*, 2008). SusC is an energy dependent β -barrel porin that is thought to actively transport the polysaccharide across the outer membrane (Koropatkin *et al.*, 2008; Reeves *et al.*, 1996).

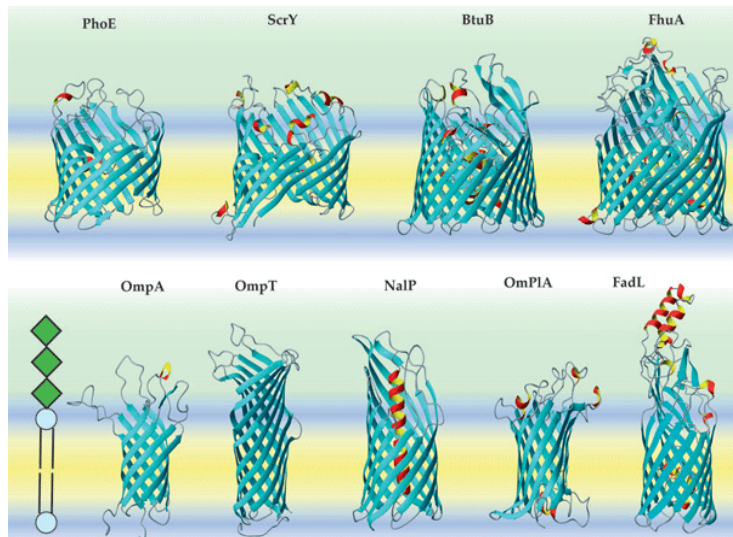


Figure 1. 21 Some known structures of β -barrel membrane proteins.
Cited from (Kleinschmidt, 2005).

Based on the protein database, an alignment of SusC to other OMPs revealed that SusC has high sequence similarity to a group of OMPs at a region of 60 amino acids, which is believed to be a domain to stabilize OMPs in the outer membrane (Figure 1. 22) (Reeves *et al.*, 1996). These OMPs function into energy-dependent iron or vitamin uptake, such as the ferrichrome receptor FhuA of *E. coli*.

<u>Protein</u>	<u>Organism</u>	<u>Sequence</u>
SusC	<i>B. theta.</i>	194 <u>LIVIDGLAID</u> N.....E GIKGMANGLS MVPPAD. <u>IET</u> LTVLKDASAT
CsuF	<i>B. theta.</i>	186 <u>LFVIDGFPIE</u> DS..... S.....AAS TLNPSD. <u>IES</u> LDFLKDASAT
IrgA	<i>V. chol.</i>	101 <u>LILVDGK...</u> .RQTSRQTRP NSDGPGEIQG WLPPLOA <u>IER</u> AEVIRGP. <u>MS</u>
BtuB	<i>E. coli</i>	97 <u>LVLIDG....</u>VRLNLA GVSGSADLSQ F.. <u>PIALVOR</u> VEYIRGP. <u>RS</u>
FepA	<i>E. coli</i>	105 <u>LILIDGKPV</u> S SRNSVRQGW GERDTRGDTS WVPP. <u>EMIER</u> IEVLRGP. <u>AR</u>
CirA	<i>E. coli</i>	100 <u>LILVDGKRVN</u> SRNAV..... .FRHNDFDLN WI. <u>PVDSIER</u> IEVVRGP. <u>MS</u>
IutA	<i>E. coli</i>	100 <u>VVLVDGVRLN</u> S..... .SRTDSRQLD SIDPFNM.HH IEVIFGA. <u>TS</u>
FhuA	<i>E. coli</i>	123 <u>LIIRGFAAEG</u> QSQNNYLNGL KIQGNFYNDA VIDPY. <u>MLER</u> AEIMRGP. <u>VS</u>
PfeA	<i>P. aerug.</i>	110 <u>LILVDGKPV</u> S SRNSVRYGWR GERDSRGDTN WV. <u>PADQ.ER</u> IEVIRGP. <u>AA</u>
		<u>LILIDG----</u> ----- -P--- <u>IER</u> -EV- <u>RGP--S</u>
SusC	<i>B. theta.</i>	<u>AIYGSRASNG</u> <u>VIIITTKK</u> GK 253
CsuF	<i>B. theta.</i>	<u>AIYGSRASNG</u> <u>VIIITTKK</u> GK 240
IrgA	<i>V. chol.</i>	<u>TYGSDAIGG</u> <u>VINIITRK</u> DQ 166
BtuB	<i>E. coli</i>	<u>AVYGSDAIGG</u> <u>VVNITTR</u> DEP 156
FepA	<i>E. coli</i>	<u>ARYGNGAAGG</u> <u>VVNIITKK</u> GS 173
CirA	<i>E. coli</i>	<u>SOYGSRCARR</u> CSEYHHQK <u>NR</u> 163
IutA	<i>E. coli</i>	<u>.LYGGGSTGG</u> <u>LINIVKKG</u> QP 157
FhuA	<i>E. coli</i>	<u>VLYGKSSPQG</u> <u>LLNMVSKR</u> PT 191
PfeA	<i>P. aerug.</i>	<u>ARYGNGAAGG</u> <u>VVNIITKQ</u> AG 177
		<u>A-YG--A-GG</u> <u>V-NI-TKK--</u>

Figure 1. 22 Conserved domain alignment of SusC to other OMPs.

The alignment shows the conserved region of 60 amino acid of a group of OMPs including SusC. The similarity to SusC is: 1) 82 % to CsuF (OMP of *B. thetaiotaomicron* for chondroitin sulfate utilisation) (Cheng *et al.*, 1995); 2) 60 % to IrgA (iron-regulated OMP of *Vibrio cholerae* (Goldberg *et al.*, 1992); 3) 66 % to BtuB (the vitamin B₁₂ receptor of *E. coli*) (Heller and Kadner, 1985), 4) 62 % to FepA (enterochelin receptor of *E. coli*) (Rutz *et al.*, 1992), 5) 59 % to CirA (the colicin I receptor of *E. coli*) (Nau and Konisky, 1989), 6) 62 % to IutA (the ferric-aerobactin receptor of *E. coli*) (Thomas and Valvano, 1992), 7) 61 % to FhuA (the ferrichrome receptor of *E. coli*) (Coulton *et al.*, 1988), 8) 63 % to PfeA (the ferric enterobactin receptor of *Pseudomonas aeruginosa*) (Dean and Poole, 1993). Cited from (Reeves *et al.*, 1996).

1.2.4.2 SusD

SusD is an outer membrane lipoprotein that is anchored in the membrane via a lipid linker covalently attached to an N-terminal cysteine. Recent data have shown that SusD is a novel form of carbohydrate-binding protein that binds directly to starch and maltooligosaccharides (Koropatkin *et al.*, 2008). The structure of SusD reveals that it forms a novel, mainly α -helical fold (Figure 1.25). Eight of these helices pack together a four tetratricopeptide repeat (TPR: a versatile all-helical structural motif) units even though SusD was not predicted to have any TPR motifs as it does not contain the amino acid signature of W₄-L₇-G₈-Y₁₁-A₂₀-F₂₄-A₂₇-P₃₂ (D'Andrea and Regan, 2003). TPRs are most commonly involved in protein-protein interactions, a fact that has significance with SusD as it is suspected to form a complex with SusC. The closest homolog of SusD is PilF, an inner membrane protein from *Pseudomonas aeruginosa* involved in pilus formation that also consists of a number of TPR motifs (Kim *et al.*, 2006). The carbohydrate binding site is composed of a number of exposed aromatic residues that form a curved pocket on the surface of the protein that matches the helical fold of starch. Thus the protein seems to specifically recognise the three-dimensional structure of the polysaccharide rather than its component sugar molecules. It is thought that SusD initially binds the polysaccharide and somehow targets it to the SusC porin, although the mechanistic details of this interaction are unknown. The gene knockout analysis of SusD revealed the essential role in growth on maltooligosaccharides >5 glucose units, such as maltohexaose (Sonnenburg *et al.*, 2006).

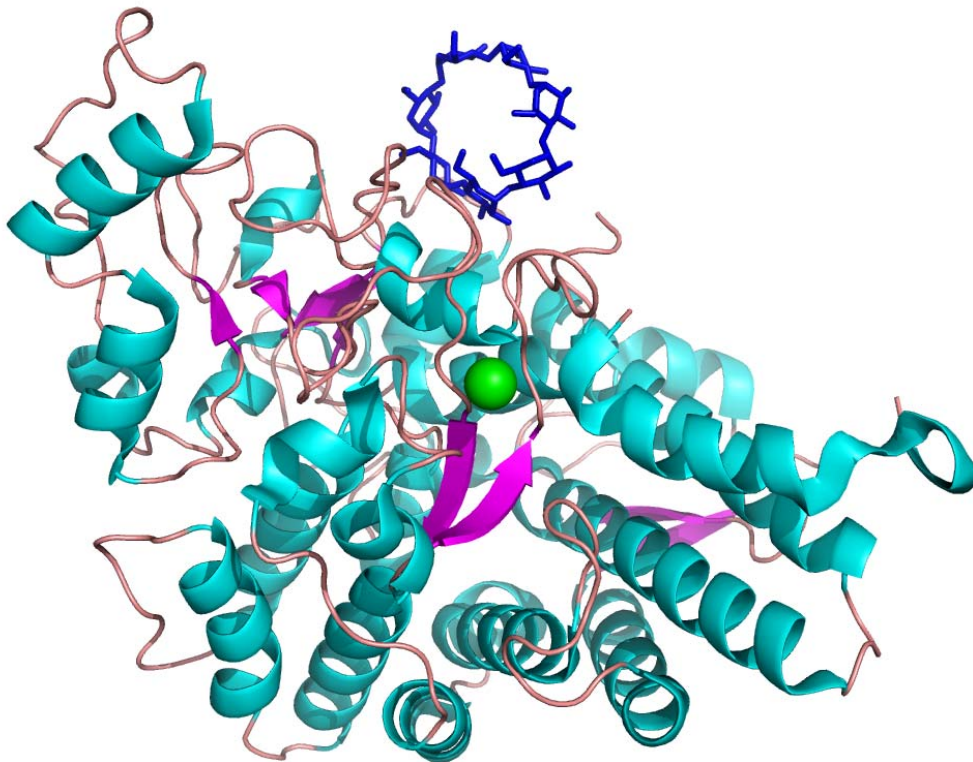


Figure 1. 23 SusD structure.

SusD has 22 α -helices (cyan) and 3 sets of two-antiparallel β sheets (purple). Calcium is shown as green ball, while the ligand maltoheptaose is colored blue. PDB ID is 3CK9 (Koropatkin *et al.*, 2008).

B. thetaiotaomicron possesses 101 pairs of SusC/SusD (Xu *et al.*, 2003). Four other human gut-associated sequenced Bacteroides species, *B. distasonis* ATCC 8503, *B. vulgatus* ATCC 8482 *B. fragilis* NCTC9343, and *B. fragilis* YCH46 have an additional 269 pairs (Xu *et al.*, 2007). These pairs also are present in soil and marine *Bacteroidetes*, such as *Cytophaga hutchinsonii*, *Croceibacter atlanticus*, *Gramella forsetii*, and *Leeuwnhoekiella blandensis* (Bauer *et al.*, 2006; Cho and Giovannoni, 2003; Pinhassi *et al.*, 2006). These pairs are linked to glycoside hydrolases and polysaccharide lyase to form PULs that target specific polysaccharides.

1.2.5 Glycoside hydrolases

Glycoside hydrolases (GHs) are enzymes which hydrolyse the glycosidic bond. The International Union of Biochemistry and Molecular Biology (IUBMB) Enzyme nomenclature of GHs (EC 3.2.1.-) is mainly based on the substrate classification which could not reflect the structural specificity and mechanism of these enzymes (Bairoch, 2000). A different classification of GHs based on amino acid sequence similarities has been established, which describes more details of the structural features of GHs, evolutionary relationships and mechanisms (Henrissat and Bairoch, 1993). The Carbohydrate-Active Enzymes database (CAZy: www.cazy.org) provides the list of the GH families. There are presently 114 families in CAZy database. Because the folds of enzymes are conserved, some families present similar fold configuration, which are grouped in larger superfamilies known as clans. For example families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, and 113 belong to clan GH-A, because they all adopt the $(\beta/\alpha)_8$ fold. There are currently a total of 14 clans (A-N) (Figure 1. 24).

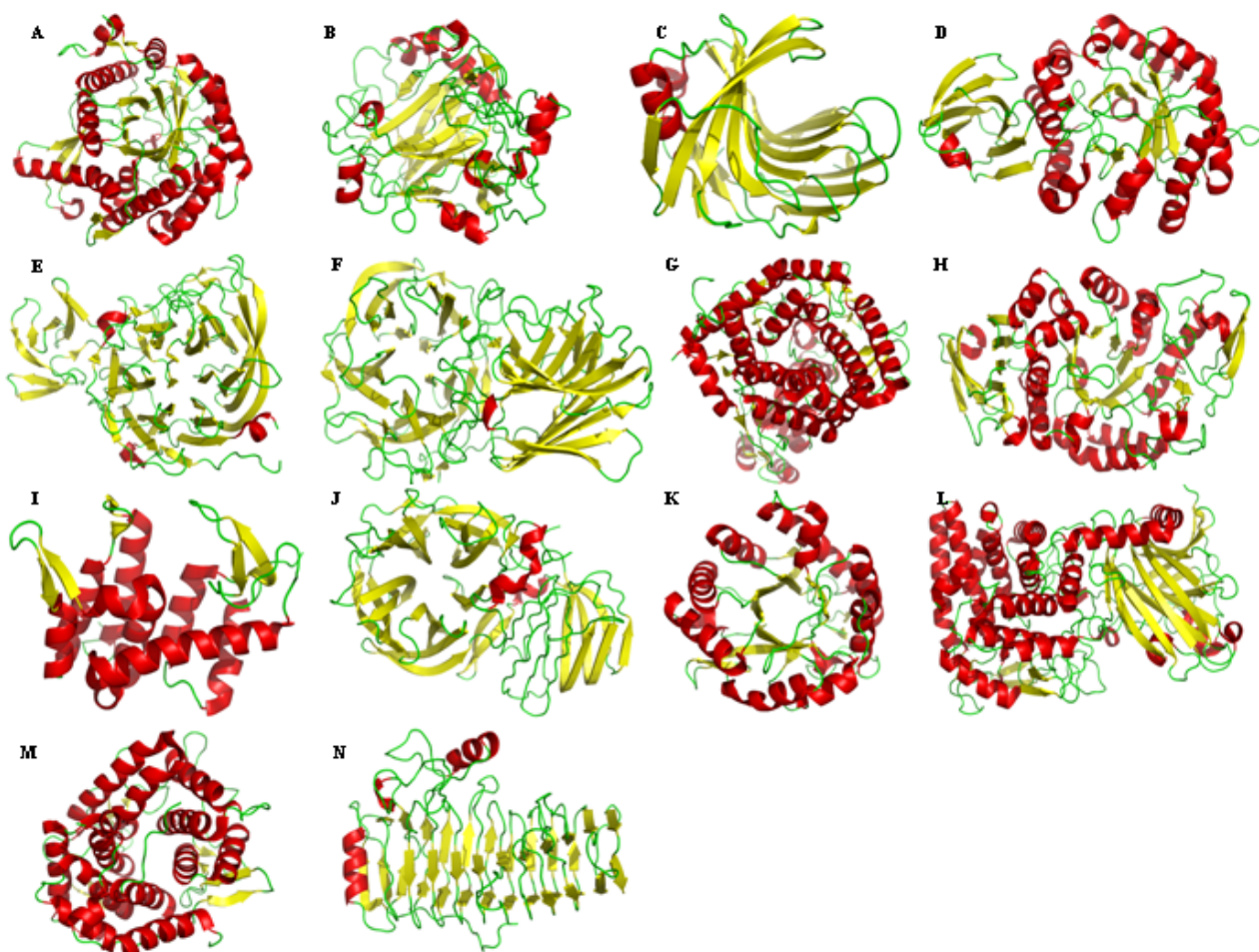


Figure 1. 24 Ribbon representation shows example structures of GH clans.

Cartoon representations show the diversity of folds of glycoside hydrolases clans A-N. β strands are shown in yellow, α helices in red and loops are in green.

Usually GHs hydrolyse the glycosidic bond using two catalytic residues: a general acid (proton donor) and a nucleophile/base, commonly an Asp or a Glu. Hydrolysis occurs with either overall retention or inversion of the configuration at the anomeric carbon (Davies and Henrissat, 1995) (Figure 1. 25). In both mechanisms, the proton donor is within hydrogen-bonding distance of the glycosidic oxygen, which is protonated by this acid catalyst. In retaining GHs, the nucleophile is close to the sugar anomeric carbon. The distance between proton donor and nucleophile/base is $\sim 5.5 \text{ \AA}$. The nucleophile/base binds to the anomeric carbon to form a glycosyl-enzyme intermediate. The sugar is released from the enzyme by a second nucleophilic substitution of a water molecule and the anomeric carbon presents the same stereochemistry as before cleavage. In inverting GHs, the distance of the two catalytic residues are $\sim 10 \text{ \AA}$, which must harbor a water molecule between the base and the anomeric carbon. This water molecule mediates the protonation and nucleophilic substitution to achieve an opposite stereochemistry of the anomeric carbon.

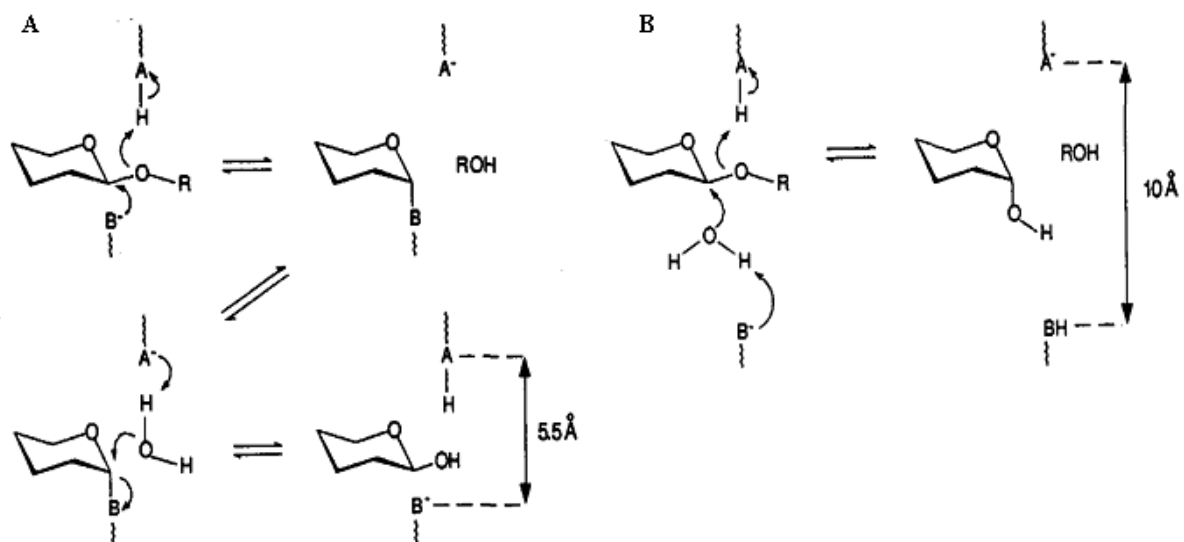


Figure 1.25 Mechanism of retaining and inverting.

A: The retaining mechanism. The glycosidic oxygen is protonated by the acid catalyst AH, while the anomeric carbon is nucleophilic substituted by a nucleophile/base B⁻. This anomeric carbon is then nucleophilic substituted by a water molecule. Finally, this anomeric carbon keeps the same stereochemistry as before cleavage. **B: The inverting mechanism.** The protonation of the glycosidic oxygen and nucleophilic substitution of the anomeric carbon are achieved by the attack of a water molecule and change the stereochemistry of the anomeric carbon. Cited from (Davies and Henrissat, 1995).

Glycoside hydrolases act via either an endo or exo mode of action. Generally, endo-acting GHs contain an open groove which binds randomly to the saccharide chains and cuts to release a mixture of different size oligosaccharides, while the exo-acting GHs commonly bind the terminal sugar residues of oligo and polysaccharides via a pocket-like active site and only release a single product from the chain end - usually a monosaccharide (Hovel *et al.*, 2003; Pell *et al.*, 2004) (Figure 1.26). A variation on the pocket topology for exo activity is provided by cellulases from families 6 and 7 bind polymeric substrates threaded through a tunnel, which provides the conditions for processivity (Figure 1.26C) (Divne *et al.*, 1998). These exo-processive enzymes release cellobiose from the reducing ends of cellulose chains while remaining firmly bound the substrate (Davies *et al.*, 1997; Rouvinen *et al.*, 1990). Processivity may be the key mechanism for the biological digestion of insoluble microcrystalline cellulose (Davies and Henrissat, 1995).

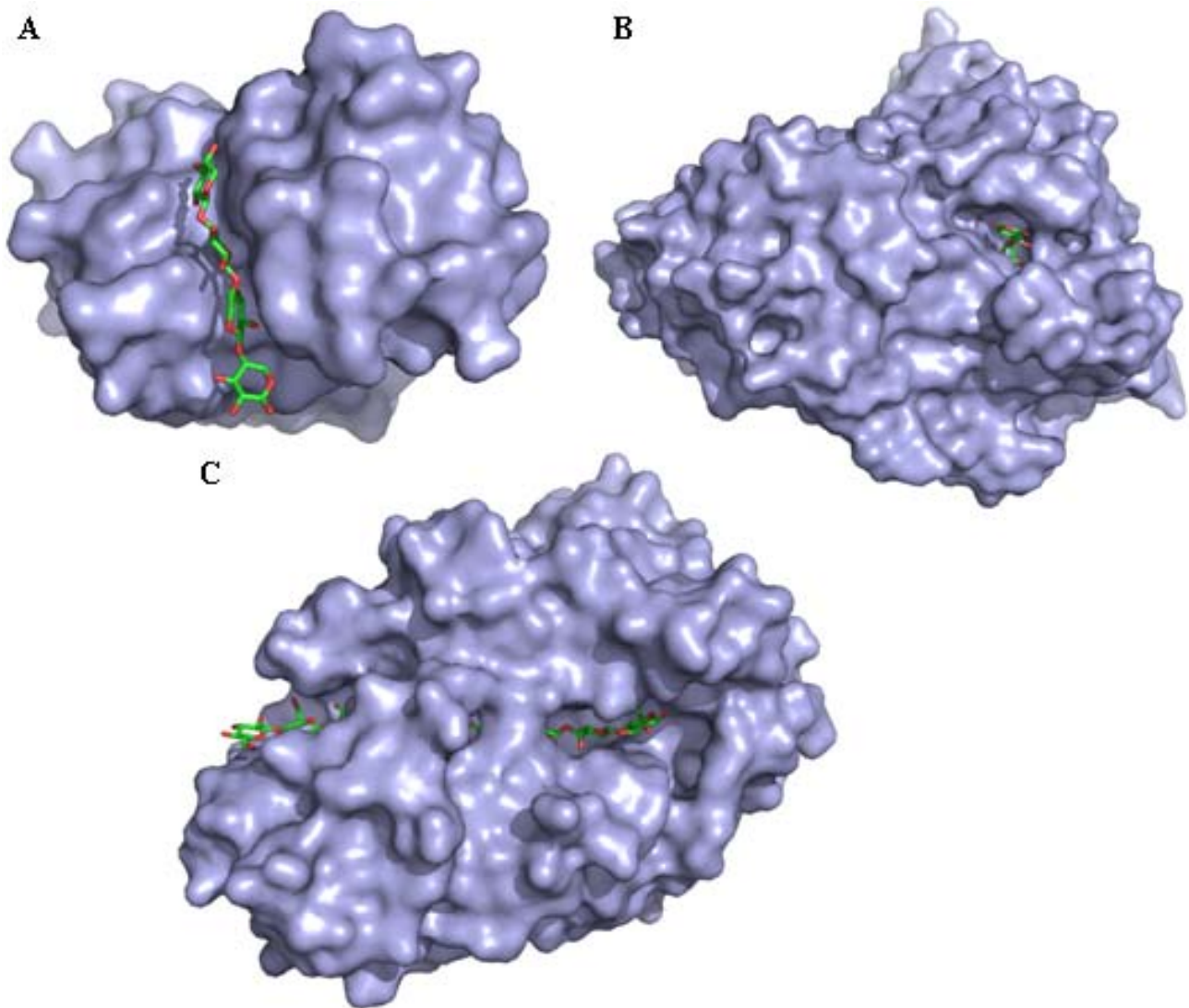


Figure 1. 26 Structural features of endo and exo-acting GHs.

A: Endo-acting GH with open cleft active site enabling the enzyme to bind randomly to the polysaccharide chain. Enzyme shown is Xylanase10C mutant E385A from *Cellvibrio japonicus* in complex with xylopentaose. PDB ID is 1US2 (Pell *et al.*, 2004). **B:** Exo-acting GH which binds sugar in a pocket like active site and releases monosaccharide from the chain end of polysaccharides and other glycans. The enzyme shown is α -L-arabinofuranosidase (GH family 51) from *Bacillus stearothermophilus* in complex with Ara- α 1, 3-Xyl. PDB ID is 1QW8 (Hovel *et al.*, 2003). **C:** Exo-processive enzyme with tunnel-like active site. The enzyme shown is cellobiohydrolase I from *Trichoderma reesei* in complex with a cellulose chain. PDB ID is 5CEL (Divne *et al.*, 1998).

B. thetaiotaomicron contains 226 glycoside hydrolases which belong to 42 families in the Cazy database and encode a wide range of different activities (Figure 1. 27). This number of GH genes exceeds that found in any other sequenced bacterium, and any other human gut habitant. This capacity enables *B. thetaiotaomicron* to digest a range of plant and host derived glycans and likely contributes to its ability to dominate in the densely populated niche of the large intestine.

Glycoside Hydrolase Family	2	3	5	13	16	18	20	23	25	27	28	29	30	31	32	33	35	36	38	42	43	51	53	57	66	67
Number of sequences in family	32	10	2	7	4	12	14	3	1	5	9	9	1	6	4	2	3	3	2	1	32	4	1	1	1	1
	73	76	77	78	84	88	89	92	93	95	97	99	105	106	109	110										
	1	10	1	6	1	4	3	23	1	5	10	1	7	3	2	2										

Figure 1. 27 *B. thetaiotaomicron* glycoside hydrolase family distribution.

Cited from www.cazy.org/geno/226186.html.

1.2.6 Other systems employed by bacteria for carbohydrate utilisation

There are other ways for bacteria to obtain and degrade polysaccharides mainly from plant cell wall. The aerobic soil saprophyte *Cellvibrio japonicus* secretes complex modular enzymes into the environment (DeBoy *et al.*, 2008). Anaerobic bacteria such as *Clostridium thermocellum* could produce multi modular enzyme complex (cellulosomes) to secrete or attach to the outside of the cell (Bayer *et al.*, 2004). Both these two mechanisms use enzymes containing multiple carbohydrate binding modules (CBM) belonging to the same or different families. 40 of 52 CBM families identified to date contain bacterial proteins, which may be because they are required to access the recalcitrant celluloses and hemicelluloses of the plant cell wall. *B. thetaiotaomicron* has just 25 CBMs suggesting that it does not access the same polysaccharides.

1.3 Objectives of this study

The main objectives of this project are to structurally and functionally characterise components of the polysaccharide utilisation system of *B. thetaiotaomicron*. Specifically these are:

1. Study the level of interaction amongst several transmembrane signalling systems from *B. thetaiotaomicron*;
2. Understand the mechanism of signal perception in BT1754 HTCS and identify the signalling molecule that activates this system;
3. Characterise the enzymic activities of the four GH32 members of *B. thetaiotaomicron* activated by fructans: BT1759, BT1760, BT1765, and BT3082;
4. Characterise recognition of polysaccharides by the SusD homolog BT1762.

Chapter 2 Materials and Methods

2.1 Bacterial strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 2.1 and Table 2.2.

Strain	Description	Use	Reference
BL21(DE3)	F' ompT, hsdS _B (r _B -m _B), gal.dcm. (DE3)	Protein expression	(Studier and Moffatt, 1986)
XL1-Blue	recA1, endA1, gyrA96, thi1, hsdR17, supE44, relA1, lac, [F' proABlac ^q ZΔ15, Tn10, (tet ^r)]	DNA manipulation	Stratagene
Tuner(DE3)	F' ompT, hsdS _B (r _B -m _B), gal, dcm, lacY1 (DE3)	Protein expression	Novagen
B834	F ⁻ ompT hsdS _B (r _B ⁻ m _B ⁻), gal dcm met (DE3)	Protein expression	Novagen
Origami™ B	F- ompT hsdSB(rB- mB-) gal dcm lacY1 ahpC gor522:: Tn10 trxB (Kan ^R , Tet ^R)	Protein expression	Novagen
Origami 2 (DE3)	(<i>ara-leu</i>)7697 □ <i>lacX74</i> □ <i>phoA PvuII phoR araD139 ahpC galE galK rpsL F'</i> [<i>lac</i> ⁺ <i>lacI</i> ^f <i>pro</i>] (DE3) <i>gor522::Tn10 trxB</i> (Str ^R , Tet ^R)	Protein expression	Novagen
C41	Protease deficient strain of BL21 (DE3) lysogen with a tighter control on protein expression than BL21 (DE3), permits the expression of more toxic proteins under control of a T7 promoter with IPTG.	Protein expression and DNA manipulation	(Miroux and Walker, 1996)
JM83(DE3)	recA ⁻ , (DE3)ara, Δ(lac-proAB), rspL, φ80, lacΔM15	Protein expression	
TOP-10	F- <i>mcrA</i> Δ(<i>mrr-hsdRMS-mcrBC</i>) ±80 <i>lacZ</i> ΔM15 Δ <i>lacX74 recA1 endA1 araD139</i> Δ(<i>ara, leu</i>)7697 <i>galU galK</i> Δ <i>λ</i> <i>rpsL nupG tonA hsdR</i>	DNA manipulation	Invitrogen

Table 2. 1 *E. coli* strains in this study.

Plasmid	Size (kb)	Genotype	Supplier
pET21a	5.5	Ampr T7 lac lacIq	Novagen
pET22b	5.5	Ampr T7 lac lacIq	Novagen
pET28b	5.37	Kanr, T7 lac lacIq	Novagen
pET32b	5.9	Ampr T7 lac lacIq	Novagen
pET39b	6.1	Kanr T7 lac lacIq	Novagen
pCR® Blunt	3.5	Kanr, Znr, ccdB	Invitrogen
minipRSET-A	2.9	Ampr N-His	Invitrogen
minipRSET-Trx	3.2	Ampr N-His-Trx	Invitrogen
pGEX_6P_1	4.9	Ampr N-GST	GE Healthcare

Table 2. 2 List of plasmids used in this study.

More details about plasmids are in Appendix C.

2.2 Growth media

Two different media were used in this study, LB with or without agar, and Se-Met medium from AthenaES™ for crystallization using selenomethionine (se-met) instead of methionine in proteins when expressed in B834 cell strain (Table 2. 3).

Medium	Composition for 1 litre
Lurie-Bertani medium	10 g Bacto®tryptone 5 g Bacto®yeast extract 10 g NaCl The pH was adjusted to 7.4 with NaOH and the final volume was made up to 1 litre before autoclaving.
LB solid medium	In Lurie-Bertani Medium, add agar to final concentration 1.5 % and autoclave at 121 °C for 20 min. Pour 25 ml into a φ90 mm sterilized Petri dish and put on bench until setting. Keep at 4 °C.
Se-Met medium	Commercial medium, used for B834 cell strain.

Table 2. 3 Growth media used in this study.

2.3 Selective media

The β-lactam antibiotics (such as ampicillin) work by inhibiting the formation of peptidoglycan cross-links in the bacterial cell wall, while kanamycin works by affecting the 30S ribosomal subunit and preventing the translation of RNA (Garrod and al., 1981). These antibiotics were used at 1000 fold dilution of stock solutions as given in Table 2. 4; they were added to the sterilized growth medium once it had cooled to ~50 °C.

Antibiotic	Stock concentration	Storage details
Ampicillin	50 mg ml ⁻¹ in water	-20 °C
Kanamycin	50 mg ml ⁻¹ in water	-20 °C

Table 2. 4 Antibiotics used in growth media.

2.4 Chemicals, enzymes and kits

Chemicals, enzymes and kits are listed in Appendix A. The water used in this study was double distilled and purified to 18.2 Ω with a Millipore Milli-RO 10 Plus Water Purification System.

2.5 Sterilisation

Solutions, media, and glassware were usually sterilized by autoclaving in either an Astell Hearson 2000 Series Autoclave or a Prestige® Medical series 2100 Clinical Autoclave at 121 °C, 151 b in²(psi) for 20 min. When necessary, solutions were filter sterilized using an 0.25 μm pore Millipore filter disc (Supor® Acrodisc® 3.2, Gelman Sciences) and a suitable sterile syringe (Plastipak®,Becton Dickinson).

2.6 Storage of DNA and bacteria

Bacterial cells were stored at -80 °C in 25 % (v/v) glycerol in cryo-vials. Bacterial colonies on agar plates were stored at 4 °C for a maximum of four weeks. DNA was stored frozen at -20 °C in Elution Buffer (EB, 10 mM Tris/HCl buffer, pH 8.5).

2.7 Centrifugation

Harvesting of bacterial cells from cultures of 50-1000 ml was carried out by centrifugation at 5,000 g for 10 min in 500 ml centrifuge tubes (Nalgene) using a Beckman J2-21 centrifuge with a JA-10 rotor. Bacterial cells from culture volumes 5-10 ml were harvested in 30 ml Sterilin universal container at 3,000 g in a MSE Mistral 3000i bench centrifuge with a swing out rotor. Centrifugation of small volumes was carried out in 1.5 ml tubes (Treff) at 13,000 g in a Heraeus Instruments Biofuge *pico*.

2.8 Chemical competent *E. coli*

E. coli strains were made competent for uptake of plasmid DNA using calcium chloride in a variation of the method described by Cohen (Cohen *et al.*, 1972).

A 1 ml aliquot of a 10 ml culture of *E. coli* grown overnight in LB with appropriate antibiotic was used to inoculate 100 ml of LB (without antibiotic) in a 1-litre non-baffled flask. The culture was incubated at 37 °C with aeration in a shaking incubator rotating at 180 rpm until log phase was reached (OD₆₀₀ -0.4). After incubating for 10 min on ice, the cells were harvested by centrifugation at 5,000 g at 4 °C for 5 min. The supernatant was removed and the cells were gently resuspended with a pipette in 8 ml of ice cold 100 mM MgCl₂. The cells were harvested again and resuspended in 4 ml ice-cold 100 mM CaCl₂. After keeping on ice for 2 h, the cells were available to be used immediately or stored at -80 °C in 100 µl aliquots with 15 % (v/v) glycerol in 1.5 ml cryovials (Eppendorf).

2.9 Transformation of chemical competent *E. coli*

Transformation was carried out as follows: 100-500 ng of the plasmid DNA to be transformed was mixed with 50 µl chemically competent *E. coli* cells in 1.5 ml Eppendorf tubes by gentle swirling and incubated on ice for 10 min. The cells were heat shocked by incubation in 42 °C for 2 min and immediately returned to the ice for 2 min. The cell was then added 0.3 ml LB Medium and incubate at 37 °C for 1 h with 180 rpm shaking. The transformed cells were spread onto LB agar plates containing the appropriate antibiotics. Plates were inverted and incubated at 37 °C for 16 h.

2.10 Purification of DNA

DNA was replicated by transformation of the plasmids into XL1-Blue cells or TOP10 cells and growth of colonies on selective LB agar plates (Clark, 1968; Joseleau-Petit *et al.*, 1987; Khachatourians and Huzyk, 1974). The DNA was purified from a small scale (5-10 ml LB) culture

using Qiagen® Plasmid mini kit according the manufacturer's instructions.

2.11 Primers

All primers used were produced using oligo 2000 synthesis technology and purified using HPSF® purification (MWG-Biotech AG, Germany). Oligonucleotide primers were designed such that complementary sequences were normally 20 - 40 bases in length having a G/C content of approximately 40 % and a melting temperature (T_m) of ≥ 45 °C and within 5 °C for the primer pair. T_m s were calculated using the “Oligonucleotide Properties Calculator” at the <http://www.basic.northwestern.edu/biotools/oligocalc.html> which uses the following calculation:

$$T_m = 64.9 + 41 * (yG + zC - 16.4) / (wA + xT + yG + zC)$$

Where w, x, y, z are the number of the bases A, T, G, C in the sequence, respectively.

If possible, primers were also designed such that one or two G or C bases at both ends of the primers, so that the primers termini anneal well to the template strand. When required restriction site sequences were added to 5'-ends of primers with the addition of the sequence CTC at the extreme 5'-end to allow for cutting of the PCR product for ligation into appropriate vectors. Primers were dissolved in sterile double distilled water to 100 μ M, freeze-thawed to ensure full solubilization and then an aliquot was diluted to 125 ng μ l⁻¹.

2.12 Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) developed by Mullis and Faloona (Mullis and Faloona, 1987) was used throughout this study to amplify specific regions of *B. thetaiotaomicron* VPI-5482 genome DNA (125 mg/ml), and to introduce amino acid mutations.

PCR requires two oligonucleotide primers, one complementary to each strand of the DNA molecule, at sites that flank the region of DNA to be amplified. A thermostable DNA polymerase catalyses the synthesis of the complementary DNA strand in the presence of dNTPs.

PCRs were performed in sterile 0.2 ml or 0.5 ml Eppendorf tubes with a final 25 μ l or 50 μ l volume as Table 2. 5.

1 µl	dNTP mix (100 mM)
1 µl	50 µM forward oligonucleotide primer
1 µl	50 µM reverse oligonucleotide primer
~60 ng	DNA template
1 µl	DNA polymerase (2.5 U µl ⁻¹)
5 µl	10x reaction buffer (200 mM Tris/HCl, pH 8.8, 20 mM MgSO ₄ , 100 mM KCl, 100 mM (NH ₄) ₂ SO ₄ , 1 % Triton® X-100, 1 mg ml ⁻¹ BSA) Add sterilized ddH ₂ O to 50 µl

Table 2. 5 General PCR working system.

The PCR enzyme was KOD Hot Start DNA polymerase, which has an antibody bound at the active site which would be displaced when the temperature is high enough. The KOD polymerase prevents DNA synthesis before the primers are annealed. It also moves fast (106-138 bases/s) with low mutation frequency. The PCR machine utilised in this study is PX2 (Hybaid). The standard PCR program was as Table 2. 6.

Step 1	95 °C for 1 min	
Step 2	32 cycles	1. 95 °C for 30 sec
		2. 50 °C for 40 seconds
		3. 68 °C for 1 min/kb in each cycle
Step 3	68 °C for 10 min	
Step 4	5 °C for 24 h	

Table 2. 6 General PCR programme.

2.13 Site directed mutagenesis

QuickChange™ Site-Directed Mutagenesis Kit (Stratagene) allows site-specific mutation in virtually any double-stranded plasmid and was used according the manufacturer's instructions (Kunkel, 1985; Sugimoto *et al.*, 1989; Taylor *et al.*, 1985).

The site-directed mutagenesis method utilised an appropriate double-stranded recombinant plasmid DNA and two synthetic oligonucleotide primers containing the designed mutation flanked by 10-15 nucleotides that fully complemented the DNA template. A high fidelity proof reading DNA polymerase *Pfu*Turbo was used to extend the primers generating a nicked plasmid containing the desire mutation. Reactions were made up in sterile 0.2 ml Eppendorf tubes to a final volume of 40 µl as shown in Table 2. 7. The amplification program for site-directed mutagenesis was described in Table 2. 8.

20 ng	dsDNA template
125 ng	Oligonucleotide primer #1
125 ng	Oligonucleotide primer #2
5 µl	10x reaction buffer
1 µl	dNTP mix
1 µl	<i>Pfu</i> Turbo DNA polymerase (2.5U µl ⁻¹)
	Add sterilized ddH ₂ O to 50 µl

Table 2. 7 Mutation PCR reaction system.

Step 1	95 °C for 1 min	
Step 2	16 cycles	1. 95 °C for 30 sec
		2. 50 °C for 1 min
		3. 68 °C for 2 min/kb of plasmid length in each cycle
Step 3	68 °C for 10 min	
Step 4	5 °C for 24 h	

Table 2. 8 Mutation PCR programme.

After completion of the amplification reaction, 1 µl of *DpnI* (10 U) was added to each tube and mixed gently by flicking several times. The mixtures were spun down briefly and incubated at 37 °C for 1 h. The *DpnI* digested the methylated template dsDNA specifically, not the unmethylated amplified DNA. An aliquot of DNA before and after *DpnI* digestion was analysed by agarose gel electrophoresis to determine the quality of DNA produced. Digested reactions were then transformed into *E. coli* competent cell TOP10. The *in vivo* plasmid replication resulted in a closed plasmid.

2.14 Agarose gel electrophoresis of DNA

The separation and determination of the sizes of DNA molecules were carried out by electrophoresis through submerged horizontal agarose gels (Meyers, 1975).

The related buffers and reagents are given in Table 2. 9.

Buffer	Ingredient
Electrophoresis Buffer-TBE (pH 8.3)	8.9 mM Tris base
	8.9 mM Boric acid
	2 mM EDTA pH 8.0
DNA loading buffer	0.25 % (w/v) Bromophenol Blue
	50 % (v/v) Glycerol
	10x TBE

Table 2. 9 DNA agarose gel running buffer and loading buffer.

Agarose (1 % w/v) was boiled in 50 ml 1x TBE buffer until fully dissolved. After cooling to ~60 °C, the gel was mixed with 5 µl 0.5 µg/ml ethidium bromide solution and then poured into a mini gel apparatus with a suitable comb and end-plates inserted. After setting, 50 ml 1x TBE buffer was poured on top of the gel and the comb and end-plates removed. The DNA samples were loaded with 1/10th volume DNA loading buffer. Electrophoresis was conducted with a constant voltage 60 V (LKB Bromma 2197 Power Supply) for approximately 45 min, when a 500 bp fragment would migrate the 4/5 full length of the gel.

2.15 Visualisation of DNA and photography of agarose gels

Ethidium bromide staining was used to visualise the DNA under UV light. DNA gel photographs were produced using a gel documentation system (Bio-Rad Gel Doc 1000, Molecular Analyst™/PC Windows Software and were printed by Mitsubishi Video Copy Processor (Model P68B) with Mitsubishi thermal paper.

2.16 Determination of DNA fragment size

The size of lineal double stranded DNA molecules can be determined because the rate of migration through gel matrices is in effect inversely proportional to the \log_{10} of the size of the nucleic acid. Therefore, the sizes of DNA fragments were determined by comparing their electrophoretic mobility with that of the DNA standards of known sizes. The standard markers used are shown in Figure 2. 1.

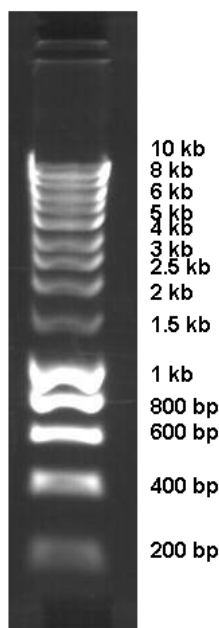


Figure 2. 1 DNA standards.
The DNA markers are from 10 kb to 200 bp.

2.17 DNA extraction from agarose gels

DNA was isolated from agarose gels with a razor blade and purified using the Perfectprep® Gel Cleanup Kit (Eppendorf) following the manufacturer's protocol (Downey, 2003; Matitashvili and Zavizion, 1997).

2.18 DNA digestion, ligation

Digestion of double strand DNA with restriction endonucleases was carried out as directed by the manufacturers (MBI Fermentas). Endonuclease restriction sites were identified using the tool online at <http://rna.lundberg.gu.se/cgi-bin/cutter2/cutter> (Bulow and Mosbach, 1982).

The required amount of DNA was mixed with the appropriate volume of 10 x concentrated reaction buffer. Endonuclease was then added at 5 -10 units μg^{-1} DNA and the digest incubated at 37 °C for 2 h (one unit is the amount of enzyme sufficient to digest 1 μg DNA in 1 h). For double digestions where both enzymes retained activity and specificity in the same reaction buffer the reactions were carried out simultaneously. When the reaction conditions for two restriction enzymes were incompatible the DNA was digested with one endonuclease, the buffer removed with the Perfectprep® Gel Cleanup Kit (Eppendorf), and DNA was then digested with the other endonuclease. Controls of undigested and single endonuclease-digested DNA were also carried out.

DNA fragments with cohesive ends were mixed to a final 3:1 insert:vector ratio, while for DNA molecules with blunt-ended termini, the ratio was at least 10:1. The reaction mixture included 1 μl 5X ligation buffer (250 mM Tris/HCl, pH 7.6, 50 mM MgCl_2 , 25 μM ATP, 25 mM dithiothreitol (DTT), 25 % (w/v) polyethylene glycol 8000), 1 μl 5U T4 ligase (Invitrogen), insert, vector, and water up to 5 μl . Ligation was performed at 16 °C for 2 h before 2 μl was transformed to competent XL1-Blue *E. coli* cells (Stratagene) or TOP10 *E. coli* cells (Invitrogen) following the manufacturer's instructions.

2.19 DNA Sequencing

DNA sequencing was conducted using the Value Read service from MWG Biotech AG, Ebersberg, Munich, Germany using ABI 3700 sequencers and Big dye technology (Applied Biosystems) (Hagemann and Kwan, 1999), each clone was sequenced in both the forward and reverse direction. Plasmid DNA (5 μl) was dried by vacuum lyophilization at room temperature in 1.5 ml tube. Plasmids were then sequenced with appropriate primers listed as Table 2. 10.

Plasmid type	Forward primer	Reverse primer
pET plasmids	T7: TAATACGACTCACTATAGGG	T7 Term: CTAGTTATTGCTCAGCGGT
pCR® Blunt	M13 uni (-21): TGTAACGACTCACTATAGGG	M13 reverse (-29): CAGGAAACAGCTATGACC
pGEX-6p-1	pGEX forward: ATAGCATGGCCTTTGCAGG	pGEX reverse: GAGCTGCATGTGTCAGAGG
minipRSETA	T7: TAATACGACTCACTATAGGG	T7 Term: CTAGTTATTGCTCAGCGGT
minipRSET-Trx	T7: TAATACGACTCACTATAGGG	T7 Term: CTAGTTATTGCTCAGCGGT

Table 2. 10 Commercial primers for sequencing.

The sequencing results were analysed using the DNASTar software (section 2.42.8). Configurations of multiple sequencing reactions, combining both sense and anti-sense directions, were generated and aligned with the wild type sequence using the CLUSTALW2 (section 2.42.9). Conflicts were resolved by reference to the sequencing trace. Nucleotide sequence was translated into amino acid sequence to ascertain the effect on the primary structure.

2.20 Purification of DNA by phenol

The removal of any impurities from DNA solutions was carried out when necessary by extraction with phenol and chloroform (Bewsey *et al.*, 1991).

An equal volume of phenol/chloroform (1:1) mixture was added to the DNA solution and mixed by vortexing. The organic and aqueous phases were separated by centrifugation at 13, 200 g for 5 min, and the upper DNA-containing aqueous phase was transferred to a fresh tube then extracted again with an equal volume of chloroform. DNA was precipitated from the aqueous layer by the addition of 0.1 volumes of 3 M sodium acetate, pH 5.2, and 2.5 volumes of ethanol (100 % v/v), followed by incubation at -20 °C for approximately 16 h. After centrifugation, the DNA pellet was washed in 70 % (v/v) ethanol, allowed to air-dry and resuspended in an appropriate volume of sterile double distilled water.

2.21 Determination of DNA concentration

DNA concentration was determined either by estimation from comparison of band size /intensity compared to quantified DNA ladders when electrophoresed on agarose gels or by scanning the absorbance of the appropriately diluted DNA (in MQ water) between 230 nm and 340 nm using a spectrophotometer (GE Ultrospec 4000 and 4300) (Hirschman and Felsenfeld, 1966; Skidmore and Duggan, 1966):

$$\text{dsDNA at } 50 \mu\text{g ml}^{-1} \text{ gives an } A_{260} = 1.0$$

2.22 Protein over expression in *E. coli*

Fresh colonies of *E. coli* containing expression plasmids were picked from plates and inoculated into 5-10 ml LB containing appropriate antibiotics and grown at 37 °C for ~16 h with 180 rpm shaking. The resultant cultures were then diluted 100-fold into 1 l fresh LB with appropriate

antibiotics in 2 l baffled flasks and cultured at 37 °C to OD₆₀₀ ~0.6-0.8. The cultures then were transferred to an incubator at the appropriate temperature and induced with required amount of isopropyl-β-D-thiogalactoside (IPTG) for protein expression (Golomb and Chamberlin, 1974; Studier and Moffatt, 1986). The flasks were incubated at this temperature for the necessary length of time before the cells were harvested by centrifugation at 5, 000 g at 4 °C for 10 min.

Recombinant protein was prepared from either the periplasmic, cell free extract or insoluble fractions as detailed below.

2.23 Preparation of periplasm protein

The cell pellets were gently resuspended in 20 ml per l of culture of ice-cold 20 mM Tris/HCl buffer, pH 8.0, containing 25 % (w/v) sucrose on ice and centrifuged at 5, 000 g for 10 min at 4 °C to remove cell debris. The supernatant comprises the periplasm. An aliquot of the periplasm, supernatant and cell debris pellet was analysed by SDS-PAGE to determine the success of the protein expression and the purity of the protein in the periplasmic fraction (Guisez *et al.*, 1998).

2.24 Preparation of cell free extracts (CFEs)

2.24.1 His tagged protein

The cell pellet was resuspended with 15 ml Talon™ buffer (20 mM Tris, 300 mM NaCl, pH 8.0) and stored in Sterilin tubes for 16 h at -20 °C. After thawing, the cell suspension was sonicated for 2 min using a B. Braun Labsonic U sonicator set at low intensity ~42 watts and 0.5 second cycling on ice then transferred to a 50 ml centrifuge tube (Nalgene). The cell debris pelleted at 30, 000 g for 30 min at 4 °C, and the supernatant comprising the CFE, was used for further study. The pellet (insoluble fraction) was resuspended in 10 ml Talon™ buffer denatured by 8 M urea.

2.24.2 GST tagged protein

The cell pellet was resuspended with 15 ml PBS (140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, pH 7.2) and transferred in Sterilin tubes for storing at -20 °C. After thawing, the cell suspension was sonicated as His tagged protein. The cell debris pelleted at 30, 000 g for 30 min at 4 °C, and the supernatant comprising the CFE, was used for further study. The pellet contained insoluble fraction.

2.24.3 Untagged protein

The cell pellet was resuspended with 15 ml Talon™ buffer and transferred in Sterilin tubes for

storing at -20 °C. After thawing, the cell suspension was sonicated as for His tagged protein. The cell debris pelleted at 30,000 g for 30 min at 4 °C, and the supernatant comprising the CFE, was used for further study. The pellet contained insoluble fraction.

2.25 Insoluble Protein

Proteins expressed as inclusion bodies (IB) were resuspended in with 20 mM Tris, pH 8.0 buffer containing 300 mM NaCl and 8 M urea and incubated at 37 °C for 30 min. After centrifugation at 30,000 g for 30 min, the supernatant was collected for further purification under denaturing conditions.

2.26 Protein Purification

2.26.1 Immobilised Metal Affinity Chromatography (IMAC)

Proteins fused to a thioredoxin-His₆-tag (Trx-His) or His₆-tag were purified using metal affinity matrix Talon™ resin (Clontech Laboratories Inc) (Hansen and Lindeberg, 1995).

A 2 ml bed volume of Talon resin was loaded onto a Fastflow™ column. The column was then equilibrated with at least 10 column volumes of Talon buffer (20 mM Tris, pH 8.0 buffer containing 300 mM NaCl). CFE or solubilised IB was loaded to the column and allowed to flow through by gravity. The column was then washed with 30 ml Talon buffer and the bound protein eluted with 5 ml Talon buffer containing 10 mM imidazole followed by 5 ml Talon buffer containing 100 mM imidazole. These fractions were collected separately and aliquots of 13 µl of each fraction were analyzed by SDS-PAGE.

Proteins fused to a GST-tag were purified using glutathione sepharose 4B resin (GE Healthcare). CFE was prepared in PBS buffer and loaded onto a 1 ml bed volume glutathione sepharose column which had been equilibrated with at least 10 column volumes of PBS buffer. The column was washed with 30 ml PBS buffer and the fusion protein was eluted with 10 ml elution buffer (20 mM reduced glutathione in 50 mM Tris/HCl, pH 8.0). Aliquots (13 µl) of each fraction were taken for SDS-PAGE analysis.

2.26.2 Ion Exchange Chromatography (IEC)

IEC was used to purify proteins fused to a His-tag after thrombin digestion. Proteins were concentrated to 4.5 ml using a centrifugal concentrator. A Q-12 anion exchange (Bio-Rad) column was equilibrated with 100 ml Talon buffer using a Bio-Rad HR workstation with a flow rate 2

ml/min. The protein was loaded using a 4 ml static loop and then washed with 60 ml Talon buffer to remove unbound protein. Finally, a 200 ml linear gradient Talon buffer containing 0-500 mM NaCl was carried out. Protein elution was detected by UV-absorbance and fractions were collected above a threshold of 0.05 AU at 280 nm. Fractions were then checked by SDS-PAGE. Fractions containing pure protein were pooled together and stored at 4 °C.

2.26.3 Gel Filtration Chromatography (GFC)

GFC was used to purify proteins by size. The protein was dialysed into buffer A (10 mM Tris, 150 mM NaCl, pH 8.0) and concentrated to 4.5 ml. The protein was loaded on a HiLoad 26/60 Superdex 75PG column connected to an AKTAprime low pressure system (GE Healthcare). The column was equilibrated with 400 ml buffer A at 3 ml/min flow rate before the protein was loaded onto the column using a 4 ml static loop. Proteins were eluted from the column using Buffer A at 3 ml/min flow rate. Proteins were collected by UV-absorbance above a threshold of 0.05 AU at 280 nm. Fractions were then checked by SDS-PAGE. Fractions containing protein were pooled together and stored at 4 °C.

2.27 Determination of protein concentration by spectrophotometer

Protein concentration was determined using published methods based on the tyrosine and tryptophan content (Gill and von Hippel, 1989; Pace et al., 1995). The absorbance difference between 280 nm and 320 nm of the appropriately diluted pure proteins was measured with a spectrophotometer and the data were used to calculate protein concentrations, using the formula:

$$A = \epsilon CID$$

Where A = absorbance 280 nm-absorbance 320 nm, ϵ = molar extinction coefficient, I = length of light path (cm), D = dilution factor, and C = molar concentration of sample.

The molar extinction coefficient value was calculated using ProtParam online software (<http://ca.expasy.org/tools/protparam.html>). The length of the light path was always 1 cm.

2.28 Protease Digestion

Proteins used in this study fused to a Trx-His or His-tag have a thrombin cleavage site between tag(s) and the start code of target proteins, which can be digested by thrombin to remove tags. The purified protein fusions were transferred to a semi-permeable membrane clipped at both terminals and dialysed into 2 l thrombin cut buffer (20 mM Tris, 150 mM NaCl, 2.5 mM CaCl₂, pH 8.4) at 4 °C overnight with slow stirring. Proteins were then transferred into a sterile tube with 1U thrombin (Novagen) per mg fusion protein to digest at room temperature overnight. The digested

proteins were then loaded onto Talon resin and the flow though containing the cleaved untagged protein collected.

Proteins fused to a GST-tag were digested by PreScission Protease in a cleavage buffer (50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5) following the manufacturer's instructions.

Proteins fused to pET32b have an enterokinase (rEK) cleavage site between N-terminal His-tag and the start code of target proteins which can be digested by rEK to remove the N-terminal extra fragments. The fusion proteins were concentrated and dialysed into rEK cut buffer (same as thrombin cut buffer), the appropriate content of rEK was added, to digest at 37 °C overnight. The digested solution was filtered for further purification by Ion Exchange.

Less specific proteases, such as protease K, trypsin and chymotrypsin, were used in this study to check the protein stability. The buffer for these protease digestions was the same as for thrombin. Digestions were performed for different time periods and different protease concentrations.

2.29 Sodium Dodecylsulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE)

The gel consists of two parts, the resolving gel and the stacking gel (Table 2. 11). The system used was the AE-6450 apparatus from ATTO Corporation, Genetic Research Instruments that has 10 cm x 10 cm glass plates that are clipped together with a rubber seal between them. The resolving gel was poured into the plates covered with water and allowed to polymerize. The water was then removed and the stacking gel poured on top of the resolving gel. This was allowed to polymerize with a 14-well comb in place. The comb and the rubber seal around the glass plated were removed when the gel was set. Protein samples were prepared by taking 13 µl of appropriately diluted protein solution, mixed with 7 µl of loading buffer and boiled for 3 min. The plates were placed in the gel tank that was filled with running buffer. An aliquot 20 µl of protein standards (SigmaTM) of known molecular weight was loaded beside protein samples into the well of the gel which was then electrophoresed at a current of 30 mA per gel.

Component	Volume/Amount	
Resolving gel	12.5 %	15 %
0.75 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	9.4 ml	9.4 ml
40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide)	5.8 ml	7.1 ml
d.d. H ₂ O	3.5 ml	2.2 ml
10 % (w/v) Ammonium persulphate	90 µl	90 µl
TEMED	30 µl	30 µl
Stacking gel		
0.25 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	3.75 ml	
40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide)	0.75 ml	
d.d. H ₂ O	3.0 ml	
10 % (w/v) Ammonium persulphate	60 µl	
TEMED	20 µl	
Loading buffer		
SDS	10 % (w/v)	
0.25 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	5 ml	
Glycerol	25 % (w/v)	
β-mercaptoethanol	2.5 ml	
Bromophenol blue dye	0.1 %	
Running buffer		
32 mM Tris/190 mM glycine, pH 8.3	350 ml	
SDS	0.1 %	

Table 2. 11 SDS-PAGE gel solutions, loading buffer and running buffer.

After electrophoresis the gel was soaked in the Stain Buffer (0.4 % comassie brilliant blue R, 40 % (v/v) methanol, 10 % (v/v) glacial acid) for 30 min with gentle shaking or used for blotting directly. The M_r of proteins separated by SDS-PAGE were estimated by the comparing their electrophoretic mobility with protein standards of known M_r (Figure 2. 2).

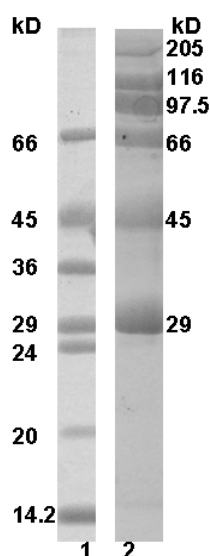


Figure 2. 2 Protein markers used for SDS-PAGE.

Lane 1: Low molecular weight marker, Lane 2: High molecular weight marker.

2.30 Affinity Gel Electrophoresis (AGE)

Native polyacrylamide gels (7.5 % w/v acrylamide) were used in the same gel apparatus (ATTO Corporation) for SDS-PAGE to detect the affinity of proteins with soluble polysaccharides based on the methods described by Takeo (Takeo, 1984) (Table 2. 12). The gel was poured into the glass plates and allowed to polymerize with a 14-well comb in place. The comb and the rubber seal around the glass plated were removed when the gel was set. Gels with and without ligand (0.1 %) were run in the same gel tank with running buffer. A 10 μ l of protein sample was mixed with 10 μ l of loading buffer before loading into a well. BSA (10 μ g) was used as a marker. Native gels were electrophoresed at 10 mA/gel at room temperature for 1.5 h. Gels were stained and destained as SDS-PAGE gels.

Component	Volume/Amount	
Native gel	Control	Ligands
250 mM Tris, 2.5 M glycine pH 8.3	2 ml	2 ml
40 % Acrylamide (BDH Electran acrylamide, 3 % (w/v) bisacrylamide)	3.75 ml	3.75 ml
d.d. H ₂ O	12.13 ml	10.13 ml
1 % ligand (w/v)	nil	2 ml
10 % (w/v) Ammonium persulphate	100 μ l	100 μ l
TEMED	20 μ l	20 μ l
Running Buffer		
25 mM Tris, 250 mM glycine pH 8.3	350 ml	
Loading Buffer		
Glycerol	25 % (w/v)	
Bromophenol blue dye	0.0025 %	

Table 2. 12 AGE gel solutions.

2.31 Protein N-terminal Sequencing and molecular weight measurement

The SDS-PAGE gel after electrophoresis was incubated in blotting buffer (10 mM CAPS containing 10 % methanol, pH 11.0) for 15 min. Four pieces of 3 mm filter paper (Whatman™) cut to the same size as the gel were soaked in blotting buffer. One piece of PVDF membrane was dampened in methanol before being moved into blotting buffer. The proteins were then blotted onto the PVDF membrane using a Bio-Rad semi-dry transfer cell in which a sandwich was constructed comprising of two pieces of dampened filter paper, a PVDF membrane, an SDS-PAGE gel, and another two pieces dampened filter paper. The transfer cell was then run at 10 V for 1 h to transfer the protein to the PVDF membrane. After blotting, the protein was visualized by staining for 2 min in Stain Buffer, destained in 40 % v/v methanol/10 % v/v glacial acid and washed in ultra pure water to remove excess dye. The protein band was then cut from the membrane and the N-terminal sequenced at the Pinnacle Lab at Newcastle University.

The same protein band was cut from an SDS gel and sent to the Pinnacle Lab for molecular mass analysis by MALDI (Matrix-Assisted Laser Desorption Ionization)-TOF (Time Of Flight) mass spectroscopy with two standards (Figure 2. 3). The error of MALDI-TOF is ± 0.1 %.

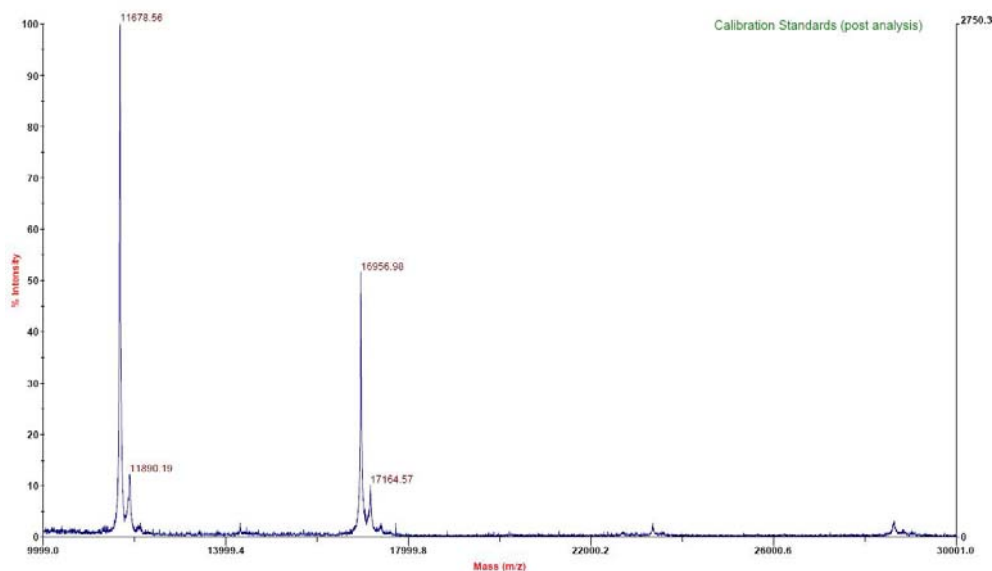


Figure 2. 3 MALDI-TOF mass spectrum standards.

The error value of MALDI-TOF mass spectrum for molecular weight measurement is 0.1 %.

2.32 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry was used to measure binding affinity of different molecules (Fisher and Singh, 1995; Leavitt and Freire, 2001; Wiseman *et al.*, 1989). ITC was carried out at 25 °C using a MicroCal Omega titration calorimeter (Figure 2. 4). All solutions were dialyzed in same buffer to minimize the noise. All solutions were thoroughly treated by vacuum stirring before loading into the syringe and the cell of the calorimeter. The samples for the cell were ~1.8 ml with lower concentration of protein while the ones for the syringe were about 0.6 ml with higher concentration of another protein. The machine was setted to inject successive aliquots of 10 µl samples from the syringe with a 200-350 seconds interval into the cell under continuous stirring at 300 rpm. The binding data were collected according to the heat effects and analysed by non-linear regression using a simple one-site binding model. The fitted data yield the association constant (K_A), number of binding sites on the protein (N), the enthalpy of binding (ΔH) and the entropy of binding (ΔS) by Origin software (version 7.0). In these titrations, T is 298 °C. Other thermodynamic parameters (ΔG , $T\Delta S$) were calculated using the standard thermodynamic equation as below:

$$-RT\ln K_A = \Delta G = \Delta H - T\Delta S$$

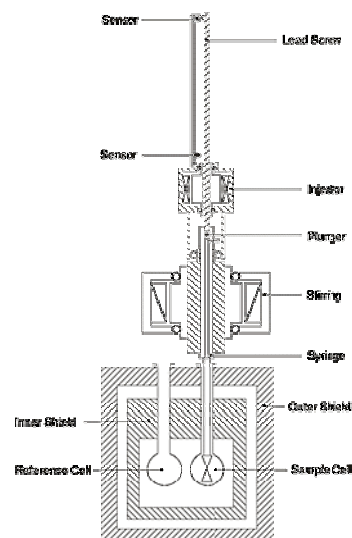


Figure 2. 4 ITC instrument used in this study.

2.33 Circular Dichroism (CD)

Circular dichroism (CD) is a form of spectroscopy based on the differential absorption of left- and right-handed circularly polarized light (Figure 2. 5). It can be used to help determine the structure of macromolecules (including the secondary structure of proteins and the handedness of DNA) (Kelly *et al.*, 2005). The secondary structures of proteins were measured using the Jasco J-815 CD. The 200 μ l 10 μ M protein in 20 mM Tris pH 8.0 were loaded in the 0.2 mm two-slide cuvettes and put into the detector with temperature controlled at 20 °C. The scan was carried out from 250 nm to 190 nm. The data were collected and analysed using the manufacturer's software.

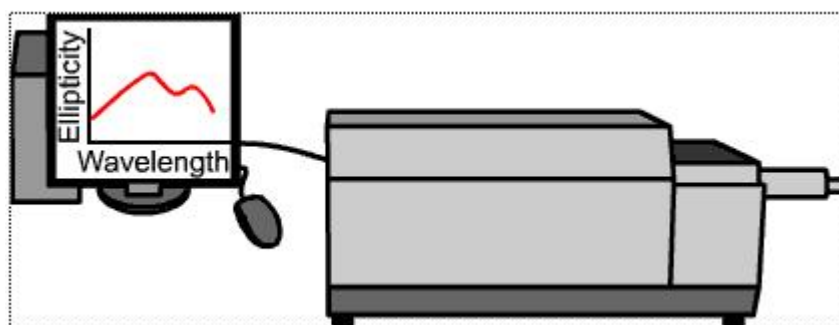


Figure 2. 5 Circular dichroism instrument used in this study.

2.34 Fluorescence Spectroscopy

Fluorescence spectroscopy was used to determine the binding of ligands to proteins, based on the intrinsic fluorescence of tryptophan and tyrosine residues. Fluorescence spectroscopy was performed using a SLM100 8100 spectrometer operating in ratio mode with an 8 nm excitation and 4 nm emission bandwidths. Proteins were diluted in 20 mM Tris buffer, pH 8.0 to a final concentration of 1 μ M. Samples were excited at 280 nm, and the emission spectra of the proteins recorded between 300 nm and 450 nm at 20 °C. Samples were corrected by subtraction of a buffer blank. The concentration of ligand varied between 0.5 mM to 50 mM.

2.35 Differential Scanning Calorimetry (DSC)

Differential Scanning Calorimetry (DSC) measures the function of temperature against target proteins by measuring the heat changes during controlled increase (or decrease) in temperature. In this study, DSC was used to measure the difference of protein with the presence and absence of ligands. One ml of 0.5 mg/ml protein in 20 mM Tris pH 8.0 buffer was degassed at 25 °C and then injected into sample cell of MicroCal Differential Scanning Calorimeter VP-DSC (Figure 2. 6). The control cell holds the same buffer. The solution was heated at a constant rate of 1 °C per min from 15 °C to 70 °C. The T_m measured indicated the heat changes. The transition midpoint T_m is the temperature where 50 % of the protein is in its native conformation and the other 50 % is denatured. The higher the T_m , the more stable the molecule.



Figure 2. 6 DSC instrument used in this study.

2.36 Protein Crystallization

Proteins for crystallization were required at >95 % homogeneity as assessed by SDS-PAGE. Proteins were concentrated to at least 12 mg/ml in ultra pure water. Crystallization of proteins was carried out initially using the sitting drop vapour diffusion technique. Drops were viewed using a Nikon SM21000 stereo zoom microscope each day in the first week. After the first week, trays were assessed periodically. The successful conditions of sitting drop experiments were then modified in hanging drop vapour diffusion technique to achieve better crystals.

2.36.1 Sitting drop vapour diffusion

For initial small scale screening, crystallization was conducted using a Mosquito™ (TTP Labtech) nanolitre pipetting robot with 96-well crystallization plates (Greiner) (Figure 2. 7). Crystallization mother liquor (50 μ l) (Newcastle, Classics, PACT, AmSO₄, or JSCG screen, see Appendix B) was dispensed into each well of the plates, and 7 μ l of protein was dispensed into each of the 8 sample reservoirs before 100 nl was aliquoted by the robot onto each of the crystallization shelves. The 100 nl of protein was then mixed with the same or half amount of crystallizing mother liquor in each two eyes of a well of a 96-well plate. The tray was then sealed using sealing film with contact adhesive and crystal conditions were viewed as described before.

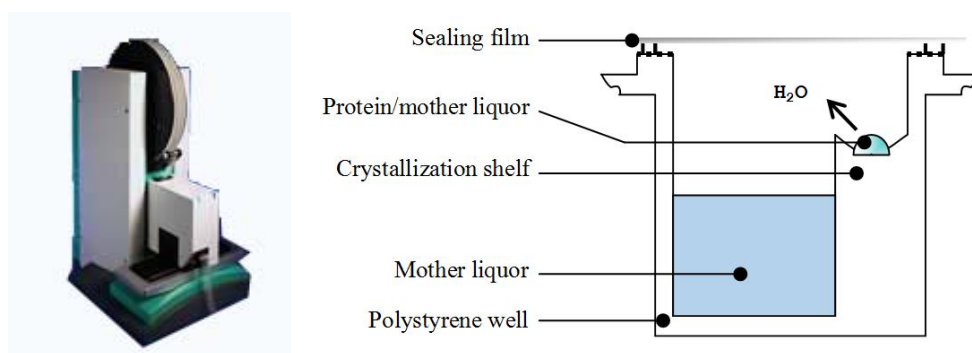


Figure 2. 7 Robot crystallization instrument and its sitting drop method.

Diffusion due to loss of water from sitting drop in vapour phase causes equilibrium to occur between drop and reservoir.

2.36.2 Hanging drop vapour diffusion

The hanging drop technique for crystal growth requires that a small aliquot of protein or protein/ligand mix (0.7-2 μ l) is mixed with an amount of crystallizing mother liquor (0.7-2 μ l) on a siliconized cover slip. Cover slips (18 \times 18 mm, Scientific laboratory supplies) were pretreated with Aqua Sil™ (Hampton Research), dried according to manufacturer's instructions and polished with a silk scarf prior to dispensing of protein or crystallization liquor. Cellstar, 24-well tissue culture, polystyrene, non-pyrogenic, DNase and RNase free plates (Greiner, Bio-one) were prepared by

applying High Vacuum Grease (Dow Corning) to the rim of each well using a 10 ml syringe with a 200 μ l Gilson tip, trimmed to allow flow of the grease. Typically 0.5 ml of mother liquor was pipetted into each well, although different well volumes could be used for alteration of diffusion rate. Once cover slips were setup, as previously described, they were immediately inverted and sealed above the corresponding mother liquor by firmly pressing down the cover slip against the vacuum grease as in Figure 2. 8. Crystallization trays were incubated at 20 °C and/or 4 °C in a Sanyo MIR-153 incubator.

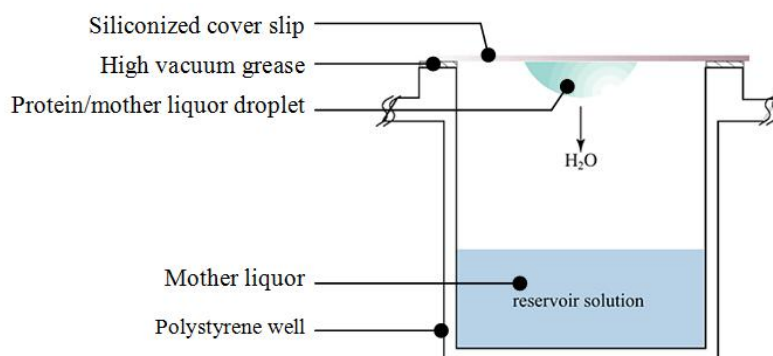


Figure 2. 8 Hanging drop crystallization method.

Diffusion due to loss of water from hanging drop in vapour phase causes equilibrium to occur between drop and reservoir. Over saturation of protein is also increased as water diffuses away from drop (adapted from Hampton Research).

2.36.3 Collection of crystal diffraction data and structure refinement

Unique crystals of a suitable form were harvested on rayon fibre loop and flash frozen in liquid nitrogen with appropriate cryoprotectant. Crystal harvesting, data collection, and structure refinement were carried out by Dr Susan Firbank, Newcastle University. Potentially useful crystals were screened on an in-house source (Rigaku rotating RU-200 X-ray generator, with a Cu 1.5418 Å target operating at 50 kV and 100 mA with focusing X-ray optics from Osmics), utilising a Rigaku RAXIS IIC or a RAXIS-IV image plate detector. Typically, two images per crystal at 0° and 90°, with an oscillation range of 0.5°/image to a resolution of 2.2 Å, were collected and reflections indexed using DENZO (Otwinowski and Minor, 1997) to determine space group and quality of crystals. Crystals with good diffraction on home source were taken to Diamond (www.diamond.ac.uk) for synchrotron diffraction.

Structures shown in this thesis were solved by using the single-wavelength anomalous dispersion method at a wavelength optimized for the f'' signal of the selenium or by molecular replacement. Prior to refinement, 5 % of the observations were immediately set aside for cross-validation analysis (Brünger, 1992) and were used to monitor various refinement strategies. The programs MOSFLM, REFMAC5, and SCALA used are part of the CCP4 software package (Project, 1994).

Models were completed with manual correction using COOT (Emsley and Cowtan, 2004) and refinement with REFMAC5.

2.37 Protein cross linking

A crosslinking protocol was used to investigate changes in the oligomeric state of proteins in the presence and absence of ligand (Jaenicke and Rudolph, 1986).

A 0.5 ml solution containing 50 µg protein in 50 mM NaPO₄ pH 7.6 was incubated 10 min at room temperature. Then 20 µl 25 % glutaraldehyde was added and incubated for 2 min at room temperature. This was quenched by addition of 25 µl 2 M NaBH₄ and incubated for a further 20 min at room temperature. A 1.5 µl aliquot of 10 % deoxycholate was added, then 22.5 µl trichloroacetic acid added slowly and incubated on ice for 5 min before centrifuging at 13, 200 g at 4 °C for 20 min. The supernatant was then discarded and the pellet washed with 600 µl ice-cold acetone, dispersing pellet using vortex or pipette tip, then centrifuging at 13, 200 g at 4 °C for 10 min. The supernatant was then removed and discarded and the pellet resuspended in 20-40 µl SDS-PAGE loading buffer. Molecular weight differences resulting from cross linking were assessed using SDS-PAGE.

2.38 Carbohydrate Digestion

Polysaccharides were partially digested into a range of smaller oligosaccharides by acid hydrolysis. HCl was added to the polysaccharide solution (1 % w/v in dH₂O) to a final concentration of 1 M. The samples were vortexed and kept at room temperature for 20 min to 1 h. NaOH was added to neutralize the pH and stop digestion. The sugar fractions were purified with a BioRad P2 column or a BioRad P4 column using degassed water at flow rate 1 ml/min. The sugar fractions were collected and freeze-dried for further use.

2.39 Thin Layer Chromatography (TLC)

TLC uses different solvent systems to separate chemicals in a horizontal direction in a wide range of temperatures and can be developed for a visualized image (Demirci, 2008; Meyers and Meyers, 2008; Tie-xin and Hong, 2008; Waksmundzka-Hajnos, 2008). A piece of 25 x 25 cm foil backed silica TLC was cut to an appropriate size and drawn a faint pencil line 1 cm from the bottom edge. Oligosaccharide samples (2-16 µl) were loaded at 0.8 cm intervals on the line using a 2 µl pipette and allowed to dry using a hair drier after each loading of 2 µl. After loading, the plate was placed in a 1 l prolate glass tank containing butanol/acetic acid/H₂O in a ratio of 2:2:1, to a depth of 0.5 cm and sealed with greased glass plate. The mixture of samples migrates through the adsorbent layer of

TLC. As the process runs, the sugars move different distances, depending on their relative affinities for the adsorbent. When the running buffer reached ~1 cm to the top of the plate, the plate was taken out and dried carefully using a hair drier and immersed for a few seconds in an orcinol sulphuric acid reagent (sulphuric acid/ethanol/water 3:70:20 v/v, orcinol 1 ‰), dried carefully and heated until sugars were revealed, at 120 °C (5 -10 min). In this study, sugar fractions were digested from fructans by acid hydrolysis. Their relative affinities are predicted same. So the movement of sugars on TLC mainly depends on the size of fractions. A group of known oligosaccharides was used as size marker. Figure 2. 9 shows the separation of hydrolyzed levan on TLC as an example.

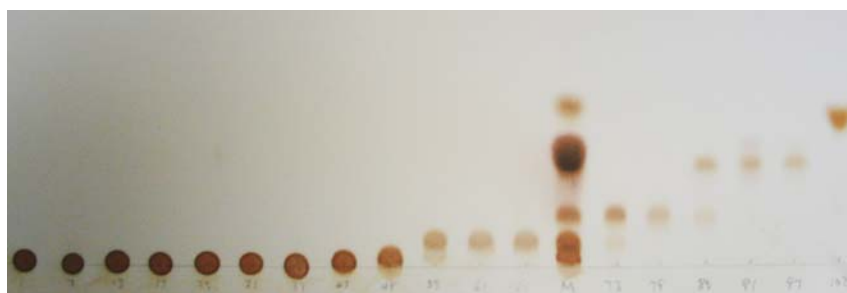


Figure 2. 9 TLC shows the isolated fragments of digested levan.

M: The size markers made by Dr. James Flint included L2 to L5, and fructose. Numbers below the pencil line were the collection tubes' numbers. Finally, No. 86-89 were identified as L2, No. 74-82 were L3, No. 63-69 were L4, No. 56-58 were L5, No. 47-51 were L6, No. 39-45 were L7.

2.40 DNSA assay

The reducing sugar content of solutions was determined by the methods of Miller (Miller, 1959). All reactions were carried out at 25 °C in 20 mM Tris buffer (pH 8.0) plus 1 mg/ml BSA. The DNSA reagent (Table 2. 13) was boiled with the sugar containing samples with a ratio of 1:1 or 3:1 depending on the volume of samples for 20 min, while the colour changed from light orange to dark red indicating the contents of reducing sugars which can be measured by a spectrophotometer at 575 nm.

Component	Concentration	Note
3, 5-Dinitrosalicylic acid (DNSA)	1 % (w/v)	
Phenol	0.2 % (w/v)	
Sodium hydroxide	1 % (w/v)	
Glucose	0.002 % (w/v)	added on the day of use
Na ₂ SO ₃	0.05 % (w/v)	added on the day of use

Table 2. 13 DNSA solution.

A series of acceptable absorbency (from 0.1 to 1.5) on at least 5 different time points were collected for final analysis. By comparing with a standard curve of reducing sugars (in a reasonable range: 0-1000 µg ml⁻¹), the enzyme activities against different sugars were calculated using the Michaelis-Menten model of nonlinear regression analysis of Prism (Version 4.0.3, GraphPad™). R² values

were checked to be higher than 95 % in each set to guarantee the quality of the nonlinear regression curves. At least two repeats were performed with acceptable R^2 values to achieve the variation of mean. The K_{cat} is calculated according the equation:

$$K_{cat} = F_{[con]}/E_{[con]}$$

Here $E_{[con]}$ means the final concentration of enzyme in the reaction solution.

Figure 2. 10 presents a graph made by Prism as an example to show how DNSA data were analysed.

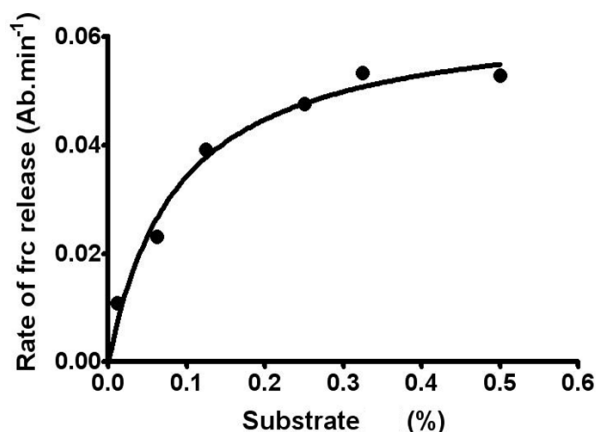


Figure 2. 10 DNSA assay example.

This nonlinear regression graph shows the change in enzymic rate within different concentration of substrate. Data were fit to a single site model to obtain V_{max} and K_m using Graphpad Prism.

2.41 High performance liquid chromatography (HPLC)

High performance liquid chromatography (HPLC) was used to determine the products made by the glycoside hydrolases (Kang *et al.*, 1990; Lohmander, 1986; Nishiura *et al.*, 1990; Parente *et al.*, 1984; Takeuchi *et al.*, 1987). The HPLC machines (Dionex DX500 and ICS3000) are fully automated and have a sample size of 215 μ l, flow rate of 1 ml min⁻¹ and run at a pressure of ~2100 psi (lb in⁻²). The sugars were detected by pulsed amperometric detection (PAD) with settings $E_1=0.05V$, $E_2=0.6V$, $E_3= -0.6V$ (Hardy *et al.*, 1988; Lee, 1990; Townsend and Hardy, 1991; Townsend *et al.*, 1988). The column is a CARBOPAC™ PA-100 anion exchange column equipped with a guard column (Dionex). Samples were prepared by removing the denatured protein from boiled enzyme assays by centrifugation (13, 200 g for 5 min) and carefully decanting the supernatant. At least two separate standards (containing glucose, fructose, and sucrose at the concentration of 10 μ g/ml for each) were used in each run, one is at the beginning and another one is at the end to control the variation in the elution times of the sugars through the run (Figure 2. 11).

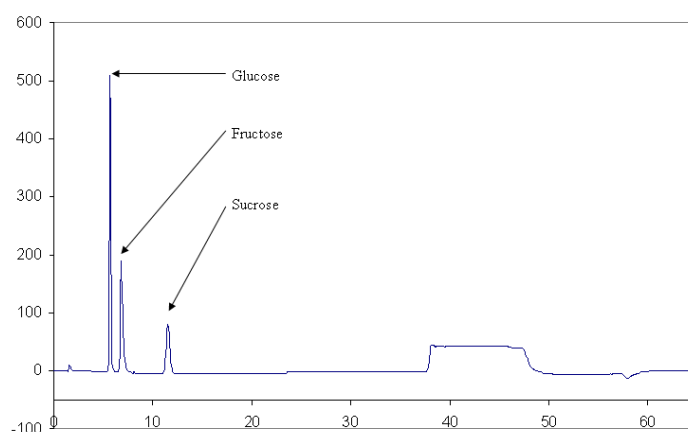


Figure 2. 11 Fructose, glucose, and sucrose as standard of HPLC.

Fructose comes out at 6.80 min, glucose comes out at 5.67 min, and sucrose comes out at 11.47min. The sugar signals are different even all their concentrations were 10 µg/ml respectively.

All the buffers used in HPLC were degassed over 30 min before poured into the tank. Each pipe linked to buffer tanks was primed ~5 min till no air bubbles came out. Prior to the first run of each batch of samples, the column was equilibrated with 66 mM sodium hydroxide at flow rate 1 ml/min for ~20 min until the baseline was flat. The programme for each sample is 65 min as described below (Table 2. 14). The gradient in step 2 varied depending on the difference of each samples. All data were analyzed through Prism.

Step	Solution	%	Time (min)
1	66 mM sodium hydroxide	100	10
2	500 mM sodium acetate in 66 mM sodium hydroxide	0-15/0-25/0-35/0-50	25
3	500 mM sodium acetate in 66 mM sodium hydroxide	100	10
4	500 mM sodium hydroxide	100	10
5	66 mM sodium hydroxide	100	10

Table 2. 14 HPLC buffers and programme.

2.42 Bioinformatics Softwares

2.42.1 NTI vector

The local software NTI vector (Version 10.0.1) was used to map any potential restriction cleavage sites of reconstructed plasmids.

2.42.2 PyMol

The local software PyMol (Version 0.99) was used to present protein structures and the accomplished ligands at three-dimensional level.

2.42.3 Tcoffee

The online software Tcoffee (http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi) was used to color protein sequence alignments.

2.42.4 SignalP

The online software SignalP (Version 3.0) (<http://www.cbs.dtu.dk/services/SignalP/>) predicts the presence and location of signal peptide cleavage sites in amino acid sequences.

2.42.5 TMHMM and TMPred

The online software Prediction of transmembrane helices in proteins (TMHMM) Server (Version 2.0) (<http://www.cbs.dtu.dk/services/TMHMM-2.0/>) and TMPred (www.ch.embnet.org/software/TMPRED_form.html) were used for Prediction of transmembrane helices in proteins.

2.42.6 SMART

The online software SMART (Simple Modular Architecture Research Tool) (http://smart.embl-heidelberg.de/smart/set_mode.cgi?NORMAL=1) was used to analyse the domain structure of proteins. Only domains in the SMART database are recognised by the software.

2.42.7 Swiss Model

The online software Swiss Model (http://swissmodel.expasy.org/workspace/index.php?func=modelling_simple1) was used to predict the protein three-dimensional structures according to the alignment with structure-known proteins.

2.42.8 DNASTar

The local software DNASTar was used to look through gene and protein sequences and to design primers.

2.42.9 ClustalW2

ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins (Jeanmougin *et al.*, 1998). It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

2.42.10 BLAST

This online software the Basic Local Alignment Search Tool Blast (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) was used to analyse the similarity of proteins and genes among different sources.

2.42.11 DALI

The online software DALI (http://ekhidna.biocenter.helsinki.fi/dali_server/) was used to search the similar secondary structures through PDB database (Holm and Park, 2000).

2.42.12 SWEET2

The online software SWEET (<http://www.dkfz-heidelberg.de/spec/sweet2/doc/index.php>) was used to predict the 3D structure of saccharides.

2.42.13 LipoP

The online software LipoP (<http://www.cbs.dtu.dk/services/LipoP/>) was used to predict the presence of lipoproteins in gram negative bacteria.

2.42.14 Multalin

Multalin (<http://bioinfo.genotoul.fr/multalin/multalin.html>) is an online multiple sequence alignment software with hierarchical clustering.

Chapter 3 Characterisation of the periplasmic sensor domain of the BT1754 hybrid two-component system

3.1 Introduction

All cells need to be able to sense and respond to their environment to survive. In this study we want to investigate the identity of the signalling molecule and the mechanism of signal perception in bacterial sensory systems and also further our understanding of bacterial survival in the human gut, as our normal microbiota play an important role in health and nutrition (section 1.2.1). To do this we are using *B. thetaiotaomicron* as a model system as it's a dominant member of the normal microbiota and the genome sequence indicates it contains a large number of extracellular sensor-regulators (section 1.2.2). These are linked to polysaccharide utilisation genes (*susC/D* homologues and glycoside hydrolases), so it is likely that they are coregulated to sense and respond to different complex carbohydrates present in the niche of the large intestine. These carbohydrates are likely to be mainly plant polysaccharides from our diet or host mucins found in the gut. Initial tests revealed that *B. thetaiotaomicron* could grow on a range of purified polysaccharides, such as pectins, storage polysaccharides (starch and fructan) and host polysaccharides (chondroitin sulphate and heparin), but not hemicelluloses (xylans and mannans; D. Bolam, unpublished data). Custom genechips (work carried out by Dr Eric Martens, Washington University, St Louis, USA) were run and the results showed that each polysaccharide activated one or more loci containing *susC* and *susD* genes as well as a number of catabolic enzymes and sugar transporters (Figure 3. 1).

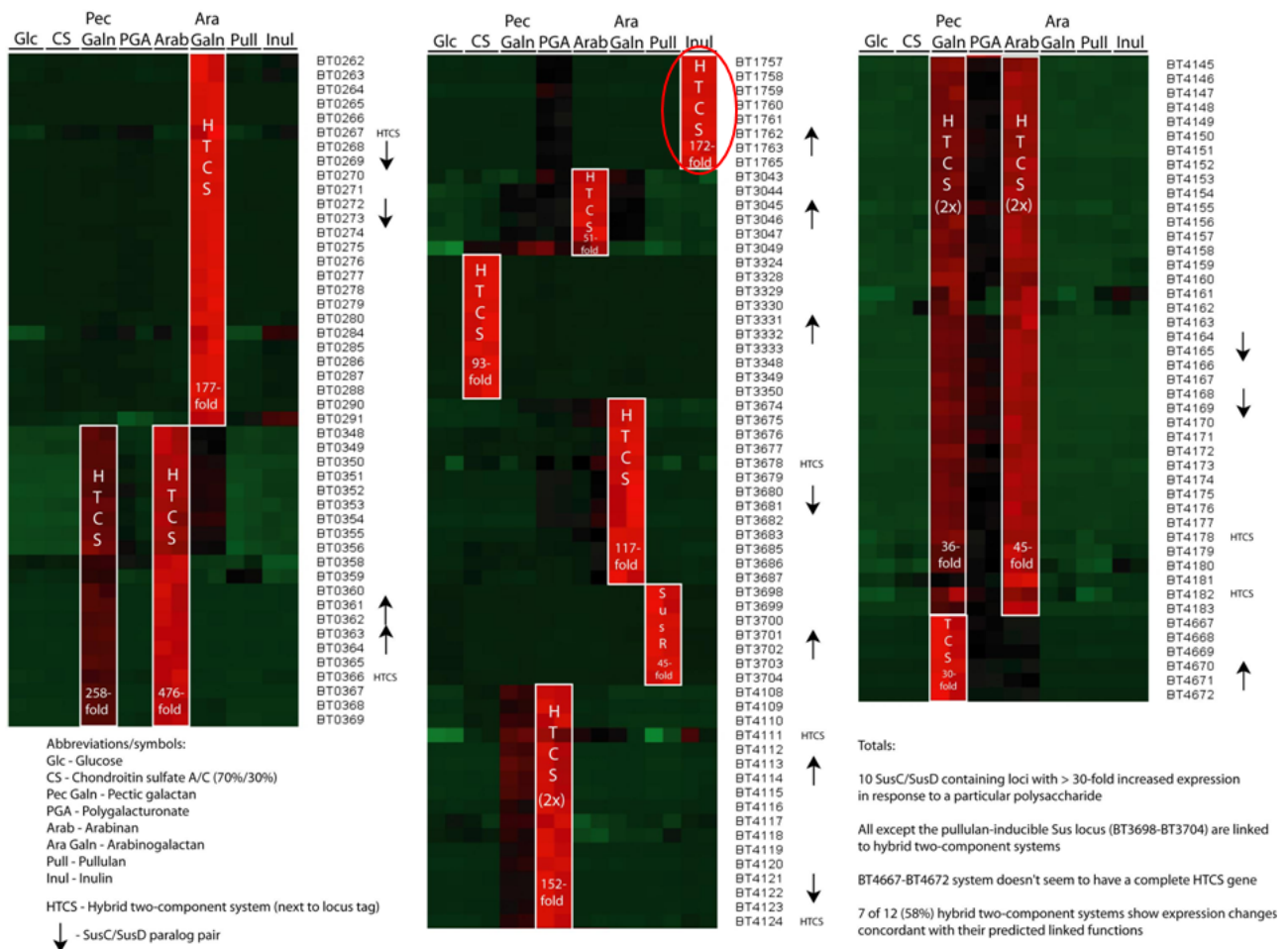


Figure 3. 1 Polysaccharide utilisation loci of *B. thetaiotaomicron*.

Heat map of unique upregulated loci containing SusC/D gene pairs linked to glycoside hydrolase and lyase genes. Cells were grown in minimal media + polysaccharide and harvested mid-log phase for GeneChip analysis. Red indicates higher levels and green lower levels of gene expression compared with the mean value for all conditions. The experiments were carried out by Dr Eric Martens (Centre for Genome Sciences, Washington University, St Louis, USA). The PUL outlined in red circle is the unique locus activated in the presence of the fructan inulin. The numbers down the right hand side are the locus tags of the genes.

The closest regulatory protein in almost all GeneChip analysis (except pullulan) is an Hybrid two-component system (HTCS). *B. thetaiotaomicron* has 32 hybrid two-component systems (HTCS) that incorporate all parts of classical two-component systems into one peptide. Most have a conserved domain architecture with a very large predicted N-terminal periplasmic domain made up of 10 or more Regulatory-propeller repeats (likely form a beta-propeller fold) (Mascher *et al.*, 2006) (Table 3. 1). However, one of the HTCS (BT1754; adjacent to the inulin activated locus, Table 3. 1 and Figure 3. 2) has a shorter periplasmic domain that displays homology to periplasmic binding proteins. Periplasmic binding proteins are components of ABC transport systems that bind small molecules like amino acids and sugars (see section 1.2.3.2.1).




Locus tag	HTCS domain architecture	Polysaccharide
BT1754		Inulin
BT4663		Heparin
BT3049		Arabinan

Table 3. 1 HTCSs of *B. thetaiotaomicron*.

The domain structures of 3 different *Bacteroides* HTCS as predicted by SMART are shown. The polysaccharide that activated the PUL to which each is closest is also shown. The red bar at the end of N-terminal is an uncleaved signal peptide, followed by either a PBPb periplasmic sensor domain or several β -propeller repeats. The blue bar indicates a transmembrane region (in BT3049 this is a pink box representing a region of low complexity as the prediction is not as strong), and then the cytoplasmic domains that make up the phosphorelay; HisKA (His Kinase A phosphoacceptor domain), HATPase (Histidine kinase-like ATPase domain), REC (cheY-homologous phosphoreceiver domain), and HTH_AraC, (helix_turn_helix, arabinose operon control DNA binding domain).

The unique locus upregulated in cells grown with inulin as carbon source comprises nine genes with the following predicted functions: a fructokinase, an inner membrane sugar transporter, a SusC homologue, a SusD homologue, four GH32 proteins, and a hypothetical protein of unknown function (Figure 3. 2) (Dr Eric Martens, unpublished data).

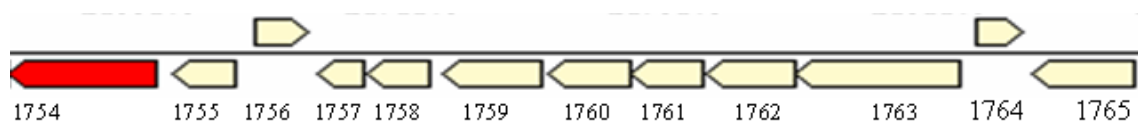


Figure 3. 2 The inulin utilisation locus regulated by BT1754.

The genes activated during growth on inulin are BT1757 (fructokinase), BT1758 (inner membrane fructose transporter), BT1759 (GH32), BT1760 (GH32), BT1761 (hypothetical protein of unknown function), BT1762 (SusD homologue), BT1763 (SusC homologue), and BT1765 (GH32). Also upregulated, but not in the locus, was the only other GH32 in *B. thetaiotaomicron*, BT3082. The arrows point in the direction of transcription.

The prediction that the periplasmic domain of BT1754 was a PBP, a class of proteins known to bind small molecules including carbohydrates, and the fact that fructans are relatively simple homopolymers initiates the study to identify the activating signal molecule and further our understanding of signal perception and transduction in these important prokaryotic regulatory systems.

3.2 Results

3.2.1 Cloning of the predicted periplasmic domain of BT1754

SMART domain prediction of the BT1754 HTCS revealed an N-terminal domain of approximately 300 amino acids flanked by two transmembrane (TM) regions (Table 3. 1 and Figure 3. 3). The C-terminus of the protein after the second TM domain contains all the domains required for phosphorelay and signal output indicating that this region is cytoplasmic and thus that the N-terminal domain is located in the periplasm (Figure 3. 4). To test this hypothesis the predicted

periplasmic region of BT1754 (residues 29-343, hereafter known as BT1754peri) was amplified by PCR from *B. thetaiotaomicron* genomic DNA using the following primers: BT1754-1: 5'-CTCCCATGGATGATACACCCCATTTTCGTATTG-3' and BT1754-2: 5'-CCGCTCGAGGACCTGTTGTGTAGCTAC-3'. The amplified DNA fragment was digested with Nco and Xho (unique sites introduced at the 5' end of each primer) and cloned into similarly digested pET22b (Novagen; encodes an N-terminal pelB leader that targets the protein to the *E. coli* periplasm and a C-terminal His₆ tag).

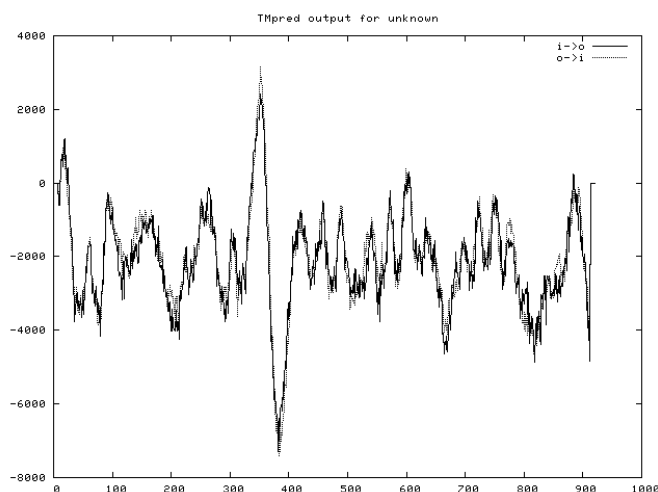


Figure 3.3 The prediction of transmembrane regions and orientation of BT1754.

TMpred was used to confirm the SMART prediction of the transmembrane regions in BT1754. Anything above zero is on the scale is considered to be a possible TM. The first peak (aa 1-28) indicates the uncleaved signal peptide (TM1). The second TM that flanks the periplasmic domain is indicated by the large peak at ~350 aa.

MMRWKTMRYLKWMLVLFGLIGMTACRQDTPHFRIGVAQCSDDSWRHKMNDEILREAMFY
 NGVSVEIRSAGDDNSKQAEVDVHYFMDEGVLLIISANEAAPMTPIVEEAYQKGIPVILVD
 RKILSDKYTAYIGADNYEIGRSVGNIIASSLKGGKGNIVELTGLSGSTPAMERHQGFMAAI
 SKFPDIKLIKADA AWERGPAEIE MDSMLRRHPKIDAVYAHNDRIAPGAYQAAKMAGREK
 EMIFVGIDALPGKGNLELVDSVLDATFIYPTNGDKVLQLAMDILEKKYPKETVMNTA
 VVDRTNAHVMQLQTTTHISELDKKIETLNGRIGGYLSQVATQQVVLYGSLIILLLVAGLLL
VVYKSLRSKNRLNKELFKQKQLEEQRDKLEEQRDQLIQLSHQLEEATHAKLVFFTNISH
DFRTPLTLVADPVEHLLADKTLSGDQHRMLMLIQRNVNILLRLVNLQILDFRKYENGKMEY
 TPVTVDLSSFEQWNESEFQAARKKHIHFSFDSMPD TDYHTLADMEKLERIYFNLLSNAF
 KFTPENKIAIRLSSLSKEDKRWIRFTVANTGSMISA EHIRNVDFRFYKIDMHHTGSGIG
 LALVKAFVEMHGGMISVESDEKQGTVFTVELPVQSC EAVAAEPD TTVLSADSR TTDVLLA
 EEEEELEKGYDSSKPSVLIIDDNEDIRSYVHTLLHTDYTVIEAADGSEGIRKAMKYVPDLI
 ISDVMPGIDGIECCRRLKSELQTCHIPVILLTACSLDEQRIQGYDGGADSYISKPFSSQ
 LLLARVRNLI DSHRRLKQFFGDGQTLAKEDVCDMDKDFVERFKSLIEEKMGDSGLNVEDL
 GKDMGLSRVQLYRKIKSLTNYSPELLRIARLKKAASLLASSDMTVAEIGYEVGFSSPSY
 FAKCYKEQFGESPTDFLKRKG

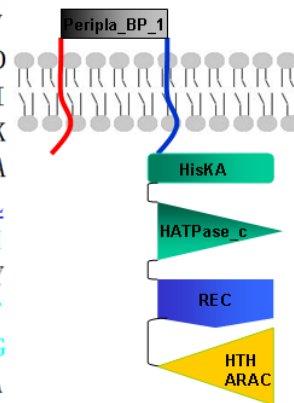


Figure 3.4 BT1754 amino acid sequence.

This fused protein has 921 amino acids, the molecular weight is 103821.24 Dalton, and the theoretical pI is 5.82. The domain prediction is made by SMART (see Table 3. 1). The first transmembrane domain (residues: 1-28) is colored in red with underline, while the second one (residues: 343-362) is blue with underline. The periplasmic domain of BT1754 is colored in black and bolded (29-343), which is used for cloning. The HisKA means the Histidine kinase (cyan square, residues: 410-476), while the HATPase_c means the histidine kinase-like ATPase (cyan triangle, residues: 523-635). The last orange triangle presents the helix_turn_helix, arabinose operon control protein (HTH-ARAC, residues: 834-917). The links of each domain are presented as black. The double-lipid layer (gray) presents the inner membrane.

3.2.1.1 Expression and purification of BT1754peri

BT1754peri was over expressed in *E. coli* C41 by induction with 1mM IPTG. The soluble CFE fraction was purified by IMAC method (Figure 3. 5). The eluate from the column was then dialysed into 20 mM Tris-HCL, pH 8.0 buffer and stored at 4 °C until required.

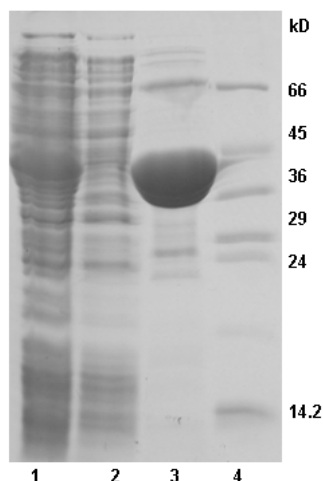


Figure 3. 5 Purification of BT1754peri.

Lane 1: CFE, Lane 2: Flow through, Lane 3: Eluate, Lane 4: Marker. The calculated MW of the protein expressed from pET22b is 36.1 kD, although this runs at ~43 kD on the SDS-PAGE gel. The pI is 5.5.

3.2.1.2 Carbohydrate binding studies with BT1754peri

BT1754 is involved in the control of inulin degradation and therefore the most likely signalling molecule this protein could recognize is inulin or one of its enzymatic breakdown products. Initial binding tests were carried out with BT1754peri against inulin and commercially available β -2,1 linked fructooligosaccharides (sucrose, kestose, kestotetraose, and kestopentaose), but no binding was observed by ITC (Figure 3. 6).

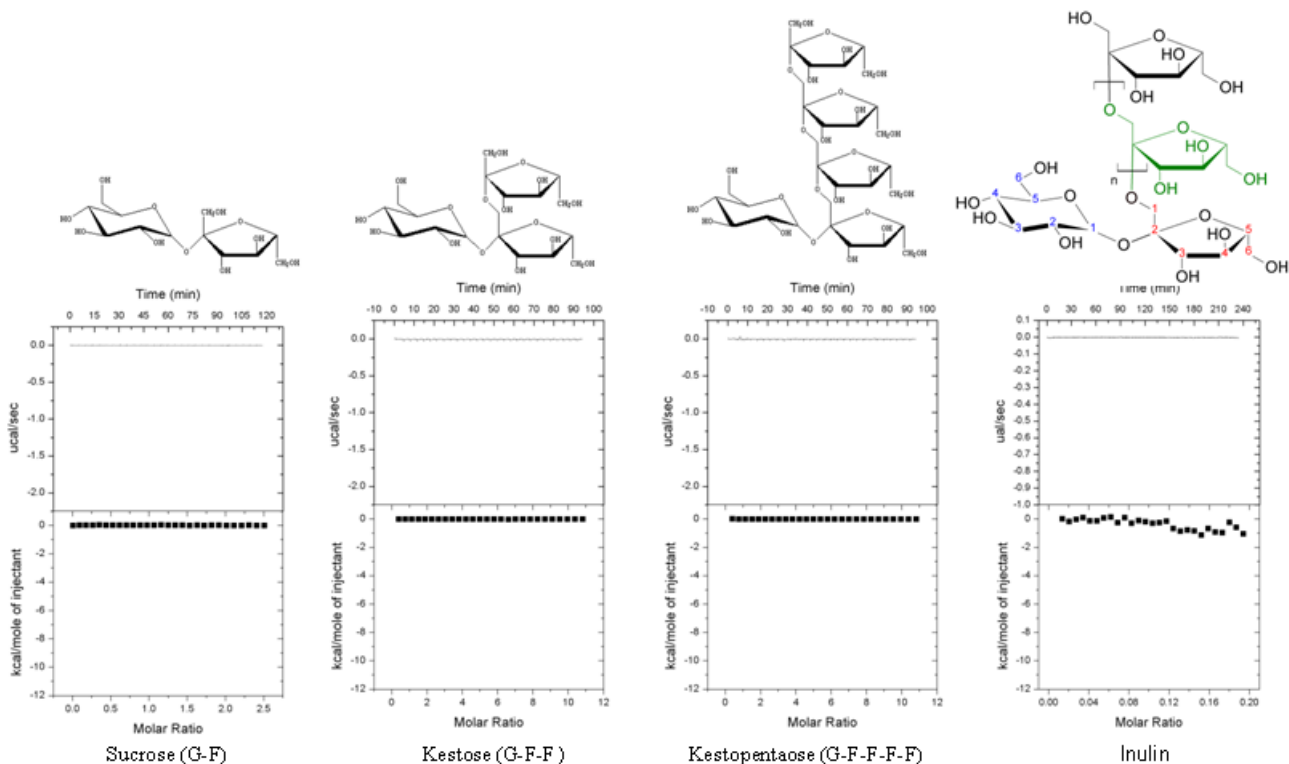


Figure 3. 6 BT1754peri does not bind inulin or fructooligosaccharides with a terminal glucose.

Sucrose, kestose (K3), and kestopentaose (K5) as oligosaccharides of inulin, none of them could bind BT1754peri. Kestotetraose (K4, G-F-F-F) was presumed that could not bind BT1754peri, because K3 and K5 could not, and K4 is longer than K3 and shorter than K5.

As the major component of fructans is β -D-fructose this monosaccharide was tested for binding and was shown to interact with BT1754peri with a K_d of $\sim 0.5 \mu\text{M}$ and a stoichiometry of 1:1. This binding event was metal independent as the affinity was unchanged in the presence of 5 mM EDTA (Figure 3. 7 and Table 3. 2). The reaction is driven by a favourable change in enthalpy ($-\Delta H$) with an unfavourable entropic contribution ($-T\Delta S$). This thermodynamic signature is common to sugar binding proteins and is thought to arise due to the desolvation of the hydroxyls and formation of new hydrogen bonds with the protein.

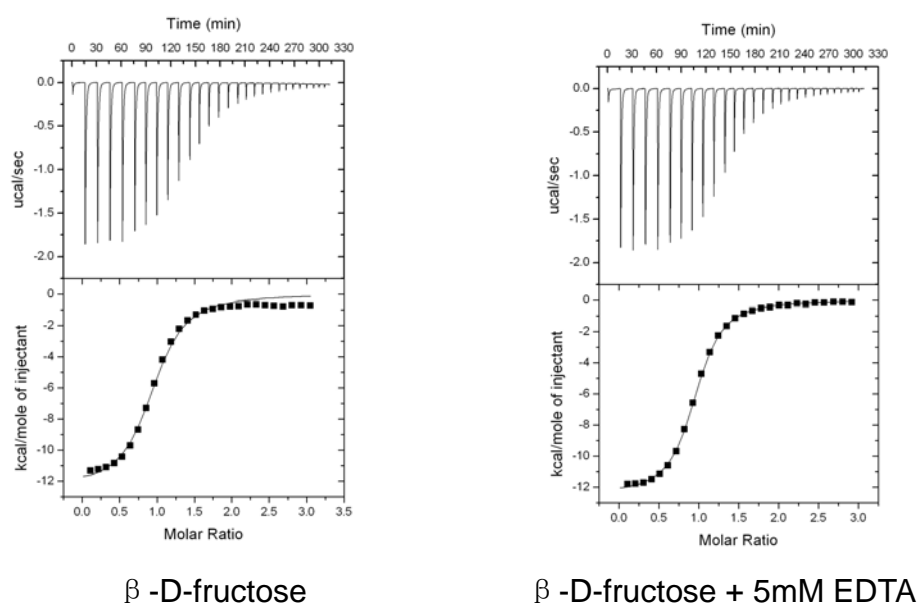


Figure 3. 7 BT1754peri binds fructose.
BT1754peri bound fructose is metal independent.

It is possible that the terminal Glc present on inulin or the commercially available kestooligosaccharides interferes with binding to BT1754peri and therefore binding was assessed against fructooligosaccharides produced by partial acid hydrolysis of inulin and separation of the mixture on a P2 bio-gel size exclusion column. No binding was observed with higher MW fractions from fractions 9-25, but fractions from 27 onwards were shown to interact with BT1754peri. As these fractions all contain monomeric fructose and the ITC data looks identical to fructose alone when n is set to 1 in the fits, it is most likely that the ligand BT1754peri is recognising in the hydrolysed fractions is fructose alone rather than any of the oligosaccharides present (Figure 3. 8).

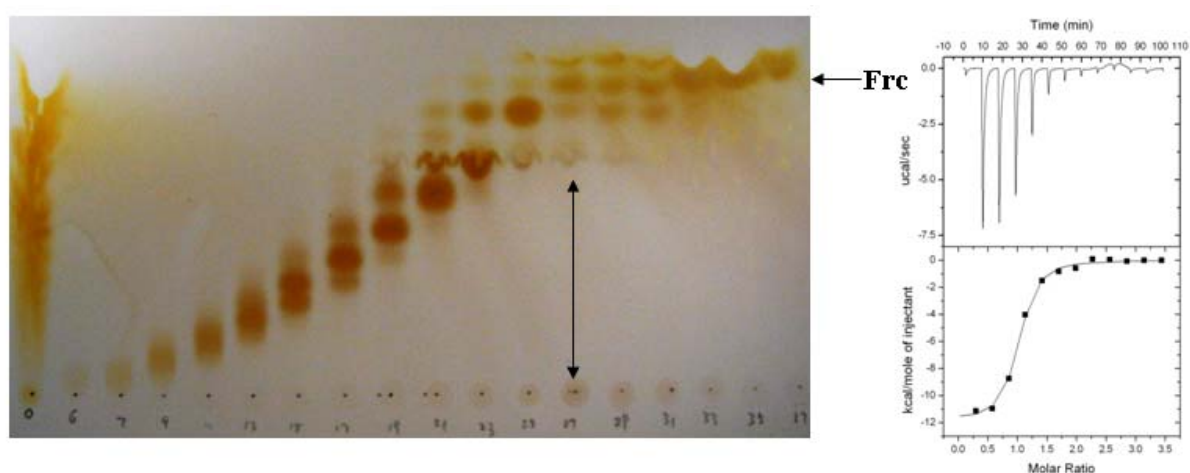


Figure 3. 8 TLC showing purification of inulin oligosaccharides produced by partial acid hydrolysis. Total hydrolysate was applied to a Bio-gel P2 size exclusion column in water and fractions collected. These were initially assayed for reducing sugar using DNSA reagent before analysis of positive fractions by TLC. **Left:** Lane 0 is the inulin after digestion and before isolation. Lanes 6-38 are the collection tube numbers. Each tube has 4 ml collection. Samples (2 μ l) from each tube were applied to the TLC plate for separation. Inulin fraction 27[#] is same size as fructose from the TLC result though a bit shift because of side-effect. **Right:** ITC of BT1754peri with fraction 27[#] at 1 mg/ml. As the molar concentration of ligand is unknown, the n -value was iteratively fitted to as close as possible to one, by adjusting the molar concentration of the ligand.

To further examine the specificity of BT1754peri, interaction of this protein with a range of small fructose containing sugars, fructose epimers and other monosaccharides was assessed by ITC (Figure 3. 9). Only sorbose showed any binding to BT1754peri, but the affinity was too low to fit accurately to a single site model in Origin (ITC data not shown). These data reveal that the periplasmic domain of BT1754 bound specifically to β -D-fructose, but not fructose containing disaccharides, oligosaccharides, polysaccharides, or analogues of fructose.

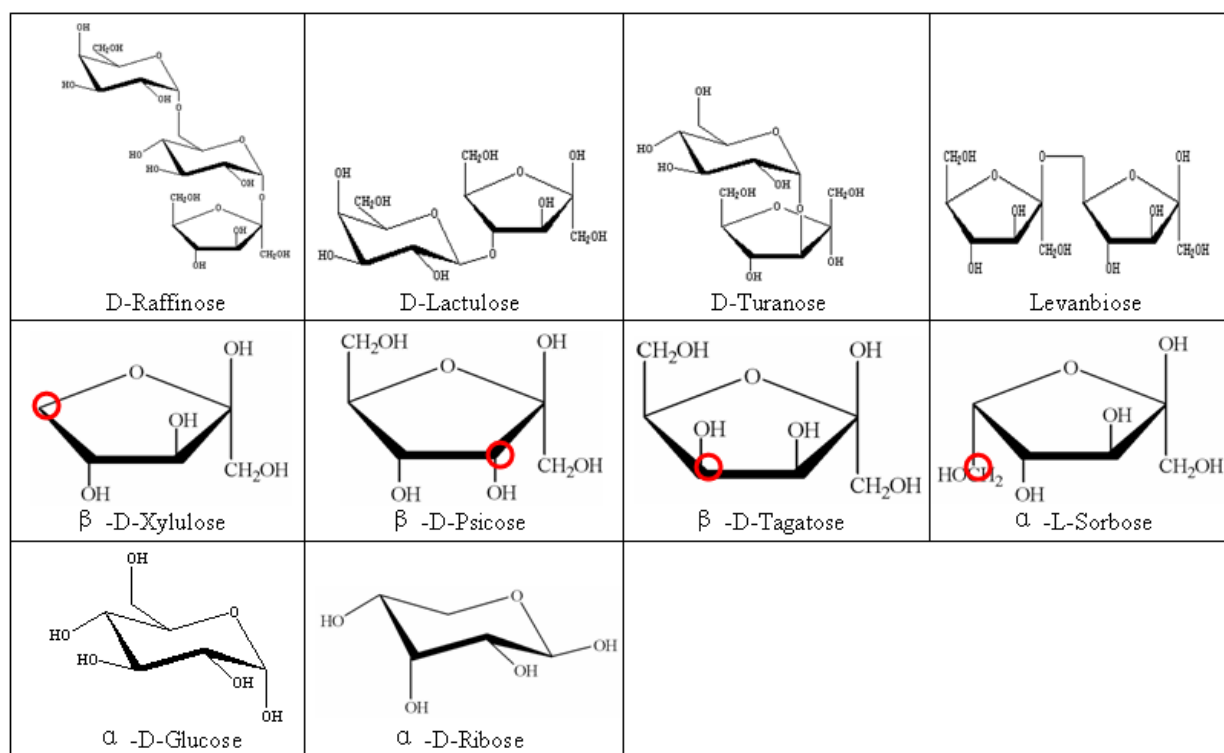


Figure 3. 9 Structures of sugars tested for binding to BT1754peri.

The top panel: Raffinose is α -D-galactose-(1 \rightarrow 6) α -D-glucose-(1 \rightarrow 2) β -D-fructose, Lactulose is β -D-galactose-(1 \rightarrow 4) β -D-fructose, Turanose is α -D-glucose-(1 \rightarrow 3)- α -D-fructose, and levanbiose is β -D-fructose-[(2 \rightarrow 6)- β -D-fructose], the disaccharide of levan without the glucose at the end. The middle panel: fructose analogs: compared with fructose, xylulose lost the C6, while psicose, tagatose, and sorbose are epimers of fructose, different on C3, C4, and C6 separately. The bottom panel: glucose is the end unit of fructans; ribose is a five carbon sugar. Only sorbose could bind fructose, albeit very weakly.

Ligand	$K_a \times 10^5 (M^{-1})$	$\Delta H (kcal mol^{-1})$	$T\Delta S (kcal mol^{-1})$	n^*
Fructose	2.6 ± 0.2	-13.9 ± 0.1	-6.5 ± 0.5	1.0 ± 0.0
Inulin-27[#]	3.4 ± 0.6	-11.9 ± 0.3	-4.3	0.9 ± 0.0
Frc + EDTA	4.0 ± 0.1	-12.5 ± 0.4	-4.8	0.9 ± 0.0

Table 3. 2 ITC data show the binding ability of BT1754peri against a group of ligands.

*: The ITC data were fitted to a single site binding model for all ligands. ITC was carried out in 10 mM Tris/HCl pH 8.0 at 25 °C. BT1754peri was 0.05 to 0.1 mM. Monosaccharides, glucose, xylulose, fructose, psicose, tagatose, and sorbose, were with a molar concentration of 1 mM respectively. Oligosaccharides: raffinose, lactulose, turanose, sucrose, kestose, kestopentaose, and levanbiose, were with a molar concentration of 1 mM, 1 mM, 1 mM, 1 mM, 5 mM, 5 mM, and 0.5 mM, respectively. Inulin fractions were at 1 mg/ml.

3.2.1.3 Biophysical studies on BT1754peri

3.2.1.3.1 BT1754peri oligomeric state in the presence and absence of fructose.

To assess the oligomeric state of BT1754peri in the presence and absence of fructose, the likely activating molecule, gel filtration on a Sephadex S200 column was carried out. Fructose was added to the protein sample as well as gel filtration buffer at a concentration of 5 mM. The protein ran as a monomer in the absence and in the presence of ligand (Figure 3. 10).

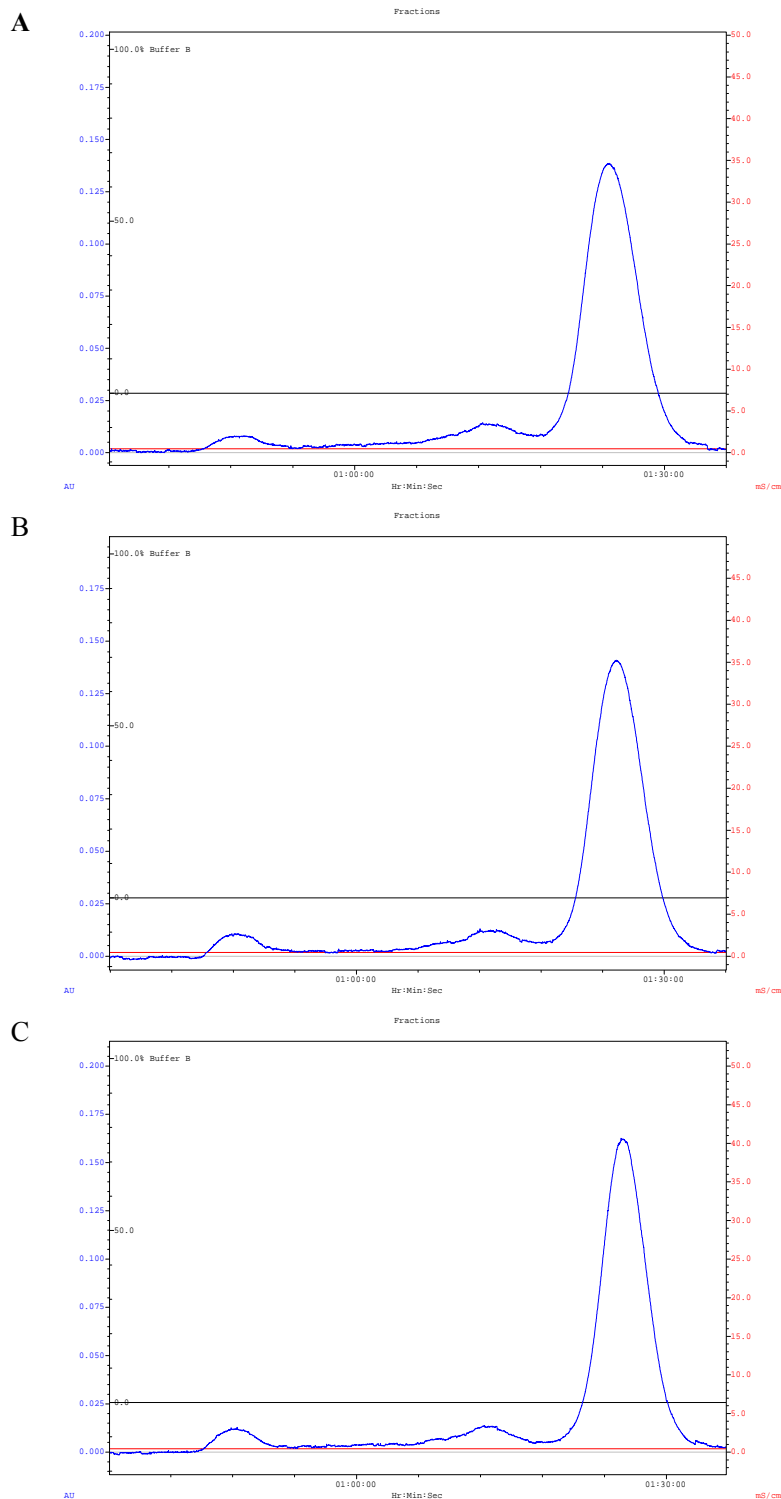


Figure 3. 10 BT1754peri is a monomer in the absence and presence of ligand as assessed by gel filtration.

BT1754peri eluted from gel filtration column at the same time point while the conditions were different. **A:** 0.115 mM BT1754peri in 10 mM Tris run in normal gel filtration buffer. **B:** 0.115 mM BT1754peri in 10 mM Tris buffer plus 1.6 mM fructose run in normal gel filtration buffer. **C:** 0.115 mM BT1754peri in 10 mM Tris buffer plus 1.6 mM fructose run in gel filtration buffer plus 5 mM fructose.

To present the possible oligomeric difference with presence and absence of fructose, the crosslinking analysis (section 2.37) was carried out. The results indicated that in both conditions, BT1754peri is a monomer, which supports the gel filtration results (Figure 3. 11).

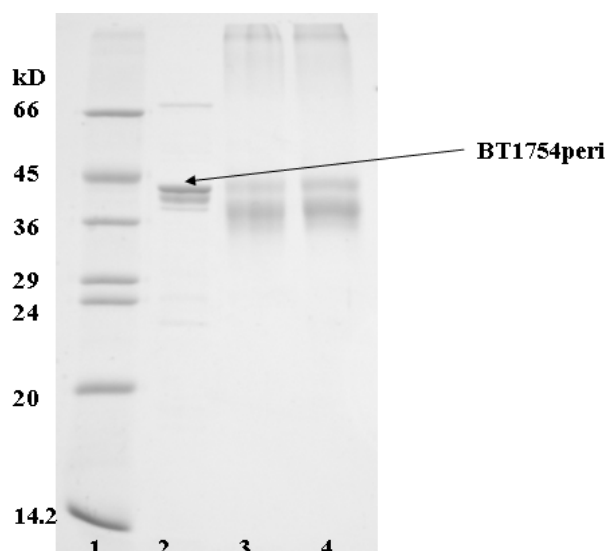


Figure 3. 11 Crosslinking analysis with BT1754peri in the presence and absence of fructose.

SDS-PAGE gel of crosslinked BT1754peri. Lane 1: low molecular weight marker. Lane 2: BT1754peri before crosslinking treatment (see arrow). This protein is slowly degraded from elution, so there are a few minor bands smaller than BT1754peri. Lane 3: Crosslinking treated BT1754peri in absence of fructose. Lane 4: Crosslinking treated BT1754peri with presence of 5 mM fructose. Lanes 3 and 4 have no difference, which indicates that in both conditions, BT1754peri is a monomer.

3.2.1.3.2 BT1754peri is thermally stabilized in the presence of fructose

Differential scanning calorimetry (DSC) was carried out to check the thermal stability of BT1754peri in the presence and absence of fructose. The data indicate that the thermal stability of BT1754peri increased in the presence of fructose, while glucose did not stabilize the protein (Figure 3. 12).

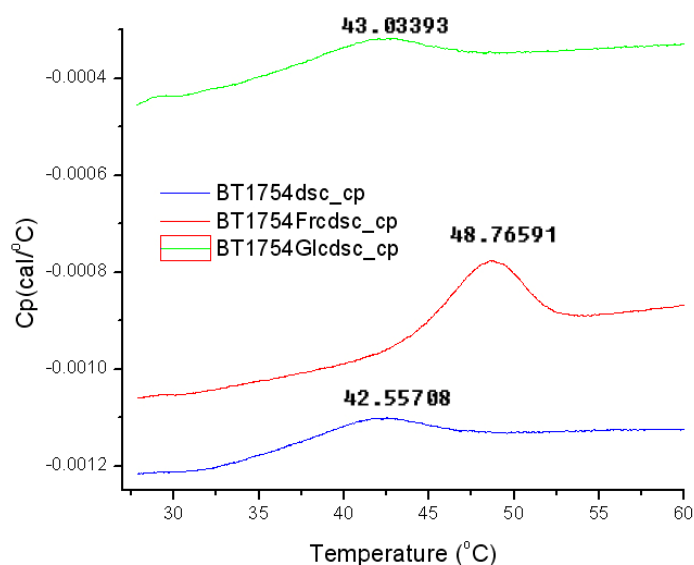


Figure 3. 12 DSC data showing the melt peaks of BT1754peri.

The 0.5 mg/ml protein was in 20 mM Tris pH 8.0 buffer. Ligand concentrations were 5 mM. The BT1754peri with the absence of fructose half unfolded at 42.55708 °C, while in the presence of fructose, it increased the half unfolding temperature to 48.76591 °C, and BT1754peri in the presence of glucose half unfolded at 43.03393 °C, which is almost no donation to the change of thermal stability.

3.2.1.3.3 BT1754peri is resistant to protease digestion in the presence of fructose

To further assess the ability of fructose to stabilize the structure of BT1754peri, the protein was incubated with protease K in the presence and absence of the monosaccharide. The data show that BT1754peri is rapidly degraded to a stable ~31 kD fragment in the presence and absence of fructose, which remains in the former sample after 19 h digestion (Figure 3. 13). However, after 19 h in the absence of fructose the protein has been almost totally digested. These results show that fructose protects BT1754peri from digestion by protease K.

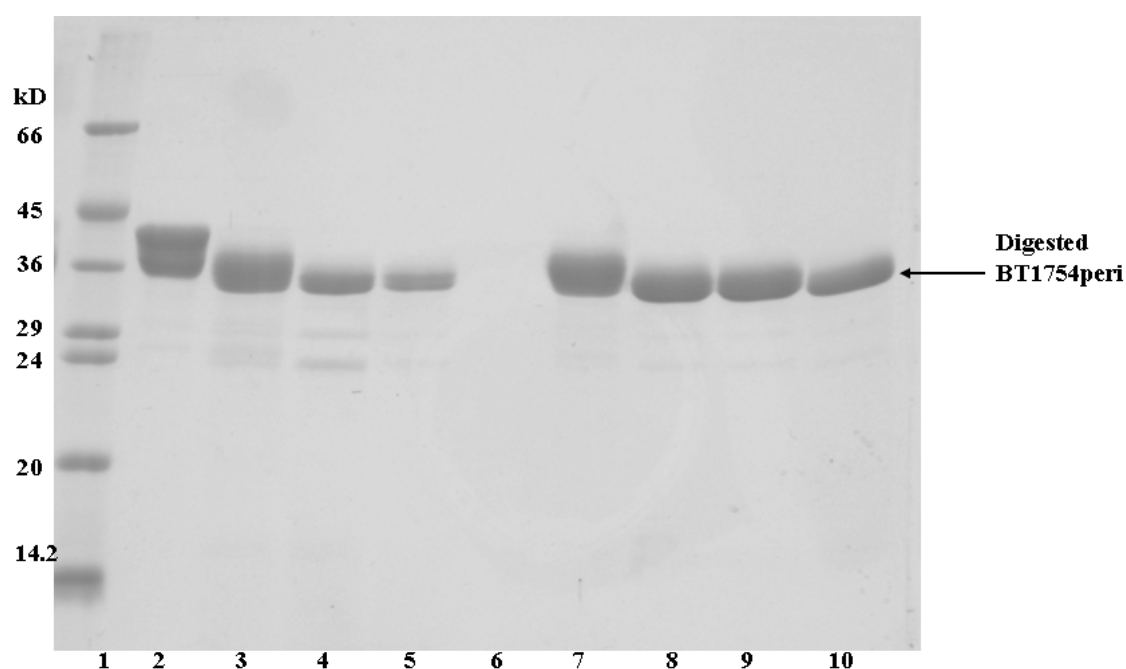


Figure 3. 13 BT1754peri protease K digestion.

160 μ l BT1754peri (1.2 mg/ml) was digested with 10 μ l 0.01 mg/ml protease K at different time points with or without 5 mM fructose. Lane 1: Marker, Lane 2: BT1754peri only, Lanes 3-6: BT1754peri digested by protease K for 10 min, 1 h, 2 h, and 19 h respectively, Lanes 7-10: BT1754peri plus 5mM fructose digested by protease K for 10 min, 1 h, 2 h, and 19 h respectively. The arrow shows the position of the stable band that was subjected to MALDI-TOF and N-terminal sequencing.

The stable band at ~31 kD after the overnight digestion was blotted for MALDI-TOF analysis and for N-terminal sequencing to identify the stable core structure. The MALDI-TOF shows that the remaining fragment has a molecular weight of 31944.12 Da. The N-terminal sequencing shows that this fragment starts from the 41st amino acid, serine, so the full length of this fragment could be calculated. It's from amino acid 41 to 329 (Figure 3. 14).

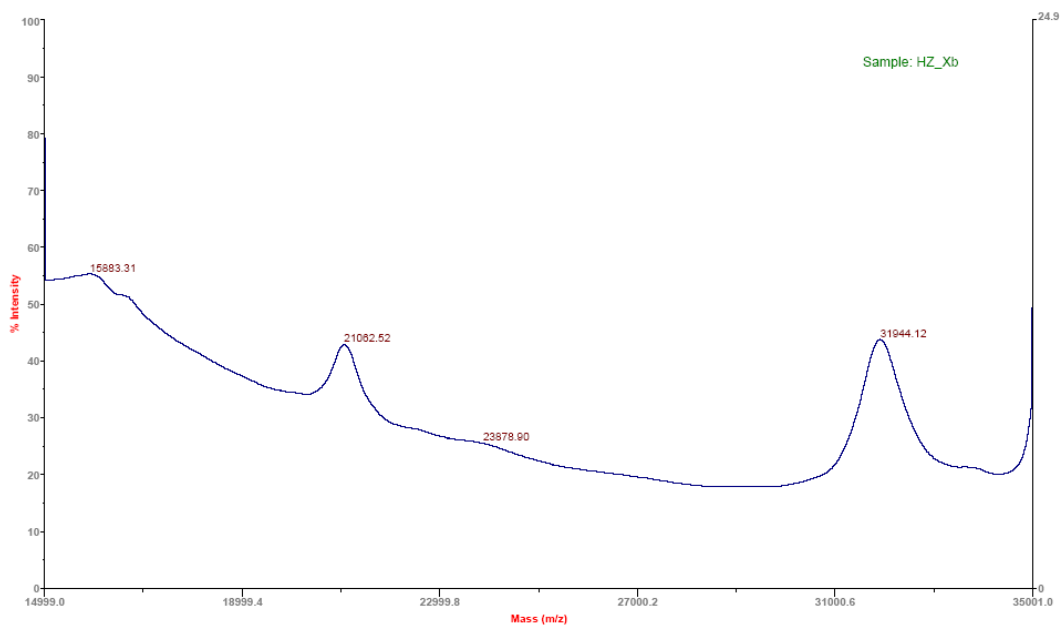


Figure 3. 14 The MALDI-TOF map of digested fragment of BT1754.
This protease K digested fragment of BT1754peri was 31944.12 Dalton.

3.2.2 BT1754peri crystallization

BT1754peri was purified by IMAC and then by ion exchange method (Q12). The purity is higher than 95 % according to the bands on SDS-PAGE gel (Figure 3. 15). Final BT1754peri concentration was 10 mg/ml in water. Four screens (PACT, JCSG, Classical, NCL, see Appendix B) were used with or without the accompaniment of 5 mM DTT or 5 mM fructose for sitting-drop crystallization. Tiny crystals formed in condition: 0.1 M HEPES sodium salt, pH 7.5, 0.8 M Sodium phosphate, 0.8 M Potassium phosphate and condition: 0.5 M Ammonium sulfate, 0.1 M tri-Sodium citrate, pH 5.6, 1.0 M Lithium sulfate. Further modification for hanging-drop crystallization was based on these two conditions. A single crystal formed in screen condition 0.1 M HEPES sodium salt pH 7.5, 0.8 M Sodium phosphate, and 0.8 M Potassium phosphate 15 days after setting up (Figure 3. 15).

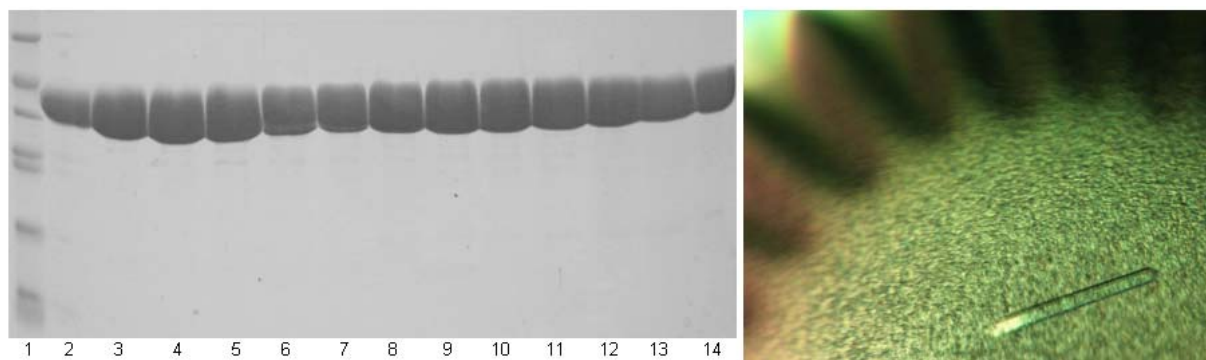


Figure 3. 15 BT1754peri for crystallization.

Left: Lane 1: Marker, Lanes 2-14: BT1754 fractions collected from ion exchange column Q12. **Right:** Crystal of BT1754peri fused in pET22b with a C-His tag formed in Classical-46.

This crystal was shot using the synchrotron (DIAMOND) at a resolution of 2.9 Å by Dr. Susan Firbank. Molecular replacement was used by Dr Firbank to solve the structure based on the structure of the most similar homologue, ribose binding protein from *Thermoanaerobacter tengcongensis* (PDB ID: 2IOY) which displays ~35 % sequence similarity (Figure 3. 16) (Cuneo, 2008).

```

RBP          K---TIGLVISTLNNPFFVTLKNG-AEEKAKELGYKIIVEDSQNDSSKELSNVEDLIQOK
1754         TPHFRIGVAQCSD-DSWRHKMNDEILREAMFYNGVSVEIRSAGDDNSKQAEDVHYFMDEG
          .      **:. .: .:: .::: .*      * .: .::: :*.**:. :*. .:::

RBP          VDVLLINPVSDAVVTAIKEANSKNIPVITIDRSANGGDVVCHIASDNVKGEMAAEFIA
1754         VLLIISANEAAPMTPIVEEAYQKGIPVILVDRKILSDKYTAYIGADNYEIGRSVGNIA
          **:*:*.. .: .:.. :** .*.***. :*. . . . :*:*:* : * . .:***

RBP          KALKGKGNVVELEGIPGASAARDRGKGFDEAIAKYPDIKIVAKQAADFDRSKGLSVMENI
1754         SSLKGGKGNIVELTGLSGSTPAMERHQGFMAAISKFPDIKLIKADAAWERGPAEIEMDSM
          .:*****:*** *: .*:.* :* :** **:*:*:*:*:* : * * :*: . . *:.:

RBP          LQAQPKIDAVFAQNDEMALGAIKAIEAANRQ-GIIIVVGF-----GTEDALKAIKEGKMAAT
1754         LRRHPKIDAVYAHNDRIAPGAYQAAKMAGREKEMIFVGIDALPGKGNLGLVLDVLDAT
          *: :*****:*:*:*.* ** :* : *.*: :*.***:* * . .:*: : . . : **

RBP          IAQQPALMGSLGVEMADKYLKGEKIPNFI PAELKLITKE-----
1754         FIYP--TNGDKVLQLAMDILEKKPYPKETVMNTAVVDRTNAHVMQLQTTTHISELDKKIET
          :      * . :*: . * : : * : : : : :

RBP          -----NVQ
1754         LNRRIGGYLSQVATQQ
          .      *

```

Figure 3. 16 The ClustalW alignment of BT1754peri with *T. tengcongensis* ribose binding protein (RBP). The amino acid sequences similarity is 35 %.

3.2.3 BT1754peri fold and ligand binding site

BT1754peri structure has been solved as a dimer at a maximal resolution of 2.9 Å. The asymmetric unit is a dimer and has cell parameters $a = 111.8 \text{ \AA}$, $b = 111.8 \text{ \AA}$, $c = 115.0 \text{ \AA}$, and $\beta = 90^\circ$, containing 4688 atoms. The visible density of residues starts from 31 (P) to 332 (G) which means two residues at the N-terminal and 11 at the C-terminal have no visible density. The N- and C-termini are on the same side, as required for insertion into the inner membrane. One molecule of BT1754peri has 10 α -helices and 12 β -sheets formed into two distinct domains and a long C-terminal helix (α_{10}) (Figure 3.17). The N-terminal domain has 4 α -helices around 6 β -sheets, while the C-terminal domain has 5 α -helices around 6 β -sheets. Interestingly, a symmetry related molecule forms a dimer via the C-terminal helical extension (Figure 3.17). A single molecule of fructose in the furanose form is bound in the middle of the protein, sandwiched between the N and C terminal domains (Figure 3.17).

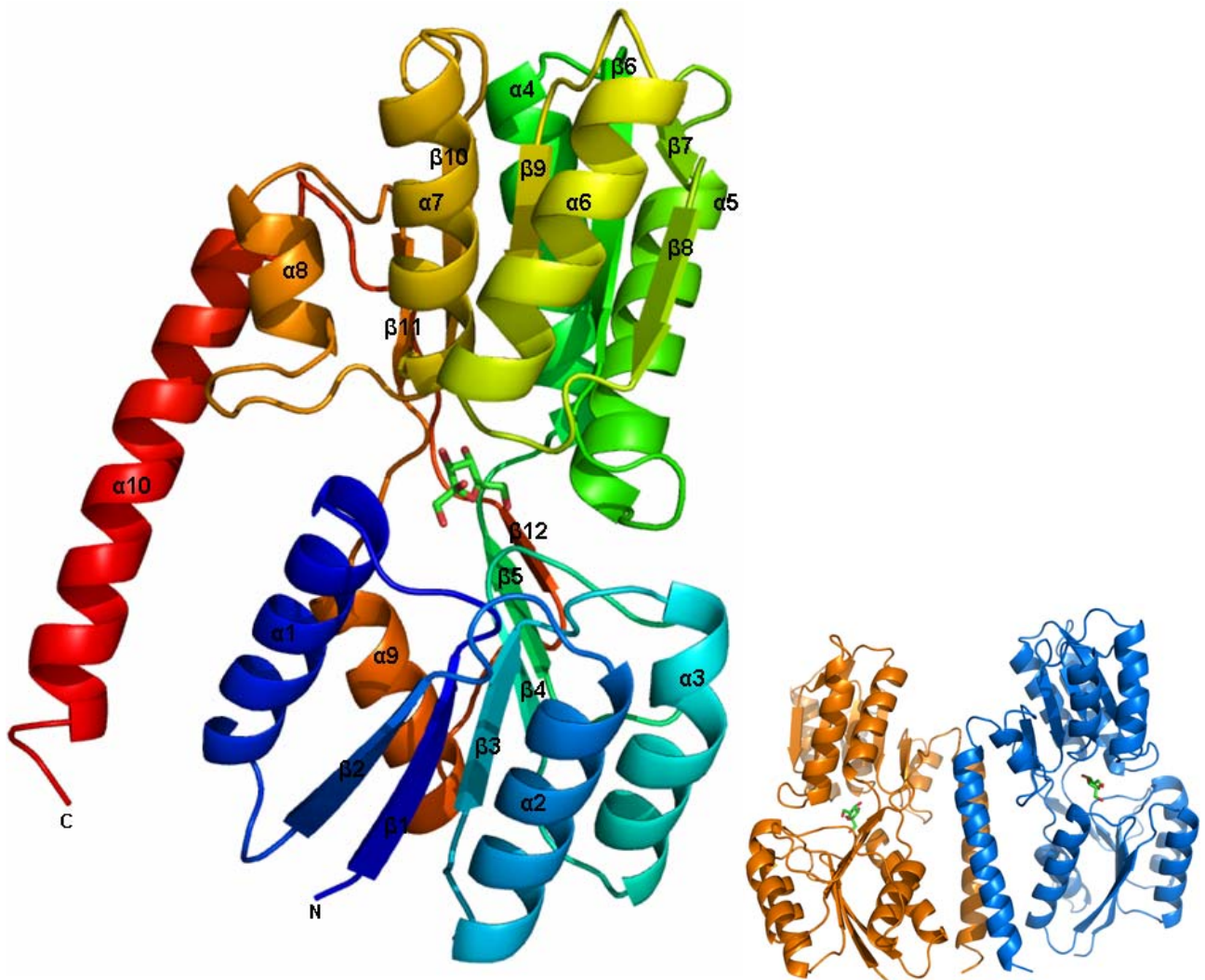


Figure 3.17 BT1754peri structure.

Left: This cartoon representation shows the protein N to C-terminal from blue to red. The two sub-domains form into a pocket that sandwiches one molecule of fructose. This protein has 10 α -helices and 12 β -sheets. The last α helix spans from one sub domain to another. The protein has a C-terminal His-tag, but the tag is not visible in the structure. **Right:** Two BT1754peri structures (orange and marine) are present in symmetry dimer form.

All the hydroxyls of fructose are involved in the hydrogen bond formation with 9 residues from both the N and C-terminal domains, especially O1, O4, and O6, which form 4 hydrogen bonds each. In addition two tryptophans (Trp45 and Trp196) from each domain, stack against each face of the sugar. Other two residues Pro168 and Tyr271 are close to one edge of the fructose and may be involved in hydrophobic interactions with the sugar (Figure 3. 18 and Table 3. 3).

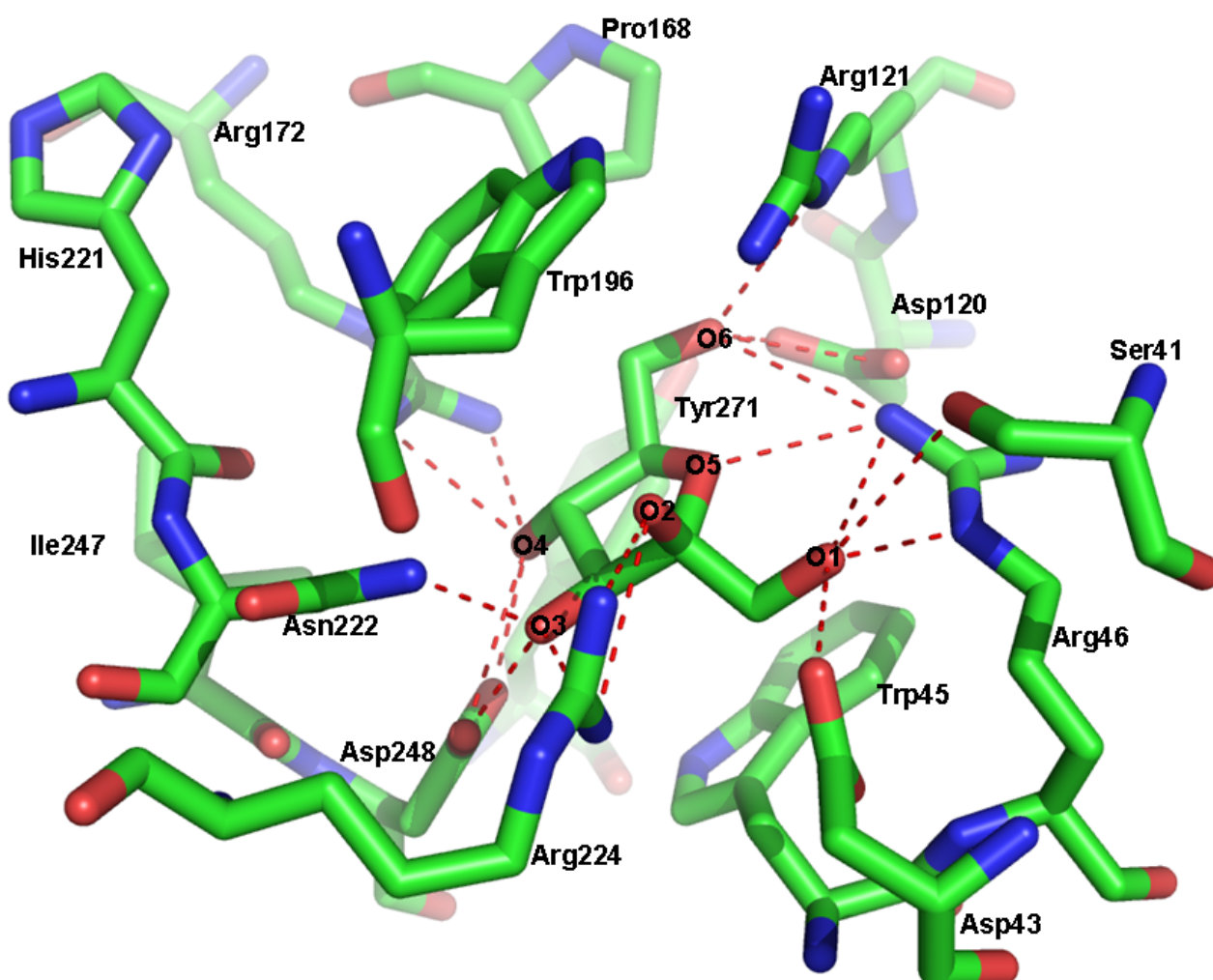


Figure 3. 18 BT1754peri binding site.

Bound fructose is shown in the binding pocket of BT1754peri. All oxygens of fructose donate to the hydrogen bond formation with residues of BT1754peri. Several -OH groups (O1, O4, and O6) make 4 Hydrogen bonds respectively. Totally, 18 hydrogen bonds are formed. The 9 residues bound fructose are Ser41, Asp43, Arg46, Asp120, Arg121, Arg172, Asn222, Arg224, and Asp248. Trp45, Pro168, Trp196, His221, Ile247, and Tyr271 are very close to fructose. Trp45 and Trp196 stack against opposite faces of the fructose molecule.

Fructose	BT1754peri	Distance (Å)
O1	Ser41-O	3.56
	Arg46-N1	3.43
	Arg46-N2	2.56
	Asp43-O	2.89
C1	Trp45-C	3.68
O2	Arg224-N1	3.04
	Arg224-N2	3.38
	Trp196-C	4.14
O3	Arg224-N2	2.84
	Asp248-O1	2.55
	Asn222-N	2.80
O4	Asp248-O1	3.35
	Asp248-O2	2.56
	Arg172-N1	3.15
	Arg172-N2	3.20
	Ile247-O	5.03
	Tyr271-C	3.49
	His221	4.69
O5	Arg46-N1	3.07
O6	Arg46-N1	3.35
	Asp120-O1	2.55
	Asp120-O2	3.19
	Arg121-N	3.14
	Pro168-C	3.57

Table 3. 3 Hydrogen bonding and close contacts between BT1754peri residues and fructose.
The hydrogen bound residues atoms are bolded. O1, O4, and O6 form four hydrogen bonds respectively.

DALI analysis was used for comparing protein structures in 3D. Figure 3. 19 shows the DALI results for BT1754peri and reveals that the most similar ten structures are periplasmic binding proteins that bind small sugars such as ribose (No. 1 and 2), glucose (No. 3 and 6), lactose (No. 4), galactose (No. 5 and 6), (2R, 4S)-2-methyl-2, 3, 3, 4-tetrahydroxytetrahydrofuran (R-THMF) (No. 7 and 8), and arabinose (No. 9).

No	Chain	Z	rmsd	lali	nres	%id	Description
1	2ioy-A	38.0	1.3	271	274	30	PERIPLASMIC SUGAR-BINDING PROTEIN
2	2fn8-A	34.9	1.7	273	292	24	RIBOSE ABC TRANSPORTER, PERIPLASMIC RIBOSE
3	2h3h-A	32.9	2.0	281	313	23	SUGAR ABC TRANSPORTER, PERIPLASMIC SUGAR-BINDING
4	3brs-A	32.1	1.6	261	271	19	PERIPLASMIC BINDING PROTEIN/LACI TRANSCRIPTIONAL
5	2rjo-A	31.6	2.3	280	322	24	TWIN-ARGININE TRANSLOCATION PATHWAY SIGNAL
6	1gca	31.3	2.2	275	309	25	GLUCOSE/GALACTOSE-BINDING PROTEIN WITH GALACTOSE
7	1tjy-A	29.5	2.0	274	316	17	SUGAR TRANSPORT PROTEIN
8	3ejw-A	29.3	2.1	275	315	17	SMLSRB
9	1abe	28.5	2.6	2.74	305	16	L-ARABINOSE-BINDING PROTEIN COMPLEX WITH L-ARABINOSE
10	3d02-A	28.5	2.3	273	300	15	PUTATIVE LACI-TYPE TRANSCRIPTIONAL REGULATOR

Figure 3. 19 DALI analysis of BT1754peri.

Dali matches the top 10 structures sorted by Z-score. Chain: PDB entry code plus chain identifier. Z: normalized Z-score that depends on the size of the structures. The program optimises a weighted sum of similarities of intramolecular distances. rmsd: root-mean-square deviation of C-alpha atoms in the least-squares superimposition of the structurally equivalent C-alpha atoms. nres: number of amino acids in the protein. %id: percentage of identical amino acids over all structurally equivalent residues. **No. 1** is RBP from *T. tengcongensis* (Cuneo, 2008). **No. 2** is ribose binding protein from *Thermotoga maritima* msb8 (Cuneo *et al.*, 2008). **No. 3** is glucose binding transporter from *T. maritima* (Tian *et al.*, 2007), **No. 4** is lactose binding transporter from *Clostridium phytofermentans* (unpublished), **No. 5** is galactose binding protein from *Burkholderia phytofirmans* (unpublished), **No. 6** is glucose/galactose binding protein from *Salmonella typhimurium* (Zou *et al.*, 1993). **No. 7** is (2R,4S)-2-methyl-2,3,3,4-tetrahydroxytetrahydrofuran (R-THMF) binding protein from *S. typhimurium* (Miller *et al.*, 2004). **No. 8** is SMLSRB from *Sinorhizobium meliloti* binds R-THMF as well (Pereira *et al.*, 2008). **No. 9** is L-arabinose binding protein from *E. coli* (Quiocho *et al.*, 1974). **No.10** is predicted to be a LacI-type regulator from *Klebsiella pneumoniae subsp.* (unpublished).

Alignment of these 10 protein sequences against BT1754peri indicates that the sequence similarity is low (Figure 3. 20). The identity of BT1754peri to other sequences is from 35 % against 2ioy (RBP) to 10 % against 1tjy.

BAD AVG GOOD

1754	1	DTPHFR-IG-VAQCS--DDSWRHKMNDEILREAMFYNG-VSVEIR-S-AGDDNS	47
2ioy	1	MK---T-IGLVISTL--NPNFFVTLKNGAEEKAKE-LG-YKIIIVE-D-SQNDSS	44
2fn8	1	MKG--K-MAIVISTL--NPNWFVVLAEATAKQRAEQ-LG-YEATIF-D-SQNDTA	45
2h3h	1	M----LTIGVIGKSV--HPYWSQVEQGVKAAGKA-LG-VDTKFFVP-QKEDIN	44
3brs	1	MSLKQYYMICIPKVLDDSSDFWSVLVEGAQMAAKE-YE-IKLEFMAPEKEEDYL	52
2rjo	1	MSLGQTTLACSFRL--TNPYYTAFNKGQSFQAKS-VG-LPYVPL-T-TEGSSE	48
1gca	1	AD---TRIGVTIYKY--DDNFMSVVRKAIEKDGKS-APDVQLLMN-D-SQNDQS	46
1tjy	1	GS--AERIAFIPKLV--GVGFFTSGGNGAQEAGKA-LG-IDVTYDGP-TEPSVS	47
3ejw	1	E----NQIAFIPKLV--GVGFFTSGGGAGAVKAGEE-VG-AKVTYDGP-TEPSVS	45
1abe	1	ENLK---LGFLVKQP--EFPWFQTEWKFFADKAGKD-LG-FEVIKI--AVPDGE	44
3d02	1	GAA-EKTVVNISKVD--GMPWFNRMGEGVVQAGKE-FN-LNASQVGP-SSTDAP	48
1754	48	KQAEDVHYFMDEGVDLLI--ISANEAAPMTPIVEEAYQKGIPVILVDRKILS-D	98
2ioy	45	KELSNVEDLIQKQVDVLL--INPVDSDAVVTAIKEANSKNIPVITIDRSANG-G	95
2fn8	46	KESAHFDALIAAGYDAII--FNPTDADGSIANVKRAKEAGIPVPCVDRGINA-R	96
2h3h	45	AQLQMLSFIAEAGVNGIA--IAPSDPTAVIPTIKKALEMGIPVVTLDTDSPD-S	95
3brs	53	VQNELIEBAIKRKPVDVIL--LAAADYEKTYDAAKEIKDAGIKLIVIDSGMKQ-D	103
2rjo	49	KGIADIRALLQKTGGNLVLMVDPNDSADARVIVEACSKAGAYVTTIWNKPKD-L	101
1gca	47	KQNDQIDVLLAKGVKALA--INLVDPAAGTVIEKARGQNVVVFNFNKEPSR-K	97
1tjy	48	GQVQLVNNFVNQGYDAII--VSAVSPDGLCPALKRAMQRGVKILTWDSDTKP-E	98
3ejw	46	GQVQFINNFVNQGYNALI--VSSVSPDGLCPALKRAMERGVLMVTDSDVFN-D	96
1abe	45	KTLNAIDSLAASGAKGFV--ICTPDKPLGSAIVAKARGYDMKVIADWDDQVFNK	96
3d02	49	QQVKIIEDLIARKVDAIT--IVPNDANVLEPVFKKARDAGIVVLTNESPQQP-S	99
1754	99	K-----YTAYIGA-DNYEIGRSVGNVIASSL-K-----GKGN-----	128
2ioy	96	D-----VVCHIAS-DNVKGGEMAAEFIAKAL-K-----GKGN-----	125
2fn8	97	G-----LAVAQIYS-DNYYGGVLMGEYFVKFL-K-----EKYPDAKEIP	133
2h3h	96	G-----RYVYIGT-DNYQAGYTAGLIMKELL-G-----GKGG-----	125
3brs	104	-----IADITVAT-DNIQAGIRIGAVTKNLV-R-----KSGKIG----	135
2rjo	102	HPWDYNNPNYVAHLSY-DGVAYGEETATQLFKSMG-----GKGG-----	138
1gca	98	AL-D-SYDKAYVYVGT-DSKESGVIQGDLIAKHWAQANQGWDLNKD-GKIQ-----	142
1tjy	99	C-----RSYYINQGTTPKQLGSMLEVMAAHQVVK-----EKAK-----	130
3ejw	97	C-----RSYYINQGTPEQLGGLLDVMAAEGVKK-----EKAK-----	128
1abe	97	G--K-PMDTVPLVMM-AATKIGERQGGELYKEM-Q-----KRGWDVKESA	136
3d02	100	A-----N-WDVEIIDNEKFAAEYVEHMAKRMG-----GKGG-----	129
1754	129	IVELTGLS-GSTPAMERHQGFMAAIS--K-FPDIKLIDKADAAM-ERGPABIEEM	177
2ioy	126	VVELEGIP-GASAARDRGKGFDEAIA--K-YPDIKIVAKQAADF-DRSKGLSVM	174
2fn8	134	YABELLGIL-SAQPTWDRSNGFHSVVD--Q-YPEFKMVAQQAQSAEF-DRDTAYKVT	182
2h3h	126	VVIGTGLS-TAMNSLQRIQGFKDAI--K--DSEIEIVDILNDEE-DGARAVSLA	173
3brs	136	V--ISFVK-NSKTAMDREGLKIGLS--D--DSNKIEAIYYCDS-NYDKAYDGT	181
2rjo	139	VVALGGIF-SNVPAIERKAGLDAAAL--KK-FPGIQLLDFQVADW-NSQKAFPIM	187
1gca	143	YVLLKGEP-GHPDAEARTTYVVVKEL--NDKGIQTEQLALDTAMW-DTAQAKDM	192
1tjy	131	VAFFYSSP-TVTDQNQWVKEAKAKIS--QE-HPGWEIVTTQFGYN-DATKSLQTA	180
3ejw	129	VAFFYSSP-TVTDQNAWAEAAKAKIA--KE-HPGWEIVTTQYGYN-DATKSLQTA	178
1abe	137	V--MAITANELDTARRRTTGSMDALKAAG-FPEKQIYQVPTKSN-DIPGAFDAA	186
3d02	130	YVIYVGS-LTVPQHNLWAD-LLVKYQ-KEHYPMHEVTRRMPVAESVDDSRRTT	180
1754	178	DSMLRR-H--PKIDA--VYAHNDRIAPGAYQAQAKMAGREK-EMI--FVGDAL	222
2ioy	175	ENILQA-Q--PKIDA--VFAQNDEMALGAIKAIEAANRQ--GII--VVGFDGT	218
2fn8	183	EQILQA-H--PEIKA--IWCNDAMALGAMKACEAAGRT--DIY--IFGFDGA	226
2h3h	174	BAALNA-H--PDLDA--FFGVYAYNGPAQALVVKNAGKVG-KVK--IVCFDPT	218
3brs	182	VELLTK-Y--PDISV--MVGLNQYSATGAARAIKDMSLA-KVK--LVCIDSS	226
2rjo	188	QAWMTRFN--SKIKG--VWAANDDMALGAIEALRAEGLAG-QIP--VTGMDGT	233
1gca	193	DAWLS-GPNANKIEV--VIANNAMAMGAVEALKAHNK-S-SIP--VFGVDAL	238
1tjy	181	EGIIKA-Y--PDLDA--IIAPDANALPAAAQAAENLKR-N-NLA--IVGFSTP	224
3ejw	179	ESILQT-Y--PDLDA--IIAPDANALPAAAQAAENLKRAE-GVT--IVGFSTP	223
1abe	187	NSMLVQ-H--PEVKHWLIVGMNDSTVLGGVRRATEGGQPKAADIIGIGINGVDAV	237
3d02	181	LDLMKT-Y--PDLKA--VVSFGSNGPIGAGRAVKEKRAKN-KVA--VYGMMP	225
1754	223	PGKGNGLLELVLDSV-LDATFIY-F-TNGDKVLQLAMDILEKKP-YP-----	264
2ioy	219	ED---ALKAIKEGK-MAATIAQQPALMGSLGVEMADKYLKGEK-IP-----	259
2fn8	227	ED---VINAIKEGKQIVATIMQFPKLMARLAVEWADQYLRGERSFP-----	269
2h3h	219	PD---ILQYVKEGV-IQATMGQRPYMMGYLSVTVLY-LMNKIGVQ--NTL--MM	263
3brs	227	ME---QIQYLEEGI-FEAMVVQKPFNIGYLGVEKALKLLKKEY-VP-----	267
2rjo	234	QP---GLVAIKSGE-LVASVDWDPFNLGGIGLSMGLQAKEKK-I-DLATLKP--	279
1gca	239	PE---ALALVKSGA-MAGTVLNDANNQAKATFDLAKNLAEGKGAADGTSWKI--	286
1tjy	225	NV---MRPYVQRGT-VKEFGLWDVVQQGKISVYVANALLKNMPPMNVGDSLDPG	274
3ejw	224	NV---MRPYIERGT-IQRFGLWDVTQGGKISVVFVADHVLKNGPMKVGEKLEIPG	273
1abe	238	SE---LSKAQATG--FYGSLLPSPDVHGYKSSEMLYNWWAKDVEPP-----	278
3d02	226	SO---AASLIKSGD-ITEGITYDPATAGYALAAVASTLLNGKTIEPGFEL--KE	273

1754	265	-----KETV-MNTAVVDRTNAHVMQLQTTHISELDKK-----	295
2ioy	260	-----NFIP-AELKLITKEN-----	273
2fn8	270	-----EIVP-VTVELVTRENI-----	290
2h3h	264	LPKVKVDG-----KVDYVIDTGVDVVTPENL-----	296
3brs	268	-----KQLD-SGCALITKDN-----	281
2rjo	280	-----DRRESF-CTATFVTKTNV-----	303
lgca	287	-----ENKIVR-VPYVGVKDNL-----	305
1tjy	275	IGKVTVSPNSEQGYHYEAKGNIGVLLPERVIFNKDNI-----	313
3ejw	274	VGTVEVSANKVQGYDYEADNGIILLPERTVFTKENI-----	312
labe	279	-----KFTEVTDVVLITRDNF-----	300
3d02	274	LGKAEVDS-----DKHIIRFHKVLVNVKDN-----	301
1754	296	-----IETLNGRIGGYLSQVATQQV-----	315
2ioy	274	-----VQGS-----	277
2fn8	291	-----GRKEEGS-----	297
2h3h	297	-----EELG-----IPIKFGS-----	307
3brs	282	-----EG-----	283
2rjo	304	ASPKAEWNNLYARVAG-----PVVYREG-----	326
lgca	306	-----ETQK-----	309
1tjy	314	-----YDF-----	316
3ejw	313	-----FDF-----	315
labe	301	-----KGLGGK-----	306
3d02	302	-----LY-----	303

Figure 3. 20 Tcoffee alignments of BT1754peri homologs from DALI.

The protein sequences are from PDB website according to their PDB IDs from DALI analysis above without the fused His tags. The proteins are sorted according Z score. The sequences vary a lot while the secondary structures are conserved. The identity of BT1754peri to other sequences is from 35 % against 2ioy (RBP) to 10 % against 1tjy.

3.2.4 The deletion analysis of BT1754peri

In an attempt to obtain better crystals to solve the structure to a higher resolution, several C-terminal deletion constructs of BT1754peri were made. The first was cut off the whole of the C-terminal helix ($\alpha 10$); and the second was to remove the extra residues after $\alpha 10$ that have no density in the structure.

Both deletions were cloned into same vector pET22b as BT1754peri with same restriction sites (*Nco* I and *Xho* I) using the primers in Table 3. 4. The 2 fragments were over expressed as soluble proteins (data not shown), and the ITC data indicated that they kept the binding ability against fructose (data not shown). They were then purified for crystallization in the same condition as for BT1754peri in the presence of 5 mM fructose. The long-slim needle crystals of these two proteins cannot diffract on X-ray to achieve structures (data not shown).

Primer ID	Sequence (5'----->3')	Note
BT1754-23	CCGCTCGAGTGC GTTGGTACGGTC	To 307 st aa
BT1754-33	CCGCTCGAGGAGCGTTTCAATC	To 327 st aa

Table 3. 4 Primers for BT1754peri deletion.

3.2.5 Mutation of residues involved in sugar binding in BT1754peri

Based on the structure analysis of BT1754peri (Figure 3. 18 and Table 3. 3), a group of residue mutants were made to understand the relative importance of each of these residues in ligand binding.

All the residues formed into hydrogen bonds with fructose: Ser41, Asp43, Arg46, Asp120, Arg121, Arg172, Asn222, Arg224, and Asp248 were site-directed mutated respectively. A further four residues (Trp45, Pro168, Trp196, and Tyr271) that are close to fructose and may form hydrophobic interactions with the sugar rings were mutated separately to alanine as well.

Based on the BT1754peri fusion in vector pET22b as template, a group of subclones was carried out using mutagenesis PCR program with the following primer pairs separately (Table 3. 5).

Primer ID	Sequence (5'----->3')	Note
BT1754-6	CGATGATTCATGGGCACATAAGATGAATGATG	R46A F
BT1754-7	CATCATTCATCTTATGTGCCCATGAATCATCG	R46A R
BT1754-10	GTTATTCTTGTAGCCCGGAAGATTCTTTC	D120A F
BT1754-11	GAAAGAATCTTCCGGGCTACAAGAATAAC	D120A R
BT1754-12	GTTATTCTTGTAGACCGGAAGATTCTTTCG	R121A F
BT1754-13	CGAAAGAATCTTCCGGTCTACAAGAATAAC	R121A R
BT1754-14	CCTGCAATGGAAGCGCATCAGGGATTTATG	R172A F
BT1754-15	CATAAATCCCTGATGCGCTTCCATTGCAGG	R172A R
BT1754-16	GCGGATGCTGCGCGGAACGTGGTCCG	W196A F
BT1754-17	CGGACCACGTTCCGCGCAGCATCCGC	W196A R
BT1754-18	GTGTATGCCCATGCTGACCGTATCGCTC	N222A F
BT1754-19	GAGCGATACGGTCAGCATGGGCATACAC	N222A R
BT1754-20	GATTTTTGTGCGGCATAGCTGCCTTGCCGGGTAAG	D248A F
BT1754-21	CTTACCCGGCAAGGCAGCTATGCCGACAAAAATC	D248A R
BT1754-25	CAATGCAGCGATGCTTCATGGCGACATAAG	D43A F
BT1754-26	CTTATGTCGCCATGAAGCATCGCTGCATTG	D43A R
BT1754-27	CAGCGATGATTCAGCGCGACATAAGATG	W45A F
BT1754-28	CATCTTATGTCGCGCTGAATCATCGCTG	W45A R
BT1754-29	GCCCATAATGACGCTATCGCTCCGGGTG	R224A F
BT1754-30	CACCCGGAGCGATAGCGTCATTATGGGC	R224A R
BT1754-34	GTGGCTCAATGCGCCGATGATTCATGG	S41A F
BT1754-35	CCATGAATCATCGGCGCATTGAGCCAC	S41A R
BT1754-36	GAGCGTTTCGACTGCTGCAATGGAACGG	P168A F
BT1754-37	CCGTTCCATTGCAGCAGTCGAACCGCTC	P168A R
BT1754-38	GCCACCTTTATCGCTCCGACCAATGGC	Y271A F
BT1754-39	GCCATTGGTCGGAGCGATAAAGGTGGC	Y271A R

Table 3. 5 Primers for mutations of BT1754peri

F: Forward primers, R: Reverse primers.

3.2.5.1 Mutated proteins expression

Site-directed mutations of BT1754peri were expressed in C41 cell strain and purified by IMAC. Among the 13 mutants, 8 (S41A, D43A, W45A, R46A, D120A, R121A, P168A, and R224A) are soluble while other 5 (R172A, W196A, N222A, D248A, and Y271A) are insoluble. The insoluble mutants were dissolved in TalonTM buffer containing 8M urea from inclusion bodies and then refolded in TalonTM buffer while bound to the Talon resin. The insoluble form of the WT (present in cells at lower levels to soluble BT1754peri) was used as a control (Figure 3. 21). Multiple protein bands are evidence of some degradation of both WT and refolded WT, which could not be inhibited in the presence of fructose (data not shown).

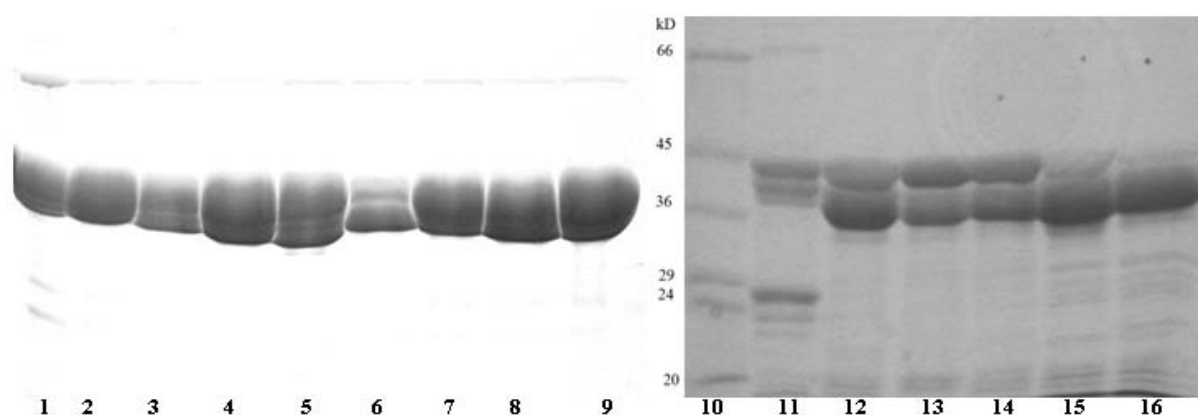


Figure 3. 21 The mutants of BT1754peri.

Lane 1: WT, Lane 2: R46A, Lane 3: D120A, Lane 4: R121A, Lane 5: D43A, Lane 6: W45A, Lane 7: R224A, Lane 8: S41A, Lane 9: P168A. Lane 10: Marker, Lane 11: BT1754peri WT (IB), Lane 12: R172A, Lane 13: W196A, Lane 14: N222A, Lane 15: D248A, Lane 16: Y271A.

3.2.5.2 ITC analysis of mutants

The ITC analysis was carried out to determine the binding ability of each mutant of BT1754peri with the WT as control for soluble mutants and refolded WT as control for refolded insoluble mutants, except for W45A (see Section 3.2.5.3 below). The ITC data show that for most of the mutants binding to fructose is abolished (Figure 3. 22 and Figure 3. 23). Only S41A and P168A retain some affinity for the sugar, although the affinity of S41A for fructose was too low to accurately quantify by Origin software (Table 3. 6).

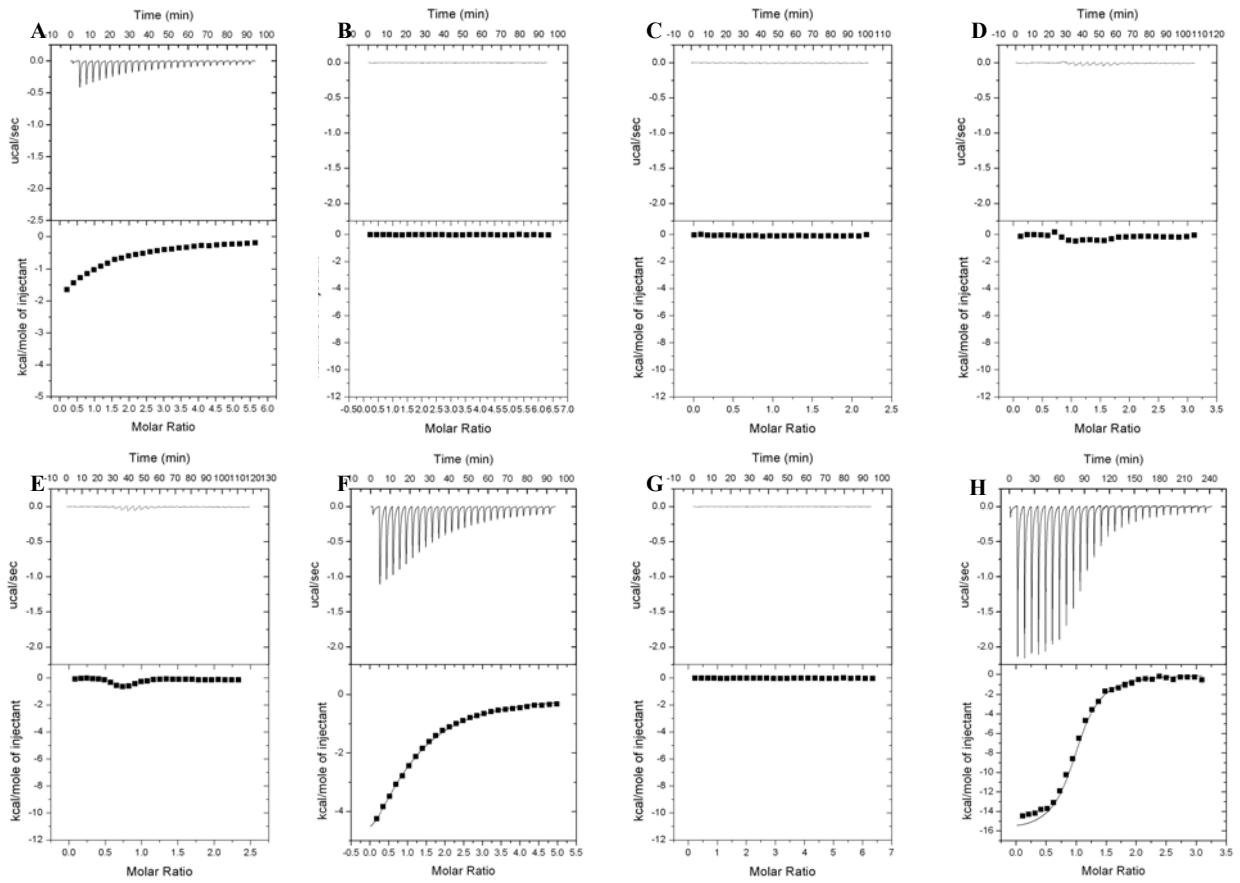


Figure 3. 22 ITC data hint the binding ability of BT1754peri soluble mutants against fructose.

A: S41A, B: D43A, C: R46A, D: D120A, E: R121A, F: P168A, G: R224A. H: WT.

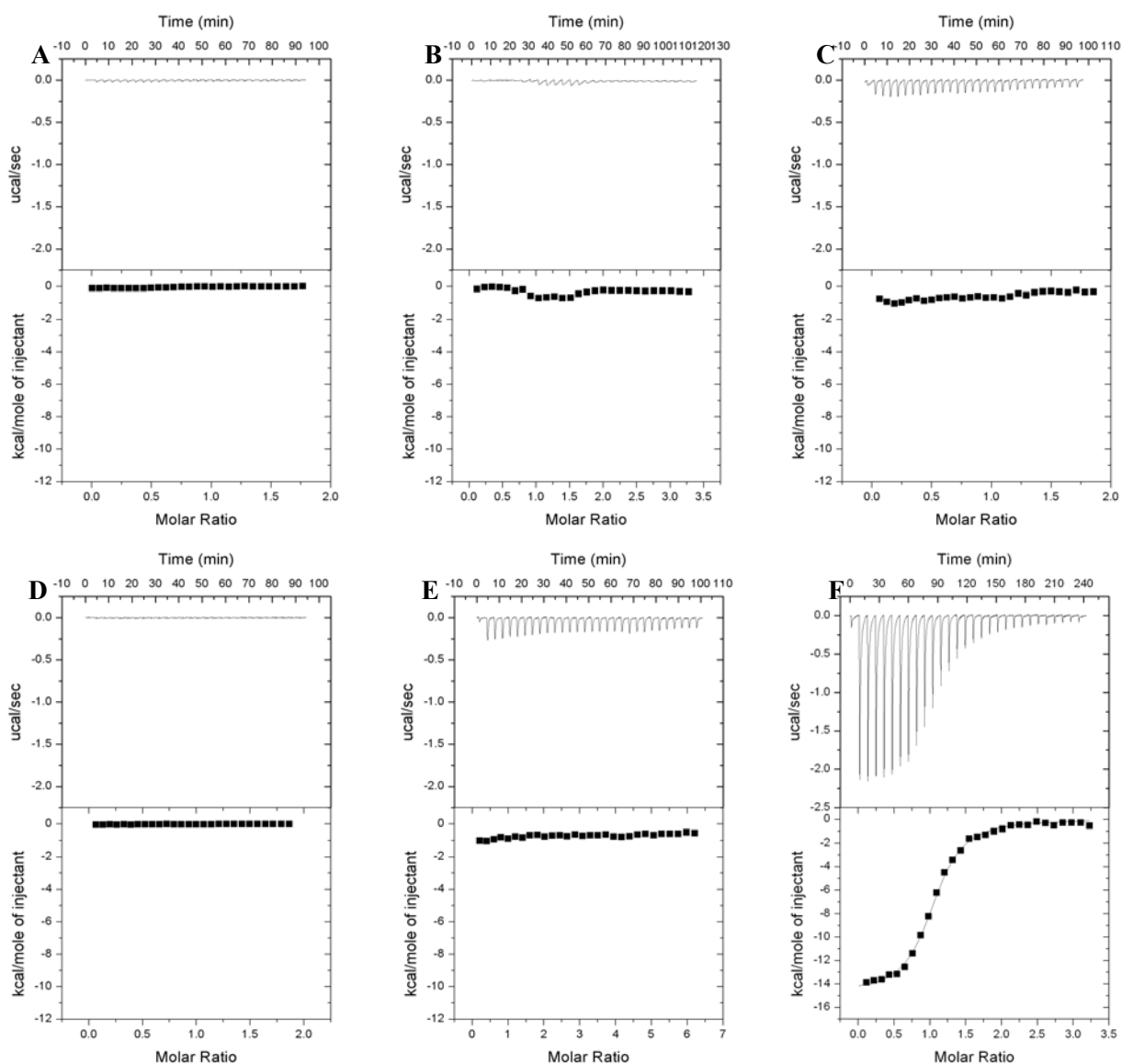


Figure 3. 23 ITC data hint the binding ability of BT1754peri insoluble mutants against fructose.

A: R172A, B: W196A, C: N222A, D: D248A, E: Y271A. F: WT. Use the refolded insoluble BT1754peri WT as control, which still has the full binding ability to fructose.

Protein	$K_a \times 10^5 (M^{-1})$	$\Delta H (kcal mol^{-1})$	$T\Delta S (kcal mol^{-1})$	n^*
WT ^a	2.6 ± 0.2	-13.9 ± 0.1	-6.5 ± 0.5	1.0 ± 0.0
P168A	0.3 ± 0.0	-11.4 ± 0.5	-5.4	0.7 ± 0.0
WT (IB)	2.8 ± 0.1	-15.0 ± 0.1	-7.3	1.0 ± 0.0

Table 3. 6 ITC data for binding of WT BT1754peri and P168A mutant to fructose.

The ITC data were fitted to a single site binding model. ITC was carried out in 10 mM Tris/HCl pH 8.0 at 25 °C. WT is soluble BT1754peri, WT (IB) is refolded BT1754peri from inclusion bodies. The S41A affinity is not high enough to be fitted by Origin. ^aThe parameters for WT are average values from 5 repeats, while other titrations are single.

3.2.5.3 Analysis of ligand binding by fluorescence

W45A was only expressed at very low levels in an insoluble form and therefore its binding could not be assessed by ITC. Instead, binding of refolded W45A to fructose was analysed using fluorescence spectroscopy with the refolded WT as control as this technique requires much less protein than ITC. Protein, either alone or in the presence of fructose, was excited at 290 nm and the emission spectra recorded from 300-450 nm. Compared with WT, W45A shows no change in the position of the emission maxima, even at 2500-fold molar excess of fructose (Figure 3. 24), indicating that the mutant protein had lost the ability to bind the monosaccharide.

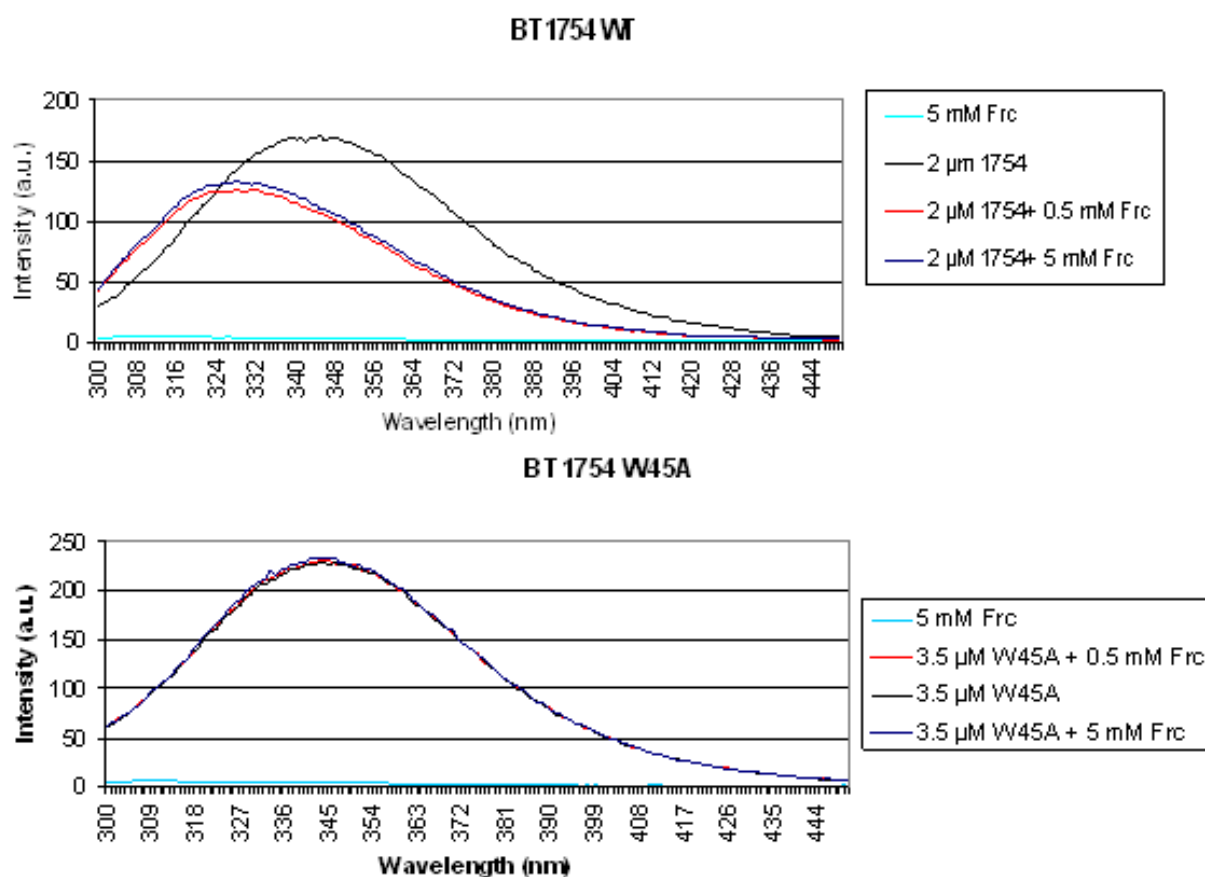


Figure 3. 24 Fluorescence spectroscopy to assess the binding of refolded WT BT1754peri and W45A to fructose.

The emission peak for BT1754peri WT in the presence of fructose shifts to left and down compared to protein in the absence of ligand, indicating an interaction with the sugar. The protein is saturated at a 250-fold molar excess of sugar. For the mutant W45A, there is no shift in the emission maxima in the presence of up to 2500-fold molar excess of fructose.

3.2.5.4 CD analysis

CD experiments were carried out to confirm that these mutants kept the secondary structure while they are abolished binding ability to fructose no matter they are soluble or refolded from inclusion bodies (Figure 3. 25 and Figure 3. 26). These data suggest the lack of binding observed with most of the mutants is not because of misfolding.

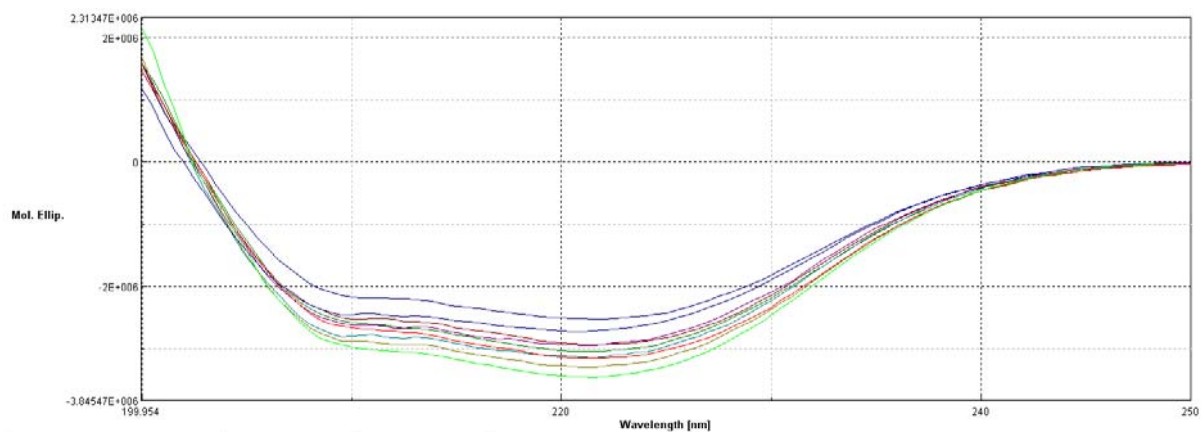


Figure 3.25 CD data showing the secondary structures of soluble BT1754peri mutants.

The curves in the diagram from 250 nm to 200 nm showed the soluble mutants and WT of BT1754peri have similar secondary structures. The voltages of these proteins were less than 600 V (data not shown).

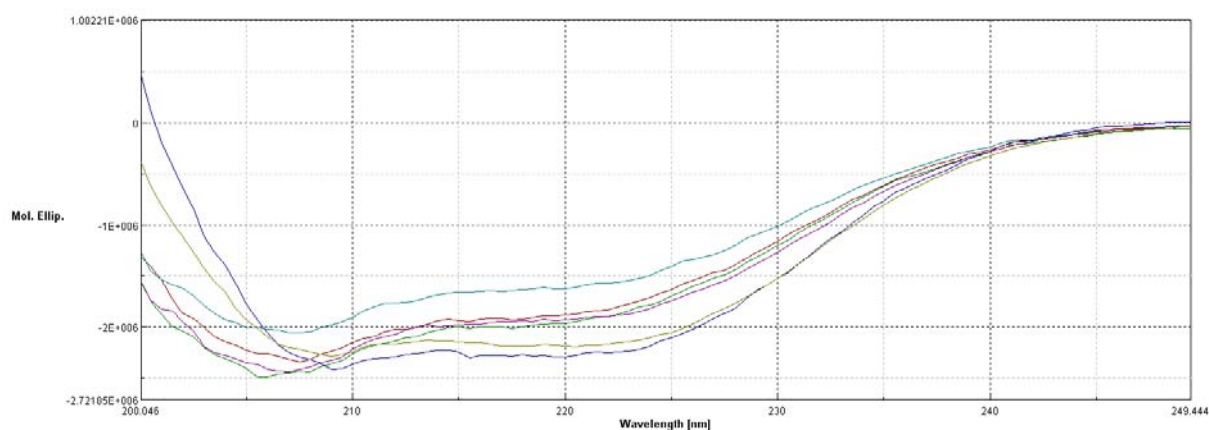


Figure 3.26 CD data showing the secondary structures of insoluble BT1754peri mutants.

All insoluble mutants and WT of BT1754peri were refolded during purification. The curves in the diagram from 250 nm to 200 nm showed the soluble mutants and WT of BT1754peri have similar secondary structures. The voltages of these proteins were less than 600 V (data not shown).

3.2.6 Structure comparisons

The ribose binding protein (RBP) from *Thermoanaerobacter tengcongensis* is structurally the most similar protein to BT1754peri (Cuneo, 2008). It was compared with BT1754peri to further understand the ligand specificity in both proteins. RBP has 9 α -helices, and 12 β -sheets formed into two sub domains, with one ribose in pyranose form inside (Figure 3.27A). There are 14 residues close to the ribose molecule, of which 7 residues formed 12 hydrogen bonds with all 5 oxygens of ribose (Figure 3.27B). O2, O3, and O4 each form 3 hydrogen bonds against residues. There is no oxygen forming 4 hydrogen bonds as fructose.

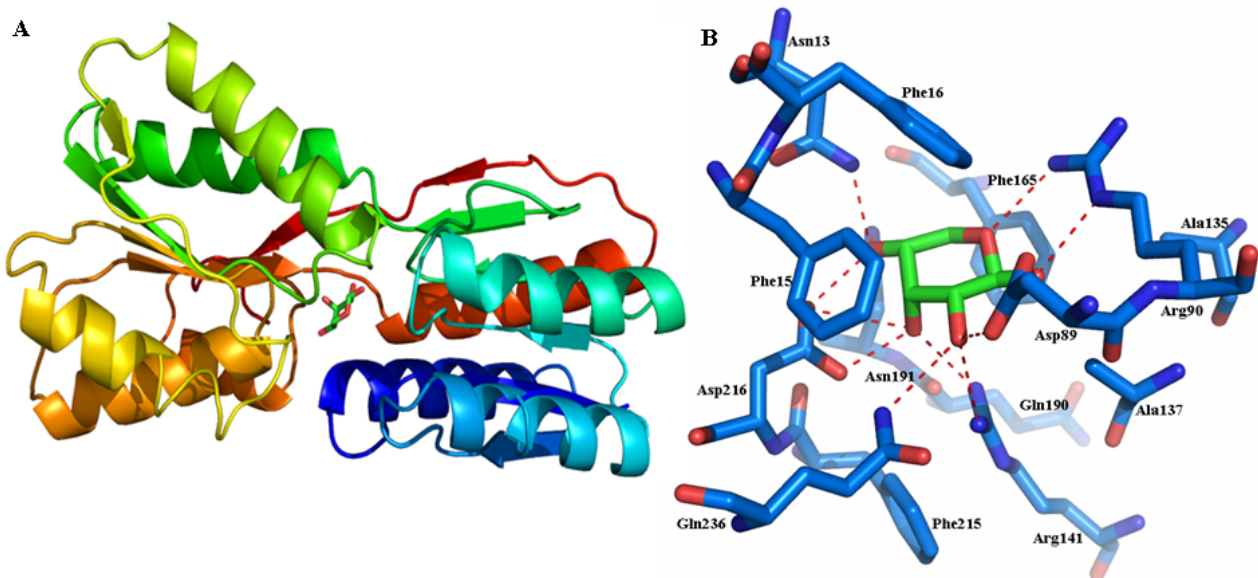


Figure 3. 27 The crystal structure of ribose binding protein from *Thermoanaerobacter tengcongensis*.

A: The RBP structure is shown as cartoon, bound ribose is shown in stick form. The resolution of this protein is 1.9 Å (Cuneo, 2008). The PDB ID is 2ioy. This protein has 9 α -helices and 12 β -sheets formed two sub domains holding the ribose molecule in the middle. The BT1754peri was modeled on this structure using SWISS-Model to enable identification of putative ligand binding residues. **B:** The 7 residues of RBP bound directly to ribose. They are N13, D89, R90, R141, N191, D216, and Q236.

Generally, the structures of BT1754peri and RBP are identical, although RBP, in common with all other (non-sensing) PBPs solved to date is missing the C-terminal helix of BT1754 (Figure 3. 28). When BT1754peri was overlaid with RBP to show the relationship of residues close to ligands, 13 of 14 residues of RBP are found at the same positions as BT1754peri (Figure 3. 29 and Table 3. 7).

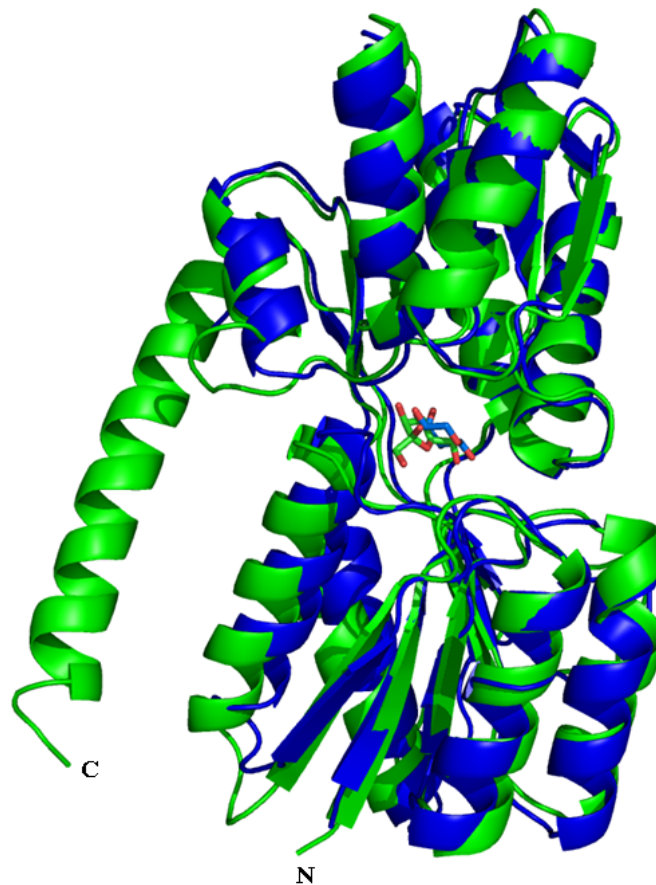


Figure 3. 28 The structures overlap of BT1754peri and RBP.

Green is BT1754peri. Blue is RBP. BT1754peri has an extra helix at the C-terminal compared to RBP.

Of the 13 residues surrounding the ligand binding site, 5 are conserved between BT1754peri and RBP: 2 Asp, 2 Arg, and one Asn. These 5 residues make 9 hydrogen bonds against O3, O4, O6 of fructose, and 10 hydrogen bonds against O1-5 of ribose, which indicate that these 5 residues are highly important for ligand binding of PBP superfamily (Figure 3. 29). Two Phe residues are facing the ring of ribose from both sides, while for fructose two Trp residues take the place of Phe (Table 3. 7). Only Ala135 of RBP is unmatched, which is far from ribose and predicted no donation for the binding activity.

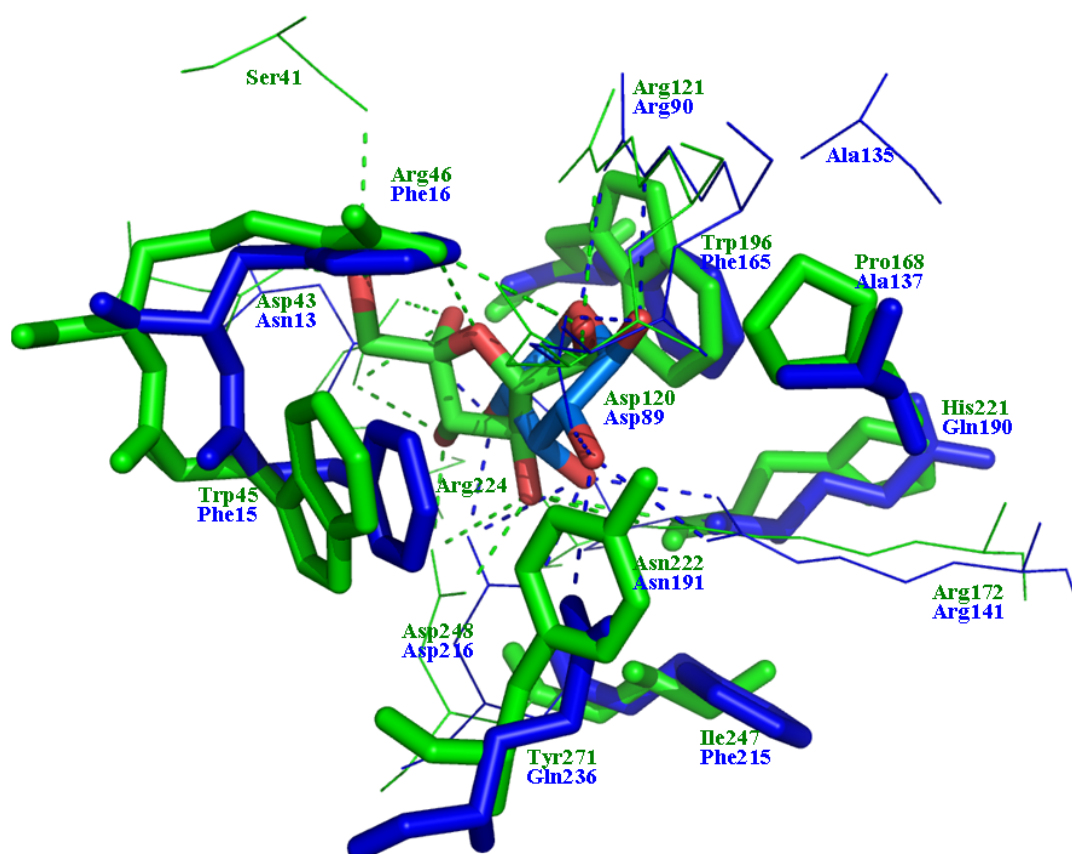


Figure 3. 29 The binding sites overlay of BT1754peri and RBP.

BT1754peri: Green, RBP: Blue. Fructose: Green and Red. Ribose: Light blue and Red. Dashes present hydrogen bonds between residues and ligand (green in BT1754peri and blue in RBP). When residues from one protein are at the same positions as in another protein, they form into pairs. When at least one residue of a pair has a ring-like side-chain, the pair is present in sticks, whilst other pairs and unmatched residues (Ser41 and Arg224 in BT1754peri, and Ala135 in RBP) are present in lines. There are 13 pairs in this overlay. 5 of them are same residues in both proteins. They are Asp248, Arg172, Asn222, Arg121, and Asp120 in BT1754peri matching Asp216, Arg141, Asn191, Arg90, and Asp89 respectively.

BT1754	Interacts with Frc at:	RBP	Interacts with Ribose at:
D43	OD1 and OD2 with O1	N13	ND2 with O4
W45	Face-to-face stack	F15	Stacks with one face of Rib.
R46	NH2 with O1	F16	Contributes to one side of aromatic sandwich formed by F15, F16 and F165
D120	OD1 with O6, OD2 with O6	D89	OD2 with O1, OD1 with O2
R121	NH2 with ring O, NE with O6	R90	NH2 with ring O, NE with O1
P168	Forms hydrophobic part of binding site near C6	A137	Forms hydrophobic part of binding site near C1
R172	NH1 with O4, NH2 with O4	R141	NH1 with O3, NH2 with O2
W196	Face-to-face stack	F165	Stacks with one face of Rib.
H221		Q190	
N222	ND2 with O3	N191	ND2 with O4
I247		F215	
D248	OD2 with O3, OD1 with O4	D216	OD2 with O3 and O4, OD1 with O3
Y271	Forms hydrophobic part of binding site near C4-C5	Q236	NE2 with O2 and O3

Table 3. 7 The 13 paired residues of BT1754peri and RBP around ligands.

The 13 pairs are sorted according BT1754peri. The 5 identical pairs are bolded.

All oxygens of both sugars are involved into the formation of hydrogen bonds. When overlay the fructose and ribose, four oxygens are found in the similar locations: O3, O4, O5, and O6 in fructose against O4, O3, O1, and O5 in ribose respectively (Figure 3. 30). But O1 and O2 in fructose form extra hydrogen bonds which do not exist in the same location of ribose case, which may explain why BT1754peri could not bind ribose.

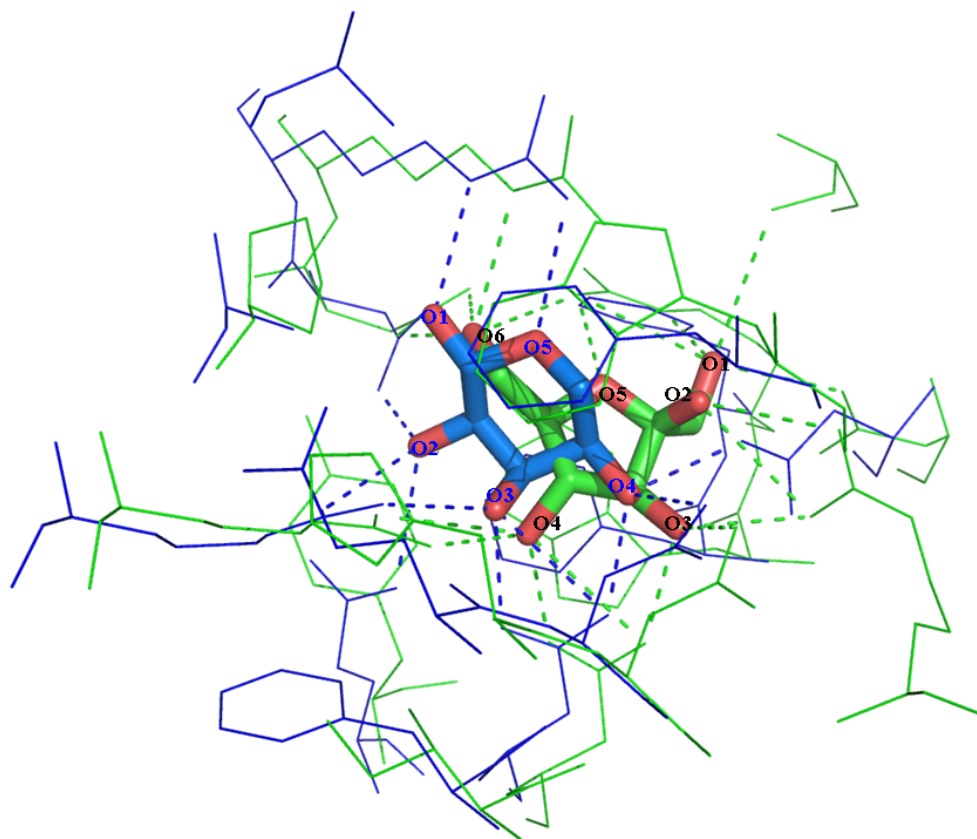


Figure 3. 30 Ribose and fructose overlay.

The black numbers O1 to O6 are labeled for fructose (green and red sticks), while the blue O1 to O5 are labeled for ribose (blue and red sticks). The residues of BT1754peri around fructose presents in blue lines, while the residues of RBP are in red lines. Dashes present the hydrogen bonds between ligand and amino acids (green for BT1754peri, and blue for RBP).

3.3 Discussion

3.3.1 BT1754peri binds specifically to fructose

The specific binding ability of BT1754peri was revealed through this study. The fructose was found to be the only candidate could bind BT1754peri. The binding affinity is relatively high (K_a 2.3×10^5 (M^{-1})) and is metal independent. Even the refolded insoluble BT1754peri still has the identical binding strength (Figure 3. 23). While other ligands, no matter how similar they are to fructose, could not bind. The 18 hydrogen bonds made by 9 residues around fructose could explain the strength of binding ability (Table 3. 3). Other sugars could not dock in the pocket of the two sub domains to form effective hydrogen bonds as fructose because of the pocket shape and sugar shape. Of fructose epimers (psicose, tagatose, and sorbose), only sorbose presented a faint binding capacity. The O6 of sorbose is different on location (Figure 3. 9), which hints the contracts between the amino acids and this hydrogen group are faint.

Most of the residues around fructose are pivotal in ligand binding as evidenced by the ITC data and CD spectra. Looking into the structure of BT1754peri, there are two tryptophans very close to fructose that sandwich the sugar and are shown here to be of pivotal importance in binding. This face-to-face stacking of aromatic residues with the sugar rings is a common feature of carbohydrate-protein recognition (Figure 3. 18) (Boraston *et al.*, 2004; Cuneo, 2008).

The ITC data show that P168A and S41A retain some binding ability to fructose (Figure 3. 22). S41 forms one hydrogen bond with fructose O1 (Table 3. 3) and S41 is the first residue of core fragment to hold fructose according to protease digestion analysis (section 3.2.1.3.3), so maybe that is why this residue is not so important for binding. P168 is at the site of the beginning of helix α_6 , which is the outside of the structure. So the main function of this residue is presumed to form hydrophobic part of binding site like a shell, not so important for binding itself, even though it is close to the fructose O6 (3.57 Å).

Specificity is important to a sensor protein as it prevents activation of the system in the presence of related sugars that cannot be utilised by the genes under its control. Fructose is a good molecule to choose to bind to because naturally fructose only occurs in inulin, levan, sucrose, and some plant derived oligosaccharides such as raffinose (all β -linked fructose). By sensing fructose, *B. thetaiotaomicron* knows the saccharides in gut to be one of these glycans. *B. thetaiotaomicron* can use all naturally occurring fructose containing sugars (D. Bolam, unpublished data), and therefore

the inulin-regulated PUL system should be able to degrade them. These data suggest that the same PUL should be activated in the presence of all fructose containing oligo- and polysaccharides, as well as monomeric fructose and recent qPCR data from our collaborators (J. Sonnenburg, personal communication) confirms this to be the case. The activation of the whole locus in the presence of fructose and sucrose would seem to be inefficient to *B. thetaiotaomicron* as the sus system and four different GH32s would not be required to utilise these small molecules. However it is likely that *B. thetaiotaomicron* has no chance to encounter these simple sugars in its niche of the large intestine as most will be removed by the host further up the digestive tract. Indeed a recent research has shown that *B. thetaiotaomicron* in mice fed a simple sugar diet upregulates genes involved in degradation of host glycans, strongly suggesting that no simple sugars reach the large intestine (Sonnenburg *et al.*, 2005).

The binding of BT1754peri to fructose and its location next to the locus upregulated in the presence of inulin indicates that this HTCS is the regulatory gene that controls fructan degradation in *B. thetaiotaomicron*. Recent gene knockout studies showed the loss of BT1754 resulted in a mutant strain that could no longer utilise inulin or levan, but whose growth was unaffected on other unrelated polysaccharides such as starch (J. Sonnenburg, personal communication).

3.3.2 Signal transduction in BT1754

BT1754peri is a member of periplasmic binding protein (PBP) superfamily (section 1.2.3.2.1). Though the sequences variation in this family is high, the general structure fold is quite conserved: two sub domains linked with a hinge area, where the ligand embeds. PBP proteins are known to bind a series of small molecule and response to the changing of environment, such as arabinose-binding protein (ABP), histidine-binding protein (HBP), maltose-binding protein (MBP), glucose-binding protein (GBP), and RBP. And PBP proteins change the conformation when binding to a suitable ligand (Marvin and Hellinga, 2001). In MBP, Ile-329 plays the role as a hinge which is closed in the global open state and open in global closed state (Marvin and Hellinga, 2001). In BT1754peri, Ile-270 located at the same position as Ile-329 of MBP, which indicated that Ile-270 may be the hinge for BT1754peri (Figure 3. 31).

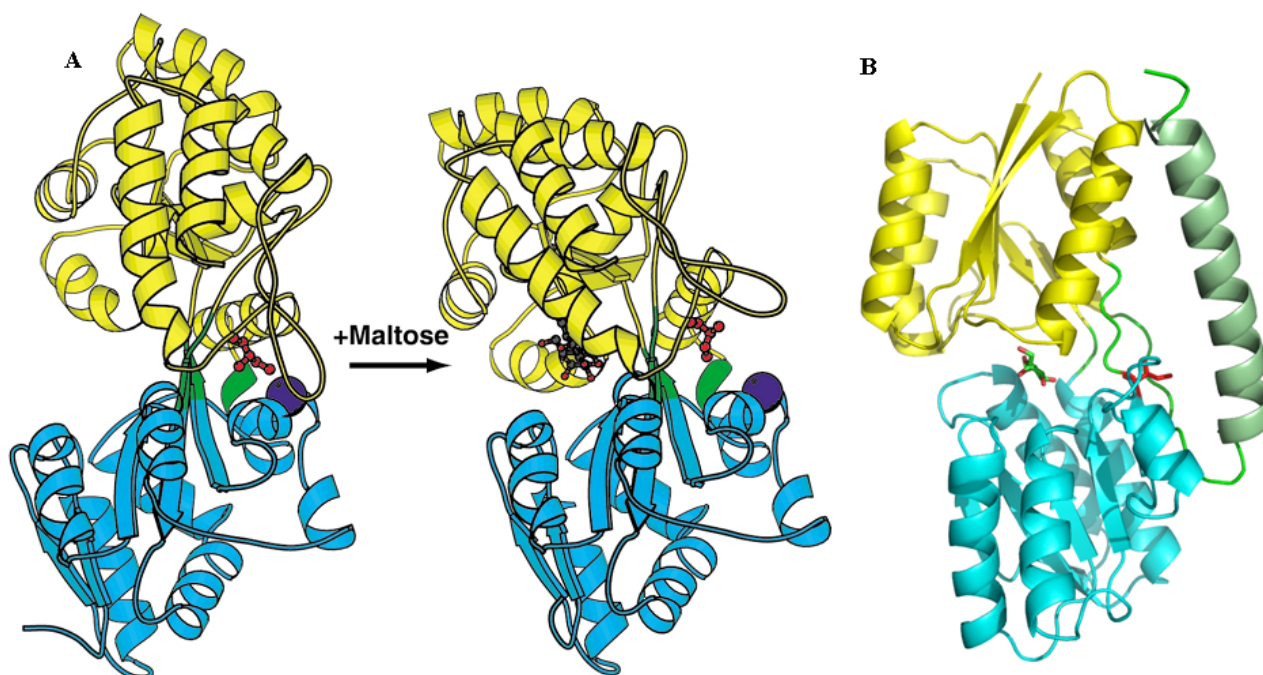


Figure 3.31 Hinge prediction of BT1754peri according to MBP structure.

A: Maltose binding protein shifts from open state (**left**) to a ligand-binding close state by bending the two sub domains (yellow and cyan) around the hinge region (green) where contains Ile-329 (red stick). Cited from (Marvin and Hellinga, 2001). **B:** BT1754peri binds to fructose with Ile-270 (red stick) as the possible hinge switch locates between two subdomains (yellow and cyan). The last helix is colored in palegreen. This structure is rotated around compared to Figure 3.17.

According to the number of β -sheets in each sub-domain, PBPs are classified into three groups: ABP represents the first class, which has 6 β -sheets, the second has 5 β -sheets in each sub domain, such as HBP, and the third has 5 β -sheets in N-terminal sub domain and 4 in C-termini, such as vitamin B12 binding protein (Karpowich et al., 2003). BT1754peri belongs to the first class. When MBP, GBP, or RBP fused to a green fluorescent protein, the fusion protein can be used for *in vivo* monitoring of cellular metabolism by fluorescence microscopy (Fehr et al., 2002; Fehr et al., 2003). The utilisation of this system will be expanded if combine computational design (Fukami-Kobayashi et al., 1999). This super family explored as engineer biosensors opens a door potential for the usage of this fructose binding protein (Dwyer and Hellinga, 2004).

Based on its similarity to other PBPs it seems very likely that conformational change on ligand binding occurs in BT1754peri. Evidence for this is provided here as we show that the protein displays increased stability against both protease action and thermal denaturation in the presence of fructose. It is likely this stabilisation is caused by tightening of the structure driven by domain closure around the sugar.

BT1754peri is present as a monomer in solution in the presence and absence of fructose by gel filtration chromatography and crosslinking analysis. But this protein forms a dimer in crystal, with the dimerisation interface formed between the C-terminal helices of each monomer. BT1754peri is predicted to be a dimer in the cell as well because most histidine kinases form homodimers (Lee *et al.*, 2008a). It is likely that the lack of association observed in solution is due to the fact that the protein is not tethered to the membrane as it would be in the cell. It is known that attachment to the membrane to hold the protein in two-dimensions can lead to an increase in the association constant for the interaction by up to 10^6 -fold (Metzger, 1992).

Modelling the structure of the dimer of BT1754peri with the known structure of the cytoplasmic domains of the TM0853 histidine kinase from *Thermatoga maritima* reveals that the entry point of the C-terminal helices of BT1754 into the membrane likely maps well onto the helices of the dimeric cytoplasmic phosphoacceptor domain that exist the inner membrane (Marina *et al.*, 2005) (Figure 3. 32).

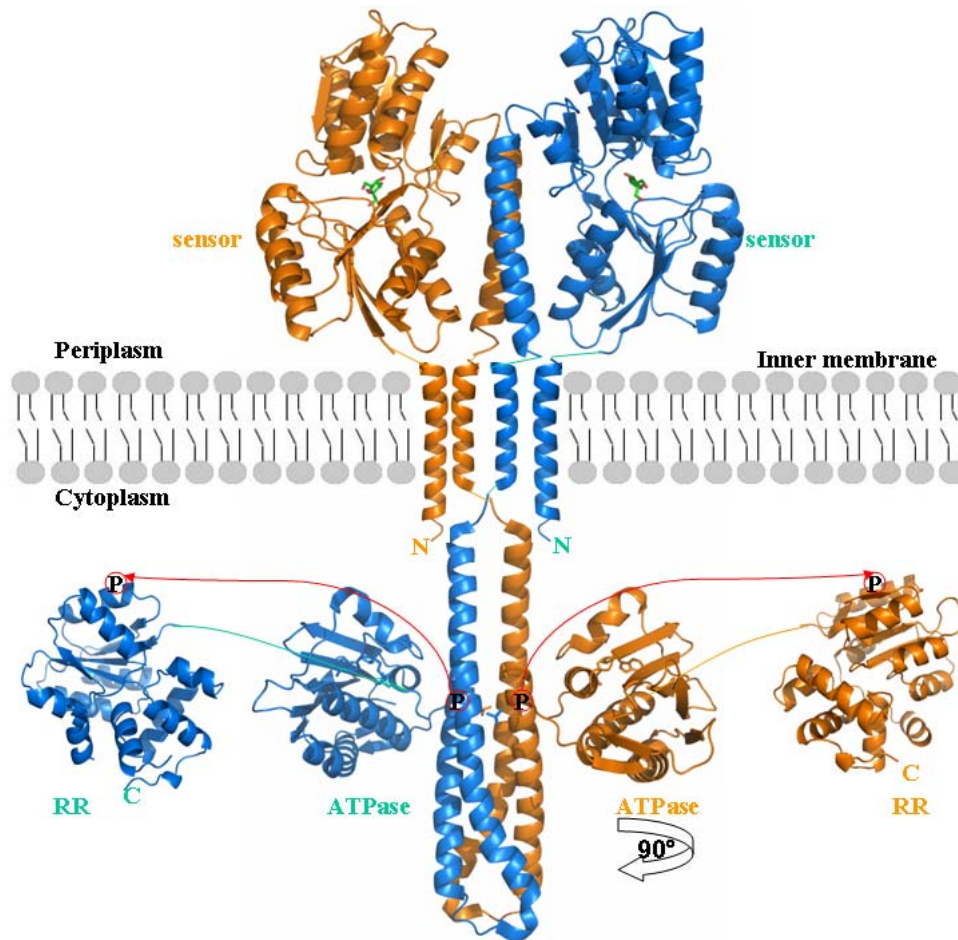


Figure 3. 32 Model for structure of BT1754 sensor HK.

Putative assembly of whole BT1754 structure based on known structure of BT1754peri, the cytoplasmic region of TM0835 sensor HK from *Thermotoga maritima* (PDB ID 2C2A) (Marina *et al.*, 2005), the entire RR protein (receiver domain and DNA binding domain) of the nitrate responsive NarL (PDB ID 1A04) (Baikalov *et al.*, 1998), and a possible transmembrane domain. The ATPase domain of each HK is linked to a RR domain as BT1754 is a hybrid two component system. The periplasmic sensor domain is predicted to form into a dimer with C-terminal helices contacting each other. When sensor binds the fructose, the conformational change is transferred across the membrane via the C-terminal helices, possible via a piston like movement, to the equivalent helices of the dimeric cytoplasmic phosphoacceptor domain. The TM helices are predicted to be dimerized into a 4-helix bundle. The phosphoacceptor domain dimer forms a 2 helix bundle that maps well onto the 2 helix bundle of the sensor C-terminal. The position on the phosphoacceptor domain of the His that is phosphorylated on activation of the HK is shown circled in red. Black arrow indicates the direction the ATPase domains would have to twist in order to phosphorylate the His of the opposite monomer as proposed by (Marina *et al.*, 2005). Precisely how a piston like movement of the sensor domain could be transduced into a twisting action is not clear. The activated phosphoacceptor then transfers the phosphate to the conserved Asp of the RR (circled in red; red arrows indicate direction of phosphotransfer) to activate the associated DNA binding domain.

Thus it appears that signal will be transduced in BT1754 by a conformational change in the sensor domain on ligand binding that is transmitted via the C-terminal helices of the dimer through the membrane helices to the dimeric helices of the phosphoacceptor domain to activate the phosphorelay. Activation of the HK is thought to be by phosphorylation of the opposite monomer, but precisely how the catalytic domain of one monomer moves to phosphorylate the conserved His of the acceptor domain is currently unknown. The movement of the helices relative to one another is also unclear but it seems likely, based on the structure of BT1754_{peri}, that ligand binding induces a ‘piston’ type movement of the helices away from the membrane (Figure 3. 33). This piston like movement has been suggested before for citrate sensing PAS domain of the citA sensor HK (Sevvana *et al.*, 2008).

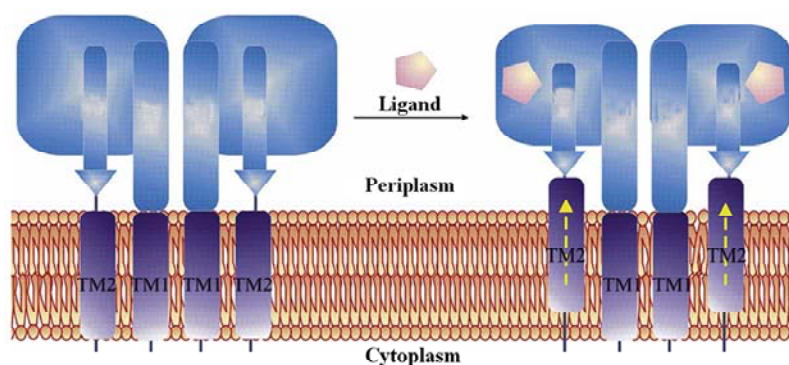


Figure 3. 33 Model of piston like motion of C-terminal helices in BT1754 on fructose binding.

The sensor linked with transmembrane domains is shown as a dimer. On ligand binding, the TM2 moves vertically towards the periplasm in a piston type fashion, thus transducing the signal across the membrane. Cited from (Sevvana *et al.*, 2008).

Overall, BT1754_{peri} specially binds fructose to initiate the signalling transduction into the cytoplasm. Its structure revealed the binding specificity and possible conformation change during the signalling process.

Chapter 4: Glycoside hydrolase family 32

4.1 Introduction

The GH32 family contains nature fructan-degrading enzymes. Fructans are β -2,1 or β -2,6 linked homopolymers of fructose found in many plants and bacteria where they play a role a storage polysaccharides, resistance to drought and in biofilm formation (section 1.2.5). According to the Cazy database of Jan 2009 (www.cazy.org/fam/gh32.html) GH32 has 833 members from archea, bacteria, fungi and plants, but not animals. Sucrose, a disaccharide of glucose and fructose is broken down in animals by an α -glucosidase, rather than a β -fructosidase (section 1.1.2). Because of this inability of humans to use fructans and the fact that they are readily utilised by health promoting symbiotic gut bacteria, these polymers are often used as prebiotics or ‘functional foods’. Most GH32s have an exo-mode of action, releasing fructose from the ends of both β -2,1-linked inulins and β -2,6-linked levans. Catalysis proceeds via a retaining mechanism, with a catalytic nucleophile (Asp) and a catalytic proton donor (Glu). They share a common 3D structure: a 5-fold β -propeller similar to the β -propeller of GH43 and GH68, with a C-terminal β -sandwich module of unknown function and are members of clan GH-J. All GH32s whose structure has been solved to date are exo acting and release fructose from their β -configured substrates (Figure 4. 1). These structures are of a GH32 from the bacterium *Thermotoga maritima* that acts on sucrose, raffinose, and inulin (Alberto *et al.*, 2004), an exo-inulinase from the fungus *Aspergillus awamori* that can cleave levan, inulin, and sucrose (Nagem *et al.*, 2004), a cell-wall invertase from *Arabidopsis thaliana* that hydrolyses sucrose preferentially but can also act on inulins and levans, albeit at a much reduced rate (Verhaest *et al.*, 2006), and the fructan 1-exohydrolase IIa from Chicory (*Cichorium intybus*) which can not digest sucrose or levan, but can break down inulin-type fructans (Verhaest *et al.*, 2005). The distance between the two catalytic residues in these structures varies between 6.1 Å to 6.9 Å, which is slightly further apart than the typical distance for retaining enzymes of \sim 5.5 Å.

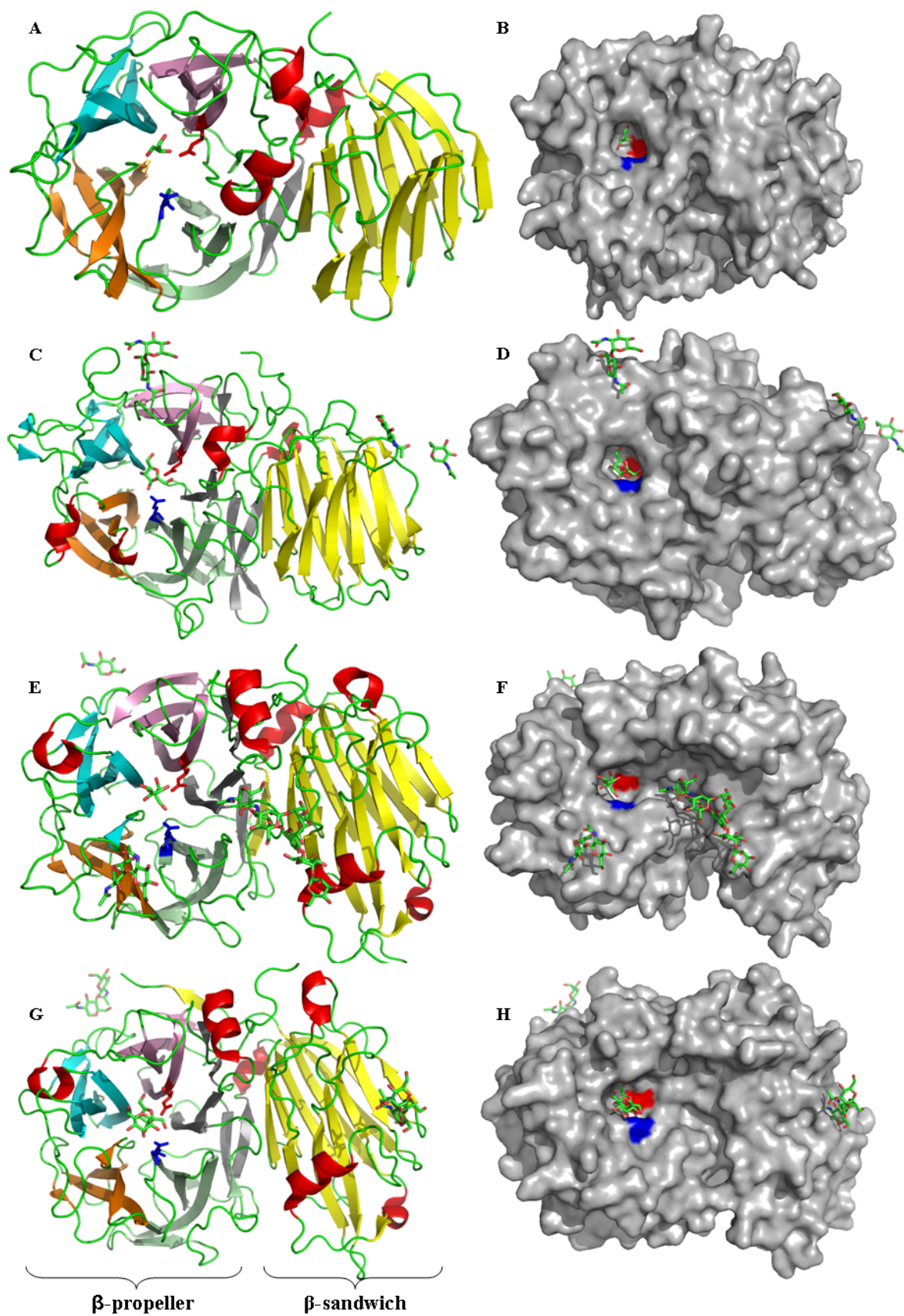


Figure 4.1 GH32 structures solved to date.
See legend at the next page.

Legend for Figure 4.1:

The four representatives of GH32 whose structures have been solved to date shown as secondary structure cartoon and surface. **A and B:** *Thermotoga maritima* (PDB ID 1UYP) (Alberto *et al.*, 2004; Alberto *et al.*, 2006), **C and D:** *Aspergillus awamori* (PDB ID 1Y9G) (Nagem *et al.*, 2004), **E and F:** Cell wall invertase from *Arabidopsis thaliana* (PDB ID 2AC1) (Verhaest *et al.*, 2006), **G and H:** fructan β -(2,1)-fructosidase, from *Chicorium intybus* (PDB ID 2ADD) (Verhaest *et al.*, 2007b). **Left:** The cartoon representation shows the GH32s have a two domain structure: a 5-fold β -propeller catalytic domain (pink, cyan, orange, palegreen, and gray), and a C-terminal β -sandwich domain (yellow). All α -helices are indicated in red, loops are in green, and binding substrates are sticks. The catalytic nucleophile residue Asp (red stick) located on the first β -blade, while the conserved acid base Glu (blue stick) located at the beginning of the fourth blade (palegreen). **Right:** Surface representation of GH32s shows the two catalytic residues located in a pocket (Asp is red, Glu is blue). In B, there is a glycerol molecule from cryo bound in the catalytic pocket. In D, there is a fructose molecule located in the catalytic pocket and a few covalently bound GlcNAc molecules around the surface as the enzyme is a glycoprotein. In F, there is a glycerol molecule located in the catalytic pocket, and a few GlcNAc molecules binding around the surface. A covalently bound (GlcNAc)₂MAN₄ molecule is located in the cleft between two domain. In the Chicory GH32 (panel H) a sucrose molecule is located in the catalytic pocket. This plant enzyme is unable to utilise sucrose so the sugar is binding as a competitive inhibitor.

Genechip data from *B. thetaiotaomicron* grown on inulin as the sole carbon source revealed a unique locus upregulated in the presence of the β -2,1-linked fructan (Figure 4. 2; E.C. Martens, unpublished data).

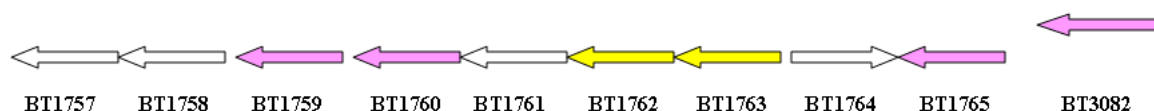


Figure 4. 2 Gene map of the unique locus upregulated in response to growth of *B. thetaiotaomicron* on inulin.

All genes are shown except BT1764 upregulated in response to inulin. BT1759, BT1760, BT1765, and BT3082 are genes encoding predicted GH32 enzymes. BT1763 and BT1762 are the SusC and SusD homologue, respectively, BT1757 is a fructokinase and BT1758 is an inner membrane monosaccharide transporter.

Within this locus were genes encoding three predicted GH32 enzymes: BT1759, BT1760, and BT1765. In addition a fourth orphan GH32 gene BT3082 was also upregulated. To further our understanding of fructan degradation by this important gut symbiont, biochemical and structural analysis of the products of these genes was carried out.

4.2 Results

There are four GH32 found in *B. thetaiotaomicron*: BT1759 and BT1760, which are in the inulin locus of BT1757 to BT1763, BT1765 next to this gene cluster and BT3082. Multiple sequence alignments of these four GH32 against the four GH32s whose structures have been solved revealed that BT1759, BT1765, and BT3082 have extra N-terminal sequence, suggesting the presence of an additional domain, which is so far unique amongst characterised GH32 enzymes (Figure 4. 3). All *B. thetaiotaomicron* enzymes also have the typical C-terminal β -sheet domain of unknown function found in all GH32s to date. The central catalytic domain contains regions of high sequence similarity including the catalytic nucleophile and acid/base which indicates the functional conservation. Also worth noting is the Arg-Glu-Pro (RDP) motif (Figure 4. 3), which is conserved

in all enzymes that act on fructosyl containing substrates, including levansucrases and sucrose-6-phosphate hydrolases, and has been suggested to play a critical role in fructose specificity (Alberto *et al.*, 2004; Nagem *et al.*, 2004).

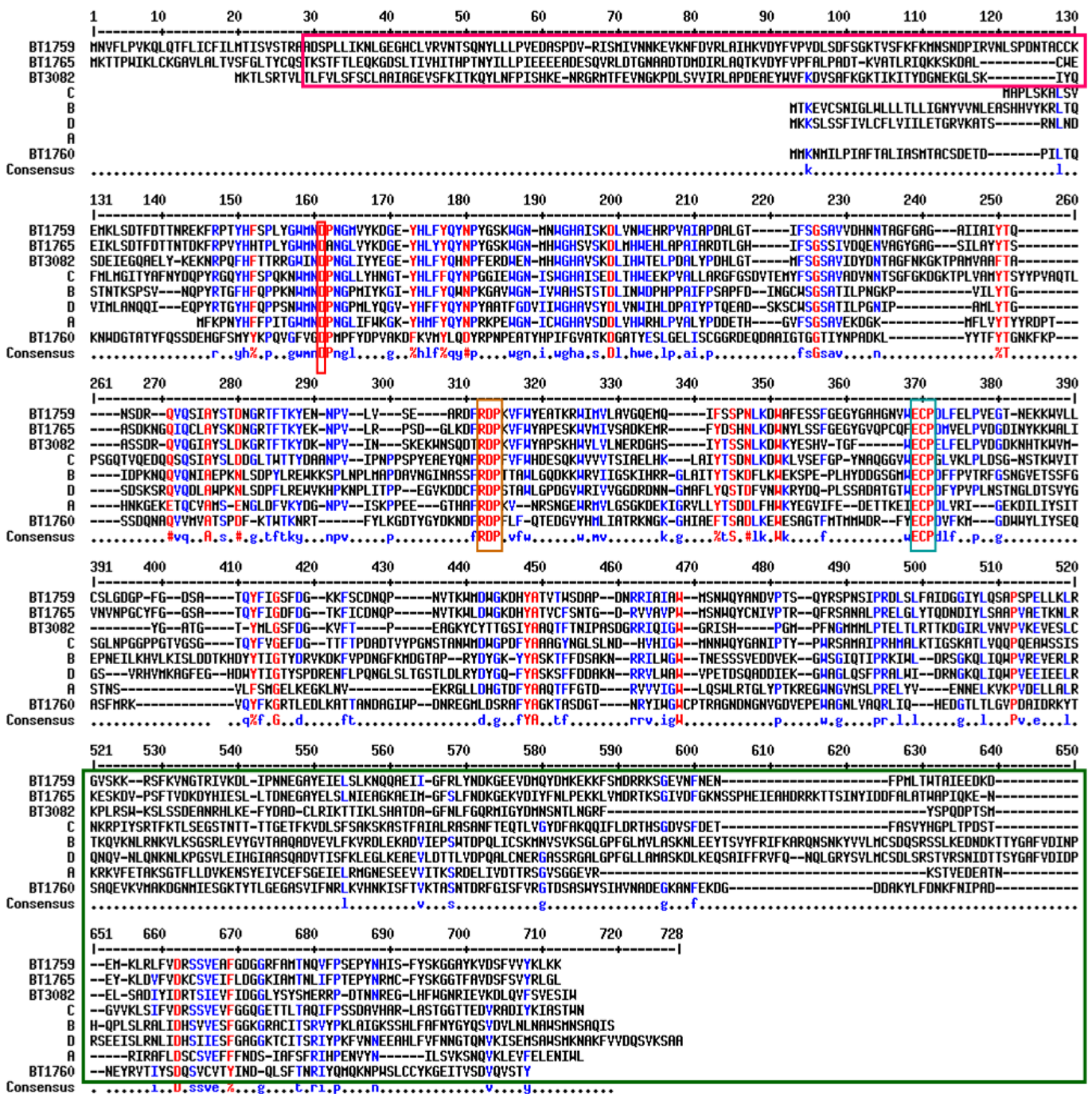


Figure 4.3 GH32 multiple sequence alignment.

Multiple alignment of *B. thetaiotaomicron* GH32s against structure solved GH32 enzymes. **A:** *Thermotoga maritima* (PDB ID 1UYP), **B:** *Arabidopsis thaliana* (PDB ID 2AC1) (Verhaest *et al.*, 2006), **C:** *Aspergillus awamori* (PDB ID 1Y9G), **D:** *Chicorium intybus* (PDB ID 2ADD). All proteins have an N-terminal signal peptide. BT1759, BT1765, and BT3082 have an extra N-terminal domain (pink box) when compared to other enzymes. The catalytic nucleophile (Asp) and acid/base (Glu) residue are boxed in red and cyan. Another motif RDP is in orange box. The C-terminal β -sandwich domain present in all GH32s is boxed in green. The similarity in C-terminal is lower than the catalytic domain. All GH32s have a RDP motif (orange box) and an Asp-Cys-Pro (ECP) motif (cyan box). This alignment is made by Multalin.

Sequence alignment revealed that BT1760 has lower identity to all other *Bacteroides* enzymes (less than 14 % for all), while BT1759 and BT1765 have 51 % identity. BT3082 is most similar to BT1759 (34 %) (Table 4. 1).

%	BT1759	BT1760	BT1765	BT3082	A	B	C	D
BT1759		5	51	34	26	18	35	20
BT1760			8	11	10	14	7	12
BT1765				30	29	16	30	18
BT3082					27	17	29	21
A						22	24	24
B							21	48
C								20
D								

Table 4. 1 GH32 identities.

Based on sequence alignment, the identities are issued. **A:** *Thermotoga maritima* (PDB ID 1UYP) (Alberto *et al.*, 2004; Alberto *et al.*, 2006), **B:** *Arabidopsis thaliana* (PDB ID 2AC1) (Verhaest *et al.*, 2006), **C:** *Aspergillus awamori* (PDB ID 1Y9G) (Nagem *et al.*, 2004), **D:** *Chicorium intybus* (PDB ID 2ADD) (Verhaest *et al.*, 2007b). BT1760 has low identities to all other sequences (less than 14 % for all), while BT1759 and BT1765 have 51 % identity. BT3082 is most similar to BT1759 (34 %).

4.2.1 Cloning and expression of the GH32s

4.2.1.1 Signal peptide prediction

The four *Bacteroides* GH32 proteins were predicted by SMART and SignalP to have signal peptides and are therefore either secreted into the environment or found in the periplasm of the gut symbiont. BT1760 is a predicted lipoprotein using LipoP and is therefore likely positioned outside the cell embedded in the outer membrane (Figure 4. 4). The SignalP prediction for a signal peptide in BT1765 is unconvincing, suggesting it may be a cytoplasmic enzyme.

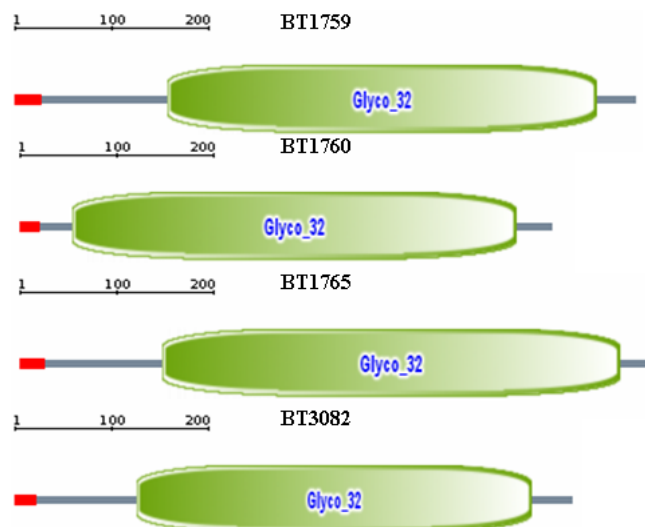


Figure 4. 4 Signal peptide prediction of GH32 members of *B. thetaiotaomicron*.

The red domain at the N-terminal of each protein represents a likely signal peptide. This was confirmed using SignalP with all but BT1765 which had a very poor prediction. The green box is the region of each protein which is conserved in all GH32 members and includes the central catalytic domain. The gray lines indicate parts of the sequence which do not contain domains recognised by SMART. The N-terminal of BT1759, BT1765, and BT3082 are longer than in BT1760, indicating the putative extra N-terminal domain of these proteins.

4.2.1.2 Primers and vectors

BT1759 [29-610] was cloned into pET22b (C-His-tag). BT1760 [22-523] was cloned into pET28b (C-His-tag) and pET32b (N-Trx-His-tag plus C-His-tag) individually. BT1765 [27-627] was cloned into pET28b (C-His-Tag). BT3082 [24-548] was expressed in minipRSETA (N-His-tag). All primers were designed in MWG, with appropriate restriction site flanked (Table 4. 2).

Primer ID	Sequence (5'----->3')	Note
BT1759-1	CTCCCATGGATGCGGATTCTCCTTGC	F Nco I
BT1759-2	CCGCTCGAGTTTCTTCAGTTTGTAAC	R Xho I
BT1760-1	CTCCCATGGATAGTGACGAGACTGAC	F Nco I
BT1760-2	CCGCTCGAGATAAGTGCTTACCTG	R Xho I
BT1765-1	CTCCCATGGATTATTGCCAGTCGACC	F Nco I
BT1765-2	CCGCTCGAGTAAACCTAATCTATACAC	R Xho I
BT3082-1	CTCGGATCCGGAGAAGTATCTTTTAAAATAACCAAGC	F BamH I
BT3082-2	CTCGAATTCCTACCAAATGGATTCTACGGAAAAGAC	R EcoR I

Table 4. 2 Primers for GH32 wild types.

The restriction sites are listed in the last column. F: Forward primers, R: Reverse primers.

4.2.1.3 Proteins expression

His-tagged GH32 enzymes were over expressed in *E. coli* BL21 cell strain and then purified in a single step by IMAC (Figure 4. 5). All proteins could be produced in a soluble form and were dialysed into 20mM Tris-HCl, pH 8.0 buffer for enzyme assays.

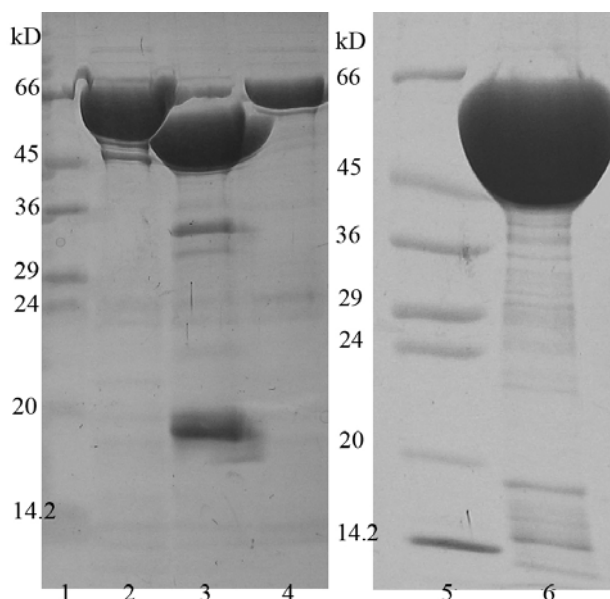


Figure 4.5 IMAC purified GH32s from *B. thalassiosira* used in this study

Lane 1: Marker, Lane 2: BT1759-pET22b (66.3 kD, pI 5.9), Lane 3: BT1760-pET28b (57.0 kD, pI 5.20), Lane 4: BT1765-pET28b (68.7 kD, pI 5.2). Lane 5: Marker, Lane 6: BT3082-minipRSETA (62.5 kD, pI 6.7).

4.2.2 Biochemical characterisation of the GH32s

4.2.2.1 Enzyme activity against fructans and fructoligosaccharides

The activity of the enzymes against a range of β -fructan oligo and polysaccharides was assessed using the DNSA method (section 2.40). The substrates include the levan (from *Serratia levanicum*), inulin (from chicory), sucrose (K2), Kestose (K3), Kestotetraose (K4), and Kestopentaose (K5). K2, K3, K4, and K5 are oligosaccharides of inulin, which are unbranched β -2,1 linked fructose with a terminal glucose (Figure 1.4). Levan oligosaccharides consist of β -2,6 linked fructose with no terminal glucose and are therefore reducing sugars which could not be used for DNSA.

The kinetic parameters of enzymic activity against a range of substrates are listed as Table 4. 3. Because the polysaccharide concentrations are in mg/ml, the results for inulin and levan are not directly comparable. The concentrations for oligosaccharides are in molar and thus the results for them are comparable. The results show that all except BT1760 could digest both β -2,1 linked, and β -2,6 linked fructans, while BT1760 could only digest levan. BT1759 and BT3082 are not sensitive to the β -2,1 chain length. BT1765 prefers short chain fructans, especially sucrose and has only very low activity on polysaccharides. Indeed the activity of BT1765 against inulin and levan was too low to accurately determine K_m for this enzyme.

	BT1759			BT1760			BT1765			BT3082		
	<i>K_{cat}</i> (min ⁻¹)	<i>K_m</i> ^a	<i>K_{cat}/K_m</i>	<i>K_{cat}</i> (min ⁻¹)	<i>K_m</i>	<i>K_{cat}/K_m</i>	<i>K_{cat}</i> (min ⁻¹)	<i>K_m</i>	<i>K_{cat}/K_m</i>	<i>K_{cat}</i> (min ⁻¹)	<i>K_m</i>	<i>K_{cat}/K_m</i>
Inulin	26.5 ± 5.3	0.07 ± 0.0	379	NA ^b			5.9	ND ^c		88.1 ± 5.3	0.1 ± 0.0	881
Levan	7.7 ± 0.5	0.05 ± 0.0	154	331.5 ± 42.8	0.3 ± 0.0	1105	1.1	ND		31.7 ± 4.5	0.4 ± 0.1	79.3
Sucrose	336.0 ± 2.0	0.9 ± 0.1	373	NA			1759.0 ± 531.1	0.2 ± 0.0	8795	121.8 ± 3.8	0.1 ± 0.0	1353
K3	34.0 ± 11.2	0.1 ± 0.0	378	NA			155.0 ± 41.4	0.2 ± 0.0	775	149.5 ± 25.6	0.1 ± 0.0	1869
K4	47.8 ± 19.6	0.1 ± 0.0	478	NA			115.9 ± 8.4	0.1 ± 0.0	1288	104.8 ± 12.8	0.1 ± 0.0	1497
K5	ND			NA			99.1 ± 17.8	0.2 ± 0.0	496	147.6 ± 14.0	0.1 ± 0.0	164

Table 4. 3 Kinetic parameters for *B. thtaiotaomicron* GH32 enzymic activities against oligo and polysaccharides determined by reducing sugar assays.

Assays were carried out at least in triplicate to calculate standard deviations for each. ^a*K_m* for polysaccharides is in mg/ml, for oligosaccharides it is mM. ^bNo activity. ^cNot determined.

4.2.2.2 Product analysis by HPLC

HPLC was used to analyse the products of enzyme digestion against inulin, levan and oligosaccharides of these fructans. Levan oligosaccharides (L2, L3, L4, and L5) were gifts from Dr James Flint and consist of β -2,6 linked fructose with no terminal glucose. The oligosaccharides were ~80-90 % pure with the main impurities being n-1 and n+1 sugars. The results for inulin and levan digestion reveal that BT1759 and BT3082 are typical exo-acting fructosidases that release only fructose from β -2,1 and β -2,6 linked fructans (Figure 4. 6A, B and D). The extra peak resulting from glucose in the BT3082 traces at 15 and 135 min (Figure 4. 6E) is the result of digestion of the polysaccharide to completion. Interestingly BT1760 displayed a distinctly different product profile from the other GH32s. Digestion of levan revealed a mixture of different size oligosaccharides characteristic of an endo-type activity, with the major products initially being L2 and L3 (Figure 4. 6C). However, unlike many endo acting enzymes BT1760 also produced significant amounts of the monosaccharide – in this case fructose.

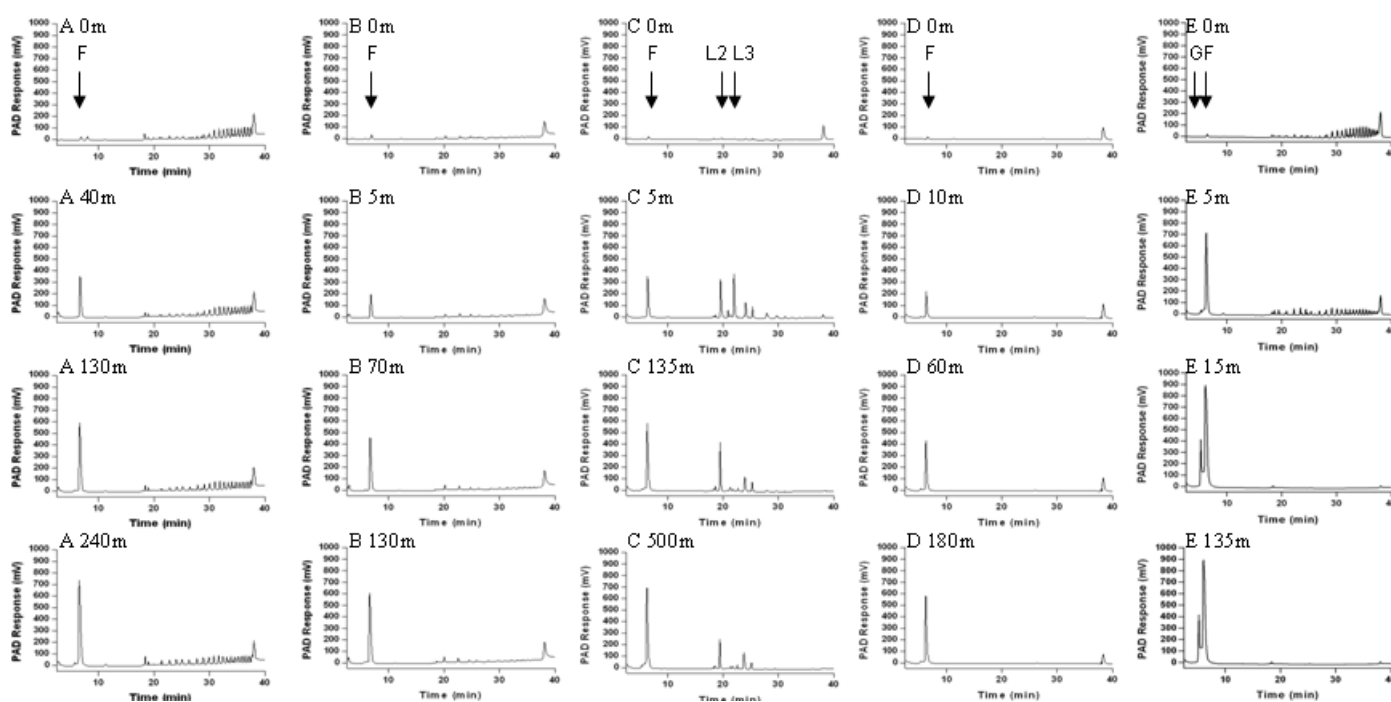


Figure 4. 6 HPLC data of *Bacteroides* GH32s against polymeric fructans.

A: BT1759 against inulin, 50 % gradient. **B:** BT1759 against levan, 50 % gradient. **C:** BT1760 against levan, 25 % gradient. **D:** BT3082 against levan, 15 % gradient (default). **E:** BT3082 against inulin, 15 % gradient. G is glucose, F is fructose, and L2-L3 are levan oligos. Reaction time points are from 0-500 min. Note: Inulin has a much shorter average chain length than levan, resulting in a large number of peaks being visible in the inulin traces.

Analysis of the product profile of the GH32s against oligosaccharides showed a similar pattern with BT1759, BT1765, and BT3082 releasing fructose from kesto- (β -2,1) and levan- (β -2,6)

oligosaccharides (example data shown in Figure 4. 7). Some minor peaks appear in the digestions that are not the linear n+1 or n-1 oligosaccharides that are the major products, but that are substrates for the enzymes (See for example Figure 4. 7B around K4). It is not clear what these peaks are, but they may be branched oligosaccharides produced by transglycosylation during the process of enzymic action.

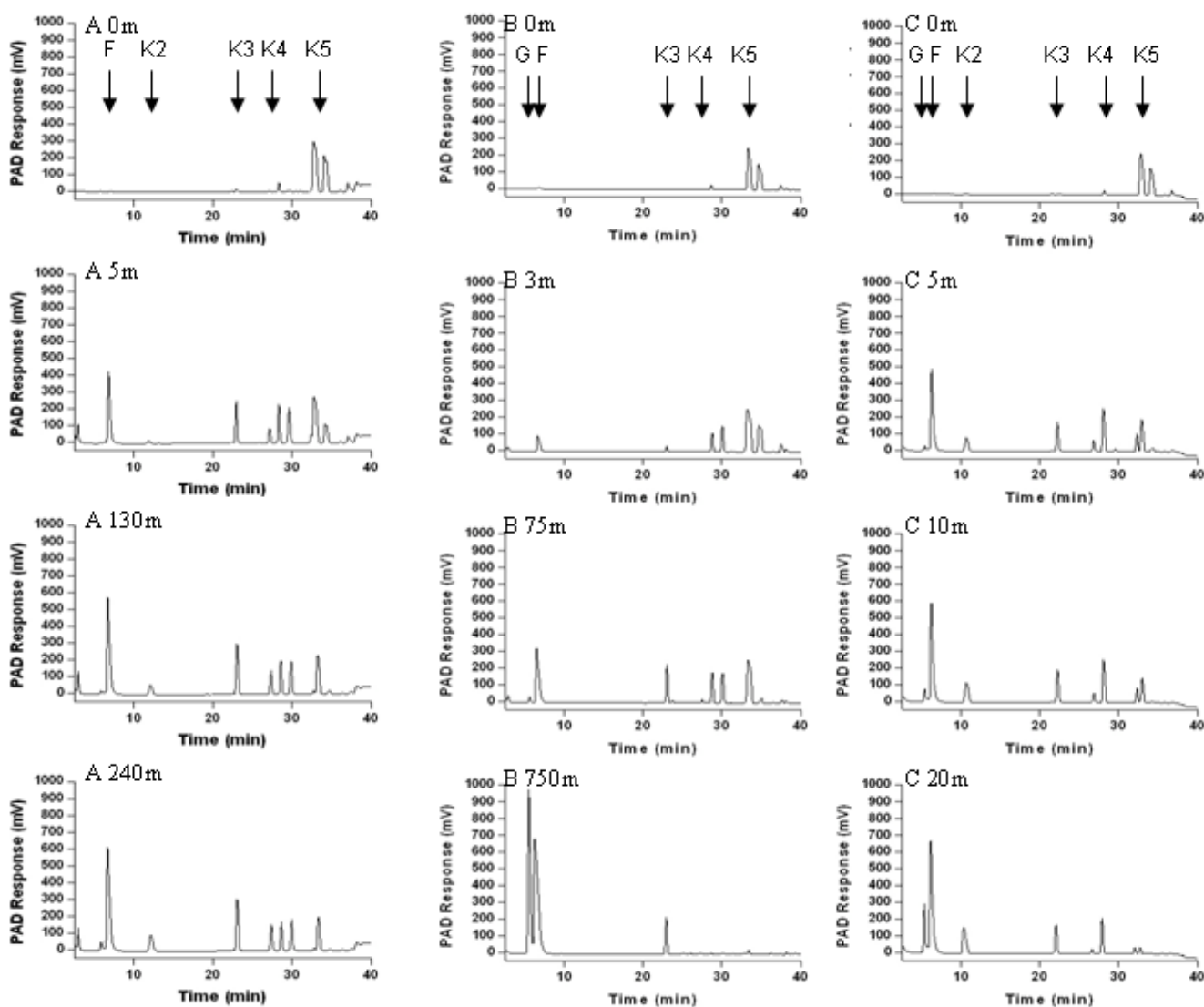


Figure 4. 7 HPLC data show product profile of BT1759, BT1765, and BT3082 against kestopentaose.

A: BT1759, **B:** BT1765, 15 % **C:** BT3082. G means glucose, F is fructose and K2-K5 are oligosaccharides of inulin. Reaction time points are present as 0 to 750 min. Only K5 with each enzyme is displayed here as an example. The peak bigger than K5 may be K6 because no K6 standard used in this study, while it is unclear what the two small peaks around K4 are.

BT1760 was able to digest all levan-oligosaccharides tested from L2-L5. Activity against levanbiose was very low as would be expected of an endo acting enzyme (Figure 4. 8A). Against L5, fructose, L2, and L3 were released from the beginning of the reaction (10 min), rather than just L4 and Frc, again supporting the mixture of endo and exo action for this enzyme (Figure 4. 8D).

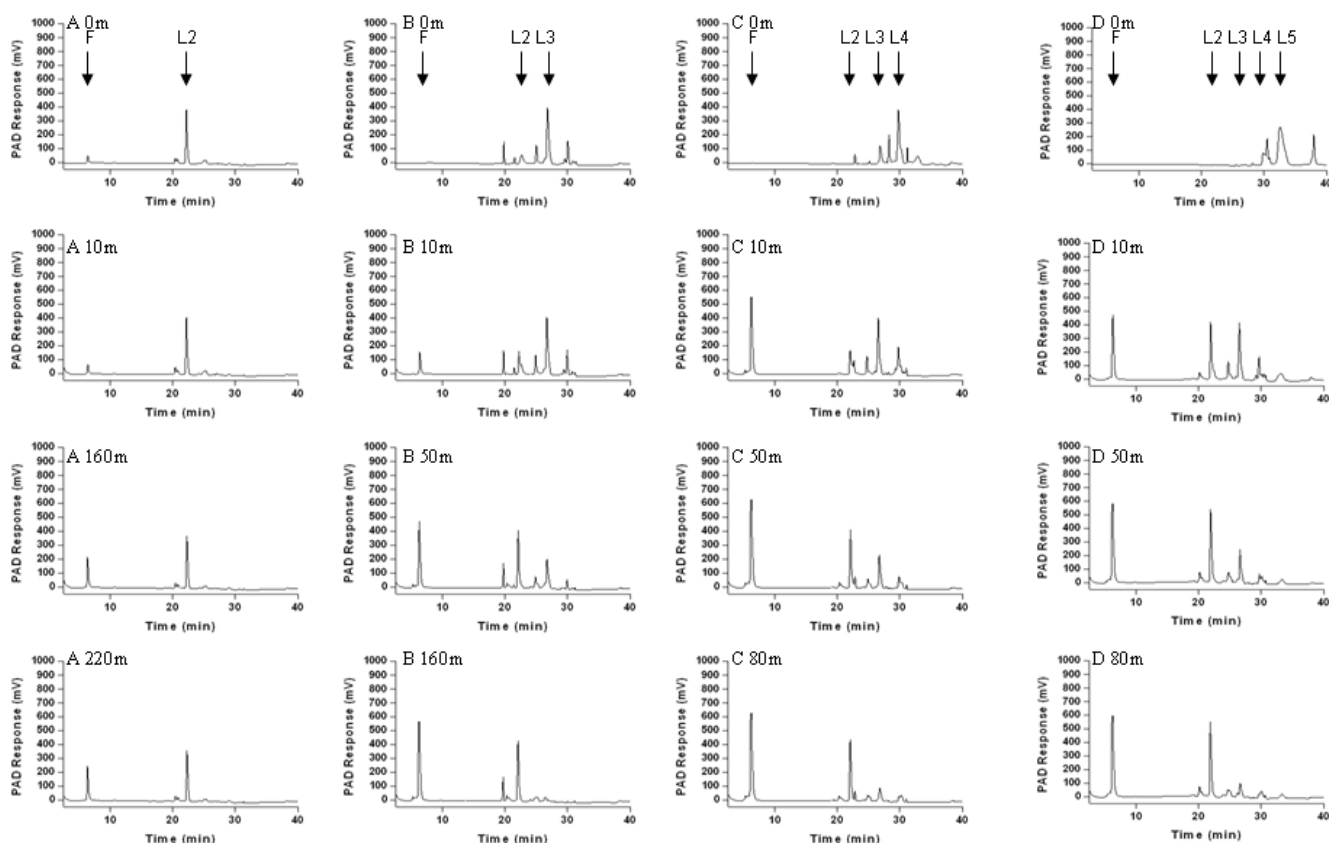


Figure 4. 8 BT1760 digest levan oligos.

A: L2, B: L3, C: L4, D: L5. F means fructose. Reaction time points from 0-220 min are shown.

4.2.2.3 TLC

TLC was carried out as alternative method of assessing the product profile of BT1760 and BT3082 against polymeric levan. From the TLC data, the difference in mode of action between BT1760 and BT3082 can be easily observed (Figure 4. 9). BT1760 could digest the levan into different lengths from the beginning, with L3 and L2 as the major products with some L4 earlier and fructose later in the reaction, while BT3082 just released fructose. These data confirm BT1760 acts as an endo-levanase. BT3082 is an exo acting β -fructosidase.

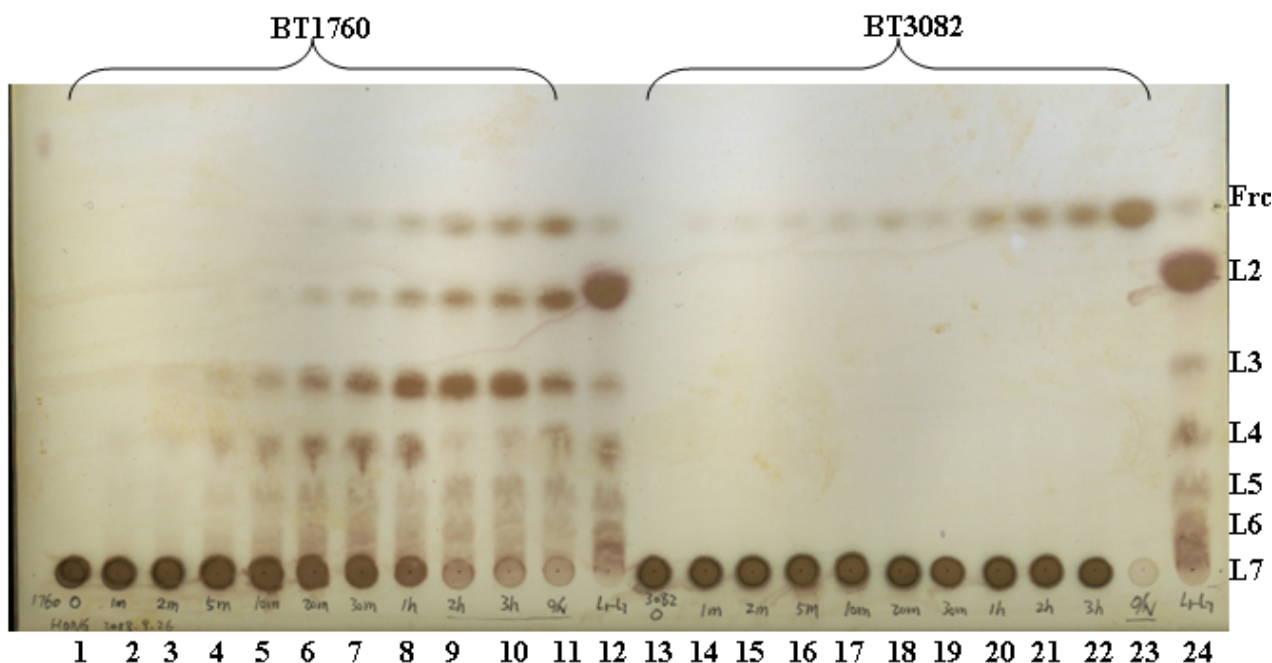


Figure 4. 9 TLC shows the process of levan digestion against BT1760 and BT3082 separately.

Lanes 1-11: BT1760 digestion at time points: 0m, 1m, 2m, 5m, 10m, 20m, 30m, 1h, 2h, 3h, overnight, Lanes 12 and 24: Levan oligosaccharides marker. Lane 13-23: BT3082 digestion at time points: 0m, 1m, 2m, 5m, 10m, 20m, 30m, 1h, 2h, 3h, overnight. The levan concentration was 1 % for both digestions. BT1760 final concentration was 0.1 μ M. BT3082 final concentration was 1 μ M.

4.2.3 Structural analysis of *B. thetaiotaomicron* GH32s

Structural studies were undertaken with all four GH32s from *B. thetaiotaomicron* to investigate the molecular details of substrate binding and catalysis. Although extensive crystallization screens were set up with BT1759 and BT1765, no shootable crystals were obtained. Structures were determined for BT1760 and BT3082 by Dr. Susan Firbank and they are detailed as below.

4.2.3.1 BT3082 fructosidase

4.2.3.1.1 Structure of BT3082

The nucleophile mutant D131A was obtained by site-directed mutagenesis with primers BT3082-3 (5'-GGCTGGATCAATGCCCTAACGGATTAATC-3'), and BT3082-4 (5'-GATTAATCCGTTAGGGGCATTGATCCAGCC-3'). This mutant lost all enzymic activity against fructans as assessed by DNSA analysis and could form shootable long-needle crystals in condition: 0.1 M MES pH 6.5, 12 % w/v PEG 20000. The structure of the D131A nucleophile mutant of BT3082 in complex with kestose (K3) was solved by molecular replacement against the invertase from *Thermotoga maritima* at a maximal resolution of 2.2 Å (Figure 4. 10). The four copy structure unit has cell parameters $a = 94.4$ Å, $b = 113.4$ Å, $c = 134.1$ Å, and $\beta = 98.87^\circ$, containing 19320 atoms.

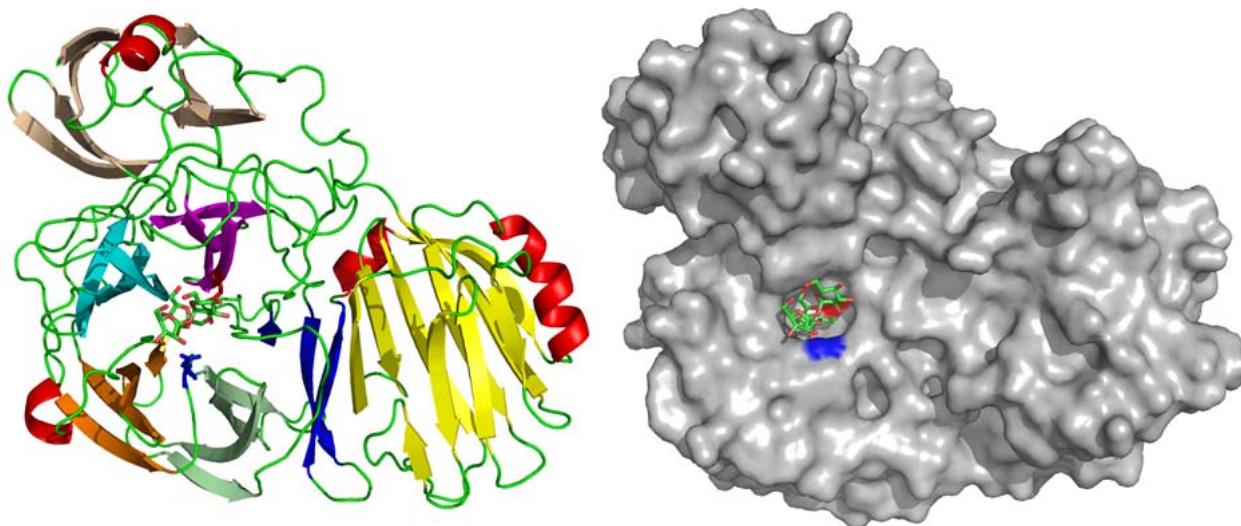


Figure 4. 10 Structure of BT3082 nucleophile mutant (D131A) with kestose bound

Left: Cartoon representation shows the structure of BT3082 nucleophile mutant D131A with ligand K3 binding in the middle of 5-blade β -propeller (purple, cyan, orange, palegreen, and blue). α -helices are colored red and loops are colored green. The β -strands of the N-terminal domain are shown in wheat. A131 (ala mutant of nucleophile) is blue stick and E301 (acid/base) is red stick. **Right:** Surface representation shows the active site of the enzyme is a pocket with bound K3 in green (A131 in blue and E301 in red).

According to the mutant structure, BT3082 wild type structure was solved at a maximal resolution of 2.5 Å (Figure 4. 11). The four copy structure unit has cell parameters $a = 94.3$ Å, $b = 114.3$ Å, $c = 135.1$ Å, and $\beta = 99.23^\circ$, containing 17012 atoms. A monomer BT3082 structure comprises three modules: an extra N-terminal module (residues 24-107), linked to a catalytic five bladed β -propeller (residues 128-408) by a 10-residue loop, and then linked to a C-terminal β -sandwich (residues 416-548) by a 7-residue loop. The N-terminal domain is completely novel within the GH32 family. The closest distance of two oxygens on side-chains of these two residues Asp131 and Glu301 is 5.89 Å.

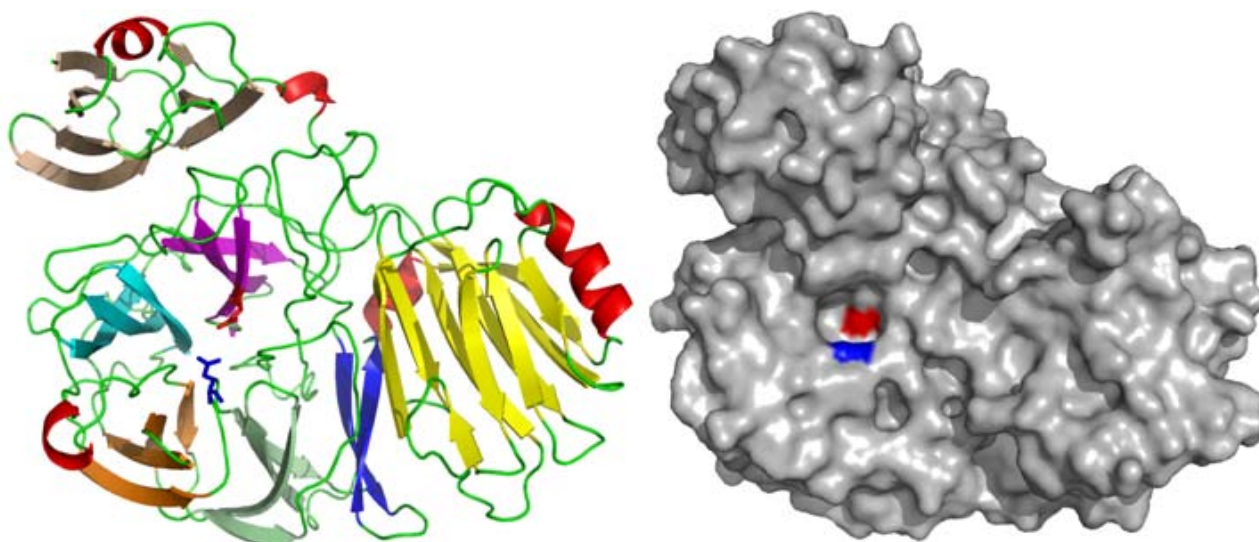


Figure 4. 11 Structure of BT3082 fructosidase

Left: This structure has an extra N-terminal of two α -helices (red) and 7 β -sheets, linked a 5-fold β -propeller (purple, cyan, orange, palegreen, and blue), and a C-terminal β -sandwich (yellow). Loops are colored green. The first helix is between the last two β -sheets, and the second helix is very small between extra module and catalytic module. The third helix is between blade 2 and blade 3, while the fourth links the catalytic domain and β -sandwich. There is an extra helix in β -sandwich on the very outside. The two catalytic residues are presented as red (D131) and blue (E301) sticks and are located on the first blade (purple) and the loop between the third blade (orange) and fourth blade (palegreen).

Right: Surface representation shows the surface of BT3082 as grey, while the D131 (red) is at the bottom of a pocket and E301 (blue) is on the sidewall of the same pocket. The closest distance of the two catalytic residues is 5.89 Å between two oxygens of side-chains of the catalytic Asp and Glu.

This wild type structure is identical to the D131A mutant except the mutated residue (Figure 4. 12).

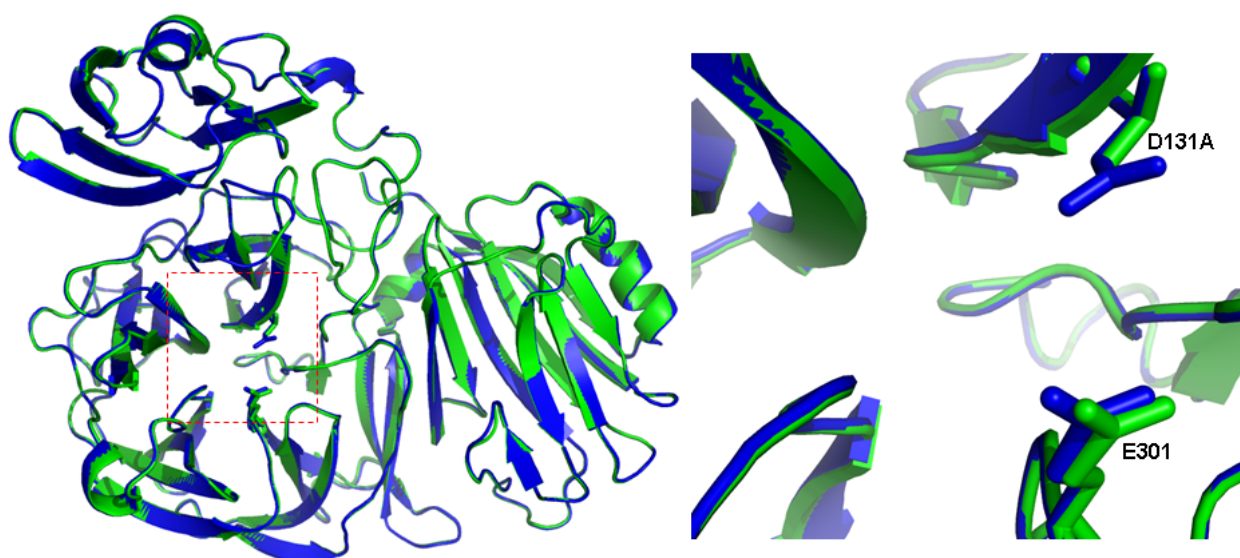


Figure 4. 12 Overlay of BT3082 WT with D131A.

Left: BT3082 WT was colored green, while D131A was colored blue. The ligand K3 in D131A was not shown. **Right:** Zoom in the red box of left panel to show the D131 (green stick) in WT was replaced by alanine (blue stick) in mutant, while E301 is same location in both.

A DALI protein structure database search with the N-terminal domain of BT3082 only indicated that the protein has little similarity to other known structures (Figure 4. 13). The most similar structures are sensory rhodopsin transducer proteins although the Z scores are low.

No:	Chain	Z	rmsd	lali	nres	%id	Description
1:	2iia-A	5.7	2.4	64	89	2	SENSORY RHODOPSIN TRSANSDUCER PROTEIN;
2:	2ii9-C	5.6	2.4	64	93	2	SENSORY RHODOPSIN TRANSDUCER PROTEIN;
3:	2ii9-B	5.6	2.4	64	92	2	SENSORY RHODOPSIN TRANSDUCER PROTEIN;
4:	2ii9-A	5.6	2.4	64	92	2	SENSORY RHODOPSIN TRANSDUCER PROTEIN;
5:	2ii8-E	5.5	2.4	64	92	2	ANABAENA SENSORY RHODOPSIN TRANSDUCER
6:	2ii8-C	5.5	2.4	64	92	2	ANABAENA SENSORY RHODOPSIN TRANSDUCER
7:	2ii7-G	5.5	2.4	64	91	2	ANABAENA SENSORY RHODOPSIN TRANSDUCER
8:	2ii7-F	5.5	2.4	64	91	2	ANABAENA SENSORY RHODOPSIN TRANSDUCER
9:	2ii9-D	5.4	2.4	64	94	2	SENSORY RHODOPSIN TRANSDUCER PROTEIN
10:	2ii8-H	5.4	2.5	64	92	2	ANABAENA SENSORY RHODOPSIN TRANSDUCER

Figure 4. 13 DALI results for BT3082 N-terminal domain.

Dali matches the top 10 structures sorted by Z-score. These 10 structures all are sensory. Chain: PDB entry code plus chain identifier. Z: normalized Z-score that depends on the size of the structures. The program optimises a weighted sum of similarities of intramolecular distances. rmsd: root-mean-square deviation of C-alpha atoms in the least-squares superimposition of the structurally equivalent C-alpha atoms. The program does not optimise rmsd, this is only reported for your information. lali: number of structurally equivalent residues. nres: number of amino acids in the protein. %id: percentage of identical amino acids over all structurally equivalent residues. Description: the COMPND record from the PDB entry.

DALI results with the C-terminal domain of BT3082 only reveal that it is related to sugar binding lectins though the highest Z score is only 11.3 (Figure 4. 14).

No:	Chain	Z	rmsd	lali	nres	%id	Description
1:	2yv8-A	11.3	2.8	117	151	8	GALECTIN-8 VARIANT;
2:	2yxs-A	11.2	2.8	116	150	8	GALECTIN-8 VARIANT;
3:	2d6k-A	11.2	2.6	114	147	7	LECTIN, GALACTOSE BINDING, SOLUBLE 9;
4:	2d6p-A	11.1	2.6	114	144	7	LECTIN, GALACTOSE BINDING, SOLUBLE 9;
5:	1y9m-A	11.1	3.0	121	517	15	EXO-INULINASE;
6:	2qqv-A	11.0	2.9	124	537	15	BETA-FRUCTOFURANOSIDASE;
7:	2d6p-B	11.0	2.6	113	147	7	LECTIN, GALACTOSE BINDING, SOLUBLE 9;
8:	2d6o-X	11.0	2.7	114	151	8	LECTIN, GALACTOSE BINDING, SOLUBLE 9;
9:	2d6n-B	11.0	2.7	115	153	7	LECTIN, GALACTOSE BINDING, SOLUBLE 9;
10:	1y9g-A	11.0	3.0	121	517	17	EXO-INULINASE;

Figure 4. 14 DALI results for BT3082 β -sandwich.

Dali matches the top 10 structures sorted by Z-score. Chain: PDB entry code plus chain identifier. Z: normalized Z-score that depends on the size of the structures. The program optimises a weighted sum of similarities of intramolecular distances. rmsd: root-mean-square deviation of C-alpha atoms in the least-squares superimposition of the structurally equivalent C-alpha atoms. The program does not optimise rmsd, this is only reported for your information. lali: number of structurally equivalent residues. nres: number of amino acids in the protein. %id: percentage of identical amino acids over all structurally equivalent residues. Description: the COMPND record from the PDB entry.

DALI searches revealed that the central β -propeller catalytic domain of BT3082 has high similarity to other solved GH32 proteins with Z score up to 39.9 (Figure 4. 15).

No:	Chain	Z	rmsd	lali	nres	%id	Description
1:	1y9m-A	39.9	1.7	299	517	42	MOLECULE: EXO-INULINASE;
2:	1y9g-A	39.7	1.8	300	517	42	MOLECULE: EXO-INULINASE;
3:	1y4w-A	39.7	1.8	300	517	42	MOLECULE: EXO-INULINASE;
4:	2ac1-A	34.9	2.1	284	537	33	MOLECULE: INVERTASE;
5:	2qqu-A	34.5	2.2	284	535	33	MOLECULE: BETA-FRUCTOFURANOSIDASE;
6:	1st8-A	34.3	2.0	283	537	31	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
7:	2aez-A	34.2	2.1	283	536	31	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
8:	2ade-A	34.2	2.1	284	537	31	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
9:	2add-A	34.1	2.1	284	537	31	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
10:	1uyp-B	34.1	1.7	266	432	38	MOLECULE: BETA-FRUCTOSIDASE;

Figure 4. 15 DALI results for BT3082 catalytic domain.

Dali matches the top 10 structures sorted by Z-score. These 10 structures all are from GH32 members. Chain: PDB entry code plus chain identifier. Z: normalized Z-score that depends on the size of the structures. The program optimises a weighted sum of similarities of intramolecular distances. rmsd: root-mean-square deviation of C-alpha atoms in the least-squares superimposition of the structurally equivalent C-alpha atoms. The program does not optimise rmsd, this is only reported for your information. lali: number of structurally equivalent residues. nres: number of amino acids in the protein. %id: percentage of identical amino acids over all structurally equivalent residues. Description: the COMPND record from the PDB entry.

4.2.3.1.2 BT3082 active site

Analysis of the complex of BT3082 D131A with kestose allowed us to identify residues that contact the substrate in the active site and may therefore be important in binding and catalysis (Figure 4. 16 and Table 4. 4). There are 11 hydrogen bonds formed between ligand K3 and BT3082 D131A. Ten of the 11 are formed with the fructose bound at the -1 position: O1 forms one with N130, O3 forms three hydrogen bonds with R256, D257, and E301 individually, O4 forms two hydrogen bonds with S188 and D257 separately, and O6 forms two hydrogen bonds with N130, one with Q147 and one with W155, with an extra bond formed by the O3 on the +1 fructose ring with catalytic residue E301 (Figure 4. 16E). There are four more aromatic residues (W120, F187, Y353, and W373) around kestose which do not form hydrogen bond with substrate, but form the hydrophobic surface of the binding pocket (Figure 4. 16D). Five other residues (A131, M158, D254, C302, and A354) around -1 fructose of kestose are around the bottom of the pocket (Figure 4. 16B). A131 is mutated from D131 which should have hydrogen bond with kestose if in wild type. So totally, there are 13 residues involved in the formation of catalytic pocket.

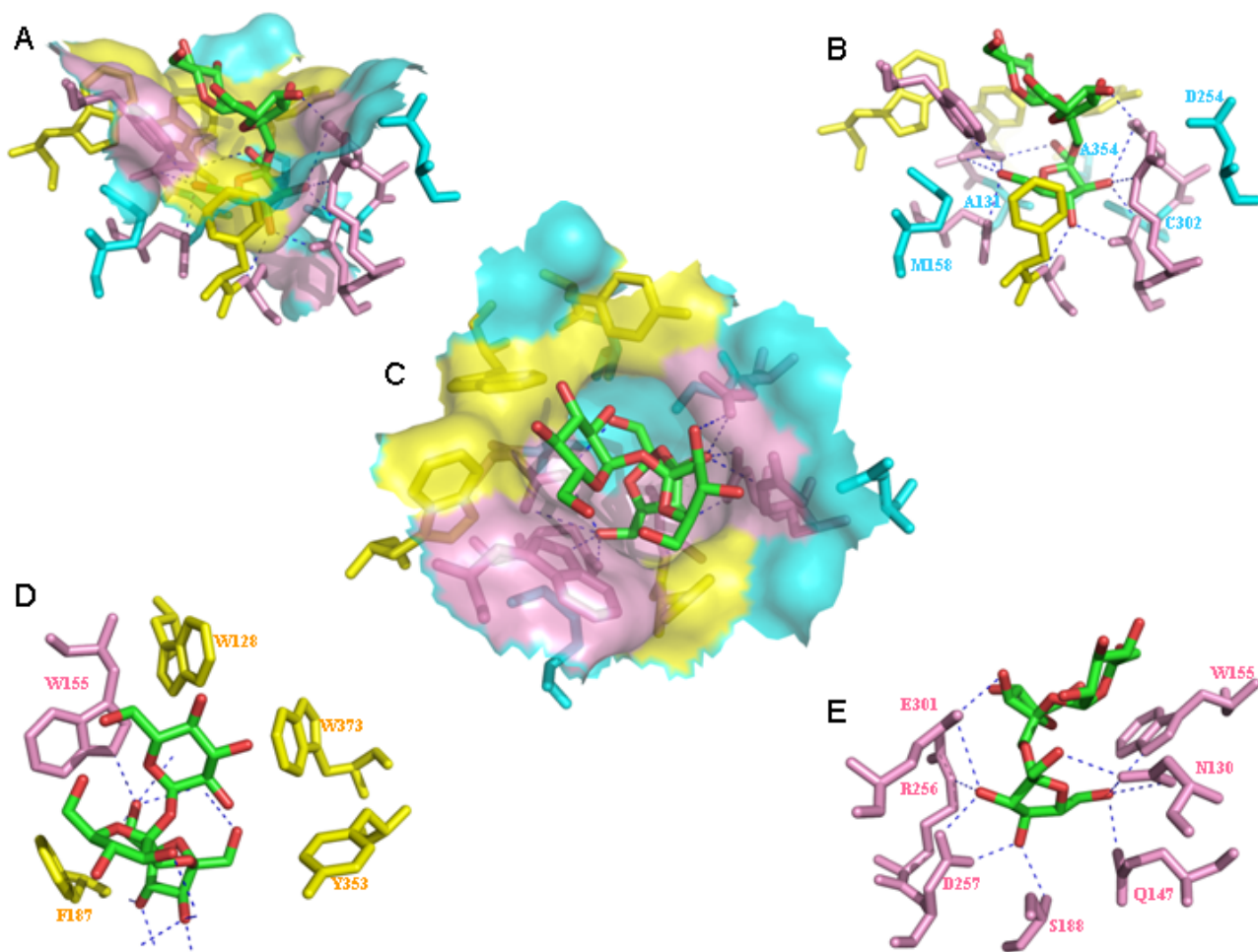


Figure 4. 16 BT3082 D131A binding site to kestose.

There are 16 residues are found around the ligand K3. Residues forming hydrogen bonds with kestose are drawn as pink sticks (panels A-E). Amino acid side chains making hydrophobic contacts with kestose are shown as yellow sticks (panels A-D) while other residues forming the active site pocket are shown in cyan (panels A-C). **A:** Surface representation from the side angle shows the binding pocket. **C:** surface representation from the top of the binding pocket. It is easy to see the pink residues are on two sides of the pocket, while aromatic residues are between the pink residues to form a whole pocket with a few cyan residues to form the pocket bottom. Panels **B**, **D** and **E** are different views of the active site to illustrate the interactions made by 3082 with kestose.

Kestose atom	BT3082 D131A residue	Distance (Å)
Frc1 O1	N130-O	3.19
Frc1 O3	R256-N	3.07
	D257-O	2.62
	E301-O	3.56
	C302-C	3.90
	W373-C	3.66
	Y353-C	3.85
Frc1 O4	S188-N	2.92
	D257-O	2.51
Frc1 O5	A131-N	5.05
Frc1 O6	N130-O	3.09
	N130-N	3.45
	Q147-N	2.76
	W155-N	2.90
Frc1 C6	F187-C	3.98
Frc2 O3	E301-O	3.06
Frc2 O4	D254-O	4.45
Frc2 O4	Q221-N	5.29
Glc O5	W128-C	3.88

Table 4. 4 Hydrogen bonding and close contacts between BT3082 active site residues and K3.

Hydrogen bonds are bolded. Frc1 is the -1 site at the bottom of the pocket, while the Frc2 is the +1 site and Glc has no hydrogen bond formed.

According to this analysis above, a group of 16 amino acids of BT3082 was chosen and mutated to alanine or phenylalanine (F) respectively to assess their importance in enzyme activity (Table 4. 5). All mutants could be expressed in soluble form. Activity was assessed by reducing sugar assay with a single high concentration of each substrate. The initial rates were calculated for each and then the molar amount of each enzyme was normalized to wild type to get a ratio of activity against each substrate compared to wild type, which was set at 100 %.

Primer ID	Sequence (5'----->3')	Note
BT3082-5	GTTACAGGCTTTTGGGCATGTCCCGAATTG	E301A F
BT3082-6	CAATTCGGGACATGCCCAAAGCCTGTAAC	E301A R
BT3082-9	GAGAGGCTGGATCGCTGACCCTAACGGATTAATC	N130A F
BT3082-10	GATTAATCCGTTAGGGTCAGCGATCCAGCCTCTC	N130A R
BT3082-11	CACCTGTTCTACGCGCACAATCCTTTC	Q147A F
BT3082-12	GAAAGGATTGTGCGCGTAGAACAGGTG	Q147A R
BT3082-13	CTTTCGAAAGAGACGCGGAGAATATGCAC	W155A F
BT3082-14	GTGCATATTCTCCGCGTCTCTTTCGAAAG	W155A R
BT3082-15	CCTTTCGAAAGAGACTTTGAGAATATGCACTGGG	W155F F
BT3082-16	CCCAGTGCATATTCTCAAAGTCTCTTTCGAAAGG	W155F R
BT3082-17	CATTTAGGAACGATGGCCTCCGGTTCGGCAG	F187A F
BT3082-18	CTGCCGAACCGGAGGCCATCGTTCCTAAATG	F187A R
BT3082-19	CAGCAGTGACAGAGCAGTGCAGGGGATAG	Q221A F
BT3082-20	CTATCCCCTGCACTGCTCTGTCCTGCTG	Q221A R
BT3082-23	CAGGACACACGCGCCCGAAAGTGTTCC	D257A F
BT3082-24	GAACACTTTCGGGGCGCGTGTGTCCTG	D257A R
BT3082-25	GTATCCAGATAGGCGCGGGAAGAATCTCC	W373A F
BT3082-26	GGAGATTCTTCCGCGCCTATCTGGATAC	W373A R
BT3082-27	CTACCCGGAGAGGCGCGATCAATGACCCTAAC	W128A F
BT3082-28	GTTAGGGTCATTGATCGCGCCTCTCCGGGTAG	W128A R
BT3082-29	CTGGATCAATGACGCAAACGGATTAATC	P132A F
BT3082-30	GATTAATCCGTTTTCGCTCATTGATCCAG	P132A R
BT3082-31	GGCTTTTGGGAAGCTCCCGAATTGTTTG	C302A F
BT3082-32	CAAACAATTCGGGAGCTTCCCAAAGCC	C302A R
BT3082-33	CTTTTGGGAATGTGCGGAATTGTTTGAAC	P303A F
BT3082-34	GTTCAAACAATTCGCGACATTCCCAAAG	P303A R
BT3082-35	GTTACAGGCTTTGCGGAATGTCCCGAATTG	W300A F
BT3082-36	CAATTCGGGACATTCCGCAAAGCCTGTAAC	W300A R
BT3082-37	GTAAAGGTCTGGGTGCAGCGATACTTCCGGTG	Y353A F
BT3082-38	CACCGGAAGTATCGCTGCACCCAGACCTTTAC	Y353A R
BT3082-39	GTAAAGGTCTGGGTGCAAAGATACTTCCGGTG	Y353F F
BT3082-40	CACCGGAAGTATCTTTCACCCAGACCTTTAC	Y353F R

Table 4. 5 Primers of 16 more single mutations of BT3082.

F: Forward primers, R: Reverse primers.

The group of 17 BT3082 mutants including the nucleophile mutant D131A was tested by DNSA method against three ligands: inulin, levan, and sucrose. Generally, the enzymic activity shifts between different mutants and different substrates (Figure 4. 17). When the two catalytic residues Asp131 and Glu301 were mutated to alanine individually, the enzyme was abolished activity completely, which happened to D257A as well. Asp257 is buried at the bottom of the pocket and forms bidentate hydrogen bonds with O3 and O4 of the fructose at the -1 site. This residue is also

highly conserved in other GH32 members and forms part of the RDP motif that is conserved in all enzymes that act on fructose containing substrates (Figure 4. 3). The Arg of this motif has not been tested here as there was a problem with the mutagenesis reaction. The residue is positioned such that it makes up a significant part of the side wall of the active site and therefore likely plays a significant role in catalysis, although it appears to make only a single hydrogen bond with O3 of the Frc at -1. The Pro of the RDP motif was not mutated here, as it is not part of the active site and may play a role in holding Arg256 and Asp257 in the correct position. N130A keeps about 10 % activity against all three substrates. Asn130 makes one hydrogen bond with O1 and two other ones with O6 of Frc1. W155F kept half enzymic activity while W155A just has 17 % ability against sucrose, 4 % to inulin, and 9 % to levan. Q221 is quite far from the active site (5.29 Å), but Q221A lost 91 % activity against sucrose, 42 % against inulin, and 67 % against levan, which indicated that this residue is involved in catalysis. Gln147 is at the bottom of the pocket, and its mutant Q147A almost lost ability against all substrates, Trp128 and Trp373 are at the same side wall of the pocket, their mutant W128A kept around half ability against all substrates, while W373 almost lost ability against all substrates. Pro132 and Pro303 are both in conserved regions close to the nucleophile and acid base residues, respectively (Figure 4. 3). P132A and P303A mutations abolished 60 % to 80 % activity compared to wild type against all three substrates indicating that while they likely play a role in holding the catalysis. Cys302 is closer to catalytic acid/base Glu301 than Pro303, while its mutant C302A lost more activity than P303A. Cys302 and Pro303 are part of the ECP motif that contains the catalytic acid base and is conserved in other GH32 members (Figure 4. 3). Pro303 is not part of the active site per se but likely holds the acid base and Cys302 in the close position for catalysis similar to the location of Pro of the RDP motif discussed above. Trp300 is on the other side of Glu301, while its mutant W300A lost activity against sucrose completely. Against levan and inulin, W300A and C302A both are at same low level (about 4 % and 10 % respectively). Tyr353 is completely conserved in GH32 proteins and is close to O3 of Frc1. Its alanine mutant Y353A is totally dead against sucrose, and has the same low activity level against levan and inulin as W300A and C302A. These data suggest the residue plays an important role in stabilizing the position of the sugar, especially glucose, in the +1 site. It is interesting to note that Y353F has 140 % activity against sucrose and keeps more than half activity against inulin and levan. The hydroxyl of the tyrosine makes no contacts with kestose, but does form a hydrogen bond with the acid base residue. The data from the F187A mutation is also very interesting as it has a slightly higher activity than wild type against inulin, while against sucrose and levan, the activities are 4 % and 18 % of wild type, respectively. It is not clear why this may be the case as the residue forms an integral part of the active site, making contacts with the C4, C5, and C6 of the Frc at -1 and so its loss would be expected to reduce activity on all substrates.

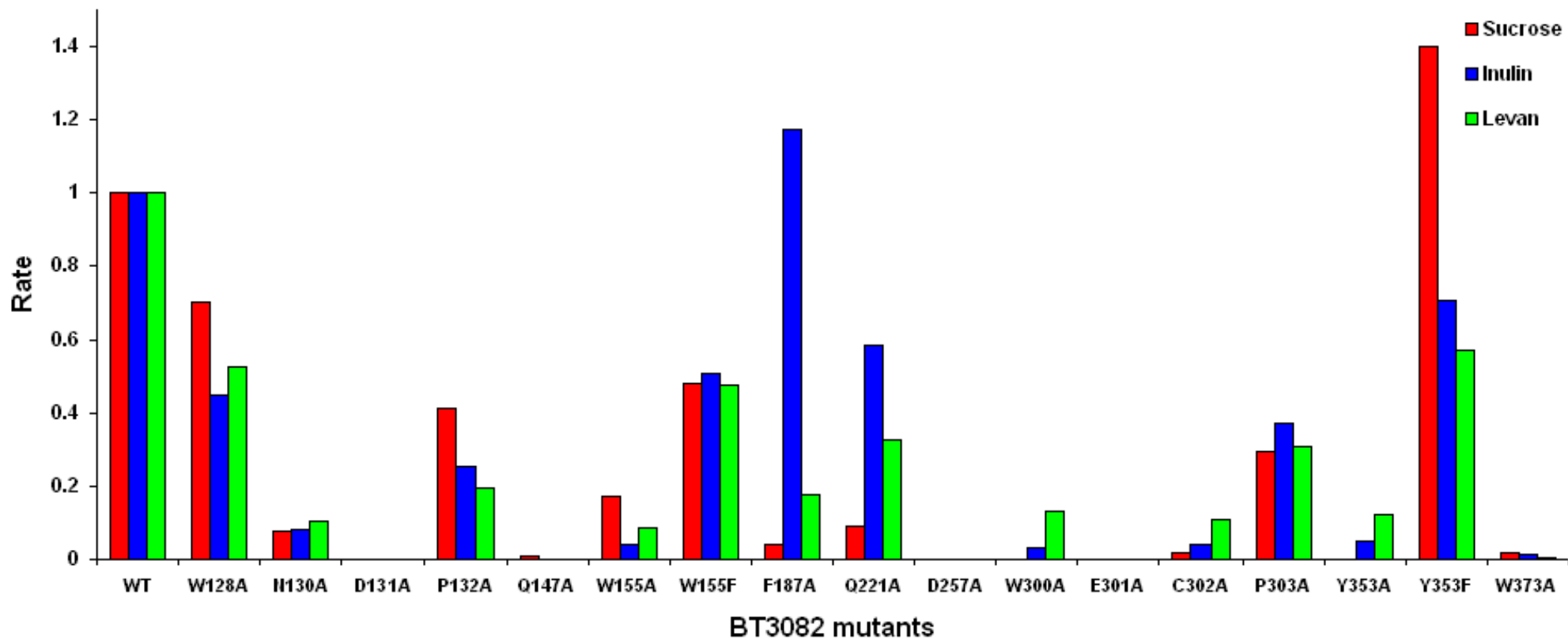


Figure 4. 17 Enzyme activity of the BT3082 active site mutants against sucrose, inulin and levan.

WT activity against each substrate is set at 100 %. Mutant enzymic activities are compared with WT to get the final values. This set of data is not repeated because of time limitation. So there is no error bars shown. More data will add in before publication.

The CD analysis revealed that the overall structures of all the mutants are very similar to wild type suggesting that any loss of activity seen were not due to major changes in conformation (Figure 4. 18).

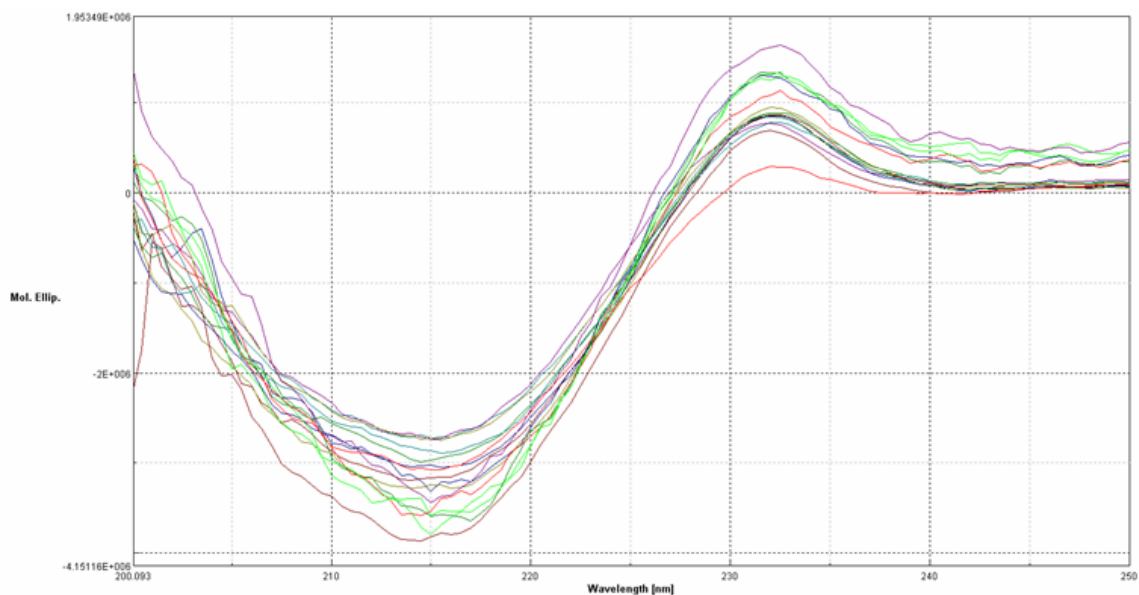


Figure 4. 18 CD spectroscopy data showing the BT3082 WT and 17 mutants. CD data indicated that BT3082 WT and 17 mutants have secondary structure.

4.2.3.1.3 Deletion analysis of BT3082

BT3082 comprises three distinct domains: the central catalytic domain and distinct β -sheet N- and C-terminal domains. A BLAST search indicated that the smaller N-terminal domain appears so far to be unique to *Bacteroides* enzymes (Figure 4. 19). The C-terminal domain is common to all structurally characterised GH32s, but it is not completely conserved in the family (Figure 4. 19).

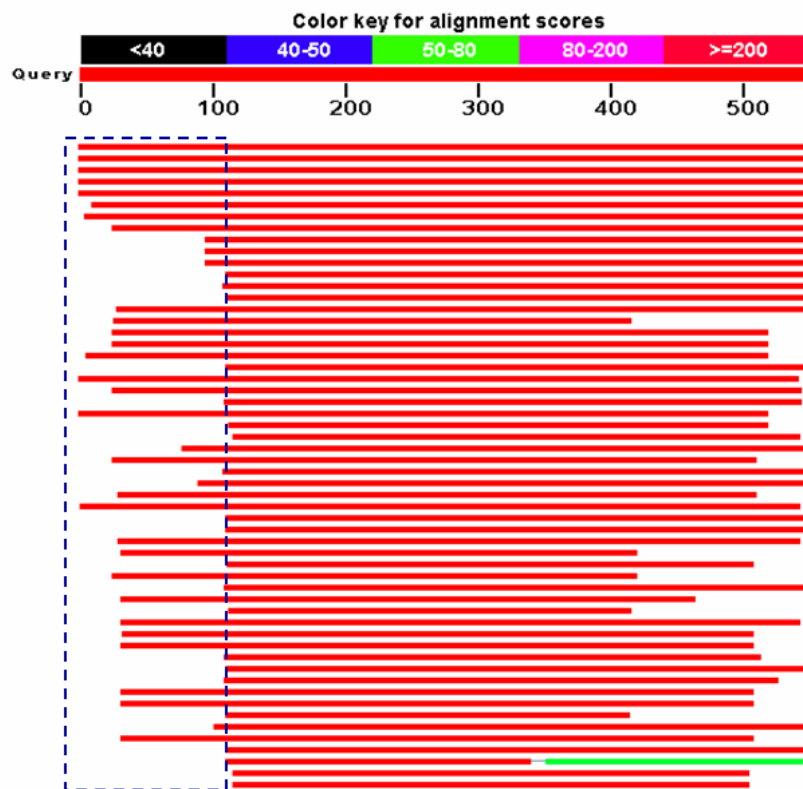


Figure 4. 19 BLAST result for BT3082.

The data revealed the most similar sequences (red lines) to BT3082. The blue box presents some of them have N-extra domain as well which are all from *Bacteroides*.

Based on the structure deletion analysis was carried out to assess the importance of the N and C-terminal domains in enzyme activity. Two primers were designed as BT3082-7: 5'-CTCGGATCCCTGTATAAGGAAAAGAACCG-3' (from 111th aa, fused to *BamH* I site), and BT3082-8: 5'-CTCGAATTCAAGAGGCTTACACAAGGATTCG-3' (to 419th aa, fused to *EcoR* I site). The Δ N BT3082 (residues 111-548) is over expressed as a soluble protein (Figure 4. 20), while Δ C BT3082 (residues 24-419) has no expression (data not shown).

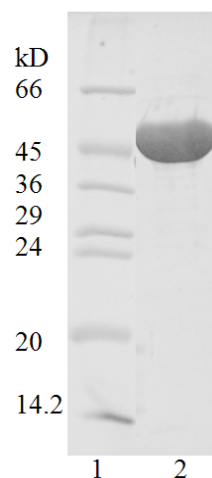


Figure 4. 20 Expression of Δ N BT3082.

Lane 1: Marker, Lane 2: Δ N BT3082. This Δ N BT3082 ([111-548]) was cloned by primers BT3082-2 and BT3082-7. The molecular weigh is 50.5 kD, and the pI is 6.33.

The ΔN BT3082 (residues 111-548) deletion abolished activity against inulin, levan or sucrose. This result indicated that the N-extra is necessary for the enzymic activity, although the function is unknown so far and the C-terminal β -sandwich is also important for the conformational stability as without this domain, the enzyme is either unable to fold or fold aberrantly and gets degraded in the cell.

4.2.3.2 Structure of BT1760 levanase

BT1760 wild type could form shootable long-needle crystals in condition: 0.2 M K/Na tartrate, 0.1 M tri-Sodium citrate, pH 5.6, 1.4 M Ammonium sulfate in the presence of 0.5 M fructose. BT1760 structure was solved as a dimer by SeMet-substituted protein at a maximal resolution of 2.6 Å (Figure 4. 21). The dimer structure unit has cell parameters $a = 176.3$ Å, $b = 176.3$ Å, $c = 215.8$ Å, and $\beta = 90^\circ$, containing 8915 atoms. One molecule of BT1760 has two individual domains: the N-terminal five-bladed β -propeller catalytic domain (residues 34-358) linked to a C-terminal β -sandwich domain (residues 368-523) by a 10-residue linker. BT1760 has no distinct N-terminal β -sandwich domain as found in BT3082. The dimer represents a symmetry bringing the β -sandwich domain of monomer A in contact with the β -propeller of monomer B and *vice versa*. This kind of dimerized conformation was seen in other members of GH32 (Alberto *et al.*, 2004). The nucleophile mutant D62A was crystallized in the presence of levan oligosaccharides to try and obtain a substrate complex, but was not successful (data not shown).

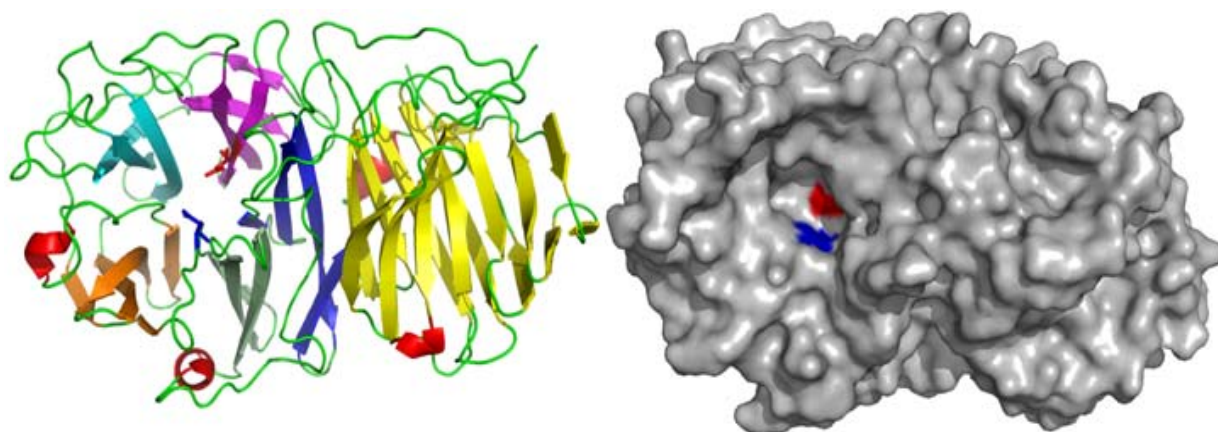


Figure 4. 21 Structure of BT1760 endo levanase.

The resolution for BT1760 is 2.6 Å. **Left:** This structure has an N-terminal 5-fold β -propeller catalytic domain (purple, cyan, orange, palegreen, and blue), and a C-terminal β -sandwich domain (yellow). Loops are colored green, while 4 α -helices are colored in red. The first helice is between blades 2 and 3, while the second is between blades 4 and 5. There is another helice between catalytic domain and β -sandwich. There is a small helice in the β -sandwich between two β -sheets. The two catalytic residues are presented as red (Asp-62) and blue (Glu-242) sticks which located on the first blade (purple) and the loop between third (orange) and fourth (palegreen) blades respectively. **Right:** Surface representation shows the surface of BT1760 as grey, while the D62 (red) is at the bottom of a pocket and E242 (blue) is on the sidewall of the same pocket. This catalytic pocket is larger and more open than the BT3082 pocket.

The DALI search for BT1760 presented a group of most similar structures, in which the top 25 are from the four solved GH32 proteins (Z scores from 33.6 to 29.5) while a few from GH43 family which includes β -xylosidases (PDB ID: 3CPN and 1YI7, Z score 24.3 and 23.6 respectively) from *Clostridium Acetobutylicum* (Figure 4. 22) (Osipiuk, 2008; Teplyakov, 2005) and several from GH68 family (PDB ID: 1PT2, Z score 23.9) (Meng and Futterer, 2003).

No:	Chain	Z	rmsd	lali	nres	%id	Description
1:	1uyp-B	33.6	2.7	400	432	16	MOLECULE: BETA-FRUCTOSIDASE;
2:	1w2t-E	33.5	2.7	394	432	16	MOLECULE: BETA FRUCTOSIDASE;
3:	1uyp-C	33.5	2.8	398	432	17	MOLECULE: BETA-FRUCTOSIDASE;
4:	1w2t-F	33.4	2.8	402	432	15	MOLECULE: BETA FRUCTOSIDASE;
5:	1w2t-C	33.4	2.8	399	432	17	MOLECULE: BETA FRUCTOSIDASE;
6:	1uyp-F	33.4	2.8	399	432	16	MOLECULE: BETA-FRUCTOSIDASE;
7:	1w2t-D	33.3	2.8	399	432	16	MOLECULE: BETA FRUCTOSIDASE;
8:	1w2t-B	33.3	2.7	399	432	16	MOLECULE: BETA FRUCTOSIDASE;
9:	1uyp-D	33.3	2.8	399	432	16	MOLECULE: BETA-FRUCTOSIDASE;
10:	1w2t-A	33.2	2.8	399	432	16	MOLECULE: BETA FRUCTOSIDASE;
11:	1uyp-E	33.2	2.8	399	432	16	MOLECULE: BETA-FRUCTOSIDASE;
12:	1uyp-A	33.2	2.8	399	432	16	MOLECULE: BETA-FRUCTOSIDASE;
13:	1y9g-A	30.6	2.8	425	517	15	MOLECULE: EXO-INULINASE;
14:	1y4w-A	30.6	2.9	424	517	16	MOLECULE: EXO-INULINASE;
15:	2qqw-A	30.5	2.8	427	534	17	MOLECULE: BETA-FRUCTOFURANOSIDASE;
16:	2qqu-A	30.5	2.8	424	535	17	MOLECULE: BETA-FRUCTOFURANOSIDASE;
17:	2ac1-A	30.5	2.8	427	537	17	MOLECULE: INVERTASE;
18:	1y9m-A	30.5	2.9	424	517	15	MOLECULE: EXO-INULINASE;
19:	2aez-A	30.1	2.8	422	536	17	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
20:	2qqv-A	30.0	2.8	426	537	17	MOLECULE: BETA-FRUCTOFURANOSIDASE;
21:	2ade-A	30.0	2.8	421	537	17	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
22:	2add-A	30.0	2.8	421	537	17	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
23:	1st8-A	30.0	2.8	421	537	17	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
24:	2aey-A	29.7	2.8	422	537	17	MOLECULE: FRUCTAN 1-EXOHDROLASE IIA;
25:	2oxb-A	29.5	2.8	425	537	18	MOLECULE: BETA-FRUCTOFURANOSIDASE;
26:	3cpn-A	24.3	2.9	262	319	13	MOLECULE: BETA-XYLOSIDASE, FAMILY 43 GLYCOSYL HYDROLASE;
27:	1pt2-A	23.9	3.1	275	440	15	MOLECULE: LEVANSUCRASE;
28:	3byn-A	23.8	3.1	271	440	15	MOLECULE: LEVANSUCRASE;
29:	3byl-A	23.8	3.1	275	440	15	MOLECULE: LEVANSUCRASE;
30:	3byj-A	23.8	3.1	274	440	15	MOLECULE: LEVANSUCRASE;
31:	1yrz-A	23.8	3.7	270	522	15	MOLECULE: XYLAN BETA-1, 4-XYLOSIDASE;
32:	1oyg-A	23.8	2.9	272	440	15	MOLECULE: LEVANSUCRASE;
33:	1yrz-B	23.7	3.7	267	522	14	MOLECULE: XYLAN BETA-1, 4-XYLOSIDASE;
34:	1yif-A	23.7	3.2	259	532	12	MOLECULE: BETA-1, 4-XYLOSIDASE;
35:	3byk-A	23.6	3.0	273	440	15	MOLECULE: LEVANSUCRASE;
36:	2exj-B	23.6	3.9	273	533	13	MOLECULE: BETA-D-XYLOSIDASE;
37:	2exh-A	23.6	3.3	260	533	14	MOLECULE: BETA-D-XYLOSIDASE;
38:	1yif-D	23.6	3.8	273	532	11	MOLECULE: BETA-1, 4-XYLOSIDASE;
39:	1yif-C	23.6	3.8	273	532	11	MOLECULE: BETA-1, 4-XYLOSIDASE;
40:	1yi7-A	23.6	3.1	258	534	14	MOLECULE: BETA-XYLOSIDASE, GH FAMILY 43

Figure 4. 22 DALI result of the top 40 structures similar to BT1760.

Dali matches the first 40 structures sorted by Z-score. Chain: PDB entry code plus chain identifier. Z: normalized Z-score that depends on the size of the structures. The program optimises a weighted sum of similarities of intramolecular distances. rmsd: root-mean-square deviation of C-alpha atoms in the least-squares superimposition of the structurally equivalent C-alpha atoms. The program does not optimise rmsd, this is only reported for your information. lali: number of structurally equivalent residues. nres: number of amino acids in the protein. %id: percentage of identical amino acids over all structurally equivalent residues. Description: the COMPND record from the PDB entry.

DALI results of structures in PDB bank presented that the most similar structure to the N-terminal β -propeller catalytic domain of BT1760 is the molecule A of PDB 2AC1, a GH32 from *Arabidopsis thaliana* with the Z score 29.6 % (Figure 4. 23), while the most similar one to the C-terminal β -sandwich of BT1760 is the molecule B of PDB 1DHK from a mammalian alpha-amylase with the Z score of 15.4 % (data not shown).

No:	Chain	Z	rmsd	lali	nres	%id	Description
1:	2ac1-A	29.6	2.7	280	537	20	MOLECULE: INVERTASE;
2:	2qqu-A	29.4	2.8	280	535	20	MOLECULE: BETA-FRUCTOFURANOSIDASE;
3:	2qqv-A	29.0	2.8	280	537	20	MOLECULE: BETA-FRUCTOFURANOSIDASE;
4:	1y4w-A	29.0	2.7	281	517	18	MOLECULE: EXO-INULINASE;
5:	2qqw-A	28.5	2.8	280	534	20	MOLECULE: BETA-FRUCTOFURANOSIDASE;
6:	2aez-A	28.1	2.8	280	536	19	MOLECULE: FRUCTAN 1-EXOXYDROLASE IIA;
7:	2aey-A	28.1	2.8	280	537	19	MOLECULE: FRUCTAN 1-EXOXYDROLASE IIA;
8:	2ade-A	28.1	2.8	280	537	19	MOLECULE: FRUCTAN 1-EXOXYDROLASE IIA;
9:	2add-A	28.1	2.8	281	537	19	MOLECULE: FRUCTAN 1-EXOXYDROLASE IIA;
10:	1st8-A	28.0	2.8	280	537	19	MOLECULE: FRUCTAN 1-EXOXYDROLASE IIA;

Figure 4. 23 DALI result of β -propeller catalytic module of BT1760.

Dali matches the top 10 structures sorted by Z-score. These 10 structures all are from GH32 members. Chain: PDB entry code plus chain identifier. Z: normalized Z-score that depends on the size of the structures. The program optimises a weighted sum of similarities of intramolecular distances. rmsd: root-mean-square deviation of C-alpha atoms in the least-squares superimposition of the structurally equivalent C-alpha atoms. The program does not optimise rmsd, this is only reported for your information. lali: number of structurally equivalent residues. nres: number of amino acids in the protein. %id: percentage of identical amino acids over all structurally equivalent residues. Description: the COMPND record from the PDB entry.

4.3 Discussion

4.3.1 GH32 enzyme activity

The DNSA assay (Table 4. 3) revealed that together the four GH32 from *B. thetaiotaomicron* could digest all fractions of inulin and levan and even the fructose containing plant oligosaccharide raffinose (data not shown). BT1760 is only active against levan and levan oligos. BT1759 acts on all oligosaccharides and polymeric substrates, but prefers sucrose. BT1765 prefers short substrates, especially sucrose, with very low activity on inulin and levan. BT3082 appears not be so sensitive to the chain length of fructans, but has a strong preference for β -2,1 over β -2,6 fructans. The different substrate preferences of the *Bacteroides* GH32s and their predicted cellular location enable us to suggest a role for each of these enzymes in fructan utilisation in this dominant gut bacterium. BT1760 may be located on the outer membrane, which is supported by the LipoP prediction that BT1760 is lipoprotein. BT1760 could cleave long-chain levan initially; then shortened levan oligosaccharides could be transferred into periplasmic domain through the β -barrel of SusC transducer BT1763 for further digestion. BT1759 and BT3082 are most likely located in periplasm to digest levan and inulin oligosaccharides coming through the SusC porin. The rationale for two fructosidases with similar specificity is not clear, but may be due to a certain level of redundancy in the system. The location of BT3082 outside the locus supports this view as it suggests the three enzymes in the locus are themselves sufficient to digest fructans and that BT3082 was acquired later by a gene duplication or horizontal gene transfer event. BT1765 is possibly located in cytoplasm because the value of signal peptide prediction was low, where it could digest smaller oligosaccharides taken up by the inner membrane transporter, BT1758 (Figure 4. 24). This kind of complementary character could exist in other GH and PL groups against different polysaccharides in *B. thetaiotaomicron* and other symbiotic gut bacteria.

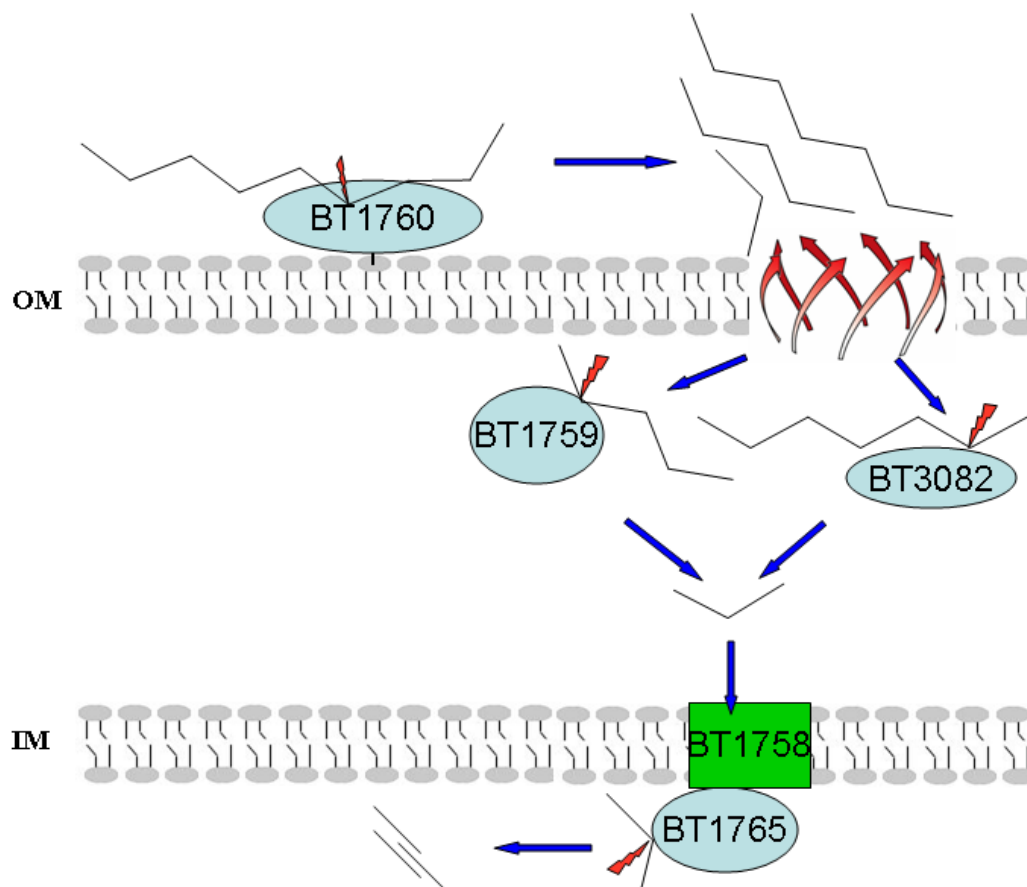


Figure 4. 24 Model for location and role of GH32s in *B. thtaiotaomicron*.

The possible location of BT1760 is on the outer membrane (OM), which digest coming levan. Fructan oligosaccharides get into periplasm through the β -barrel of SusC (red arrows), and then digested by BT1759 and BT3082. The smaller sugars get into the cytoplasm via the inner membrane transporter BT1758 (green box) and are digested into monosaccharides by BT1765.

BT1760 demonstrates the endo-activity while other three are exo-acting from the HPLC data (Figure 4. 8) and the TLC analysis (Figure 4. 9). And there is no minimum length of oligo-levan required for BT1760; it could cut even the shortest levanbiose. By this means, BT1760 is mixed endo and exo acting, which is different with the typical endo-acting GH proteins (Pell *et al.*, 2004). And BT1760 is different from other two reported endo-levanase, β -2,6-fructan fructanohydrolase from *Bacillus sp.* (strain L7) which needs at least 3 fructose units and β -2,6-fructan levanbiohydrolase from *Bacillus subtilis* which remove levanbiose from the end of the levan chain (Daguer *et al.*, 2004; Miasnikov, 1997). The enzyme to digest sucrose is called invertase (β -D-fructofuranosidase), a key enzyme to control the sucrose concentration in plants in response to the nutritional status, which is different to its function in bacteria (Sturm and Tang, 1999). Many plant GH32s cannot cleave sucrose, probably because this molecule plays an important role in signalling in plants (Alberto *et al.*, 2004; Verhaest *et al.*, 2007a). This class of enzymes is known as fructan exohydrolases (FEH). FEHs are either specific for inulin type fructans (1-FEH) or levan type fructans (6-FEH). The structural basis for this specificity is currently unknown as to date there are

no structures of 6-FEH, however an overlay of BT3082 (which cleaves levan, but ~10-fold less well than inulin) and 1-FEH from Chicory, both with kestose bound, suggests the structural differences that drive this specificity must be very subtle as the active sites of BT3082 and 1-FEH are very similar. The only obvious difference is the position of the +1 Frc and Glc of the kestose which adopt significantly different positions in the active site of BT3082 compared to 1-FEH (Figure 4. 25). Whether this is relevant to the specificity difference seen between these two enzymes is not clear and it would require a complex with a β -2,6 linked oligosaccharide to understand the molecular details of this specificity.

Sucrose is a substrate of invertases such as BT1759, BT1765, and BT3082, but an inhibitor of FEH, such as 1-FEH (PDB ID 2ADD) from *Chicorium intybus* (Verhaest et al., 2007a). This inhibition occurs because in the 1-FEH-sucrose complex a hydrogen bond is formed between the O2 of the glucose and the acid base catalyst. This interaction obviously does not occur in BT3082 or other invertase type enzymes, but as there are no invertase-sucrose complexes available to study it is not possible to see the alternative ‘non-inhibiting’ interaction between the sugar and the enzyme.

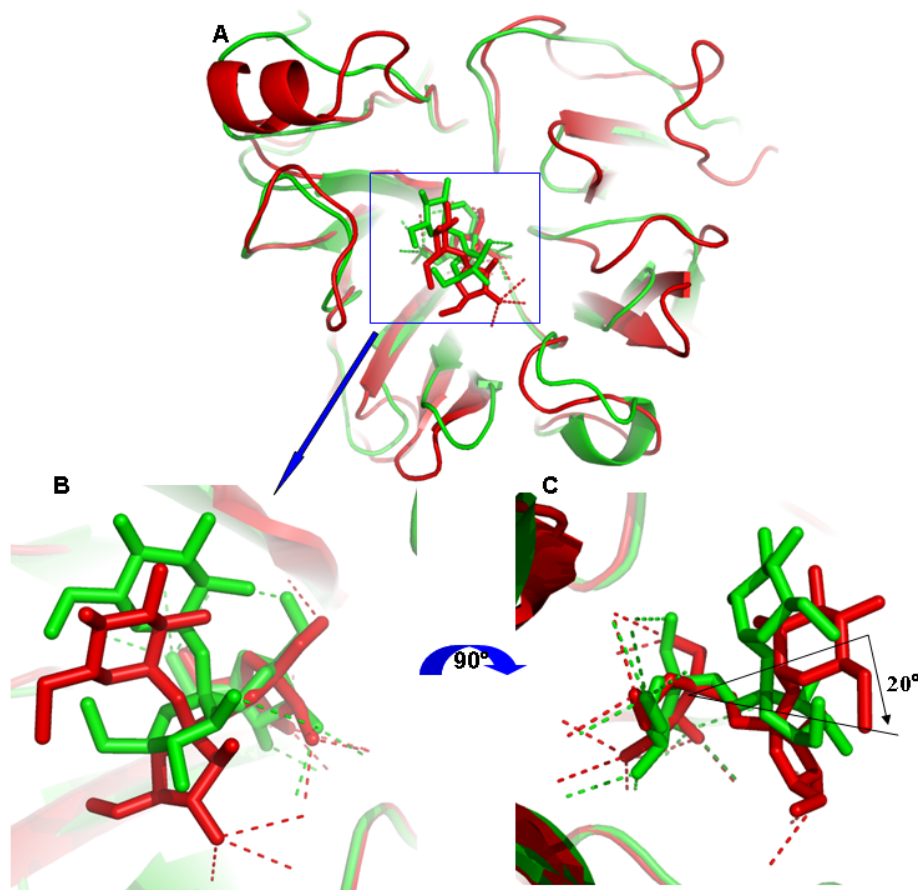


Figure 4. 25 Overlay shows kestose difference of BT3082 and 1-FEH from *Chicorium intybus*.

A: Overlay of BT3082 D131A (green) with 1-FEH (red, PDB ID 2AEZ) shows the kestose locations are very similar. **B:** The zoom in of the blue box of panel A reveals the positions of the Frc at +1 and the Glc of kestose are in different positions in the two enzymes. The hydrogen bonds (red dash lines) formed by the +1 fructose (red stick) with enzyme 1-FEH and solution, while the +1 fructose in kestose just formed one hydrogen bond. **C:** The 90° rotation of panel B. From panel C, the kestose in 1-FEH rotated about 20° around the -1 fructose of kestose in BT3082 D131A.

4.3.2 Role of the N- and C-terminal domains in enzyme activity

BT3082 has three domains. The catalytic domain is flanked with two smaller, mainly β -sheet domains (Figure 4. 26). The number of β -sheets in the N-terminal domain is less than in the C-terminal, which forms a β -sandwich and is highly conserved in GH32 members, while the N-terminal is unique, at least amongst structurally characterised members

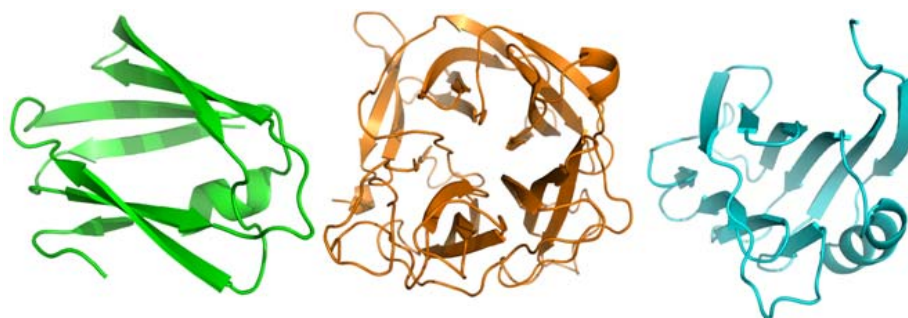


Figure 4. 26 Three domains of BT3082.

The N-terminal domain (green) has two sets of β -sheets and one α -helix. The middle domain (orange) is the catalytic β -propeller domain. The C-terminal β -sandwich (cyan) has two sets of β -sheets and one α -helix as well.

The most similar structures to the N-terminal domain of BT3082 are rhodopsin transducer proteins, which hints that BT3082 N-extra may be a degraded sensor (Figure 4. 13). Deletion analysis indicated that this N-terminal domain is necessary for catalytic function of BT3082 (section 4.2.3.3.3). The BLAST analysis of BT3082 revealed that the N-terminal domains are only conserved in other *Bacteroides* species and so could be an adaptation to the gut environment.

The DALI analysis of C-terminal of BT3082 (Figure 4. 14) and the alignment of these 4 GH32 proteins against structure solved GH32 members (Figure 4. 3) indicated that the C-terminal similarity is very low except two residues Asp-498 and Phe-505 of BT3082 are highly conserved. All these results suggested that though the C-terminals of GH32s vary highly in sequence identity they have significant structural similarity. The role of these domains in enzyme function is unclear. Previous studies with Chicory 1-FEH have suggested the domain plays a role in inulin binding, as introduction of an N-glycosylation site in the cleft formed between the catalytic domain and the C-terminal domain reduced the activity of the mutant enzyme against high DP inulins. However, although this may be the role in 1-FEH, the presence of the C-terminal domains in BT1760 and BT1765, which has no, or very low activity against inulins, respectively, indicates it is not the function of the C-terminal domain in all GH32s. Deletion from BT3082 results in a loss of expression in the heterologous host. This may be due to an inability to fold correctly without the C-terminal domain or that the truncated enzyme is unstable and is degraded by proteases. These data

suggest the domain plays a key role in enzyme stability, as suggested previously (Alberto et al 2004). Indeed their ubiquitous presence in all known GH32s would support a general stabilisation role for this domain as opposed to a substrate specific function.

4.3.3 Active site of BT3082

BT3082 prefers inulin-type fructans, while it can digest levan at a low activity (Table 4. 3). BT3082 F187A digests inulin more effectively than wild type. F187 does not form hydrogen bonds with kestose module (Table 4. 4). But this aromatic residue forms one side of the catalytic pocket (Figure 4. 16C). W128, Y353, and W373 are aromatic residues on the other side of the pocket with which have no hydrogen bond with kestose as well, but their mutants W128A, Y353A, and W373A have different enzymic activity against inulin, especially W373A which loses activity against all substrates and Y353A which loses activity against sucrose.

The mutagenesis results also indicated the docking of inulin has special orientation which is predicted to be from the angle of F187. So the location of F187A makes the catalytic site wider and gives more flexibility to the access of inulin. It is interesting to find the higher activity of Y353F against sucrose than wild type, which indicated the hydroxyl from Y353 could restrict the enzymic activity in some level. It is possible that the released fructose from sucrose slips away from the orientation of Y353, so the removal of the hydroxyl of Y353 could make the product move out of the catalytic pocket more effectively while the hydrophobic surface keeps identical. W155 is the only aromatic residue forming hydrogen bond with kestose. W155F was abolished half enzymic ability against all three substrates, which is predicted still close enough to form one hydrogen bond with kestose, while W155A is too far to do so, which is supported by the low enzymic activity (Figure 4. 17).

When the residues which form hydrogen bonds with kestose (N130, Q147, D257, and E301) were mutated to alanine individually, the enzymic activity was abolished to less than 10 % or completely lost. E301 is the catalytic acid residue of BT3082. Other two such residues (R256 and S188) were not mutated as alanine in this study for time limitation, but mutants of both residues would be predicted to lose activity as well. The closest distance to kestose to Q221 is 5.29 Å which is quite far from catalytic pocket (Table 4. 4). The Q221A mutant has more than half activity against inulin (Figure 4. 17), which is predicted to be caused by subtle changes to the shape of the binding pocket. P132, C302, and P303 are close to two catalytic residues and completely conserved in GH32 proteins, so it is not surprising to see their alanine mutants are abolished some activity against all three substrates. It is surprising to find W300A totally lost activity against sucrose because the

closest distance of W300 to kestose is 5.01 Å and it is on the outside of the pocket edge. But this residue is highly conserved in GH32 proteins, suggesting it is important in other GH32 members as well.

R256 and D257 both are in the RDP motif, which are highly conserved in GH68 family (retaining mechanism, levansucase) as well (Meng and Futterer, 2003). GH68 members possess same activity and same substrate (sucrose), which have 5-blade β -propeller as catalytic module as well. This kind of similarity reveals the structural relationship.

4.3.4 BT1760

BT1760 is predicted to be the first endo-acting levanase with solved structure. The DALI analysis revealed it is similar to GH43 and GH68 enzymes. Though there is structural similarity, GH43 are inverting mechanism, which is opposite to GH32 and GH68 (retaining mechanism). And GH43 uses different substrate, while GH32 and GH68 use same substrate (sucrose) and both have the RDP motif (Alberto *et al.*, 2004).

Even BT1760 and BT3082 which have different activity and low amino acid similarity, have structural similarity, though BT3082 has an N-terminal function-unknown essential domain (Figure 4. 27). Unfortunately, the BT1760 D62A was not solved, so the accurate binding conformation of levan oligosaccharides in the catalytic pocket is still unknown.

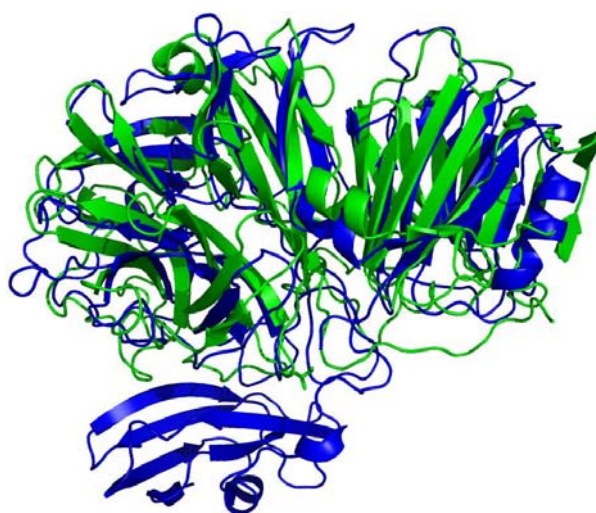


Figure 4. 27 BT1760 WT overlays BT3082 WT.

Blue: BT3082, Green: BT1760. Compared with BT1760, BT3082 has an extra N-terminal domain, which mainly is β -propellers. And at the C-terminal, BT3082 has one more α -helix.

But overlay the catalytic domain of BT1760 with other three GH32s presents the active site difference (Figure 4. 28). The catalytic pocket in BT1760 is larger and more open than exo-acting enzymes, and the space above the pocket of BT1760 is flatter than others (Figure 4. 28, circled in red), which may provide binding sites for the levan chain to allow endo-activity.

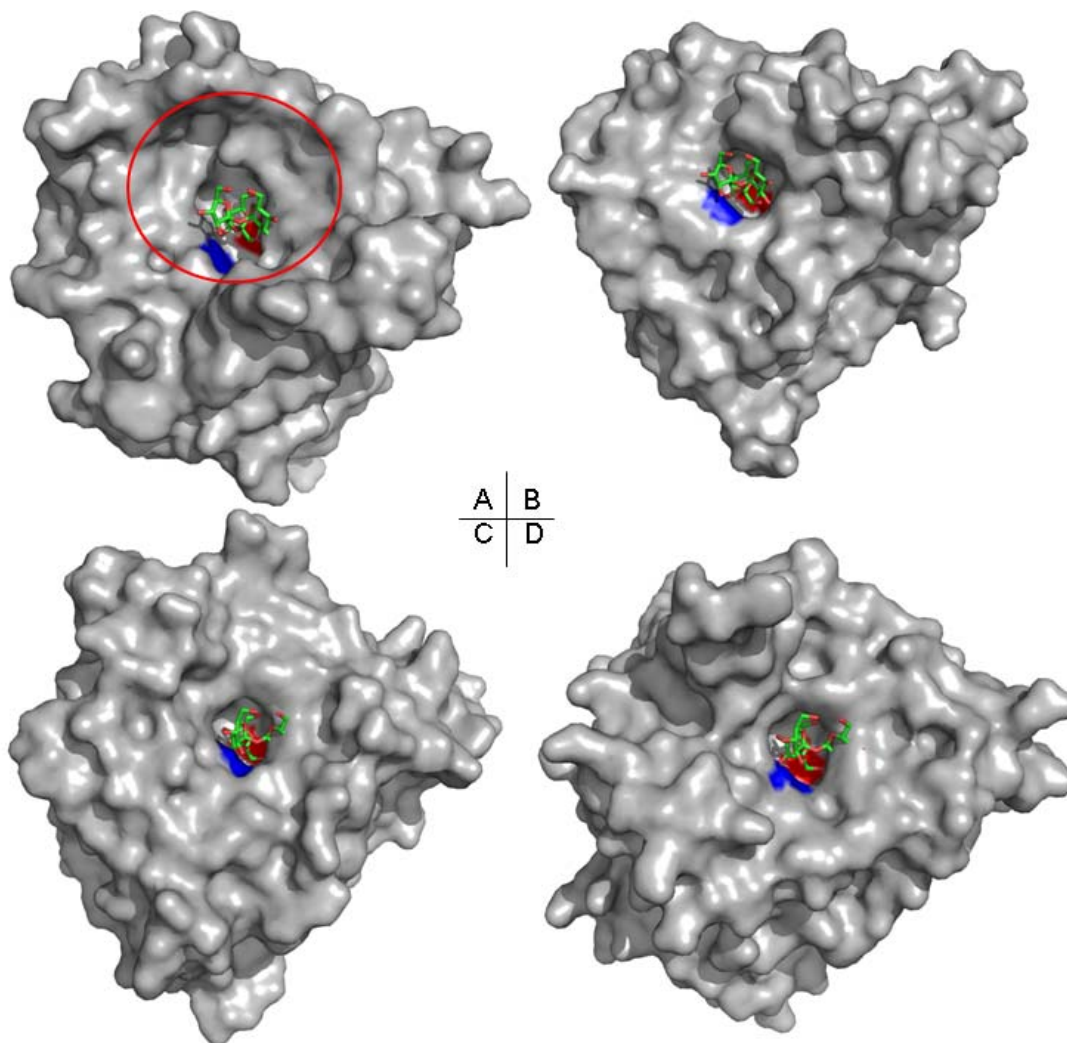


Figure 4. 28 Surface representation shows the overlay GH32s β -propeller catalytic domains with kestose in the active site.

A: BT1760, **B:** BT3082 wild type, **C:** invertase of *Aspergillus awamori* (PDB ID 1Y9G), and **D:** invertase from *Thermotoga maritima* (PDB ID 1UYP). The catalytic residues Asp (red) and Glu (blue) is highlighted to show the active pocket. **Note:** The kestose in the enzymes shown in panels A and C is not bound, but is positioned from an overlay with BT3082.

Comparing the modeled structures of levan-hexaose and inulin-hexaose reveals that the β 2,1-linked inulin-hexaose has a tighter twist than the β 2,6-linked levan-hexaose (Figure 4. 29). BT1760 may cut the fructose labeled in the red circle of Figure 4. 29, while other joined fructose units bind the flatter surface around the catalytic pocket. An inulin molecule may not get to the bottom of the catalytic pocket based on the possible location of kestose from the BT3082 structure (Figure 4. 30A).

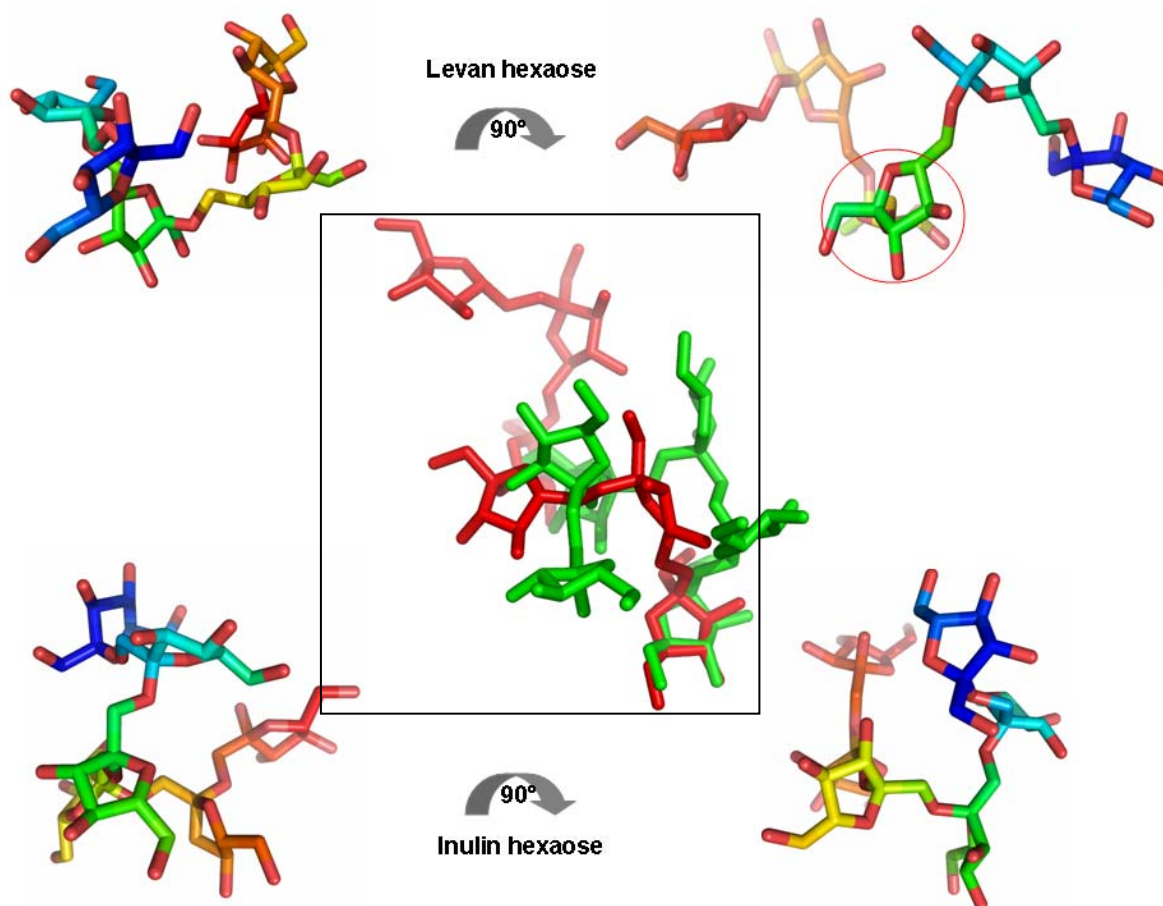


Figure 4. 29 Modeled structures of fructooligosaccharides

The two structures of levan-hexaose and inulin-hexaose were generated using the program SWEET. Both are present from two views with 90° rotation and colored from blue to red. Inulin is much tighter twist than levan. The middle panel (boxed) presents the overlay of levan (red) and inulin (green) to show the different twist of the chains with the first fructose units overlaid.

When looking into the residues around the BT1760 catalytic pocket, it is interesting to find that all aromatic residues (Y90, W338, W300, and Y303) and a histidine H91 gather on one side, while other residues (A122, I123, G124, T125, N144, Q155, R189, D190, C243, S258, D260, R265, A304, and E335) are on the other side (Figure 4. 30A). This kind of residue arrangement makes the pocket more spacious for BT1760 to dock levan oligosaccharides from endo while for smaller pocket of BT3082 substrates can only drop in by terminal fructose. 8 residues (T142, Q155, R189, D190, C243, Y303, A304 and W319) of BT1760 are completely conserved in other GH32s (Figure 4. 3), which suggests the mechanism of GH32 is conserved between endo-acting enzymes such as BT1760 and exo-acting enzymes such as BT3082.

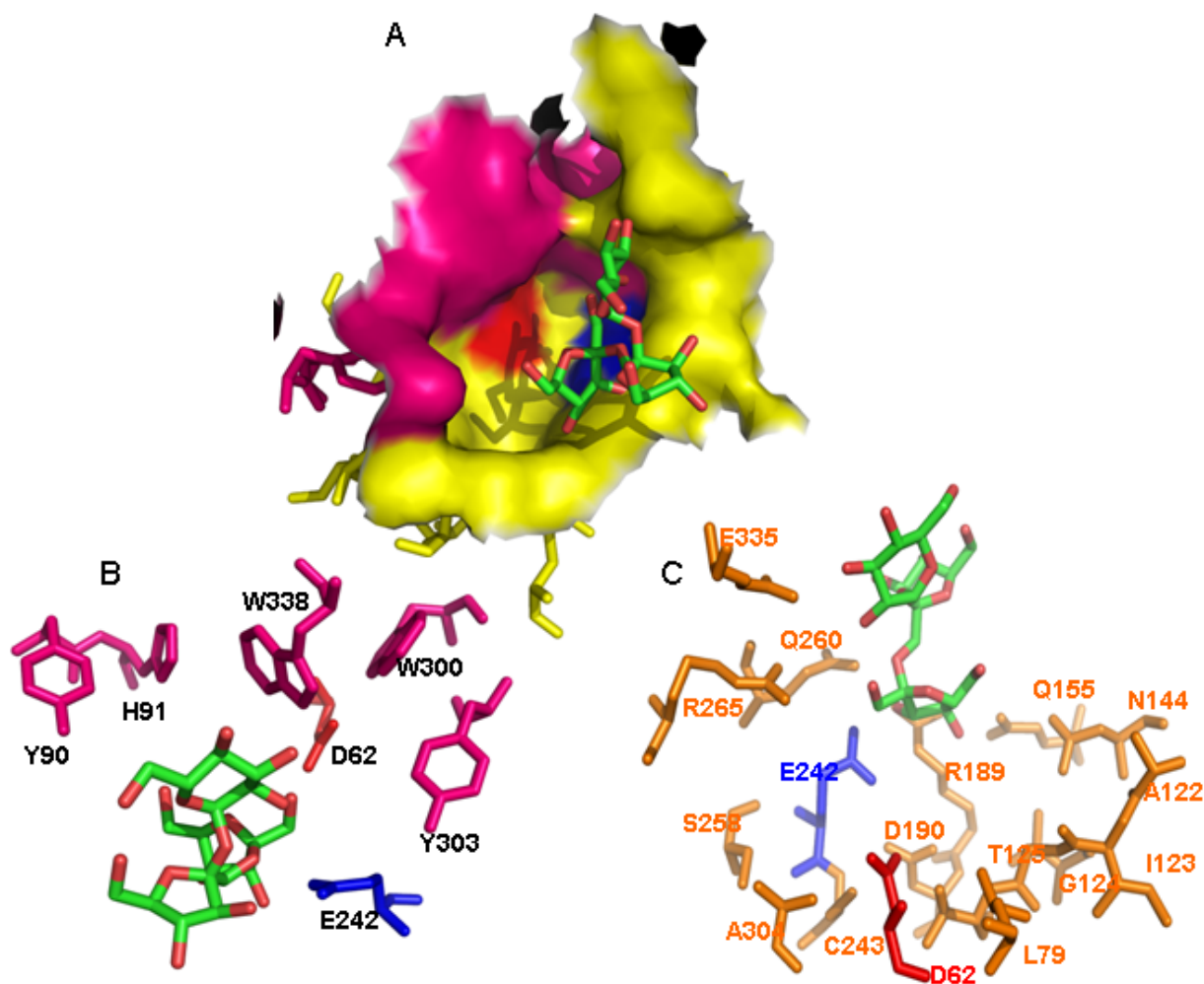


Figure 4.30 BT1760 catalytic pocket with kestose positioned from BT3082.

A: Surface representation shows the catalytic pocket of BT1760 with kestose (green stick) from BT3082 D131A via overlay method of PyMol. The pink surface indicates the aromatic residues (see labels in panel B), other residues except base D62 (red) and donor E242 (blue) are presented in yellow (see labels in panel C). **B:** The aromatic residues Y90, W338, W300, Y303 and H91 are presented in pink, while D62 in red stick, and E242 in blue. **C:** Residues A122, I123, G124, T125, N144, Q155, R189, D190, C243, S258, D260, R265, A304, and E335 are presented in orange sticks, while E242 in blue and D62 in red. Totally, there are 22 residues involved in the formation of catalytic pocket.

4.3.5 Applications of GH32s

GH32 enzymes are used widely in the food industry to make fructose and glucose from sucrose as the monosaccharides are sweeter and don't crystallize as easily as the disaccharide (VanDamme and D.G.Derycke, 1983). It is also possible to use these enzymes to digest levan and inulin to make ethanol, the biofuel (Fortman *et al.*, 2008; Groom *et al.*, 2008). Fructose is a good source of dimethylfuran because of the furan ring, which could be a potential liquid fuel with 2.5 fold higher energy per unit weight than ethanol (Roman-Leshkov *et al.*, 2007).

Chapter 5 SusD homologue BT1762

5.1 Introduction

There are 101 SusC/D paralog pairs in *B. thetaiotaomicron* genome which are ubiquitous components of PULs; indeed the presence of a susC/D gene pair is one of the defining features of a PUL (Bjursell *et al.*, 2006). Previous studies have revealed that SusD associates with the β -barrel porin SusC to bind starch and transport the polysaccharide into the periplasmic space (Shipman *et al.*, 2000). The structure of SusD has recently been solved and reveals it has a novel fold of 22 α -helices and 3 sets of two-antiparallel β sheets, of which 8 helices formed four tetratricopeptide repeat (TPR) units that are thought to be the interaction surface for the complex with SusC (Koropatkin *et al.*, 2008) (Figure 5. 1). SusD has been shown to be essential for growth on maltooligosaccharides larger than maltopentaose, a feature mirrored in its ligand specificity (Koropatkin *et al.*, 2008). The preference for maltooligosaccharides larger than ~ 5 sugar units is likely due to its curved binding site that complements the natural helical structure of starch. Only sugars larger than maltopentaose adopt this helical structure as a stable form and therefore smaller sugars bind with much reduced affinity (Koropatkin *et al.*, 2008). Recently the structure of another SusD homologue, BT1043, was solved in complex with the disaccharide *N*-acetylglucosamine (Koropatkin *et al.*, 2009). BT1043 is involved in mucin utilisation, but the interaction with the mucins component (*N*-acetylglucosamine) is very weak ($K_d \sim 12$ mM) indicating this is not the true ligand or is only a small fragment of it. So far, however, the functions of other SusD homologues are not well understood at the molecular level. Indeed it is not clear why *B. thetaiotaomicron* contains so many homologues of this gene. One possibility is that each SusD variant has a distinct specificity related to the carbohydrate the PUL has evolved to degrade. To further our understanding of the role of these proteins, experiments were carried out to characterize the ligand specificity and the structure of the SusD homologue BT1762 from the fructan activated PUL spanning genes BT1757-BT1765 (see Figure 3.2).

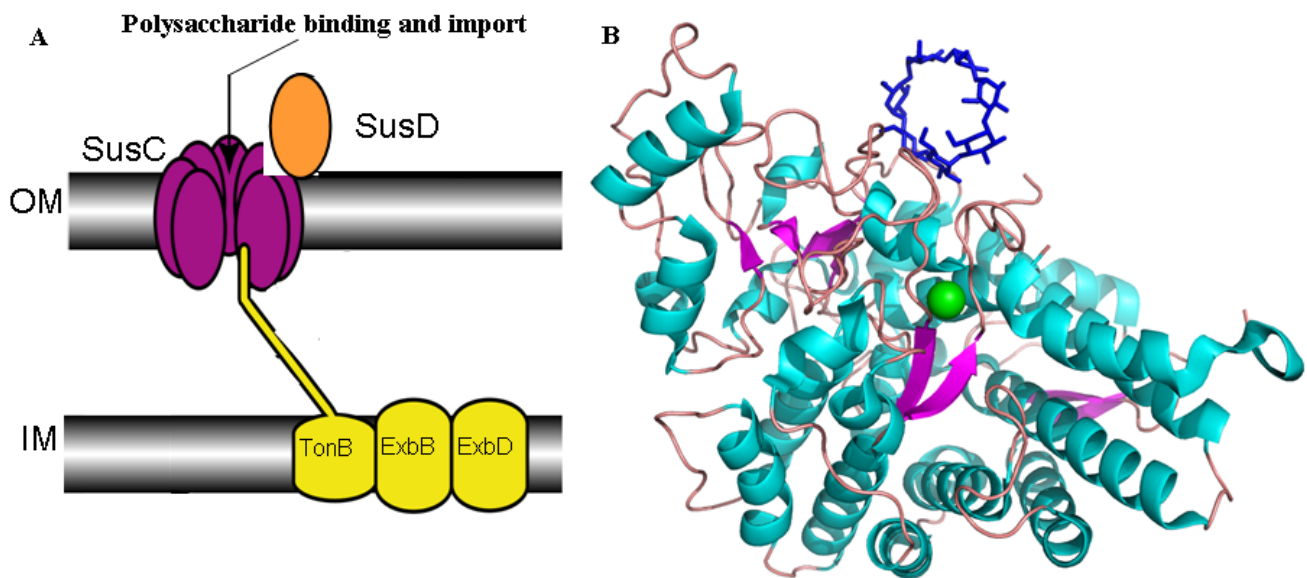


Figure 5. 1 SusD location and structure.

A: SusD locates on the outer membrane and associates with SusC to bind and import polysaccharide. This kind of complex is energy dependent and relies on the TonB complex which locates on the inner membrane.

B: The structure of SusD with maltoheptaose bound (blue ring). PDB ID is 3CK9 (Koropatkin *et al.*, 2008).

5.2 Results

5.2.1 Sequence analysis

Several SusD homologues (BT2625, BT3789, BT3855) were randomly chosen to align with BT1762, SusD (BT3701), and one recently-solved SusD homologue BT1043 involved in mucin degradation (Koropatkin *et al.*, 2009). SusD is predicted to be an outer membrane lipoprotein with a conserved cysteine at the N-terminal end of the Type II signal peptide that is covalently attached to a lipid tail embedded in the membrane. The alignment reveals that all the SusD homologues chosen, including BT1762, have a cysteine at the same position near the N-terminus, indicating they are all outer membrane lipoproteins (Figure 5. 2). Although all the sequences have a conserved region near the N-terminus which is a TPR helix domain of SusD and BT1043 (boxed in Figure 5. 2), the sequence identity on the whole is low (only ~7 % between BT1762 and SusD), especially BT1043 is the least similar of all aligned sequences.

BAD AVG GOOD



BT1043	1	MKK-YIIL-LLA--LCS-----MSACDYEAVNTN--PYG-VSDGELGPL-KYG	40
BT1762	1	MKKIYYI-ATIGI-TLL-----TSCDD-FLDRQ--VPQGIVTGDQIASPEYV	43
BT2625	1	MKNKIIIFIQFGM-LLC-----MSSCN--FLNVEDEFKDMLSYDSVPANKRNV	45
BT3701	1	MKTKYIKQ-LPSA-ALIAVLSSGVTSICN-DLDIS--PIDPQTGGSPDQQ----	45
BT3789	1	MKKLVYY-ILLSISALS-----FSACTD-YINVDKYFYDQVSLDSAFSKRIYV	46
BT3855	1	MKI-YSYL-LLGA-MFL-----LGGCSD-LLDID--PKNKIPADELFSTPEGV	42
BT1043	41	-ARFMNQQRVIP---I-GSPSLTTPGNDI-Q----NTDLISSGNYIGYFGNN	84
BT1762	44	DNLVISAY-AIWA--TGD--DI-N--SSPSLW-----NYDV-----	71
BT2625	46	EKYLWATA-ANFPDEGKLF--QYPY--TPGPL-----	72
BT3701	46	-GVFVKGY-AMLG--VTGQKIDG--SPDL-D----GQDEGESGFYRTTFNCN	87
BT3789	47	DGWLSSAY-SVMN--KLG--EY-K--EPFRW-----	69
BT3855	43	QAHMANLY-GRLP---I-E--DFTY--SPNRGFNVGVGTDVNNAGFMAAHF-CD	86
BT1043	85	NN-----WGFFN---EANWNFTDSRMNY--AYQNFYSQIFLPW	117
BT1762	72	--RSDDCYKG-SGTEDGGVFNALEISKGINTT--DWNINDIWKRLYQCI TRAN	120
BT2625	73	--ASDEAFTTFSGDQFLMGFVLGRETDPNSG----GLGNWGTMYKIIRKTN	118
BT3701	88	ELPTDECLWAW-QENQDIPQLTISISWSPSSQ---RTEW--VYVRLGYDITQYN	134
BT3789	70	--ASDDLYHPD-MKEYVEGYSADKPKGDENA---GESRLWK-YYEGIRKAS	114
BT3855	87	E-----AIHPEYNDF--GEWFDYWEDGYKLIRDLN	115
BT1043	118	NEI--YEIAKSDSPSEQAILEIANIVRNIAWLRATDVFGP IAYN-----S	161
BT1762	121	TALQSLDQMDKTYPLKNQRIAREMFRPLRGHAHFMLKQLFKKIVIVNDENMEPD-	173
BT2625	119	LIFRRIDEAKDLKTTKELELLAYTRFMRAYAYNLLMDFGPIVILGDEVMDNNE	172
BT3701	135	FFL---DQTEGMDAETLRQRMEIRFLRALHYWYFLDLFGKAPFK--EHFSNDL	183
BT3789	115	TFLDNVDRCPELTMDEIADMKGQARFLRAYCYWALIRVFCPVPIIPIEGLDADL	168
BT3855	116	SLLVTIPTLTSITEQQKNEINAETHFLRAYTYFALAKRYGGVPIIK-EPQEYNG	168
BT1043	162	AGDGSIAPKFDSQEVVYRSM LADLSKSVELLNT-ISISVMAQYDLIYNGNVQNW	214
BT1762	174	AYNELSNTTYTNDEQ-WQKIADDFQFAYDNLPE-VQI-EK----GRPAQAAAA	219
BT2625	173	TIDYYNAHRSTYDES-VEYVCGELEAAAVNLPITVPISQF----GRPTKGAAY	220
BT3701	184	PVEKK-----GTEL-YTYIQNELNEIEADMYE-PRQAPF----GRADKAANW	224
BT3789	169	SYEELSPLRTHFDEI-VSPIDTELAETARLLPMRRTVNNL--GRPTRGAAL	216
BT3855	169	NIEELRVPRNTEKDT-WDFVLEECDQAVSLPGD-ANENDV----LRANKWVAL	215
BT1043	215	VKLANSMLRIARVHFIDETLAK---EYIT--KALDPKNGGVI-----	253
BT1762	220	AYL-AKTYLYKAY---RQD--GADNALTG--IN-----EE	246
BT2625	221	GLI-ARLR LQHAS---PLYNGGSVAKTYFGSWKRSVDNAH--YISQTENVE	265
BT3701	225	LLR-ARLYLNAGV---YTGQT--DYA-----KA--EE	248
BT3789	217	GLR-ARVLLYAAS---PLYNGNLD--FF--NVVDNKGNG--LISQTYDES	256
BT3855	216	ALK-SRAALYAAS---VAKFTHQ--PYVSFSGPAVDQKLVGIEVISA---DH	258
BT1043	254	--EDISSEAKIKS---SDKMPLLSM-----LASV-NEY-----	281
BT1762	247	DLKQVVKYTDPLIMMA--KGGYGL-----	267
BT2625	266	RWALAAAACARIID---MGLYELHTVKS DINT-----PKLPQGVSTEPFPAG	309
BT3701	249	YASKVIGSA-----YKLC-----	261
BT3789	257	KWAKAAAAAKDVIELAKTSGLYELYTIAPKIGTLDMYRPPVHP-EYSTKDYPDG	309
BT3855	259	YYDECISASQEI MN---SGKFGLYKPS-----PATP-EEAT-----	290
BT1043	282	-----NETRMGATI W-GYLDGYKDPRI-----SAYFTEG	309
BT1762	268	-----ETDYSMNFLPQYE---NGA-ESVWA-IQYSIND--GTYNGNLNW--G	305
BT2625	310	AGGIDPFRS YSDMFTGETV-PQTNP-EYIWRMSSSIVDITRHSFNG	354
BT3701	262	-----TNYSELF MADNDENENAMQEIILPI---RQD--GVK--TRNYGGS	299
BT3789	310	WANIDPLLSYKSNFDG SVQ-GSKNP-ELIFT--RTSD--GTG--TIND--W	350
BT3855	291	-----TNYQKLFEPQPF-QCLDGLKEPIF-----MKA YAAN	319
BT1043	310	TY-----GSGS-----	315
BT1762	306	MGLT-----TPQILGCCDFHKPS-----QNLVNAFKTDSQG---KPL--	339
BT2625	355	-----DILGGFNGMCPV-----QKVIDKYKMVD-G--RSIQN	383
BT3701	300	TYLVCGTRVAGMPRMGT TNGW SCIF---ARAAMVQKFFSNLEDVPMLPADVEI	349
BT3789	351	MYQA-----LPRTISGNNRLCVT-----QKQVNAYAMND-G--RTISE	385
BT3855	320	TILAHNYDVWFSPRQMILDF-NLYPGRMNE TLDFVDSFEDYTDDGTGTPKPIST	372

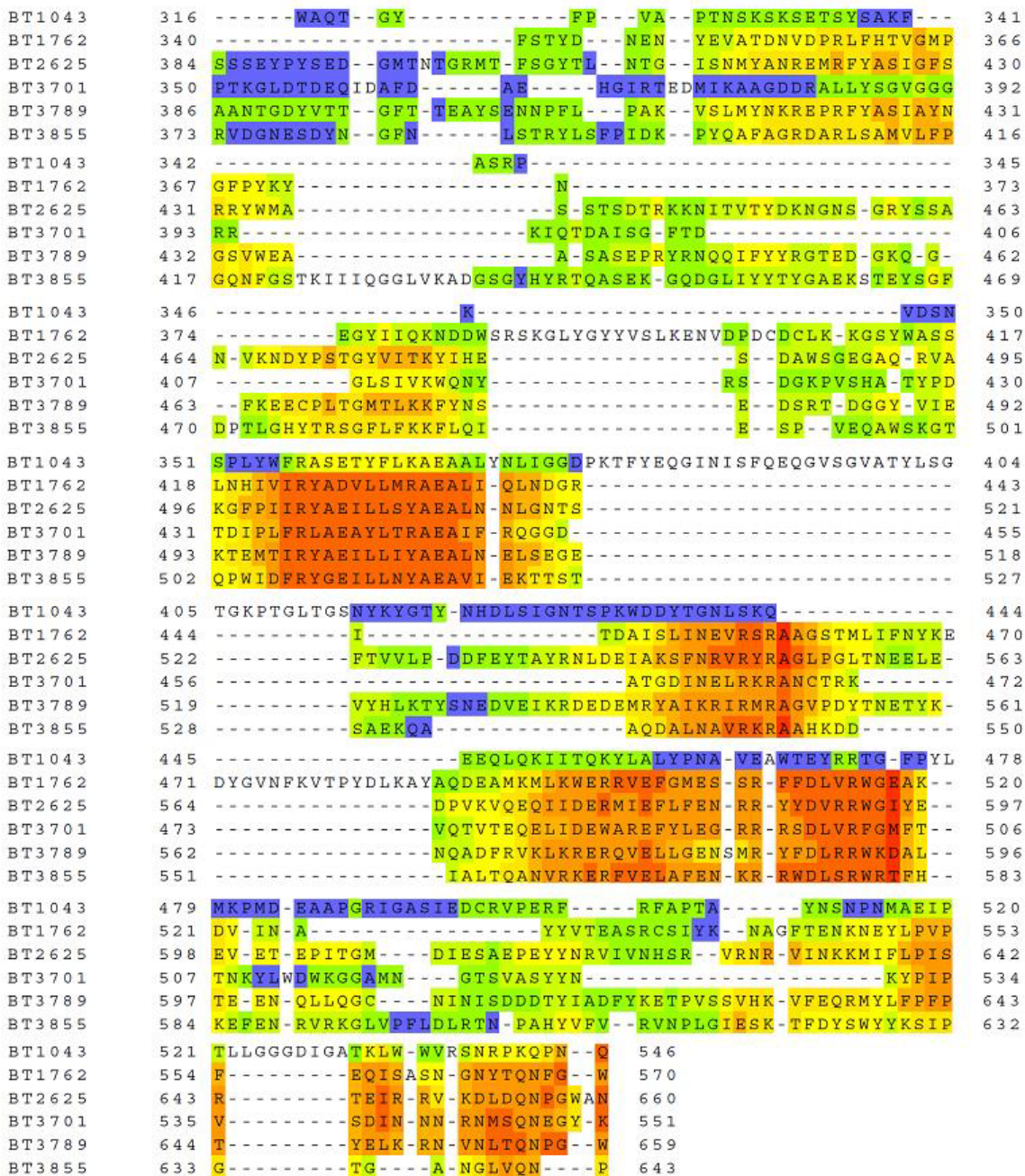


Figure 5. 2 Alignment of SusD, BT1762, and four other SusD homologues from *B. thetaiotaomicron*. All proteins are from *B. thetaiotaomicron*. BT3701 is SusD. The conserved cysteine at the end of the signal peptide that is lipidated and attached to the outer membrane is indicated by an arrow. The boxed region is highly conserved in SusD homologues, which is a TPR helix domain of SusD and BT1043. The alignment was produced using Tcoffee.

5.2.2 Protein expression

Native BT1762 is a predicted outer membrane lipoprotein (LipoP server) with the cleaved type II signal peptide spanning residues 1-19 (Figure 5. 2), so, only residues 19-570 of BT1762 were expressed in *E. coli*. The gene of BT1762 was cloned into minipRSETA between *BamH* I and *EcoR* I. The forward and reverse primers used are BT1762-F: 5'-CTCGGATCCTGTGACGATTTTTTGGAC-3', and BT1762-R: 5'-

CTCGAATTCTTACCAACCGAAATTCTGTG-3', respectively. This generated a recombinant protein with a His tag fused to its N-terminus when expressed in BL21 (DE3) *E. coli* cell strain. This was then purified in a soluble form from the CFE by IMAC (Figure 5. 3). The calculated MW of recombinant BT1762 is 65.1 kD.

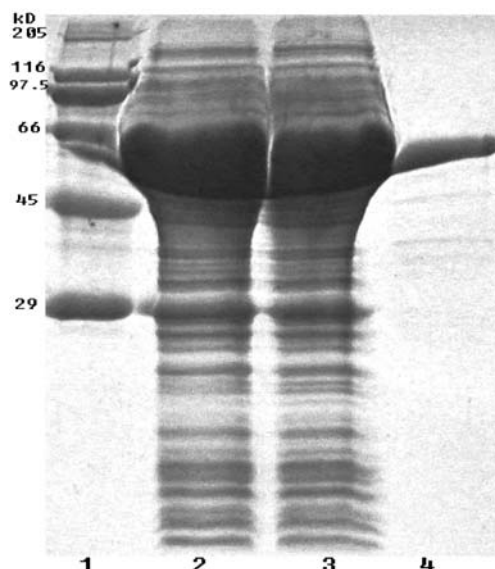


Figure 5. 3 N-terminal His tagged BT1762 purification.

Lane 1: Molecular weight marker, Lane 2: CFE, Lane 3: flow through from IMAC column, Lane 4: material eluted from IMAC column with 100 mM imidazole. The molecular weight of BT1762 is 65.1 kD.

5.2.3 Ligand binding studies

Considering the ligand specificity of BT1754 recognition (chapter 3) and the enzymic activity of GH32 proteins (chapter 4), only fructans were used to analyse the binding ability of BT1762. ITC data revealed that BT1762 could not bind inulin, which means BT1762 could not recognize the β -2,1 linkage of fructans. So inulin oligos, such as sucrose and kestose, were not analysed further. BT1762 could bind levan with a K_a of $\sim 1.1 \times 10^4 \text{ M}^{-1}$ (Figure 5. 4 and Table 5. 1), and this binding is metal independent (data not shown). Binding was driven by a favorable change in enthalpy, with an unfavorable entropic component; a feature common to many protein-carbohydrate interactions. Levanoligosaccharides smaller than levanheptaose (L7) displayed no binding to BT1762 (data not shown). The affinity of BT1762 for L7 was low but detectable by ITC. However the binding was too weak to be accurately fitted to a single site model (Figure 5. 4).

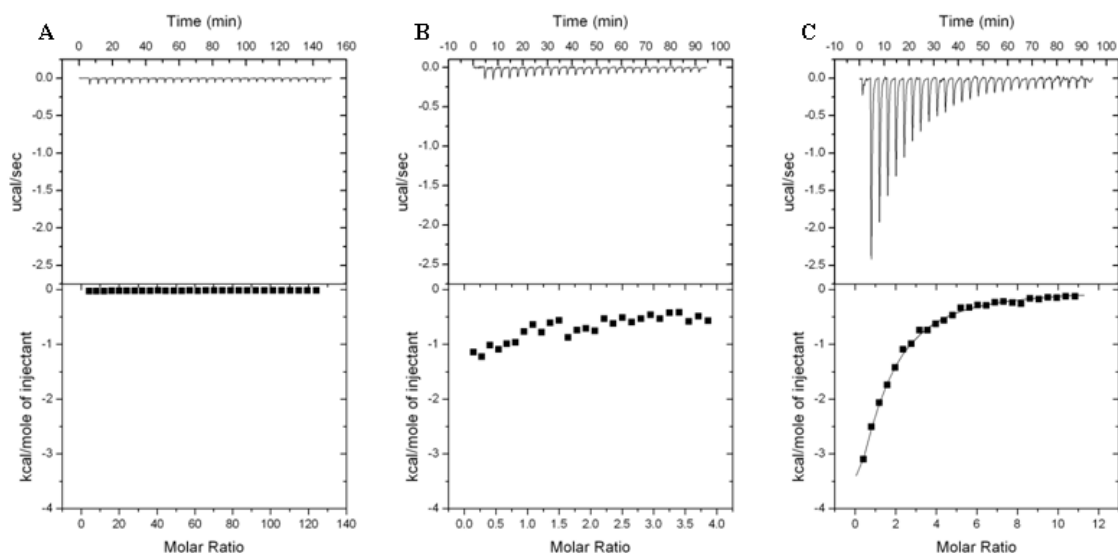


Figure 5. 4 ITC data show the binding ability of BT1762 to different fructans.

A: Inulin, B: L7, C: Levan. The top part of each figure shows the raw binding heats and the bottom the integrated heat data. In panel C the integrated data is fitted to a single site binding model to generate values for n , K_a , ΔH , and $T\Delta S$. Levan was at 10 mg/ml and proteins were used at 100 μ M. To obtain the molar concentration of levan the WT was fit such that $n=1$ (number of binding sites on the protein). This assumption is deemed valid due to the structural and mutagenesis data that indicates only a single binding site on the protein (see later).

Ligand	$K_a \times 10^3$ (M^{-1})	ΔH (kcal mol^{-1})	$T\Delta S$ (kcal mol^{-1})	n^*
Levan	9.1 ± 0.8	-7.7 ± 1.5	-2.8 ± 0.3	1.0 ± 0.1

Table 5. 1 ITC data show the binding ability of BT1762 to levan.

5.2.4 BT1762 crystallization

In order to crystallize BT1762, it was over expressed in BL21 (DE3) *E. coli* and purified by IMAC primarily (Figure 5. 3), followed by ion exchange purification (Figure 5. 5A). The purified protein was dialyzed against 10 mM Tris pH 8.0 buffer, and then concentrated down and buffer exchanged into water. After extensive screening, rod-like crystals formed in three conditions: 2.0 M Ammonium sulphate; 0.1 M Sodium acetate, pH 4.6, 8 % (w/v) PEG 4000; and 0.2 M Ammonium acetate, 0.1 M tri-Sodium citrate, pH 5.6, 30 % (w/v) PEG 4000 (Figure 5. 5B).

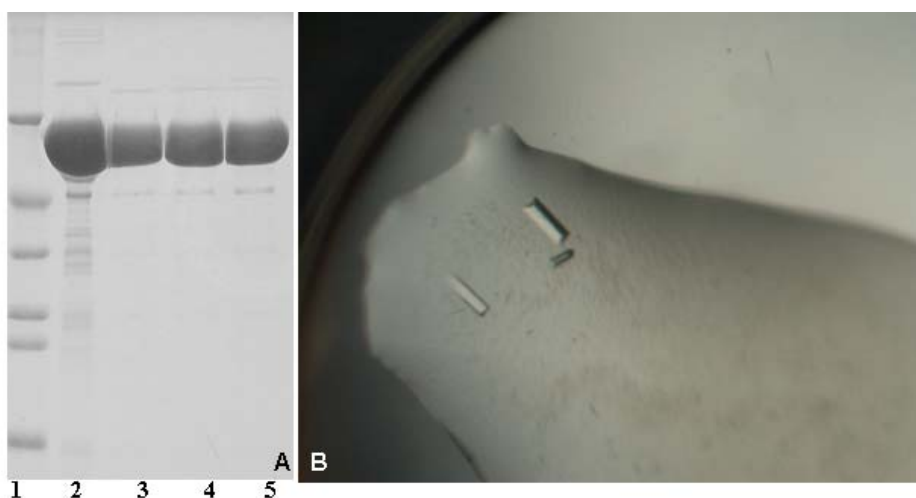


Figure 5. 5 Native BT1762 for crystallization.

A: Lane 1: Marker, Lane 2: BT1762 after IMAC purification, Lanes 3-5: Collected fractions of BT1762 from ion exchange Q12 column. **B:** The condition for this drop is 2.0 M Ammonium sulfate, from robot tray. This rod-like crystal gave a good diffraction, which is enough to solve the structure.

The Se-Met method was employed to solve the structure of BT1762 as the sequence similarity to SusD was too low to use molecular replacement. Se-Met BT1762 was purified by IMAC and ion exchange methods and then crystallized in the same conditions as native protein crystals (Figure 5. 6). When crystals grew up, the drop was soaked with ligand L7 to a final about 10 mM concentration in order to find the binding sites. Co-crystallization was also employed with 5 mM L7 final in the protein solution.

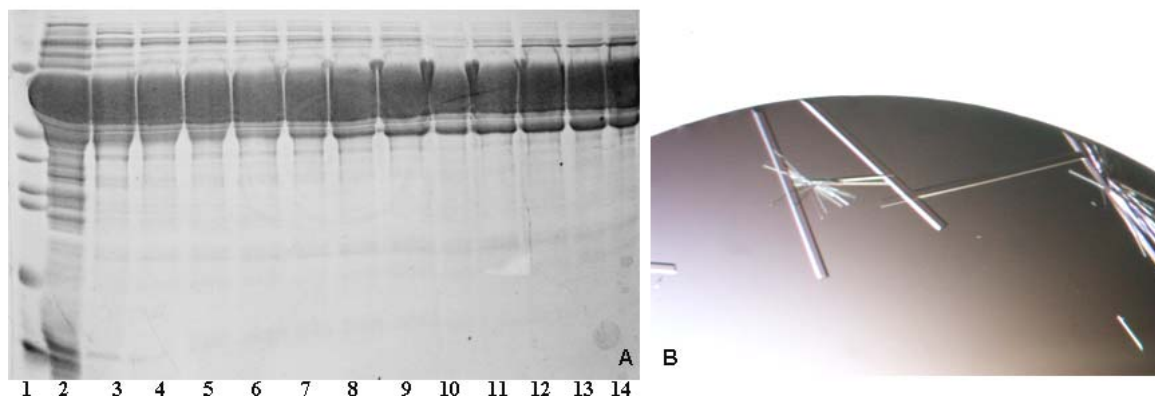


Figure 5.6 Se-Met BT1762 for crystallization.

A: Lane 1: Marker, Lane 2: Se-Met BT1762 purified by IMAC purification, Lanes 3-13: Collected fractions of Se-Met BT1762 from Q12 column. **B:** BT1762 formed rod-like crystals in condition: 0.2 M Ammonium acetate, 0.1 M tri-Sodium citrate, pH 5.6, 30 % (w/v) PEG 4000.

5.2.5 BT1762 structure

BT1762 native structure was determined to a resolution of 1.9 Å by Dr. Susan Firbank. This monomer-like structure unit has cell parameters $a = 107.77$ Å, $b = 129.64$ Å, $c = 86.73$ Å, and $\beta = 90^\circ$, containing 4958 atoms. Residues 32-294 and 300-570 were defined in the native protein. Residues 295-299 are DGTYN were not visible in electron density map. There is no density for the first 13 residues of the structure, which is not surprising because they likely form a flexible extension from the lipid anchor. One BT1762 molecule has 22 α helices, 2 pairs of antiparallel β sheets, and two glycerol molecules (Figure 5. 7). Inside of the 22 helices, 8 helices: $\alpha 1$ (40-56), $\alpha 4$ (103-128), $\alpha 5$ (137-159), $\alpha 6$ (185-204), $\alpha 7$ (215-230), $\alpha 8$ (245-255), $\alpha 14$ (424-439), and $\alpha 15$ (447-455) formed into four TPR units.

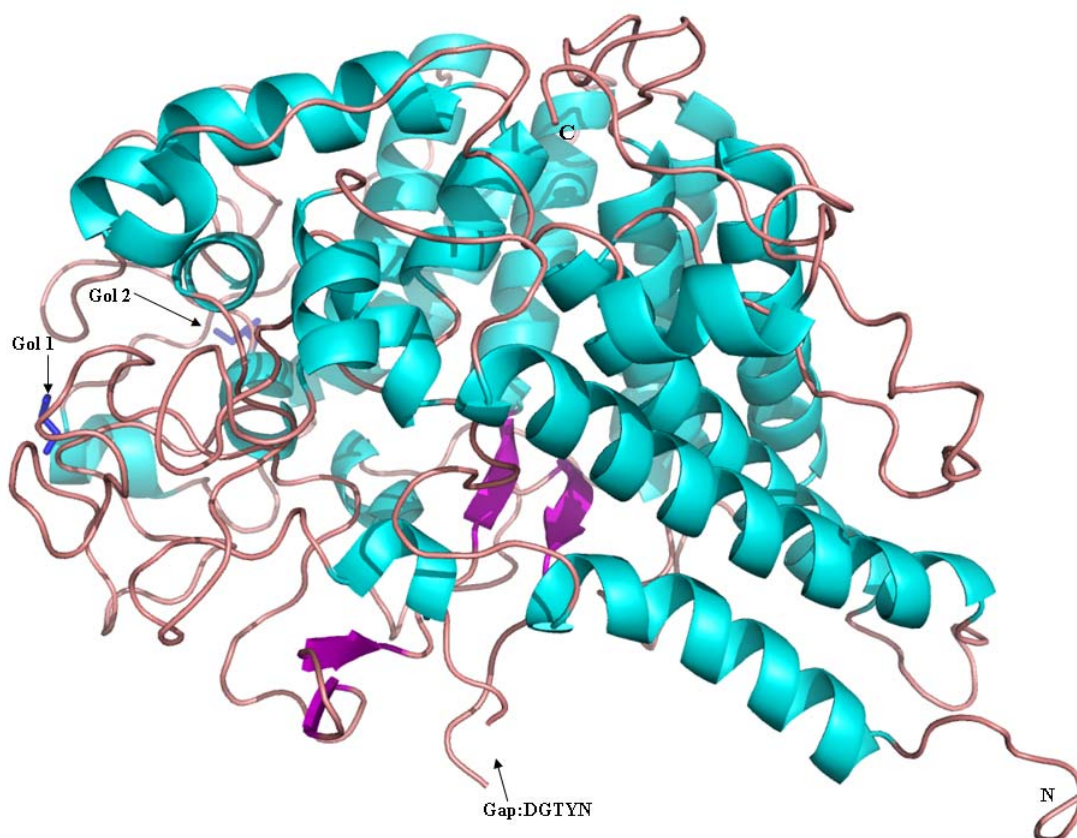


Figure 5. 7 Structure of BT1762.

Cartoon representation shows the structure of BT1762 of residues 32-570, which has 22 α helices, 2 pairs of antiparallel β sheets, and two glycerol molecules bound from the cryo (blue sticks) There is a unstructured gap which has no electron density, suggesting the region is highly flexible in the unliganded structure (residues 295-299: DGTYN - see arrow).

A highly unusual disulphide ring is formed by two adjacent cysteine residues C316 and C317 in BT1762 (Figure 5. 8). This ring is not present in the SusD or BT1043 structures. The presence of reducing agent TCEP (Tris (2-carboxyethyl) phosphine hydrochloride) in the ITC buffer abolished the ability of BT1762 to bind levan (data not shown), suggesting that this disulphide ring is functionally important. There are another two cysteine residues C405 and C407 in BT1762 but they do not form a disulphide in the native structure.

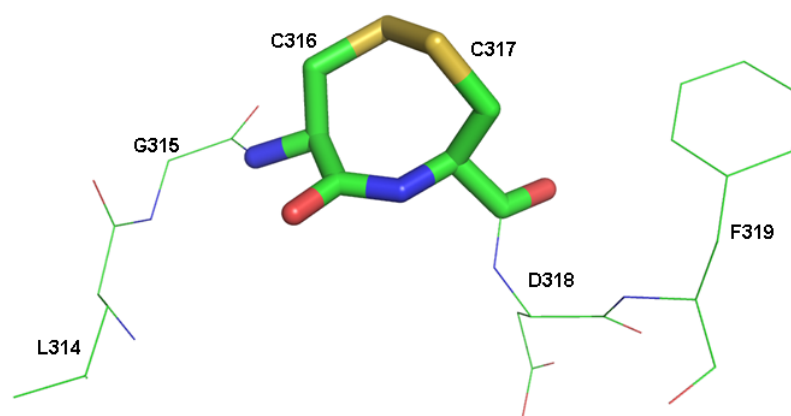


Figure 5. 8 BT1762 disulphide ring.

Stick representation shows the disulphide ring formed by C316 and C317 which is between L314, G315, D318, and F319.

DALI analysis of the 3D structure of BT1762 indicates that the most similar structure is SusD (Figure 5. 9). Unfortunately, the complex structure of BT1762 with L7 was not solved as only very small crystals could be produced, so the binding site is not confirmed.

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	3ck7-A	30.3	2.7	426	496	21	PDB	MOLECULE: SUSD;
2:	3ck8-B	30.2	2.8	429	501	21	PDB	MOLECULE: SUSD;
3:	3ck8-A	30.2	2.8	429	501	21	PDB	MOLECULE: SUSD;
4:	3ck7-C	30.2	2.7	426	497	21	PDB	MOLECULE: SUSD;
5:	3ck7-D	30.2	2.7	426	498	21	PDB	MOLECULE: SUSD;
6:	3ckc-B	30.1	2.7	424	500	20	PDB	MOLECULE: SUSD;
7:	3ck7-B	30.1	2.7	426	497	21	PDB	MOLECULE: SUSD;
8:	3ckc-A	30.0	2.7	425	501	20	PDB	MOLECULE: SUSD;
9:	3ckb-B	30.0	2.7	425	501	20	PDB	MOLECULE: SUSD;
10:	3ck9-A	29.8	2.7	428	508	20	PDB	MOLECULE: SUSD;
11:	3ckb-A	29.8	2.7	425	504	20	PDB	MOLECULE: SUSD;
12:	3ck9-B	29.5	2.7	428	515	20	PDB	MOLECULE: SUSD;
13:	3fdh-A	20.1	3.3	355	472	14	PDB	MOLECULE: SUSD HOMOLOG;
14:	3cgh-A	19.7	3.3	381	507	11	PDB	MOLECULE: SUSD HOMOLOG;
15:	3ejn-A	18.5	3.6	358	450	10	PDB	MOLECULE: SUSD HOMOLOG;
16:	2pl2-B	9.0	3.9	150	194	9	PDB	MOLECULE: HYPOTHETICAL CONSERVED PROTEIN TTC0263;
17:	2qfc-B	8.9	5.2	165	284	9	PDB	MOLECULE: PLCR PROTEIN;
18:	2fbn-B	8.9	3.1	140	151	11	PDB	MOLECULE: 70 KDA PEPTIDYLPROLYL ISOMERASE, PUTATIVE;
19:	2qfc-A	8.8	5.2	165	284	9	PDB	MOLECULE: PLCR PROTEIN;
20:	2hr2-E	8.7	3.4	149	157	8	PDB	MOLECULE: HYPOTHETICAL PROTEIN;

Figure 5. 9 DALI analysis for BT1762 structure.

The most similar structures so far are different versions of SusD, with the Z scores around 30 %. Other structure is less similar with Z score around 20 % (PDB ID 3fdh), which is a SusD homolog from *B. thetaiotaomicron* solved by a structural genomics consortium.

Overlays of BT1762 with SusD and BT1043 reveal that BT1762 matches all TPR units of SusD and BT1043 individually (Figure 5. 10). These high similar TPR units of SusD homologues are on one side spanning helices while another side of SusD and BT1043 are the ligand binding sites. BT1762 has one pair of anti-parallel β -sheets matching one of SusD.

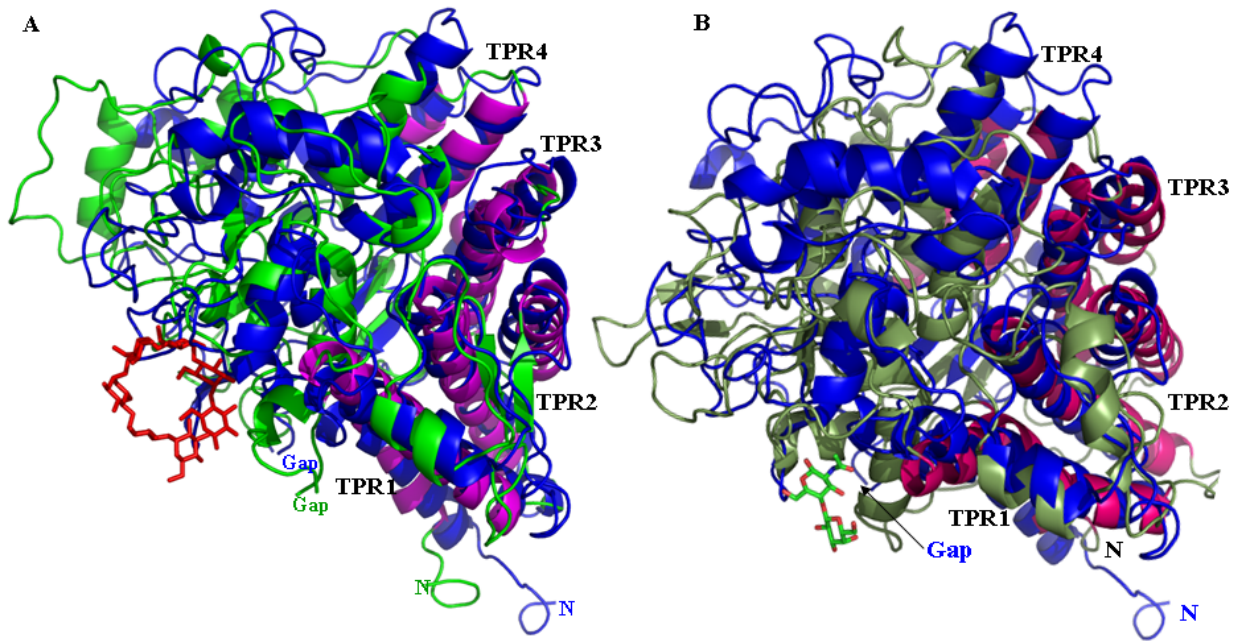


Figure 5. 10 Cartoon representations show the overlay of BT1762, BT1043, and SusD.

A: BT1762 (blue) presents conserved structure when comparing with SusD (green) (PDB ID 3CK9). BT1762 has one pair of anti-parallel β -sheets matching one of SusD. SusD has 8 α -helices (purple) formed into four tetratricopeptide repeat (TPR) units on one side of the structure. BT1762 has 8 α -helices to match all TPR units of SusD: α 1 (40-56), α 4 (103-128), α 5 (137-159), α 6 (185-204), α 7 (215-230), α 8 (245-255), α 14 (424-439), and α 15 (447-455). Maltoheptaose (red stick) binds to SusD on the other side of these TPR repeats. **B:** BT1762 (blue) overlay with BT1043 (khaki) (PDB ID 3EHN) shows the conserved 8 α -helices (pink) TPR units. A molecule of *N*-acetyllactosamine is bound to BT1043 and is shown in green sticks.

The alignment of BT1762 against SusD also revealed that the TPR locations are parallel to each other (Figure 5. 11).

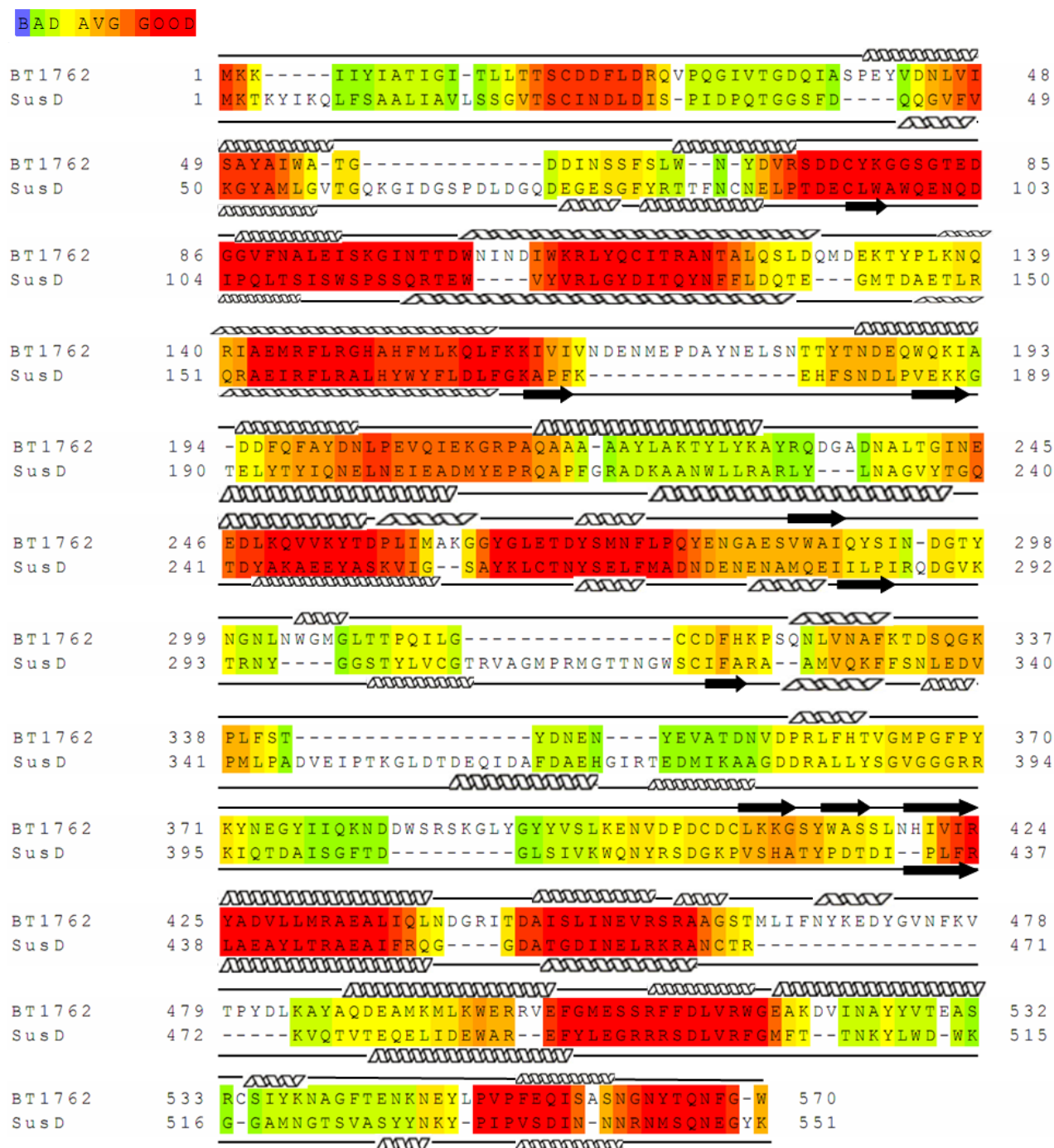


Figure 5. 11 Alignment of BT1762 and SusD.

Tcoffee alignment of BT1762 and SusD. Waves indicate α -helices, black lines indicate loops, and arrows indicate β -strands.

Overlays of BT1762 with SusD and BT1043 also reveal that the location of the ligand binding sites in SusD and BT1043 is conserved and on the opposite side of the protein to the TPR domains. The levan binding site of BT1762 may be around this site as well (see the red circle of Figure 5. 12). However, the ligand binding sites of SusD and BT1043 are both not visible in the structure of BT1762 when overlaid, which indicated that the binding residues are different (Figure 5. 12).

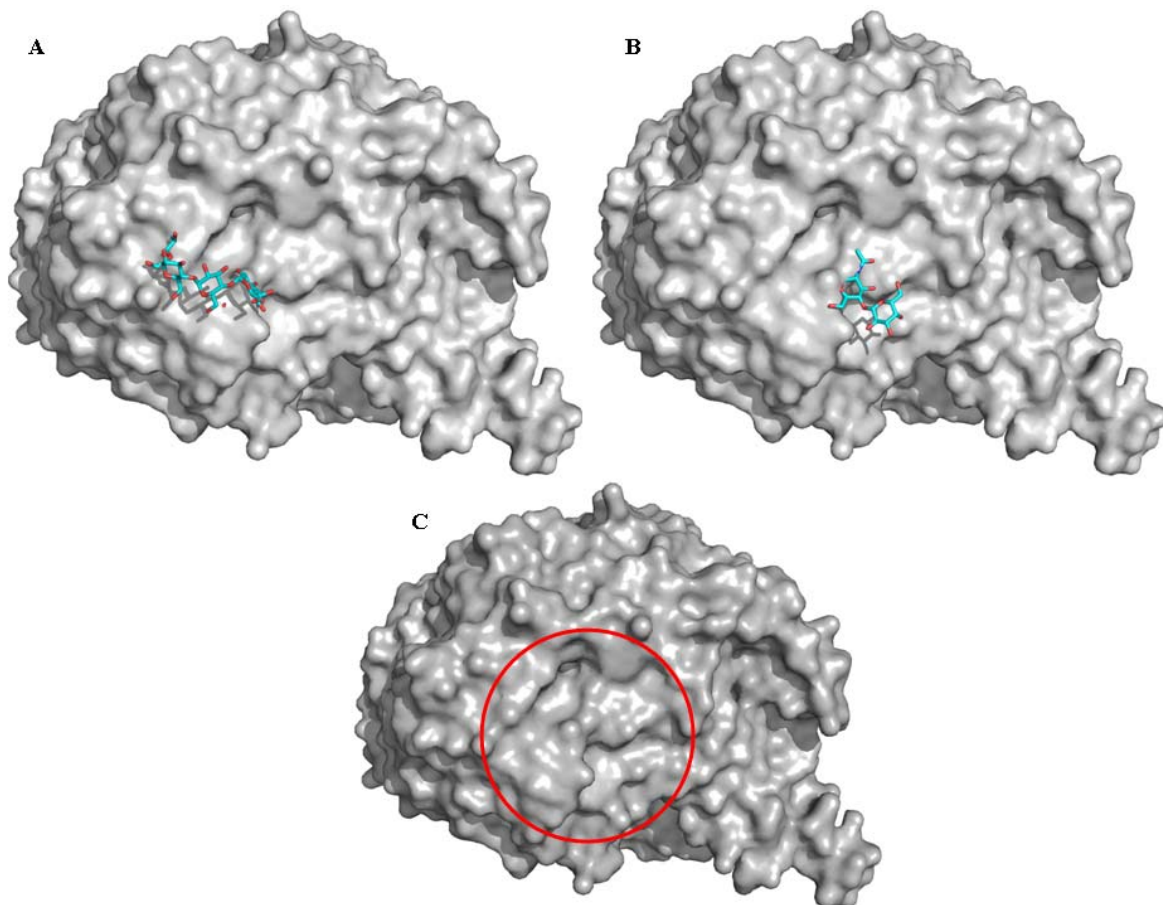


Figure 5. 12 Surface representation shows the possible binding site of BT1762.

BT1762 is shown in all three panels with maltoheptaose from SusD (**panel A**) and LacNac from BT1043 (**panel B**) overlaid. **Panel C** shows BT1762 alone with the location of the putative binding site based on the other SusD structures highlighted by a red circle

Looking at the binding sites of BT1043 and SusD, there are several aromatic residues around ligand. In SusD four aromatic residues form an arch that mirrors the helix of starch (Figure 5. 13). According to this analysis, the aromatic residues around the predicted ligand binding site of BT1762 are potential key residues.

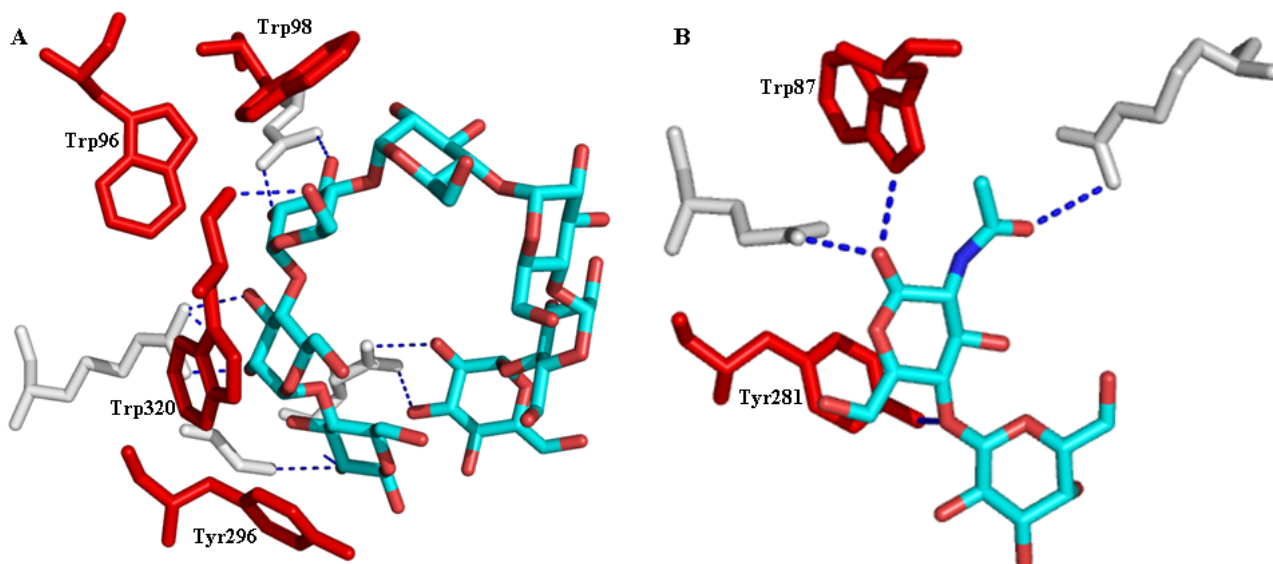


Figure 5. 13 Key residues around ligands of SusD and BT1043.

A: Key residues of SusD around maltose (cyan stick), in which four aromatic residues W96, W98, Y296, and W320 (red sticks) formed one arch around half ring of maltose, while other four residues formed hydrogen bonds with ligand. **B:** Key residues (red sticks) of BT1043 around *N*-acetyllactosamine (cyan sticks), in which have two aromatic residues W87 and Y281 (red sticks), and two charged residues.

Using PyMol, the aromatic residues around the potential binding site are pointed out as W103, W304, Y413, and W414. And the disulphide ring formed by C316 and C317 is also in this area, and it could be looked as an aromatic residue as well (Figure 5. 14).

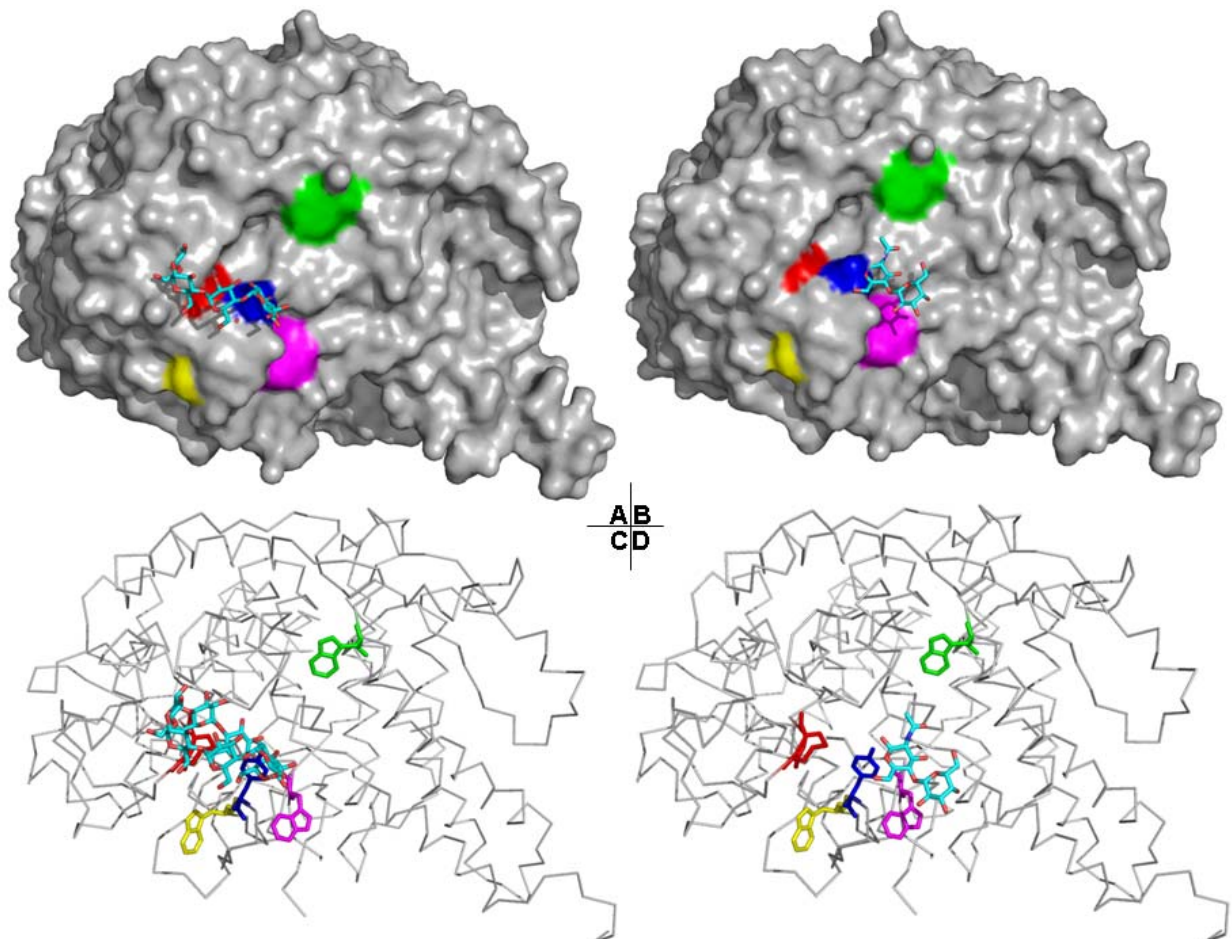


Figure 5. 14 Surface and ribbon representation show the aromatic residues of potential binding area.

A: Surface representation of BT1762 with maltoseheptaose (cyan sticks) positioned from the SusD structure to indicate the possible position of the levan binding site in BT1762. The disulphide ring of C316 and C317 (red), W304 (pink), Y413 (blue), and W414 (yellow) are near maltoseheptaose, which is half-buried inside of the surface of BT1762. W103 (green) is quite far from the maltoseheptaose, but may be still important for ligand recognition in BT1762 as levan forms a more extended structure to starch (Figure 5. 19) and so may bind across more of the protein surface. **B:** Same residues of BT1762 as panel A around *N*-acetyllactosamine of BT1043. **C:** Ribbon representation of panel A. **D:** Ribbon representation of panel B.

The disulphide ring (C316 and C317) and aromatic residues W103, W304, Y413, W414 form a potential ligand binding site of BT1762 (Figure 5. 15). C405, C407 and the gap (residues 295-299: DGTYN) are close to this area as well. In addition a semi-exposed Trp-497 on the other side of the protein has a glycerol molecule bound from the cryo, suggesting a possible involvement in carbohydrate binding.

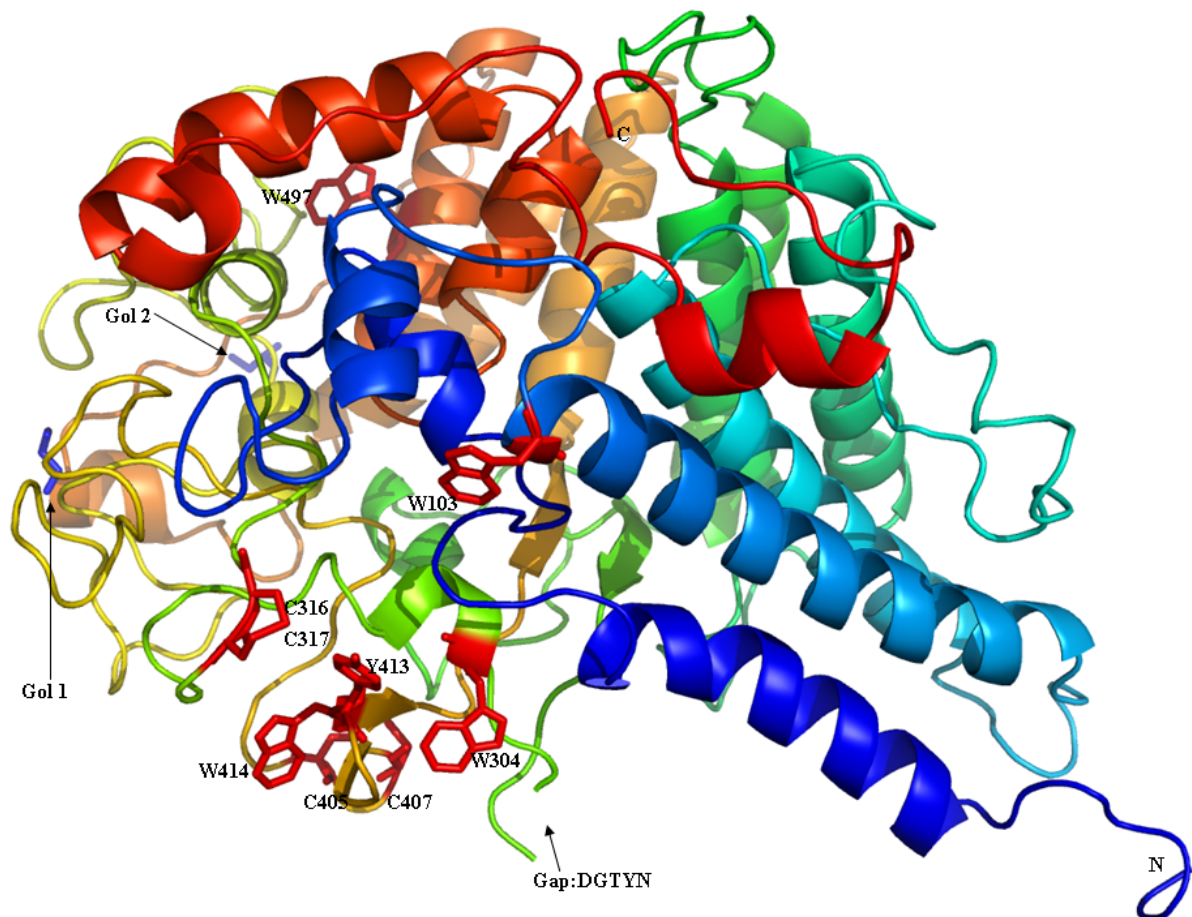


Figure 5. 15 Putative ligand binding site of BT1762.

Several residues are highlighted in red sticks: W103, W304, C316, C317, C405, C407, Y413, and W414. These residues may be involved in ligand binding as they are solvent exposed and in the same region as the binding sites of SusD and BT1043. C405 and C407 are in a loop near W304, Y413 and W414. Y298 in the unstructured loop may also be involved in levan recognition as it is near the same region. W497 is on the opposite side of the protein, but has glycerol molecule bound in the crystal structure. Glycerol can sometimes mimic a carbohydrate due to its hydroxyls and may indicate that W497 is involved in levan binding despite its location.

5.2.6 Mutational analysis of predicted binding site of BT1762

5.2.6.1 BT1762 mutant protein production

Alanine mutants of W103, Y298, W304, Y413, W414, and W497 from BT1762 were produced to assess the importance of these residues in levan binding. Solvent exposed aromatic residues were chosen as they are a common feature of carbohydrate binding sites where they stack against the sugar rings. In addition, one cysteine residue, C316, from the disulphide ring was mutated to alanine and serine separately to analysis the function of the disulphide bond. C405 in the other pair of cysteine residues was mutated to alanine and serine separately as well. Site directed mutagenesis was carried out using primers in Table 5. 2 and BT1762-minipRSETA as template. All mutants could be produced in the cytoplasm of *E. coli* in a soluble form (Figure 5. 16).

Primer ID	Sequence (5'----->3')	Note
BT1762-3	CAATACCACCGACGCGAATATCAATGATATATG	W103A F
BT1762-4	CATATATCATTGATATTCGCGTCGGTGGTATTG	W103A R
Bt1762-5	GAAAATGCTGAAAGCGGAACGCCGCGTAG	W497A F
BT1762-6	CTACGCGGCGTTCCGCTTTCAGCATTTTC	W497A R
BT1762-7	GGTAACCTGAATGCGGGAATGGGACTG	W304A F
BT1762-8	CAGTCCCATTCCCGCATTTCAGGTTACC	W304A R
BT1762-9	GAAGAAAGGTTCTGCTTGGGCCAGTTCTC	Y413A F
BT1762-10	GAGAACTGGCCCAAGCAGAACCTTTCTTC	Y413A R
BT1762-11	GAAAGGTTCTTATGCGGCCAGTTCTCTG	W414A F
BT1762-12	CAGAGAACTGGCCGCATAAGAACCTTTC	W414A R
BT1762-13	CAGATACTGGGCGCCTGCGACTTCCAC	C316A F
BT1762-14	GTGGAAGTCGCAGGCGCCCAGTATCTG	C316A R
BT1762-15	CAGATACTGGGCACCTGCGACTTCCAC	C316S F
BT1762-16	GTGGAAGTCGCAGGTGCCAGTATCTG	C316S R
BT1762-17	GTAGATCCCGATGCTGACTGTCTGAAG	C405A F
BT1762-18	CTTCAGACAGTCAGCATCGGGATCTAC	C405A R
BT1762-19	GTAGATCCCGATAGTGACTGTCTGAAG	C405S F
BT1762-20	CTTCAGACAGTCACTATCGGGATCTAC	C405S R
BT1762-21	CAACGACGGTACTGCCAACGGTAACCTG	Y298A F
BT1762-22	CAGGTTACCGTTGGCAGTACCGTCGTTG	Y298A R

Table 5. 2 Primers for BT1762 mutations.

F: Forward primers, R: Reverse primers.

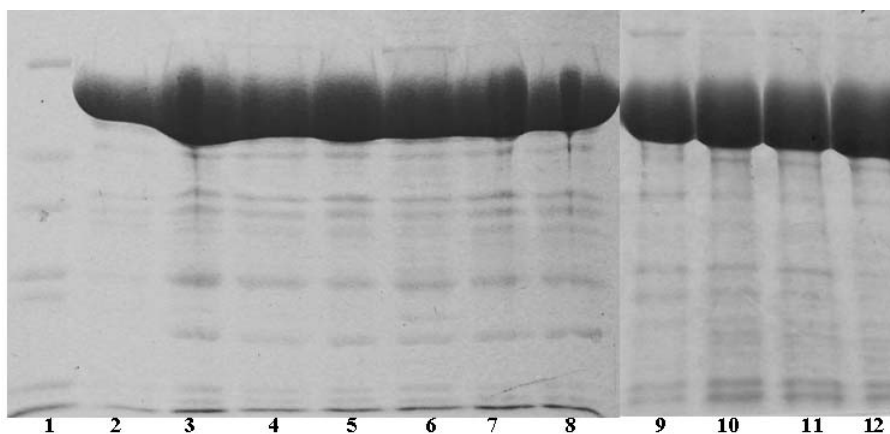


Figure 5. 16 SDS-PAGE showing purified BT1762 mutants.

Lane 1: Marker, Lane 2: BT1762 WT, Lane 3: W103A, Lane 4: W497A, Lane 5: W304A, Lane 6: Y413A, Lane 7: W414A, Lane 8: Y298A, Lane 9: C316A, Lane 10: C316S, Lane 11: C405A, and Lane 12: C405S, All mutants were purified by IMAC in a single step from the CFE of the BL21 expression strain.

5.2.6.2 Ligand binding analysis of BT1762 mutants.

All the 10 mutants of BT1762 were analysed for ligand binding ability against the levan using WT as control. (Figure 5. 17 and Table 5. 3). ITC revealed that C316A and C316S abolished binding affinity, indicating that the disulphide ring is essential for the ligand recognition. W103A shows no binding to levan, whereas Y413A displayed only very weak binding, which suggest W103 and Y413 also play key roles in levan recognition. W304A and W414A showed some reduced but still significant affinity, whereas the Y298A mutant has the same affinity as wild-type indicating that this residue is not involved in ligand binding. The ITC data also revealed that W497 and C405 are not involved in levan recognition in BT1762. Parameters for mutants are from single titrations only. Although the affinities determined for each of the levan binding mutants is similar to wild-type, there are significant differences in thermodynamics and stoichiometries (Table 5. 3). These may be actual differences driven by the mutation, but are more likely to be errors of the fit caused by using the estimated molar value of levan calculated from the wild-type titrations for all mutants.

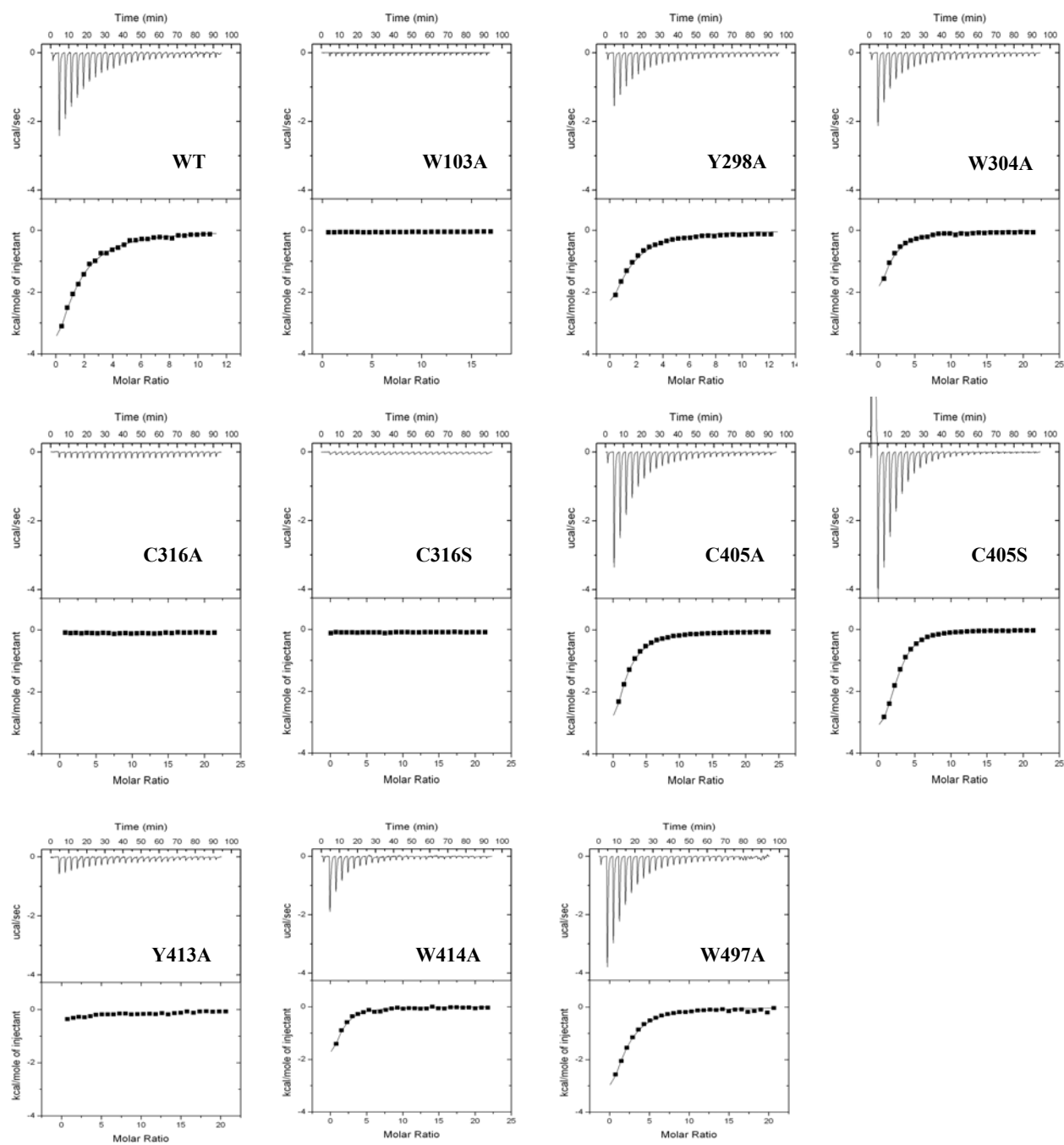


Figure 5. 17 ITC data show the BT1762 mutants binding ability against levan.

The upper part of each panel shows the raw heats, the lower part is the integrated data. This is fit to a single site model to derive K_a , ΔH , and stoichiometry. Levan was at 10-20 mg/ml and proteins were used at 100-200 μ M. To obtain the molar concentration of levan, the WT was fit such that $n=1$ (number of binding sites on the protein). This assumption is deemed valid due to the structural and mutagenesis data that indicates only a single binding site on the protein. All mutants were then fit to this same molar concentration (10 mg/ml = 3 mM).

Mutant	$K_a \times 10^3 (M^{-1})$	$\Delta H (kcal mol^{-1})$	$T\Delta S (kcal mol^{-1})$	n^*
WT ^a	9.1 ± 0.8	-7.7 ± 1.5	-2.8 ± 0.3	1.0 ± 0.1
Y298A	13.3 ± 1.7	-5.5 ± 1.0	0.2	1.0 ± 0.2
W304A	12.1 ± 1.0	-4.5 ± 0.6	1.1	1.0 ± 0.3
C405A	14.5 ± 0.2	-4.5 ± 0.5	1.2	1.9 ± 0.0
C405S	32.3 ± 2.4	-3.6 ± 0.9	2.4	2.4 ± 0.0
W414A	24.3 ± 5.1	-2.7 ± 0.5	3.2	1.2 ± 0.3
W497A	15.4 ± 1.1	-4.3 ± 0.2	1.3	2.0 ± 0.1

Table 5. 3 ITC parameter of the BT1762 mutants binding to Levan

*: The ITC data were fitted to a single site binding model for all titrations. The n-value was iteratively fitted to as close as possible to one, by adjusting the molar concentration of levan from 10 mg/ml to 3 mM. a: The parameters for WT are average values from three repeats. Parameters for mutants are from single titrations individually. The K_a values are quite similar except for C405S and W414A, which may be caused by the limitation of prediction of software Origin.

Circular dichroism spectroscopy revealed that none of the mutations had caused significant changes in the secondary structure compared to wild-type (Figure 5. 18).

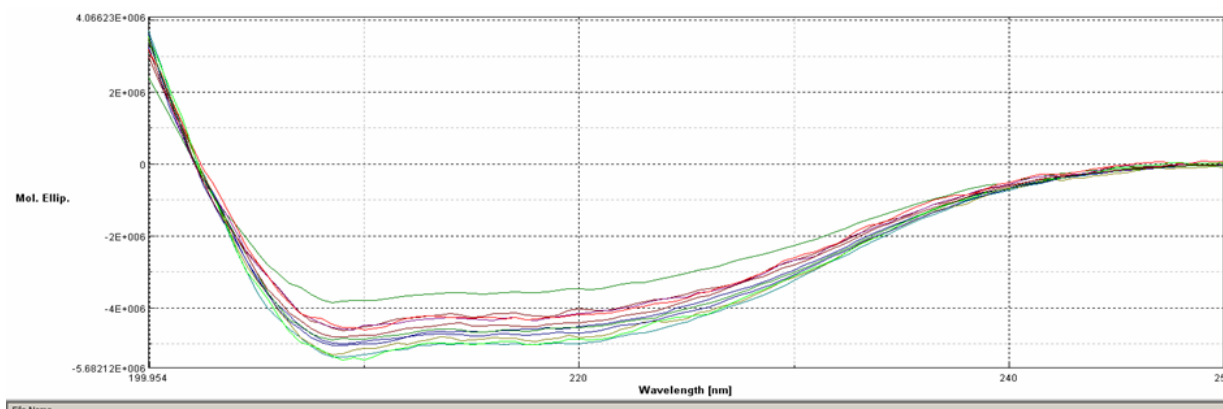


Figure 5. 18 CD spectroscopy BT1762 wild type and mutants.

All mutants and WT were checked at 10 μ M concentrations in 10 mM Tris pH 8.0 buffer. Data revealed these proteins keep the secondary structure as wild type.

5.3 Discussion

The sequence similarity of SusD homologues is low (only ~7% between SusD and BT1762), but the structural comparison here reveals that the general fold, TPR repeats and general location of the binding sites are highly conserved. TPRs are mainly involved in protein-protein interactions and it has been suggested that the TPR domains are involved in the formation of a complex with the SusC porin and possibly SusE, F and G (Koropatkin *et al.*, 2008). It seems likely that the TPR units of BT1762 play the same scaffold role and bind the SusC homologue BT1763. It is also possible that the TPRs in BT1762 bind the predicted outer membrane GH32 enzyme BT1760 to form a multiprotein complex, although no interaction between these two proteins was detected on a native gel (data not shown). However, the proteins were expressed in a soluble form rather than tethered to the membrane and this is known to have a significant affect on the ability of normally membrane associated proteins to interact (Moore *et al.*, 2008).

SusD could bind a range of malto-oligosaccharides, cyclodextrins and starch, while BT1762 can just bind polymeric levan but not to levan-oligosaccharides smaller than seven sugars, suggesting the protein has a larger binding site than SusD. In addition BT1762 is specific for β -2,6 linked fructan, showing no affinity for β -2,1 linked inulin. Levan is predicted to have a secondary structure which comprises 5 fructose rings in one left-hand twist of the helix with the distance of 12.74 Å between sugars n and $n+4$ (Figure 5. 19A). Inulin has a much tighter helical structure, thus explaining the lack of interaction with BT1762 (Figure 5. 19B). Starch has a helical secondary structure with the helix repeated every 7-glucose rings with an average diameter of 10.51 Å (Figure 5. 19C). The structural difference between starch and levan explains the difference in the residues involved in the binding sites (Figure 5. 15).

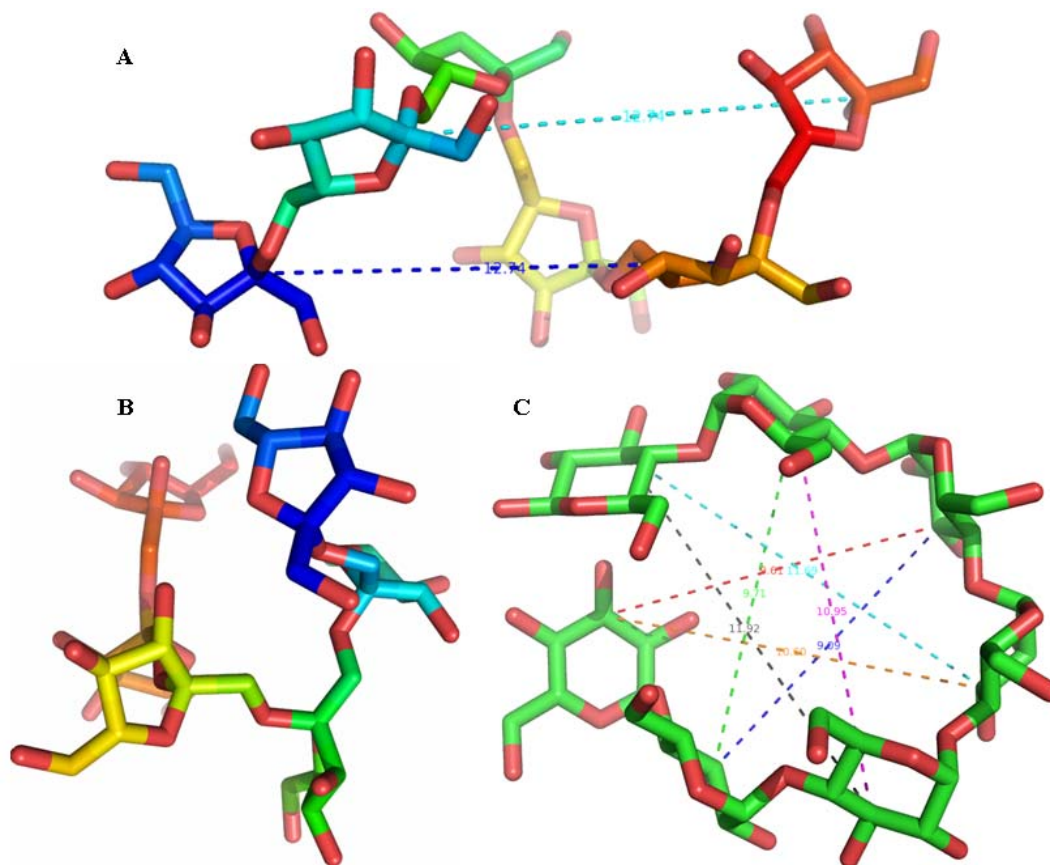


Figure 5. 19 Structures of levan hexaose, inulin hexaose, and maltoheptaose.

A: Levan hexaose modelled using SWEET2 forms a left-hand helix with the repeat unit of 5 fructose rings. The distance of C2 of frc1 (blue) to C2 of frc5 (orange) is 12.74 Å, which equals the C2 of frc2 (cyan) to frc6 (red). **B:** Inulin hexaose modelled using SWEET2. **C:** Maltoheptaose from the SusD complex forms a much tighter twist.

SusD has a crescent shaped binding site composed of W96, W98, Y296, and W320, that docks exactly with the tight helical twist of polymeric starch (Figure 5. 20A). In contrast, BT1762 has much more extended binding site, composed of a unique functional disulphide ring (C316 and C317), a tyrosine (Y413), and a tryptophan (W103) which fits the more extended model of a levan chain in two possible orientations (Figure 5. 20B/C).

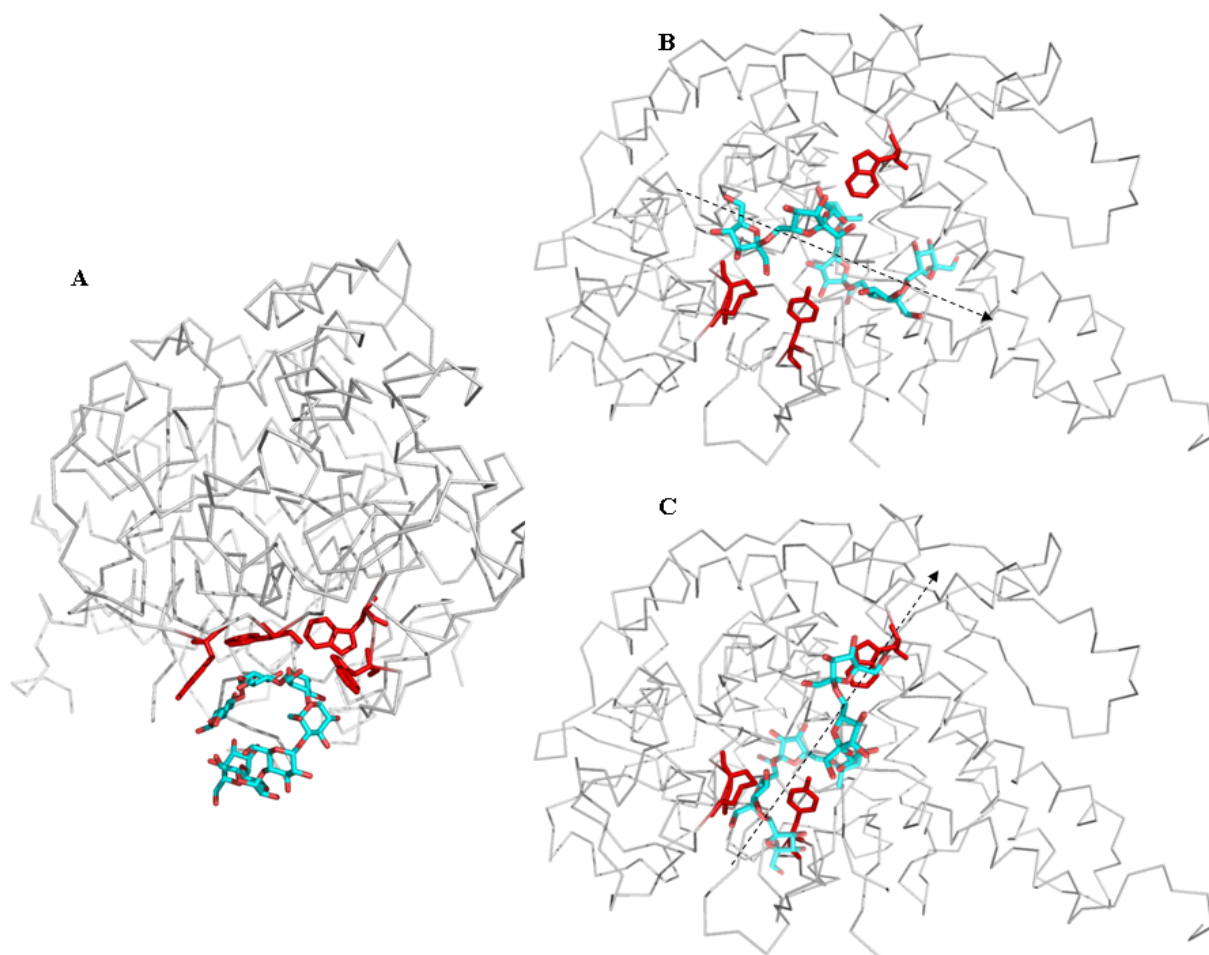


Figure 5. 20 SusD and BT1762 ligand binding sites.

A: Ribbon representation shows SusD complex with maltose (cyan sticks) in the arch formed by four aromatic residues (red sticks). **B and C:** Ribbon representation shows BT1762 with aromatic residues W103, Y413, and disulphide ring (C316 and C317) highlighted as red sticks. Levan hexaose (cyan sticks) is put among these residues to mimic the possible binding orientations (see arrows).

The disulphide ring in BT1762 formed between adjacent cysteines (C316/317) is a very unusual feature of any protein and has only been seen before in the active site of a quinoxinoprotein alcohol dehydrogenase (Figure 5. 21) (Oubrie *et al.*, 2002). However, the disulphide ring in this enzyme was shown to be functionally nonessential, unlike in BT1762 where it seems to be essential for levan binding (Kay *et al.*, 2004). It is possible the disulphide ring has a structural, rather than a carbohydrate-binding role, but as CD spectra recorded for the C316A and C316S mutants are indistinguishable from the wild-type protein spectra indicates that this is not the case (Figure 5. 18). Precisely how this disulphide ring could interact with levan is currently unclear, however it may stack against the fructose rings in the same way as an aromatic residue.

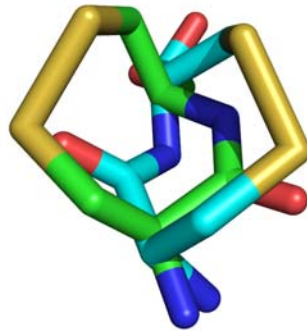


Figure 5. 21 Overlay the disulphide rings from BT1762 and quinohemoprotein alcohol dehydrogenase.

The green ring is from BT1762, while the cyan ring is from the alcohol dehydrogenases of *Comamonas testosterone* (PDB 1KB0). The disulphide bonds are present in yellow while oxygen is red, and nitrogen is blue.

BT1762 may be essential for the utilisation of levan long chains, as SusD is essential for use of oligosaccharides bigger than 5 glucose units. The decreased binding affinity to L7 also suggests BT1762 mainly binds long chains, which is also partly due to that fact that L7 has only one possible helical unit and it is not a standard shape of levan because no other helices before and after it to form internal hydrogen bonds to sustain the secondary structure. Combined with the enzyme BT1760 on the outer membrane which can cleave long levan chains, BT1762 can transfer digested levan oligosaccharides through SusC homologue BT1763 to the periplasm. BT3082 and BT1759 can degrade levan oligosaccharides efficiently in the periplasm. A model for the role of BT1762 in levan degradation system of *B. thetaiotaomicron* is shown in Figure 5. 22.

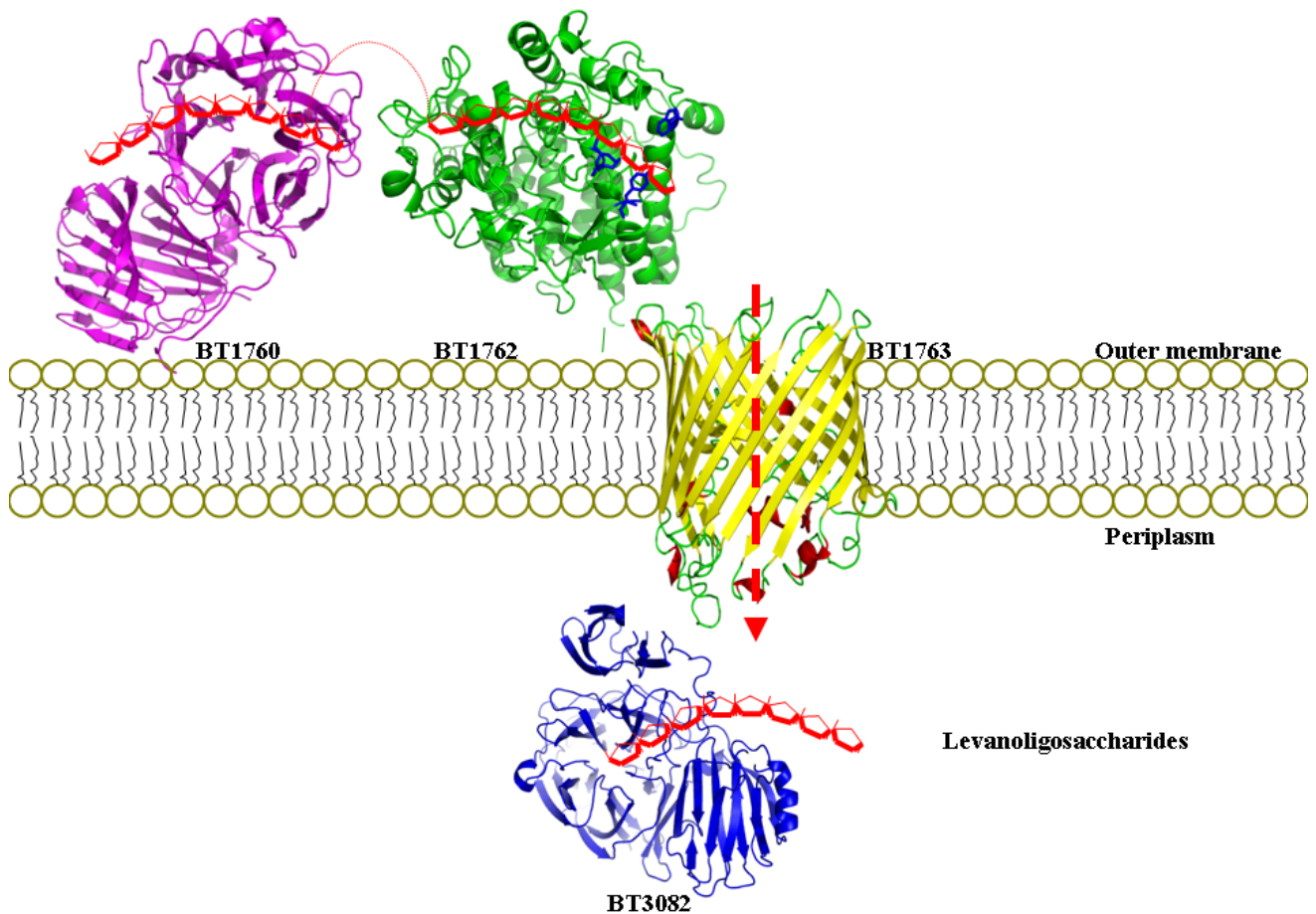


Figure 5. 22 Model for the function of BT1762.

BT1762 (green) binds levan (red rings) in a binding site (blue sticks), which anchors levan in close proximity to the GH32 levanase BT1760 (purple) and enable the enzyme to hydrolyze the levan. The cleaved levanoligosaccharides bound SusD are then presented to the entrance of SusC homologue BT1763 porin channel (yellow; structure shown is a β -barrel porin from *Salmonella typhimurium*, (PDB ID 1A0T) (Forst *et al.*, 1998), as the structure of BT1763 has not been solved yet). In this way, BT1762 binds levan tightly to enable enzymic hydrolysis, while levanoligosaccharides interact weakly with BT1762 enabling its efficient transfer to transporter BT1763. Levanoligosaccharides in periplasm are digested by BT3082 (blue).

Chapter 6 Analysis of the trans-envelope signalling apparatus of *B. thaitaomicron*

6.1 Introduction

Many ECF sigma factors are known to signal via trans-envelope signaling. The best studied similar signaling system is the Fec system of *E. coli*. *E. coli* employs the Fec system to transport ferric citrate into the bacterium, which has seven genes: *fecIR*, *fecABCDE* (Figure 6. 1). First, cells respond to iron deficiency by synthesizing the regulatory proteins FecI and FecR. But the low iron concentration is not enough to induce transcription. The presence of ferric citrate is also needed for this transcription (Braun *et al.*, 2003b). The system is activated by ferric citrate in the form of dinuclear ferric citrate (Fe^{3+} citrate)₂ binding to the OMP FecA which has a β -barrel domain, a plug domain and an N-terminal periplasmic extension. The signal from FecA is transmitted to the C-terminal periplasmic domain of the transmembrane protein FecR through its N-terminal extension. The activated N-terminal cytoplasmic domain of FecR then stimulates the ECF sigma factor, FecI, to bind the RNAP to transcribe the *fecABCDE* operon which directs the synthesis of proteins that mediate ferric ion transport (Angerer *et al.*, 1995; Enz *et al.*, 2003; Kim *et al.*, 1997; Ochs *et al.*, 1995). An energy-transducing system is associated with the process of iron up-taking which contains TonB, ExbB, and ExbD. The C-terminal region of TonB interacts with a specific region of the N-terminal extension (TonB-Box) of FecA (Braun and Mahren, 2005; Koebnik, 2005). It is believed that the binding with TonB changes the structure of FecA which results in the dissociation of ferric citrate from FecA.

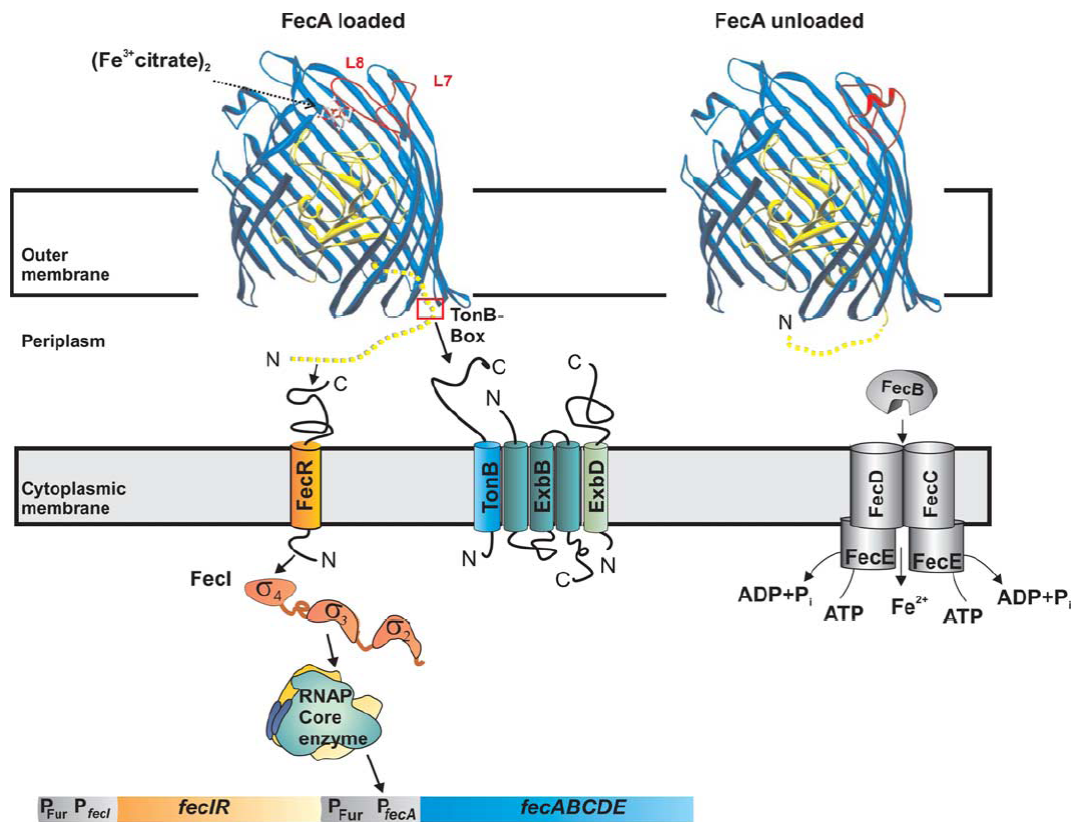


Figure 6. 1 Schematic of the FecAIR transmembrane signaling apparatus from *E. coli*.

In FecA, part of the barrel is not drawn to show the plug domain (yellow). The model depicts the movement of loops L7 and L8 upon binding of dinuclear ferric citrate. FecA signals its occupation by ferric citrate into the periplasm, where it interacts with FecR, which transmits the information across the cytoplasmic membrane and activates the FecI sigma factor. The TonB–ExbB–ExbD complex is involved in transport and signaling across the outer membrane. Cited from (Braun and Mahren, 2005).

The interacting site of the FecR with FecA is located between residues 237-317 of FecR within which region has a leucine heptad motif that is conserved in other predicted FecR like proteins. Mutations around this region reduce most of the FecR induction activity and its ability to bind *in vitro* to the N-terminal periplasmic extension of OMP FecA (Figure 6. 2) (Braun *et al.*, 2003b). That FecI does not induce the transcription of *fec* genes without FecR would suggest the FecR is not or not only an anti-sigma factor, but is in fact an activator of FecI.

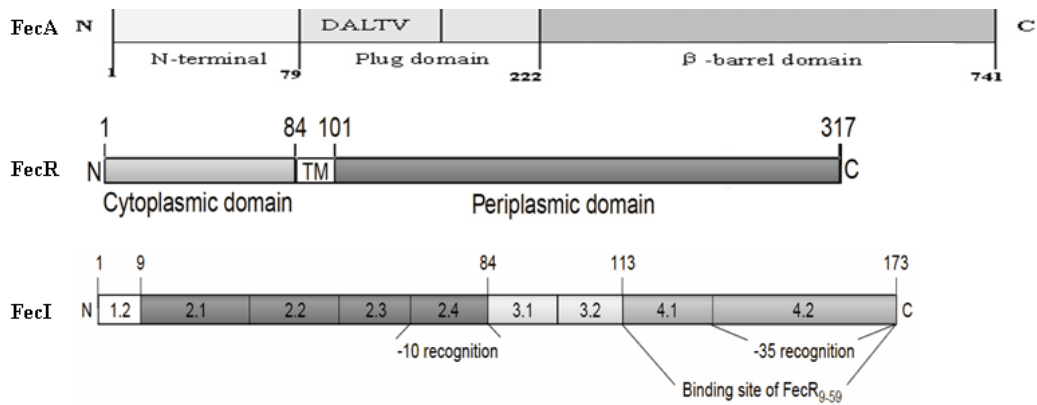


Figure 6. 2 Domain structures of FecAIR.

The bottom panel shows the sigma factor FecI. Domains 1 to 4 are shown as is the binding site for the cytoplasmic domain of FecR. The middle panel shows anti-sigma FecR which has a single cytoplasmic membrane spanning domain (TM). The N-terminal is the cytoplasmic domain which interacts with the sigma factor; the C-terminal periplasmic domain interacts with the N-terminal extension found on FecA (The top panel). Cited from (Braun *et al.*, 2003b; Garcia-Herrero and Vogel, 2005).

TonB is a cytoplasmic membrane protein which utilises the proton motive force to provide energy for active transport of large molecules across the outer membrane by the TonB dependent OMP. The mechanism for this is via a contact in the periplasm between the TonB and the OM receptor (Figure 6. 3).

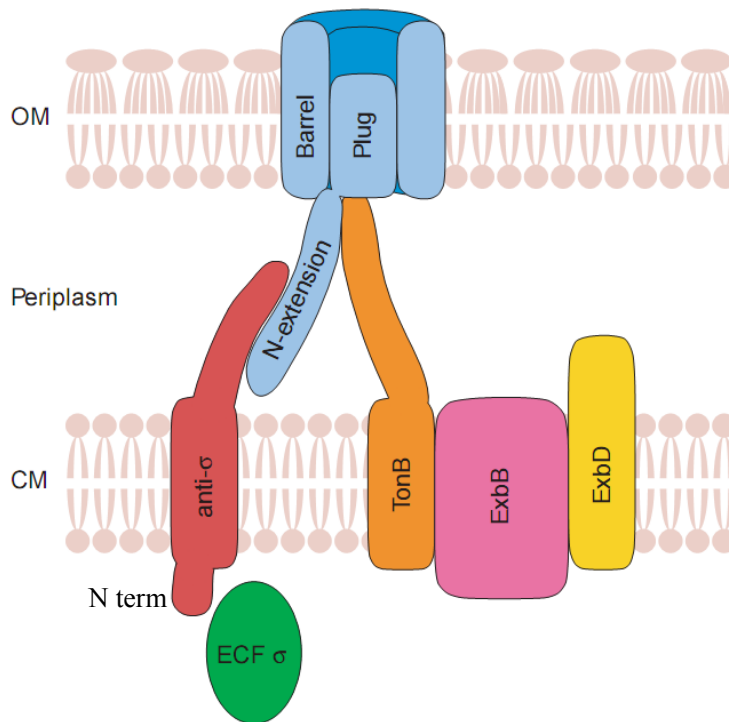


Figure 6. 3 TonB-dependent system.

OM means the out membrane. CM means the cytoplasmic membrane. TonB/EXbB/ExbD complex binds SusC homolog Plug domain, while the N-terminal tail of SusC homolog (blue) interacts with anti-sigma factor in periplasm. Cited from (Koebnik, 2005).

TonB normally forms into a complex with cytoplasmic membrane proteins ExbB and ExbD. TonB-dependent signaling mechanism appears to be common. TonB-dependent proteins are found to have four different types depend on the structure analysis with presence or absence of signal transduction (Figure 6. 4). Without the function of signal transduction, TonB-dependent proteins are named receptors (Koebnik, 2005).

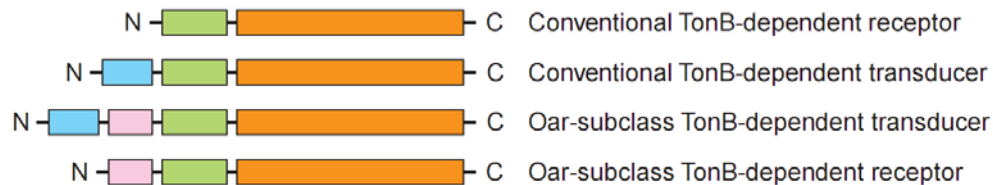


Figure 6. 4 Diversity of TonB-dependent proteins.

All TonB-dependent proteins have a C-terminal β -barrel (orange) and a plug domain (green), TonB-dependent transducers have an extra N-terminal periplasmic extension (blue) of about 70 amino acids. Oar-subclass proteins often have another additional domain in the N-terminal region (pink) which include *Myxococcus*, *xanthus*, *Bacteroides*, *Xanthomonasm* and *Xylella* species. Cited from (Koebnik, 2005).

B. thetaiotaomicron contains at least 12 ECF σ /anti- σ polysaccharide utilization loci (PUL), which have all essential components for trans-envelope signaling. The anti-sigma factors have homology to FecR, the sigma factors have homology with FecI, and the SusC homologues are likely TonB dependent transducers as they have the N-terminal extension (Koebnik, 2005). The sequence alignments of PUL elements to Fec proteins reveal more details about the conservation. ECF sigma factors of *B. thetaiotaomicron* are aligned to FecI of *E. coli* (Figure 6. 5) while anti-sigma factors are aligned to FecR (Figure 6. 6 and Figure 6. 7) and the SusC transducers are aligned to FecA (Figure 6. 8). The ECF sigma factors' alignment indicates the N-terminal is variable; the middle is conserved and then followed a flexible C-terminal domain. Anti-sigma factors cytoplasmic domains (N-terminal) (Figure 6. 6) have low similarity, while the periplasmic domains (C-terminal) including FecR are highly conserved. The barrel domain of SusC transducers are conserved, whilst the N-terminal extensions are diverse. Furthermore, the alignment of SusC transducers also indicates that non-ECF sigma PULs have no N-extension tail (Figure 6. 8). In addition, *B. thetaiotaomicron* possesses 8 TonB genes and the SusC transducers contain TonB boxes, which also suggests the similar energy-dependent process exists in this organism involved in the transport of polysaccharides (Xu *et al.*, 2003).

BAD AVG GOOD

BT0188	1	RYATDTLMDERELVLRLLIGDEDAFCGLYAAYKRNLLYFAMKFM	KSREPAEDI	53
BT0752	1	MNLPEV--SEK-LIEQLNHGSTKAPDKRIYHTVYLYLCAIAVYVY	NDNRVAGEI	50
BT1278	1	MTPNQ---DER-LL-----AQQFETISTKYYSVVVYFALMLL	KSERDAKDI	42
BT1617	1	METFD---EKQLLKAISEGDEKAFKTFFLYVYFRIKGFINGLL	QSQRKAEDE	49
BT1877	1	MENT---ETL-IVEQLKIGNENAYRYIYONHYALLCYVANGYL	FDQFLSEFI	48
BT2198	1	MIHSLP--SSE-PL-----TAQKLFSELYVSYARLVRFASLYM	GAMGDAENI	45
BT2463	1	MPAKL---T-----RPENSPUNIIYRYYKNSFLVRYEYI	HSEMVAEDE	40
BT3517	1	NASDIHTLSDSLLWKRFLLEGDSSAYSQIYNQTVQELFRYGLLYT	SDRELVKDC	53
BT3748	1	MTA-----INFSIYTTYRRAFLFTLSYV	HADLVAEDE	34
BT4636	1	MFTQDE--SYIKWKLFLLEGDDQAYSNIYTHYIQVLYNYGLQIT	PDSEIVKDC	50
BT4705	1	MKISF---SRQ-----TKERAFFQLYEDYVAFPCLYAERFY	PDKEVREDE	42
BT4722	1	IVDSIN--EKR-LLELTKNGSFOAFERLYMYSGLYVNFIMRIS	GNQYMAEEV	51
FecI	1	MSDRAT--T-----TASLTFESLYGTHHGWLKSWLTKLI	SAAPDADDI	41
BT0188	54	FQDAPFVYVQSRFFIN---PDASFSYLYTIVENHILRQIRDMANEDK	LKEHIL	104
BT0752	51	VNDVPVSVKQNRHHIT---Y--PALPFLRRATQNASISYLRSSAFNER	IMTEQM	99
BT1278	43	FQDVFTKLWTKPELWT---EVPNPTFYIYTLTSTTLNFIKHKKVELA	YQEKII	93
BT1617	50	SQDIFLTLWNRSSLH---TINNLKPYLFRISENAVYRHIERALLFRNY	QQKET	100
BT1877	49	VGDTIPHLWEIRETLD---ISVSIRSVELLRAVENRRCINVLNHSERE	KREIAF--S	97
BT2198	46	VQDFFLYLWERKEILP---ELQQPDAVLFSAVENRCLNHLRSQLSIV	DRRQPLS	96
BT2463	41	ASESLIKLWQWIQDNE---VENIEPNLLSILRHKALDYLRRHESMK	QQVITRIS	90
BT3517	54	INDVPVKIYTNRAKLT---ETDNIIAAYLMVALENTLFRALKKTS	SDSFSLDEADE	104
BT3748	35	VSEAIYYLWELSEKQE---IPSIEAVLITYIRKSELYLKKHLQVQEN	VYQNL	84
BT4636	51	FQDVFPKIKYAKKKLT---VPQHPKIVYLNIALENNIYNTFNQERL	QKNYAFSLY	101
BT4705	43	VQDVFTSLWDRLOTDSFDLQSETALGYIKOVENSCLNFKKHQYEW	SYAENIQ	96
BT4722	52	VQSAPIRVKEVREVE---PESSFISFLCTIAENLLMMYQRQTV	VEY--VY--N	98
FecI	42	AQDTFLRVMVS-ETLS---TIRDPRSFLCTIAKRVNVDLFRRNALE	K--AY--L	87
BT0188	105	S-----HAVDSA-NE-TNNKILFDDLKDVLSRALEQLTYRQREV	FNMSKDLQ	149
BT0752	100	EEIWAFLNHILS--SDNPLQALESSEMNEIILRKVEELPAKCRAV	FNASLYEG	151
BT1278	94	EKS--LIDE-LFQ--SEDTLNPIYYKEAQLIIRLVLEELPEQRN	IFEMSRFEN	142
BT1617	101	EK--YSP---PQES-NE-TDDTIHLKLELELLVTMVVERNPPQR	QKIYENBRESG	147
BT1877	98	ALMPDEITDDKIILSDSHPLGILLERELENEIYKATCQLPDECR	EVFAKSRFEG	151
BT2198	97	DIMEQEFKLLKLYSLQLLD--DSQMSIDEVEKQICRAIDSLPER	CREIFVMSKLG	149
BT2463	91	EKQNEELALRLSSL-EDCNPNEIFPSKEVMDIVQRTLQSLPEQ	TSEIFTLSEFGN	143
BT3517	105	KE--DQSEEHFSTP-ETIYINKEQEKNTHMKVHAMSSLTTRQR	REIVYRYIKD	155
BT3748	85	DKGQRELEIRISTL-EACDPKEVMSEELRSEVKTLLAGNPEK	TRIAFISDRLOG	137
BT4636	102	QT--E---EQLTVK-NE-FIDQEARHEEMNNIKRMMKILTFRQ	REVIYYRFIEE	148
BT4705	97	KKAP-L-----Y---ETDTSVYTLDELRYMELYETLNEEL	PEMYRTVFMKSPFEG	141
BT4722	99	EYLNKTGVD-----RDSQTEESIDLRFLENYIDSLAELPAQR	KKIYFILSERQK	147
FecI	88	EMLALM-PEGG-A---PSPEERESQLETQLLDSMLDGLNGKTR	EATFLSGLDG	136
BT0188	150	NSHREIAREALQVSVRTVQERISVSLKVIHAYLTKYSGTSAD	ILLILLC--LNF	200
BT0752	152	KSYREIAREEQINVATVAVQNHIALTLRESLGTPI-----	MIAIL--MFL	195
BT1278	143	NSNLEIARELWIKRRTVEHRIYTLLEMKKI---E-----	FAPFL--LFF	183
BT1617	148	NSNLEIARELGINRRTVEHRLSQALDIRKIL-----	FITFI--LFF	187
BT1877	152	KSYREIARELQISINTVKYHINKVALASLNAHLSKY-----	LISLLLF--FFI	196
BT2198	150	MKYREIARESLQISQNTVEGQNAIALRELREELRHC-----	MPLLLL--LSV	193
BT2463	144	ETUREIARELNISIEDVEYHISFSLKALEKTLKDYL---	E-LFYFFF--YHF	189
BT3517	156	MSIDEISKITOMRYQSVENSIQRALGRVRNLF-----	-----KRE	190
BT3748	138	KSHREIARELQISVEGVEYHISRAVELLRDNLKEYA---	E-F--LI--FFI	180
BT4636	149	LSYDDICQINGLNLYQSAYULLQPSLQKIREAYGVTG-I-W	MLILHQLT--YLN	197
BT4705	142	KTHAREIAREMNLSEVESINRYFKQKTMELLANLKDYL---	P-LFLLFLSPEQL	189
BT4722	148	YTHAREIARENHGISBSTVATQLSLAVKFWREQLMKHY---	DKIVALLFA--FFC	195
FecI	137	LYSREIARELQVSVRYVYKAVEH-----	CLLFRLE--YG	173

Figure 6. 5 Alignment of ECF sigma factors to FecI.

Tcoffee presents the score of alignment. From blue to red, the conservation score goes up. Residues in white lettering on red boxes are completely conserved. Color codes for other alignments are identical. The diversity of the N-terminal is high. The middle domains are quite conserved with variation at some region as well. The C-terminal 10-15 residues are diverse.

BAD AVG GOOD

BT0189	1	MKDTKTEK-----N--	KLRRYL--DDM-YI--	REEASQLLESMRDADH	36
BT0753	1	MIEYYDKE-----ELEL-	QISNYL--SGN-ST--	DEEKETLLAFL-ASNE	38
BT1279	1	MKNYFQK-----	IITLFT--GND-YP--	ESTQQDFYKWL-VDEE	33
BT1618	1	MSNAH-----K-----	I IKNFT--SDK-FS--	PELKEKLWKWL-VDSS	32
BT1876	1	MEEENKHI-----DE-----	LIANYL--TEG-LD--	KNALDELKTWI-AASA	36
BT2197	1	MMNE-----	KWNQYF--FGE-PT--	EEEKRELFOEL-EKNE	30
BT2462	1	MDE-----	LLQKYI--AGN-AS--	EKESQRIMEWL-REDE	29
BT3518	1	MINSELKHTD FSRYTFEE--	FLQNDF--FIS-SVKYP	TEEIQEFWDRFEKSTP	48
BT3749	1	MNQN-----	LLLKYI--SGK-AS--	QREKEEVAAWI-DADA	30
BT4635	1	MDIKNK--YTHYSADQ--	LLNDFY--FLE-SEQHP	SPESIAPWETMQQKDK	44
BT4706	1	MENMNPE-----S-----	LLRKAQALGED-IK--	EMESI-----	26
BT4723	1	MDKIHY-----KE-----	LIEKYF--EGN-IA--	DTEIKELSDWI-KNDR	34
FecR	1	MNPLLTDS--RRQALRSAS	HWYAVL--SGERVS--	PQQEARWQQWY-EQDQ	44
BT0189	37	KDILDEL-SAEVWEES-VSQQPVTDL	E--REKYKKE--ARQL--	LK-HIEHKK	80
BT0753	39	AAARTFREMSAVWALS-SVPSFA-EIE-	N---SNL---VRI--	KE-RMTAPA	78
BT1279	34	HTSEKDEALQKLWDEA-HKQRTATDMQ	---EAYELL-KKNAG--	IP-PI-QRK	77
BT1618	33	EQAEKKEEALMELWEDQ-NFKADAG---	TERSYQNPRRKIA-----	PRQRKA	74
BT1876	37	ENQQYFIRQREIWFSVA-SVREAASVYD	---KDKAFENFRNRVES--	QKEIQSTS	84
BT2197	31	DMKREFAEQMNIIVGLSGLLPREDDSLK	---GE---RNL---EAM--	MN-RQEKKL	73
BT2462	30	QHLREYKRQRKLYDIT-LWQT-----	---KSP-----	VDIQQ-EKK	60
BT3518	49	SNIDDYFAAREYIETI-ST-SEGDLLS	---DQELGELW-ADIQ--	TT-NI---K	90
BT3749	31	ANLKEFMSLRKSYDAL-VW-QDADELK	---TGRDK--LL-S--LRT	66	
BT4635	45	ALAKEIKIARSFLHDI-RH-IPVSHLP	---QNEVETVW-KNISR--	LN-NI---E	87
BT4706	27	-----DVM	---GAYQQ---AQIQ--	IK-TNRRRS	46
BT4723	35	QLQ-----NWWEQE-FTKSDAAIDP	ILRDKLFARIKEGT---	LKHTSPRTK	76
FecR	45	DNQWAWQQ-----	---VENLRNQLGGVPG	GDVASRALH	73
BT0189	81	RTWFRRV		87	
BT0753	79	SSKPVRK		85	
BT1279	78	RTIRPIH		84	
BT1618	75	TARYTLR		81	
BT1876	85	RRGFSL		91	
BT2197	74	RRKRVLQ		80	
BT2462	61	DPLRRVL		67	
BT3518	91	NDKIKHK		97	
BT3749	67	FTMKAMR		73	
BT4635	88	KQRKKKI		94	
BT4706	47	MYNQLMR		53	
BT4723	77	GVRTLPM		83	
FecR	74	DTRLTRR		80	

Figure 6. 6 The alignment of the cytoplasmic domain (N-terminal) of anti-sigma factors to FecR.

Totally, the cytoplasmic domain of all anti-sigma factors including FecR are high diverse. Not all ECF sigma factors are presented here, but the rule is general.

BT0189	82	KFDVKSYSSTDEIVSVSVESGKVVQVDLE---	EA-MMRLTAKEQVLINTVSGEYS	130
BT0753	91	VFNVRAYDDDDNYVMVSLLEGRVNLSA---	SANSVMKLFENEQALYNKNTGRME	140
BT1279	82	EFNVSAYPENPVLAATLISGSVVLVEYND---	LKSQVILKFPNEQLAYNKNTHYHS	132
BT1618	84	KFNVRHAYAEDERTITTTLESGSVVQKAN---	NEDIITLTFNEQLEYDNPSGFEFN	134
BT1876	85	KFNFRDYPEDHEVVVSLLEGKVGGLNNLL---	REEKEAVLSPDERAVLNKANGLLT	136
BT2197	83	EFNVYAYSNSKFPETDLLSGKVRVSSTG---	FPEESVNLLPDEKVS LV--DGKLV	132
BT2462	83	EFNVLAYEKDSIWETALLEGAVEILQKK---	SEVSLMKLKPFGDMARLSKN--QLT	132
BT3518	144	KFNVTAYESEPIRSVVLAQGCQVETTQ---	T-P-KAILAPNQMF--SSVEGKEN	191
BT3749	82	KFNVKAYRDESEITTTTLEGGKVFEFNNTA	QKPKQYITMAPGQKLIYYSQSGKTE	135
BT4635	151	TFNVSAYENKQTQSVVLVSGKIKVKTSK---	N-E-SKTLSENLLSYNEQEGI-H	199
BT4706	83	KFNVMAYEDDNSIETVLEKGVNVI SPD---	GKTTVQLAPGERLLYNKVDQKLL	133
BT4723	84	SFNVSAYEDAKDITVVLLLEGKVGIIYTQ---	ET-SRMMKPGGKIEYNKTTTHQLV	132
FecR	84	EFTVREQQDNFTQLDV--QQHAVEVLLASA-	PAQKRIVNAGESLQFSASEFG--	131
BT0189	131	KKK--EERGVAVWINGSLEFNSSTPIRDVAKELER	VYNCQITFASGQEFDNLITGE	183
BT0753	141	KLKTNASKACDNLGGGLTFENASPADIAHRLSRK	FQVKISIESERLKAEHFSGS	194
BT1279	133	LDHPDMKEVTAMQRGELVPREMSVKDIITILERK	YPYTFEYQLKTLKDDRYSPR	186
BT1618	135	KKIIDASVYSGWTRGELNFAAMTSLDIFITIERI	YDIHIIVPPLATTDVYTIK	188
BT1876	137	VESVTASNASQWTDGGLFFDEELLFDIAKELERS	YNVKIHIANDSLKTFRFYGN	190
BT2197	133	KSSSHFGGKEYREQGIYDFEELPLGEVLERLKQ	WYDIHFTVDDPSLLSKIISGK	186
BT2462	133	KEKIHTTDYFRNKEGLICFNDISLRDIMEKRLK	LYDVNFVINMQILDHYTK	186
BT3518	192	ISQVDVEQMISWVNGLYCFNSADLGIIVLRKLS	TYGIVNVEFDSA-LSKIKCSGK	244
BT3749	136	LYTTSGEGELANKDGIIVFKQTSLQDALEILAD	RYDVEFIVRRNVPDDDLPSGT	189
BT4635	200	IQSVDVQKYIAWKDGFYLFQTEKLEKDIATKLS	DVYGGKIMIDSP-LKTITCSGK	252
BT4706	134	KGKVDVYEKVAWKDGLIFRNABELGEIFKRLAR	HFNVDIQFNNISGKEYKYRAT	187
BT4723	133	ATQVHPNDYIEWTRGNIIYFEMASLENIMKTL	SRIDVEIAPDSNKLKPEYFTGT	186
FecR	132	AVKPLDDESTSNTKDLISFSORPLGEVIATL	TRRNGVLRCDPA-VAGLRLSGT	184
BT0189	184	HD-N-KSLESVLRKSEIETI-SDINYKKEGR---	-----NILLY-----R	215
BT0753	195	FD-SNQNIYDILHEINVE-KQYTWKVS GDTIE	---ITD-KRK-G-----V	232
BT1279	187	FK-DQAPLSEVMDVIVNVVGMNYKIKGD---	-----RCYLI-----P	220
BT1618	189	FK-QKAPIKEIMNIVTKTIGNIDYKVEDE---	-----NILLIYSPLNKKGG	230
BT1876	191	FVRREQNIQEVLEALAST-EKMQYKIE---	-----ERNI-T-----I	222
BT2197	187	FR-QSDQIETILKRAISRA-DLFEYKILSQ---	-----RE-IT-----I	218
BT2462	187	FR-THDGIHVRVRLALN-NKFTYIKDNE---	-----SNTIT-----I	219
BT3518	245	ID-LKDNFETVINGLTFV-APISYAYDGQY---	-----KTYRVV-----R	279
BT3749	190	FT-S-RSLEQILNYIEAS-SKIRWRYLNSVQG	---SKEKMKIEI-----P	228
BT4635	253	LD-LKEDLDEVLQTLIRT-VPARIEKS-DG---	-----IIHIYV-----K	286
BT4706	188	FR-N-ETLPQILLYLAKS-AALKWRTEEAVQQ	ADDTETKKKIIVD-----L	230
BT4723	187	IE-S-GGIQNALNMLMLT-SPFYEMDGS---	-----VIVLK-----E	218
FecR	185	FP-LK-NTDAILNVIAQT-LPVKIQSITR---	-----YWINIS-----E	217

Figure 6. 7 The alignment of the periplasmic domain (C-terminal) of anti-sigma factors to FecR.

Generally, the C-terminal of anti-sigma factors are much conserved, especially in the middle area, while some of them (BT3518 and BT4635) have longer N-terminal tails.

These data suggest that the ECF-sigma PULs in *B. thetaiotaomicron* signal in the same way as the Fec system, but *E. coli* has one system whereas *B. thetaiotaomicron* has at least 12. A lot of sensory systems of *B. thetaiotaomicron* present high sequence similarity (for example, TCS HK and RR receive domain, and anti-sigma factor periplasmic domains), which increase the chance of unexpected cross-talk. Some research has detected the cross-talk of TCS in *E. coli* (Siryaporn and Goulian, 2008). My objectives are to investigate the pathway of signaling of ECF-PULs in *B. thetaiotaomicron* and possibility of specificity in the ECF-sigma linked PULs.

6.2 Results

Four ECF-sigma linked gene clusters (or PULs) which have an ECF-sigma factor, an anti-sigma factor, SusC transducer homolog, SusD homolog, and at least one glycoside hydrolase were chosen for analysis (Figure 6. 9 and Table 6. 1). All of these systems have been shown to be involved in either dietary or host polysaccharide utilisation (Bjursell *et al.*, 2006).

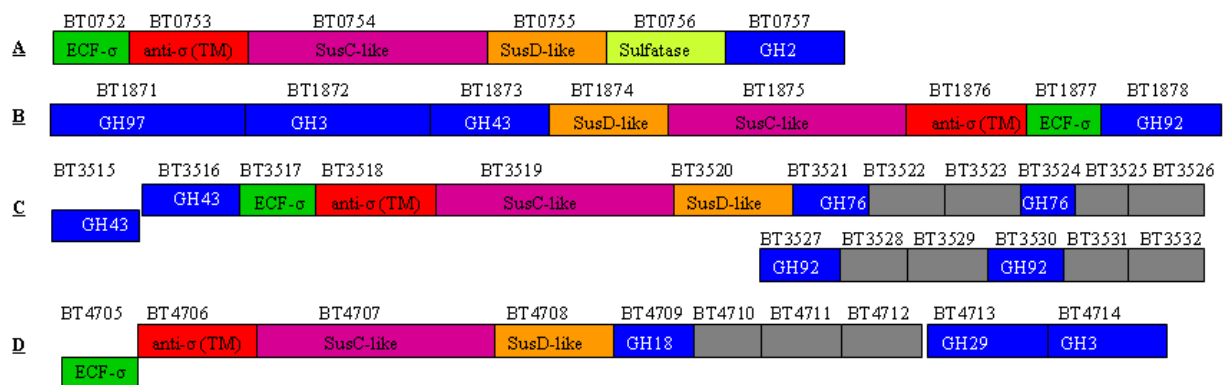


Figure 6. 9 ECF-PUL gene clusters for this study.

There are four gene clusters (A, B, C, D) studied in this chapter. Green box shows the ECF- σ factor, red presents the anti- σ factor with transmembrane (TM), pink means SusC-like OMP, and orange means SusD-like OMP. Light green in A represents the sulfatase. Blue boxes means GHs and grey boxes represents a protein of unknown function.

Gene Cluster	Sigma factor	Anti-sigma factor	SusC homolog
A	BT0752	BT0753	BT0754
B	BT1877	BT1876	BT1875
C	BT3517	BT3518	BT3519
D	BT4705	BT4706	BT4707

Table 6. 1 Gene clusters expressed in this chapter.

Gene clusters, A, C, and D are transcribed from the sense strand. Gene cluster B is transcribed from the anti-sense.

Based on the *E. coli* Fec system model it was hypothesized that the ECF-sigma, anti-sigma and SusC transducer from each cluster were able to interact to form a trans-envelope signalling system whereby binding and transport of a specific complex glycan to the SusC transducer would initiate the signalling cascade that ultimately led to the upregulation of the genes required for its degradation by the cognate ECF sigma factor (Figure 6. 10).

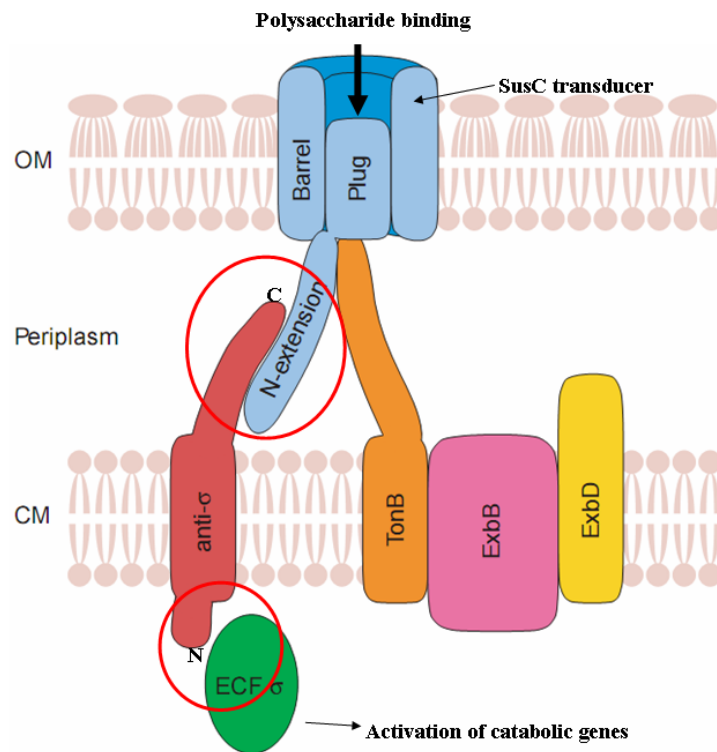


Figure 6. 10 Model of the predicted transenvelope signalling system in *B. thetaiotaomicron* based on the *E. coli* Fec system.

Polysaccharide binding to the SusC homologue (blue) in the outer membrane induces conformational change that allows the N-terminal periplasmic extension to contact the periplasmic domain of the cognate anti-sigma factor (red). This signal is then transduced across the membrane to the cytoplasmic ECF sigma factor (green) which is activated and initiates expression of the genes under its control. The inner membrane TonB apparatus which provides the energy required for active uptake of the polysaccharide is also shown.

6.2.1 Protein Expression

In total, four proteins were expressed in a recombinant form from each ECF-PUL. They are the sigma factor, anti-sigma cytoplasmic domain, anti-sigma periplasmic domain, and N-terminal periplasmic extension of SusC transducers after the signal peptide (Figure 6. 11 and Table 6. 1). Signal peptides were predicted using SignalP (Figure 6. 13) and the transmembrane domains were predicted using TMHMM (Figure 6. 15).

N	ECF- σ		anti- σ cyto		TM	anti- σ periplasmic domain		N-tail		SusC-like		C
	Gene ID	Start	Gene ID	End		Gene ID	Start	Gene ID	Start	Gene ID	End	
1	BT0752	195	1	BT0753	85	109	BT0753	341	1	80	BT0754	
1	BT1877	196	1	BT1876	91	115	BT1876	337	1	80	BT1875	
1	BT3517	190	1	BT3518	97	121	BT3518	400	1	80	BT3519	
1	BT4705	189	1	BT4706	53	77	BT4706	307	1	80	BT4707	

Figure 6. 11 Domains expressed in this study.

The 16 genes are listed from N-terminal to C-terminal. The second gene cluster is conversed because the gene cluster order is different to other three. The anti- σ cyto in the first red box means the cytoplasmic domain of anti- σ factors. The N-tail in the purple box means the N-terminal extension of ~80 residues after the signal peptide of the SusC transducer. The numbers between each gene ID present the protein start and end residue numbers.

The proteins used in this chapter are listed in Table 6. 2.

Protein ID	Note
BT0752	ECF sigma factor of PUL A
BT1877	ECF sigma factor of PUL B
BT3517	ECF sigma factor of PUL C
BT4705	ECF sigma factor of PUL D
BT0753N	Anti-sigma factor cytoplasmic domain of PUL A
BT1876N	Anti-sigma factor cytoplasmic domain of PUL B
BT3518N	Anti-sigma factor cytoplasmic domain of PUL C
BT4706N	Anti-sigma factor cytoplasmic domain of PUL D
BT0753C	Anti-sigma factor periplasmic domain of PUL A
BT1876C	Anti-sigma factor periplasmic domain of PUL B
BT3518C	Anti-sigma factor periplasmic domain of PUL C
BT4706C	Anti-sigma factor periplasmic domain of PUL D
BT0753CA	Anti-sigma factor C-terminal of periplasmic domain of PUL A
BT1876CA	Anti-sigma factor C-terminal of periplasmic domain of PUL B
BT3518CA	Anti-sigma factor C-terminal of periplasmic domain of PUL C
BT4706CA	Anti-sigma factor C-terminal of periplasmic domain of PUL D
BT0754N	N-extension of SusC homolog of PUL A
BT1875N	N-extension of SusC homolog of PUL B
BT3519N	N-extension of SusC homolog of PUL C
BT4707N	N-extension of SusC homolog of PUL D

Table 6. 2 Proteins used in this chapter.

There are 20 proteins used in this study. The anti-sigma facto C-terminal of periplasmic domain is about the half length of periplasmic domain after a cleavage site (see Figure 6.20). More details about this will be described later in this chapter.

Target genes were amplified by PCR using the genome of the *B. thetaiotaomicron* as template (Figure 6. 12).

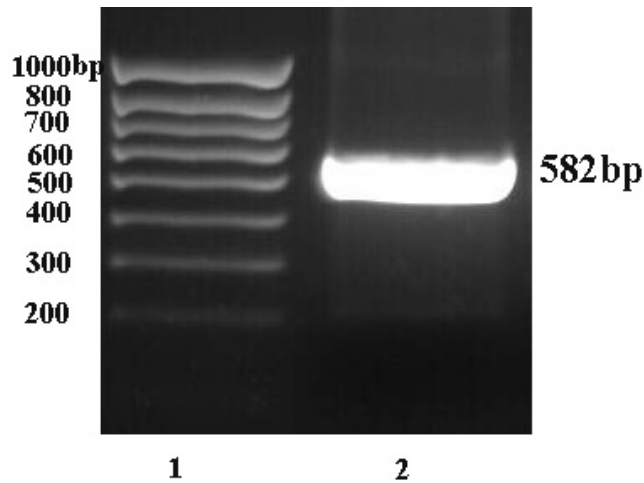


Figure 6. 12 Amplification of BT4705 sigma factor from *B. thetaiotaomicron* genomic DNA.
Lane 1: marker, lane 2: BT4705 PCR product.

6.2.1.1 Expression of the N-terminal periplasmic domain of the SusC transducers

According to the research on the function and structure of OMP FecA, the periplasmic N-terminal 81 residues of FecA after the signal peptide form a distinct domain that plays an important role in information transmission (Figure 6. 3 and Figure 6. 8). The N-terminal 80 residues after the signal peptide of the *B. thetaiotaomicron* SusC transducers used in this study were cloned. The position of the signal peptide cleavage site was determined by SignalP (Figure 6. 13).

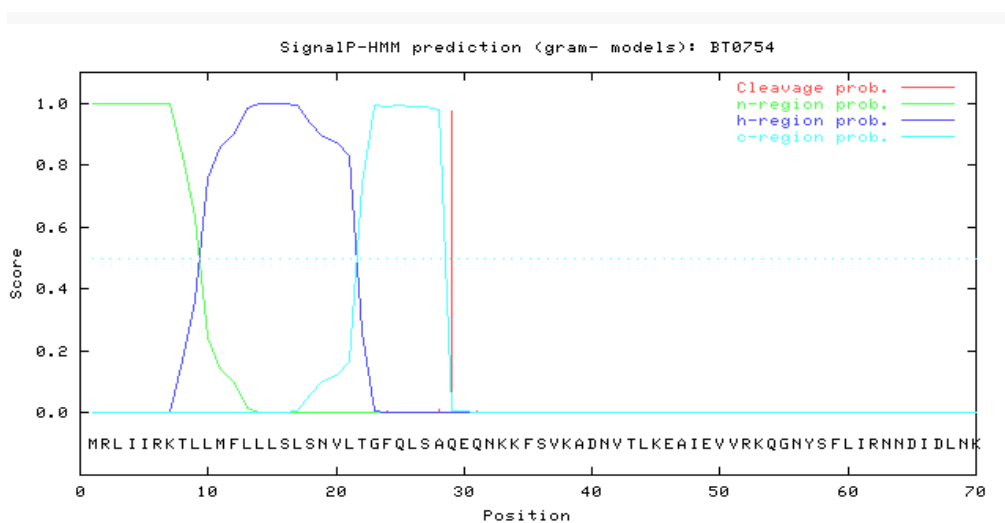


Figure 6. 13 A map shows the predicted signal peptide cleavage site of BT0754 SusC transducer.
The red line indicates the cleavage site of BT0754 which is between residues 28 and 29 (LSA-QEQ).

The primers for SusC transducers used in this study are shown in Table 6. 3.

Primer ID	Sequence (5'----->3')	Note
BT0754-1	CGCGCTAGCCAGGAACAAAACAAG	F Nhe I
BT0754-2	CCGCTCGAGGGCATGGAATATGACC	R Xho I
BT0754-3	CATGCGGGCCGCATGCAGGAACAAAAC	F Not I
BT1875-1	CGCGCTAGCGTTTCATTGAAAATGG	F Nhe I
BT1875-2	CCGCTCGAGCTTTGAAGTATTCTGCC	R Xho I
BT3519-1	CGCGCTAGCAATAACAATAATTATTC	F Nhe I
BT3519-2	CCGCTCGAGAAAACTTGTCTGTC	R Xho I
BT3519-3	CGCGCTAGCCAGCTCAAACTCTCTCGG	F Nhe I
BT3519-4	CCGCTCGAGAGCTGACTTTCCTGAG	R Xho I
BT4707-1	CGCGCTAGCCAACGGACGCATGTTG	F Nhe I
BT4707-2	CCGCTCGAGCATTTTATTAGTCGATAAAATAATC	R Xho I

Table 6. 3 Primers for the N-terminal extension of SusC transducers studied.

F: Forward primer, R: Reverse primer.

The N-terminal 80 residues of the SusC transducers were cloned into pET21a which has a C-terminal His tag, expressed in *E. coli* C41 cell strain. The IMAC purified proteins were all produced in a soluble form (Figure 6. 14).

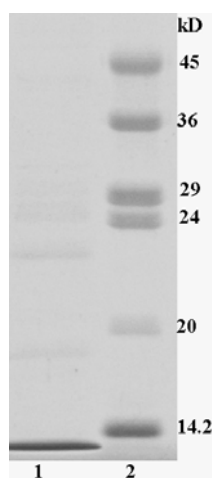


Figure 6. 14 SDS-PAGE of purified BT4707N SusC transducer.

BT4707N was expressed as a C-terminal His tag construct and purified by IMAC in a single step from the cell free extract of lysed *E. coli* C41 cells. Lane 1 is eluted pure BT4707N, lane 2 are molecular weight markers.

6.2.1.2 Expression of the periplasmic domain of the anti-sigma factors

The periplasmic domain, cytoplasmic domain and transmembrane domain of anti-sigma factors were determined using the software Prediction of transmembrane helices in proteins (TMHMM). All anti-sigma factors' N-terminals are in cytoplasm, while C-termini are in the periplasm. Figure 6. 15 shows an example of a TMHMM transmembrane location prediction.

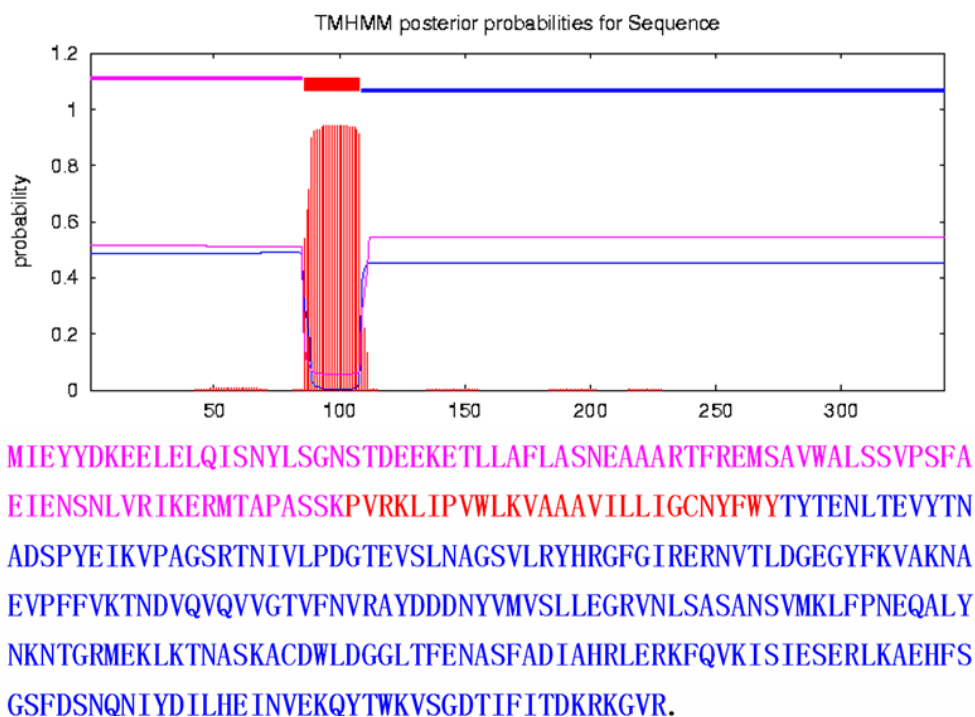


Figure 6. 15 Transmembrane prediction for BT0753 anti-sigma factor.

Residues 1-85 (purple) are cytoplasmic domain, residues 86-108 (red) are transmembrane domain and 109-341 residues (blue) are periplasmic domain.

The transmembrane regions and orientation of each anti-sigma factor predicted by TMHMM are listed in Table 6. 4.

Gene Cluster	cytoplasmic domain (N-terminal)	Transmembrane	periplasmic domain (C-terminal)
BT0753	[1-85]	[86-108]	[109-341]
BT1876	[1-91]	[92-114]	[115-337]
BT3518	[1-97]	[98-120]	[121-400]
BT4706	[1-53]	[54-76]	[77-307]

Table 6. 4 The transmembrane regions and orientation of anti-sigma factors used in this study.

All the prediction for transmembrane regions and locations of anti-sigma factors are made by TMHMM. Generally, the N-terminal domains are in cytoplasm, which are less than 100 residues. The transmembrane domains are 22 residues in each anti-sigma factor, while the lengths of C-terminal periplasmic domain shift from 222 (BT1876) to 279 amino acids (BT3518).

The primers originally designed for full length of periplasmic domain of anti sigma factors are listed as Table 6. 5.

Primer ID	Sequence (5'----->3')	Note
BT0753-1	CGCGGATCCACGTATACGGAAAATCTGACG	F BamH I
BT0753-2	CCGGAATTCCTTATCTTACCCCCTTTCTTTTATCG	R EcoR I S
BT0753-3	CCGCTCGAGTCTTACCCCCTTTCTTTTATCG	R XhoI
BT0753-4	CTAGCTAGCACGTATACGGAAAATCTGACGG	F Nhe I
BT0753-5	CGCGGATCCTTATCTTACCCCC	R BamH I
BT0753-12	CTAGCTAGCACGTATACGGAAAATCTGACGG	F Nhe I
BT0753-13	CGCGGATCCTTATCTTACCCCCTTTC	R BamH I
BT1876-1	CGCGGATCCGAGGTGAATGTGAAAG	F BamH I
BT1876-2	CCGGAATTCATAGATAGTTATATTAC	R EcoR I S
BT3518-1	CTCCCATGGCGAATTATCAGGCGGTTTTGGC	F Nco I
BT3518-2	CTCCTCGAGTTTTTTTCACGACCCGATAAG	R Xho I
BT3518-3	CCGCTCGAGCTATTTTTTCACGACCCG	R Xho I S
BT3518-7	CCGGAATTCCTATTTTTTCACGACCCG	R EcoR I

Table 6. 5 Primers for the full length of periplasmic domain of anti-sigma factors.

F: Forward primer, R: Reverse primer, S: with Stop code on. The primers for full length of BT4706 were not designed.

The periplasmic domains of anti-sigma factors (BT1876C, BT3518C) in pET32b (see Appendix C; recombinant protein expressed as a fusion with thioredoxin, Trx) were over expressed in *E. coli* C41 cell strain. All IMAC purified proteins were produced in a soluble form in the *E. coli* cytoplasm. The Trx-BT1876C fusion should be 47 kD, and the purified two bands are 16 kD and 31 kD, which indicated that BT1876C fusion was cut at one point by unknown protease (Figure 6. 16). The digestion could not be inhibited by protease inhibitors or 8 M urea added to the cell suspension prior to sonication indicating it was taking place in the cells during expression.

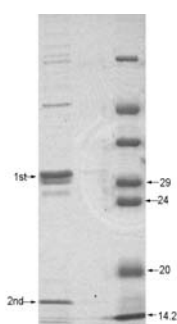


Figure 6. 16 SDS PAGE of purified BT1876C fused to thioredoxin.

BT1876C expressed in pET32 as two separated bands. One is predicted to be about 16 kD including a His tag, while another is about 31 kD. BT1876C is 25.5 kD itself, while the N-terminal tags are 21.5 kD. So the total length of BT1876C fusion is 47 kD and equals the length sum of the two bands, which indicated this fusion is cleaved at some point. The small band was predicted to be the second fragment (C-terminal), while the big band was the N-terminal.

The Trx-BT3518C is about 50 kD. After IMAC purification, there were two fragments shown on the SDS-PAGE gel (Figure 6. 17). The length sum of these two bands is about 50 kD, which indicated that BT3518C fusion was cleaved at one point by unknown protease, which could not be inhibited by protease inhibitor either.

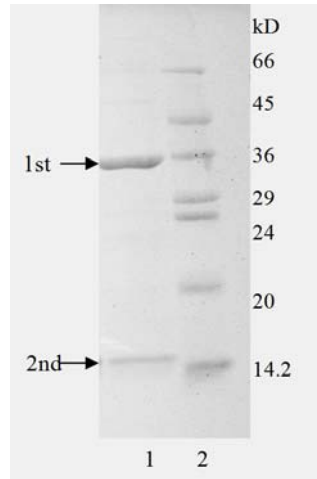


Figure 6. 17 BT3518C express two bands in pET32b.

Lane 1: BT3082C. Lane 2: Marker. The small band is about 15 kD, while the big band is about 35 kD. The length sum of these two bands equals 50 kD, which is the full length of BT3518C fusion. This result indicated that BT3518C fusion was cleaved at one point. And the small band is predicted to be the second fragment (C-terminal), while the big one is the first fragment.

BT0753C was cloned into minipRSETA (N-terminal His-tag, 1.7 kD), minipRSET-Trx (N-terminal His tag, plus thioredoxin tag, about 14 kD), and pGEX-6P-1(N-terminal GST tag, about 25 kD) respectively (Appendix C). BT0753 fusions were expressed in a range of *E. coli* strains and purified by IMAC method. In all cases two bands were produced (Figure 6. 18) and all have one fraction at the same size. That would suggest a specific protease cut at a special site of this protein, so the sizes of the C-terminal (second fragments) would not be influenced.

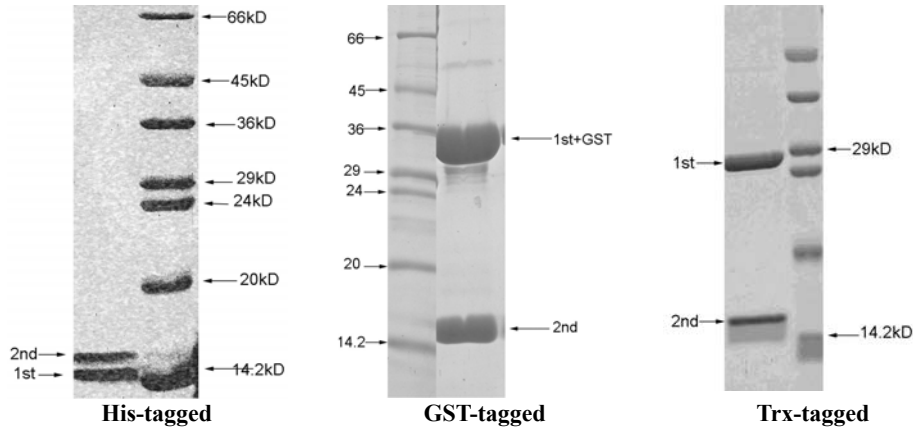


Figure 6. 18 BT0753C expressed as two bands when fused to three tags.

The BT0753C is 26.4 kD, which expressed two bands in three different vectors: minipRSETA, minipRSET-Trx, and pGEX-6p-1 which have tags at N-terminal respectively. There is always one band kept at the same size position (about 16 kD), which was predicted to be the second band (C-terminal) because the N-terminal is shifting the length with the length of tags in different vectors.

In order to check out the cleavage site, the small band of BT0753C in minipRSETA-Trx was blotted to PVDF membrane for N-terminal 6 amino acid sequencing, and cut from SDS gel for molecular mass analysis by MALDI-TOF mass spectrum (Figure 6. 19). The data from MALDI-TOF and N-terminal sequencing show that BT0753C had been cleaved by an unknown protease between Gly-197 and Thr-198 in the full length protein specifically (Figure 6. 20). So the small fragment of BT0753C started from residues: TVFNRR and had the molecular weight at 16.5 kD.

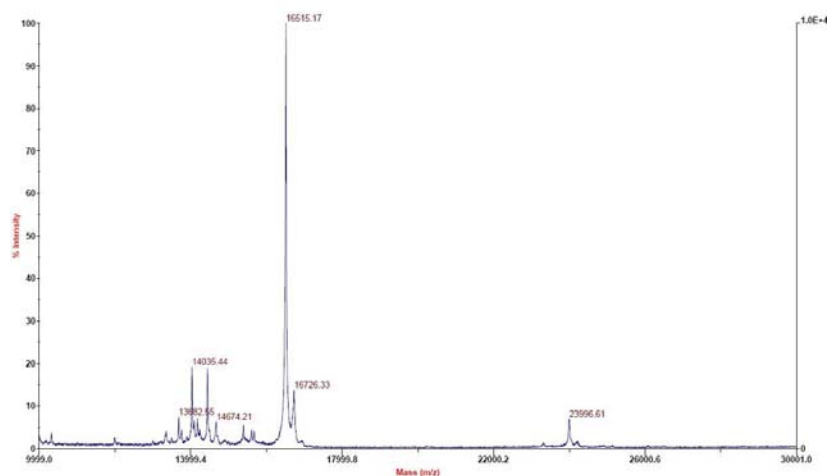


Figure 6. 19 MADL-TOF map shows the C-terminal fraction of BT0753C peri.

BT0753C was fused to minipRSETA-Trx, over expressed in *E. coli* and purified by IMAC and the molecular weight of the smaller band determined by MALDI-TOF. The molecular weight is 16515.17 Dalton.

This cleavage event occurred in the cell as evidenced by the presence of two bands even when the cells were resuspended in 8 M urea before lysis (data not shown). A similar phenomenon was observed with FecR in *E. coli* (Wriedt *et al.*, 1995). According the alignment of FecR and 19 anti-sigma factors, we found that both G and T sites are conserved in anti-sigma factors (Figure 6. 20).



Figure 6. 20 Alignment of partial periplasmic domains of anti-sigma factors from *B. thetaiotaomicron* and *E. coli* FecR showing position of conserved cleavage site.

A presents the location of conserved Gly-197 and Thr-198 (red box) in BT0753. B shows conserved site in all anti-sigma factors as well as FecR. All anti-sigma factors in gene cluster which has an ECF type sigma factor adjacent to homologues of *susC*, *susD*, and glycoside hydrolases. Star symbol indicates the residues are completely conserved.

In the following experiments, only the portions after this conserved residue glycine were cloned. They are BT0753CΔ ([198-341]), BT1876CΔ ([198-337]), BT3518CΔ ([263-401]), and BT4706CΔ ([198-307]). Primers for C-terminal of periplasmic domain of anti-sigma factors were designed as Table 6. 6. Subsequently, only one band was seen on SDS-PAGE gel (Figure 6. 21).

Primer ID	Sequence (5'----->3')	Note
BT0753-6	CGCGGATCCACGGTATTCAATGTCCG	F BamH I
BT1876-3	CGCGGATCCACAAAATTCAATTTCCG	F BamH I
BT1876-4	CCGGAATTCCTAATAGATAGTTATATTACG	R EcoR I
BT3518-4	CGCGGATCCACTAAATTCAATGTGAC	F BamH I
BT3518-7	CCGGAATTCCTATTTTTTTCACGACCCG	R EcoR I S
BT4706-3	CGCGGATCCACCAAATTTAATGTG	F BamH I
BT4706-4	CCGGAATTCCTAATATAAATCTACTATTATC	R EcoR I S

Table 6. 6 Primers for the second part of periplasmic domain of anti-sigma factors.

F: Forward primer, R: Reverse primer, S: with Stop code on.

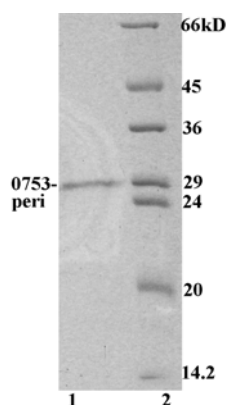


Figure 6. 21 SDS-PAGE shows purified BT0753CA.

This protein fusion (lane 1) (30.5 kD) has an N-terminal His tag and a Trx tag (14 kD) and the C-terminal BT0753CA (16.5 kD, [198-341]).

6.2.1.3 Expression of the cytoplasmic domain of the anti-sigma factors

The cytoplasmic domains of anti-sigma factors were determined as described before. Their genes were cloned into pET21a by PCR with primers from MWG (Table 6. 7), expressed in *E. coli* C41 cell strain. The IMAC purified BT0753N, BT1876N, and BT3518N are all expressed in a soluble form in the cytoplasm of *E. coli*. BT4706N has no expression in pET21a, but it expressed as soluble protein in pGEX-6p-1 which has an N-terminal GST tag. Figure 6. 22 shows BT0753N as an example of cytoplasmic domain of the anti-sigma factors.

Primer ID	Sequence (5'----->3')	Note
BT0753-10	CGCGCTAGCATAGAATATTACGATAAAG	F Nhe I
BT0753-7	CGCGGATCCATAGAATATTACGATAAAGAAGAACTGG	F BamH I
BT0753-11	CCGCTCGAGTTTCCGGACAGGC	R Xho I
BT1876-5	CGCGCTAGCGAAGAAGAAAATAAAC	F Nhe I
BT1876-6	CCGCTCGAGTGACAGGCTAAAAC	R Xho I
BT3518-5	CGCGCTAGCATTAAATTCAGAACTG	F Nhe I
BT3518-6	CCGCTCGAGTTTATGTTTAATCTTATC	R Xho I
BT4706-1	CGCGCTAGCATGGAAAATATGAATCCGG	F Nhe I
BT4706-2	CCGCTCGAGACGCATCAACTGATTATAC	R Xho I
BT4706-8	CGCGGATCCATGGAAAATATGAATCCGG	F BamH I
BT4706-9	CCGGAATTCACGCATCAACTGATTATAC	R EcoR I

Table 6. 7 The primers for cytoplasmic domain of anti-sigma factors.

F: Forward primer, R: Reverse primer.

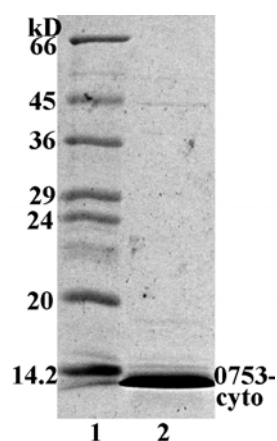


Figure 6. 22 SDS-PAGE of purified N-terminal cytoplasmic domain of BT0753 anti sigma factor (BT0753N). BT0753N (lane 2) is 9.6 kD and was purified in a single step by IMAC.

6.2.1.4 Expression of sigma factors

There are four sigma factors BT0752 ([1-195]), BT1877 ([1-196]), BT3517 ([1-190]), and BT4705 ([1-189]) cloned by PCR using primers from Table 6. 8 and the *B. thetaiotaomicron* genomic as template, ligated into minipRSETA which has a thrombin cleavage site immediately after the N-terminal His tag. The fusions of sigma factors were over expressed in *E. coli* C41 cell strain. Sigma factors BT0752, BT3517, and BT4705 are IMAC purified from CFE and soluble in Tris pH 8.0 buffer while BT1877 has no expression in a range of *E. coli* cell strains, such as BL21, JM83, and Tuner and so could not be studied. Figure 6. 23 shows purified BT0752 as an example.

Primer ID	Sequence (5'----->3')	Note
BT0752-1	CGCGGATCCAATTTACCAGAAGTC	F BamH I
BT0752-2	CCGGAATTCCTTAAAGAAACATCAAATAGC	R EcoR I S
BT0752-3	CGCGCTAGCAATTTACCAGAAGTC	F Nhe I
BT1877-1	CGCGGATCCGAGAATACTGAAAC	F BamH I
BT1877-2	CCGGAATTCCTTACCTGAAAAAATAGTAATAG	R EcoR I S
BT3517-1	CGCGGATCCGCCTCAGATATAC	F BamH I
BT3517-2	CCGGAATTCCTATTCTCGCTTG	R EcoR I S
BT4705-1	CGCGGATCCAAAATATCTTTTCCAGAC	F BamH I
BT4705-2	CCGGAATTCCTTACAGTTGCTCGGGAG	R EcoR I S

Table 6. 8 Primers for sigma factors.

F: Forward primer, R: Reverse primer, S: with Stop code on.

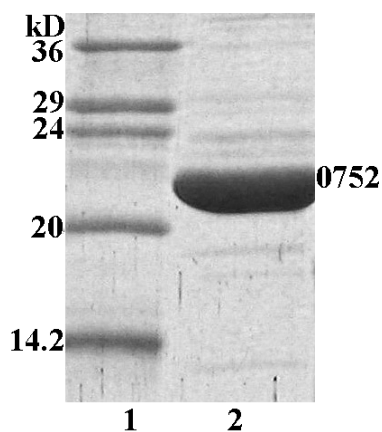


Figure 6. 23 SDS-PAGE of purified N-terminal cytoplasmic domain of BT0752 sigma factor.
Sigma factor BT0752 is 22.5 kD and was purified in a single step by IMAC.

6.2.2 Protein Interaction Studies

6.2.2.1 The N-terminal periplasmic extension of SusC transducers binds specifically to the periplasmic domain of their cognate anti-sigma factor

AGE was used to evaluate the binding of N-extension of OMP SusC transducers and the C-terminal of anti-sigma periplasmic domain initially. The native gels reveal the specific binding of anti-sigma peri to N-terminal of SusC transducers from the same gene cluster with the presence of a third band when mixed (Figure 6. 24). No cross binding was seen between anti-sigma peri and SusC transducer N-[1-80] derived from different gene clusters (data not shown).

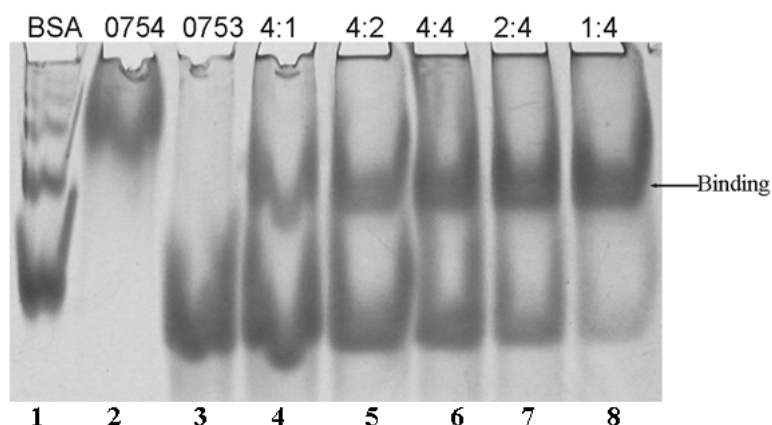


Figure 6. 24 Native gel of SusC transducer N-extension and periplasmic domain of anti-sigma factor interaction from same PUL.

Lane 1: BSA, lane 2: BT0754N, lane 3: BT0753C Δ , lanes 4-8 are the mixture of BT0753C Δ and BT0754N at ratio 4:1, 4:2, 4:4, 2:4, and 1:4. The arrow shows the new band resulting from the complex formed between BT0753C Δ and BT0754N.

ITC was then used to quantify the binding of the C-terminal of anti-sigma factor periplasmic domain to N-extension of SusC transducer. The data show the interaction has a K_a of between 1.7 and $14.5 \times 10^6 \text{ M}^{-1}$ for the gene pairs analysed. Interestingly, the thermodynamics of the interaction vary significantly between cognate gene pairs. In all cases binding involves a favourable change in enthalpy ($-\Delta H$), but the entropic contribution varies from being strongly favourable for BT1875C Δ /BT1876N ($T\Delta S +6.1 \text{ kcal mol}^{-1}$) to slightly unfavourable for BT0753C Δ /BT0754N ($T\Delta S -1.1 \text{ kcal mol}^{-1}$) (Figure 6. 25 and Table 6. 9).

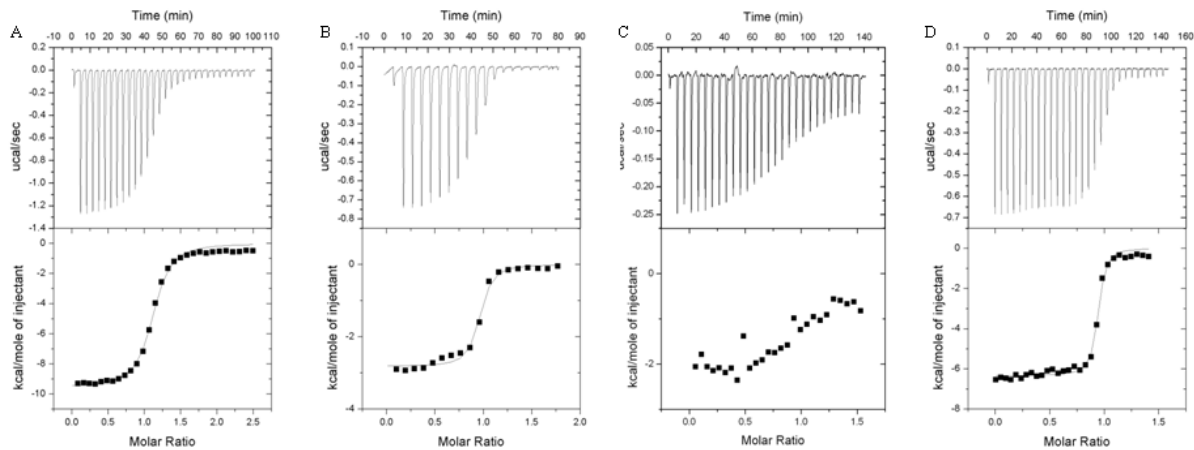


Figure 6. 25 ITC shows the binding of SusC N-extension and periplasmic domain of anti-sigma factors from same PUL.

A: BT0753C Δ vs BT0754N, **B:** BT1876C Δ vs BT1875N, **C:** BT3518C Δ vs BT3519N, **D:** BT4706C Δ vs BT4707N.

The ITC data supports the results from native gels: binding just occurs between proteins from same gene cluster and does not happen between proteins of different gene clusters (Figure 6. 26).

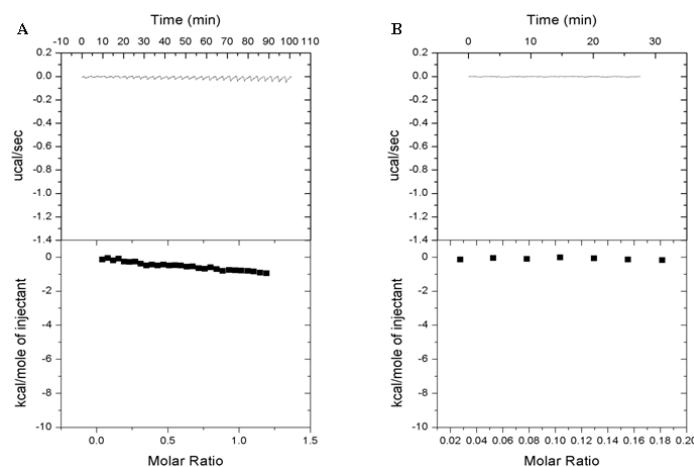


Figure 6. 26 ITC shows no cross binding of SusC N-extension and periplasmic domain of anti-sigma factors from different gene clusters.

A: No cross-binding of BT1875N to BT0753C Δ . **B:** No cross-binding of BT3519N to BT0753C Δ .

No change in binding affinity was observed in the presence of 5 mM EDTA, indicating that the interaction of anti-sigma peri with the SusC N-extension is metal independent (data not shown).

ITC pair		$K_a \times 10^6$ (M ⁻¹)	ΔH (kcal mol ⁻¹)	T ΔS (kcal mol ⁻¹)	n*
BT0753C Δ BT0754N	vs	1.7 \pm 0.2	-9.6 \pm 0.1	-1.1	1.1 \pm 0.0
BT1876C Δ BT1875N	vs	3.4 \pm 0.9	-2.8 \pm 0.5	6.1	0.9 \pm 0.0
BT4706C Δ BT4707N	vs	14.5 \pm 3.0	-6.3 \pm 0.5	3.4	0.9 \pm 0.0

Table 6. 9 ITC data of SusC N-extension and periplasmic domain of anti-sigma factors.

*: The ITC data were fitted to a single site binding model for all titrations. ITC was carried out in 10 mM Tris/HCl pH 8.0. Protein concentrations are from 50 to 500 μ M. The titration of BT3518C Δ vs BT3519N could not be calculated correctly using Origin software. This set of binding looks like enthalpy decrease, while entropy varies in each. The free energy change in each case is around -9 kcal mol⁻¹.

6.2.2.2 The cytoplasmic domains of anti-sigma factors bind specifically to their cognate ECF-sigma factors

To evaluate the binding ability of sigma factor and cytoplasmic domain of anti-sigma factor native gel electrophoresis was used initially. The data revealed that the sigma factors bind the cytoplasmic domain of anti-sigma factors of a same PUL (Figure 6. 27), but do not associate with other anti-sigma cytoplasmic domains from different gene clusters (data not shown).

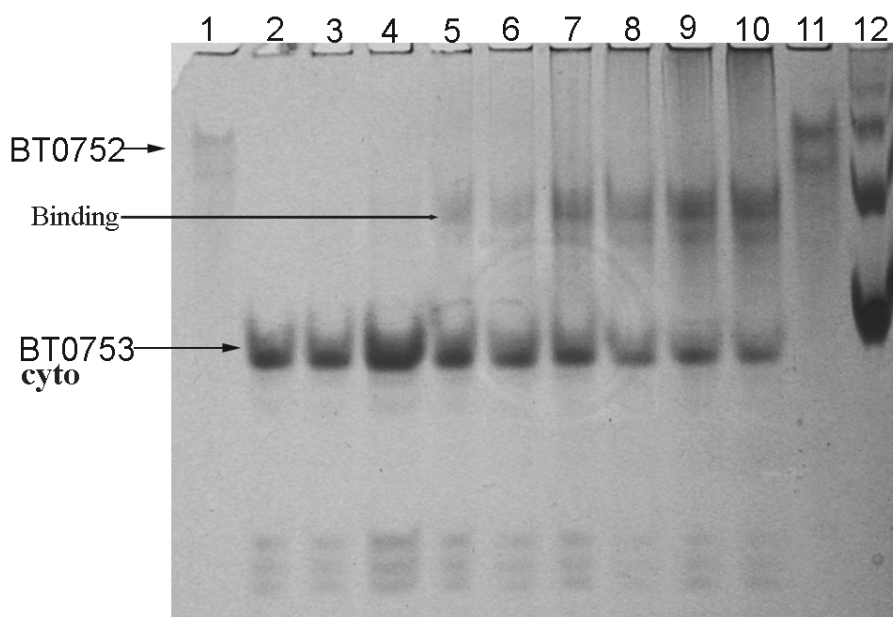


Figure 6. 27 Native gel shows binding of sigma to cytoplasmic domain of anti-sigma factor from the same PUL.

The ratio of BT0753N/BT0752 in lanes 1-11: 0:4, 4:0, 4:1, 4:2, 4:3, 4:4, 4:6, 4:8, 4:10, 4:12, and 0:8, Lane 12: BSA. The binding arrow points to the new band produced by the complex of BT0753N and BT0752.

ITC was then used to quantify the affinity of sigma factor to anti-sigma cytoplasmic domain. BT1876N was not tested against BT1877 because of no expression of the latter. The results show that the association constant varies from 9.5 to $34.4 \times 10^6 \text{ M}^{-1}$ (Figure 6. 28 and Table 6. 10). In addition, the thermodynamics of the interactions vary significantly; both are dominated by a favourable enthalpy, but whilst BT0752/BT0753N has a favorable entropic component, BT3517/BT3518N has significantly unfavorable entropy. Addition of 5 mM EDTA had no effect on the magnitude of the interaction indicating that the binding is metal independent (data not shown).

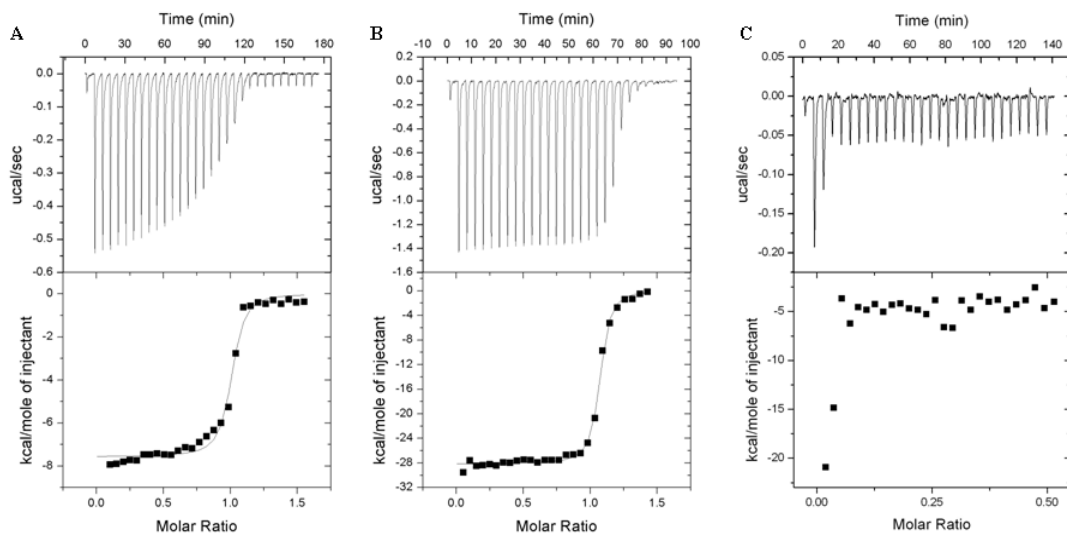


Figure 6. 28 ITC of sigma factors and anti-sigma factor cytoplasmic domains from the same PUL.

A: BT0752 vs BT0753N. **B:** BT3517 vs BT3518N. **C:** BT4705 vs BT4706N.

As expected, no cross binding was seen between the BT0752 sigma factor and BT3518N anti-sigma cytoplasmic domain (Figure 6. 29).

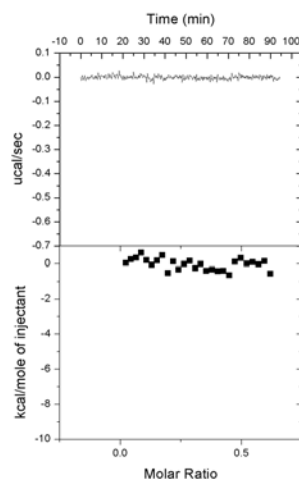


Figure 6. 29 ITC shows no binding of BT0752 to BT3518N.

No cross-binding of sigma factor BT0752 to anti-sigma BT3518-cyto which belong to different clusters.

ITC pair	$K_a \times 10^6$ (M ⁻¹)	ΔH (kcal mol ⁻¹)	TAS (kcal mol ⁻¹)	n*
BT0752 vs BT0753N	9.5 ± 0.2	-7.6 ± 0.9	2.0	1 ± 0.0
BT3517 vs BT3518N	34.4 ± 4.6	-28.2 ± 0.1	-17.9	1.1 ± 0.0

Table 6. 10 ITC data of the interaction between cognate sigma and anti-sigma factors.

*: The ITC data were fitted to a single site binding model for all titrations. ITC was carried out in 10 mM Tris/HCl pH 8.0. Protein concentrations are from 50 to 500 μ M. Titration of BT4705 vs T4706N could not be fitted correctly by the Origin software.

6.2.3 Protein Crystallization

The specific binding of sigma/anti-sigma factors and the anti-sigma factors with the N-extension of SusC transducers led us to further research the interaction using protein crystallography to understand the specificity at a molecular level. We used different ways to approach producing complexes of BT0752/BT0753N and BT0753C Δ /BT0754N.

First we used digestion by thrombin to remove the Trx tag from BT0753C Δ fusion, but this cleaved protein could bind IMAC nonspecifically, which would suggest this digested protein could not be used to co-purify with BT0754N. Then we modified pET21a to enable co-expression of BT0753C Δ and BT0754N (Figure 6. 30 and Figure 6. 31).

```

      NheI
      ~~~~~
M   A   S   T   Y   T   E   N   L   T   E   V   Y   T   N   A   D   S   P   Y
1  ATG GCT AGC ACG TAT ACG GAA AAT CTG ACG GAA GTA TAT ACC AAC GCC GAT TCT CCT TAC
   E   I   K   V   P   A   G   S   R   T   N   I   V   L   P   D   G   T   E   V
61 GAA ATC AAG GTT CCT GCC GGA TCA AGA ACG AAT ATC GTT CTT CCT GAC GGG ACG GAG GTT
   S   L   N   A   G   S   V   L   R   Y   H   R   G   F   G   I   R   E   R   N
121 TCT CTG AAT GCA GGT TCT GTA TTG CGC TAT CAC CGT GGG TTT GGT ATC CGC GAA CGT AAT
   V   T   L   D   G   E   G   Y   F   K   V   A   K   N   A   E   V   P   F   F
181 GTG ACG TTG GAT GGC GAA GGA TAC TTT AAA GTT GCG AAA AAT GCA GAG GTT CCC TTC TTC
   V   K   T   N   D   V   Q   V   Q   V   V   G   T   V   F   N   V   R   A   Y
241 GTG AAG ACG AAC GAT GTG CAG GTG CAG GTA GTC GGC ACG GTA TTC AAT GTC CGT GCT TAT
   D   D   D   N   Y   V   M   V   S   L   L   E   G   R   V   N   L   S   A   S
301 GAT GAC GAT AAT TAT GTG ATG GTG TCT CTG CTG GAA GGC AGG GTG AAT TTA TCG GCG TCT
   A   N   S   V   M   K   L   F   P   N   E   Q   A   L   Y   N   K   N   T   G
361 GCT AAT TCT GTG ATG AAA TTA TTC CCG AAT GAG CAG GCA CTG TAT AAC AAG AAC ACC GGA
   R   M   E   K   L   K   T   N   A   S   K   A   C   D   W   L   D   G   G   L
421 CGA ATG GAA AAG CTG AAG ACC AAT GCA AGT AAA GCC TGC GAT TGG CTG GAT GGC GGG TTG
   T   F   E   N   A   S   F   A   D   I   A   H   R   L   E   R   K   F   Q   V
481 ACC TTC GAG AAT GCT TCA TTC GCT GAT ATT GCC CAT CGT CTG GAA CGT AAA TTC CAG GTG
   K   I   S   I   E   S   E   R   L   K   A   E   H   F   S   G   S   F   D   S
541 AAG ATC AGC ATA GAA AGC GAA CGT CTG AAG GCC GAA CAC TTC TCC GGT TCT TTC GAC AGC
   N   Q   N   I   Y   D   I   L   H   E   I   N   V   E   K   Q   Y   T   W   K
601 AAT CAG AAT ATT TAT GAT ATA TTA CAC GAA ATC AAC GTT GAG AAA CAA TAC ACG TGG AAA
                                             BamHI
                                             ~~~~~
V   S   G   D   T   I   F   I   T   D   K   R   K   G   V   R   *
661 GTC AGT GGT GAC ACC ATC TTT ATT ACC GAT AAA AGA AAG GGG GTA AGA TAA GGA TCC AAG
      NotI
      ~~~~~
M   Q   E   Q   N   K   K   F   S   V   K   A   D   N   V   T   L   .
721 GAG GCG GCC GCA TGC AGG AAC AAA ACA AGA AGT TCA GTG TCA AAG CAG ATA ACG TTA CTT
   . K E A I E V V R K Q G N Y S F L I R N N .
781 TGA AAG AAG CCA TTG AAG TAG TCA GGA AGC AGG GAA ATT ATT CTT TTC TGA TTC GCA ATA
   . D I D L N K K V S V N V D K G T I N D V .
841 ACG ACA TTG ACC TGA ATA AGA AAG TCT CTG TGA ACG TGG ACA AGG GTA CAA TTA ATG ACG
   . M A Q L L T G T G I S Y E V N G N R V V .
901 TGA TGG CCC AAC TGC TGA CTG GTA CGG GGA TCA GTT ATG AAG TGA ACG GAA ACC GGG TGG
      XhoI
      ~~~~~
. I F H A L E H H H H H H *
961 TCA TAT TCC ATG CCC TCG AGC ACC ACC ACC ACC ACC ACT GA

```

Figure 6. 30 Sequence of both the BT0753CA and BT0754N genes in pET21a for co-expression.

DNA map was made by NTI vector to show the structure of BT0753CA (green) and BT0754N (Red) in mutated pET21a. This mutated pET21a has a second ribosome binding site (AAGGAG, blue box) before Not I site and after BamH I site, while other restriction sites after BamH I and before Not I were removed by site directed mutagenesis.

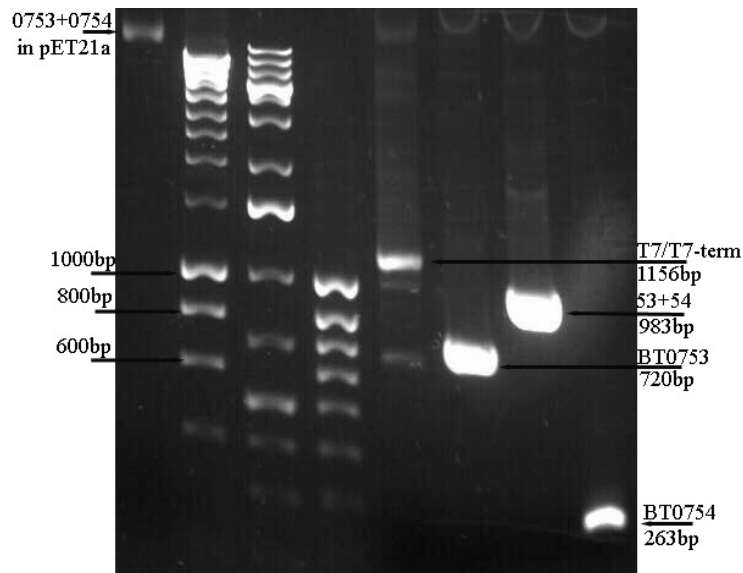


Figure 6. 31 Agarose gel shows co-expression structure of BT0753CA and BT0754N in modified pET21a.

Using different pairs of primers and a same plasmid as template to run PCR, different products demonstrate different parts of insert. Using primers T7 and T7-term, a 1156 bp band is produced which is between T7.tag and T7 terminator including BT0753CA and BT0754N. Using primer BT0753-12 and primer BT0753-13, a 720 bp band is produced which is BT0753CA. Using primer BT0754-3 and primer BT0754-2, a 263 bp band is produced which is BT0754N. Using primer BT0753-12 and BT0754-2, a 983 bp DNA band is produced which are BT0753CA and BT0754N. Lanes 2, 3, and 4 are size markers.

The co-expressed complex BT0753CA and BT0754N was purified by the IMAC method, but is not pure enough for crystallization (Figure 6. 32). Even with gradient elution increasing the imidazole from 10 mM, 30 mM, 50 mM, 75 mM, up to 100 mM, the impurity remained (Figure 6. 33).

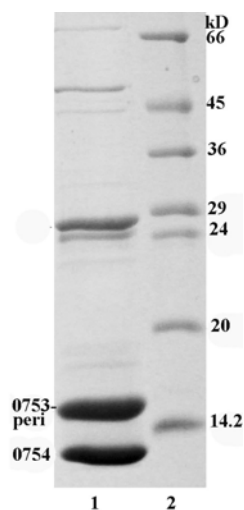


Figure 6. 32 SDS-PAGE shows the co-expression of BT0753CA with BT0754N.

This complex is not pure enough for crystallization. BT0753CA is the periplasmic domain of anti-sigma factor BT0753, while BT0754N is the N-terminal tail of SusC transducer BT0754.

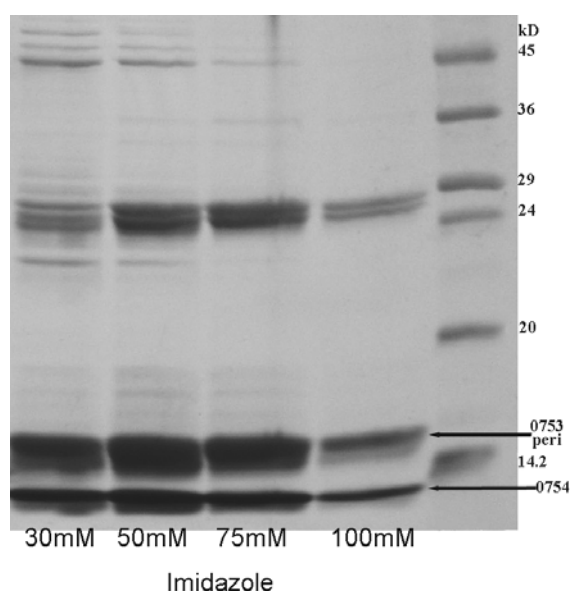


Figure 6. 33 SDS-PAGE shows the co-expression of BT0753CA and BT0754N.

Following the IMAC purification for the co-expression of BT0753CA Δ and BT0754N, a contamination could not be removed by gradient elution, which influenced the purity of the co-expressed protein and crystallization afterwards.

Finally, a fishing approach was used to achieve the complex. BT0754N was bound to IMAC column before loading the CFE of untagged BT0753CA Δ . The latter binds BT0754N to form a complex because of the high affinity, while other proteins in CFE flow away. Then, the complex was eluted from the column, which was not pure enough for crystallization because of extra bands (Figure 6. 34 Lane 1). Then gel filtration was used to further purify the two-component complex of BT0753CA Δ and BT0754N. Unfortunately, this complex was dissociated through the process of purification (Figure 6. 34).

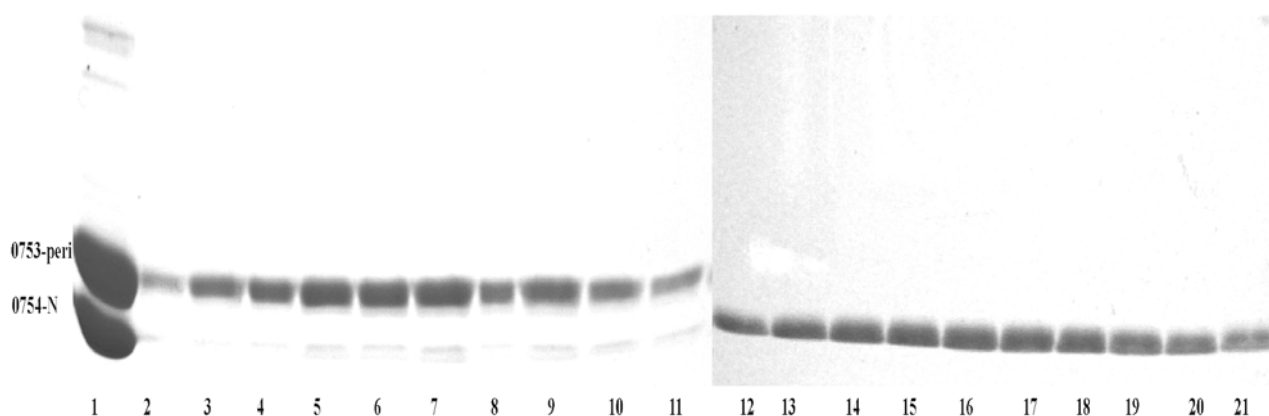


Figure 6. 34 Gel filtration for the complex of BT0753CA and BT0754N.

Lane 1: the complex of untagged BT0753CA Δ and BT0754N after IMAC purification. Lanes 2-21 are the fractions collected from gel filtration column. In lanes 2-11, most BT0753CA has been eluted while BT0754N still binds to the column. In lanes 12-21, the remained BT0754N was eluted.

A fishing approach was used to achieve the complex of BT0752 and BT0753N as well, by using the binding ability of BT0753N against BT0752 to fish out untagged BT0752 from CFE solution when BT0753N was bound on IMAC column (Figure 6. 35).

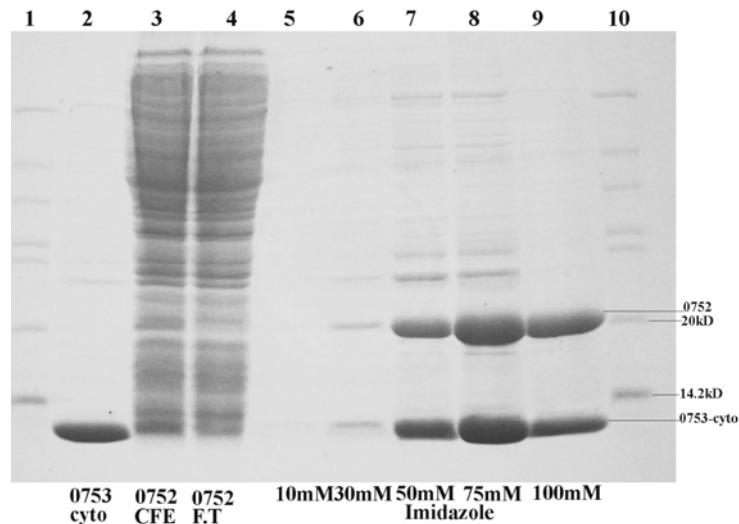


Figure 6. 35 SDS-PAGE shows untagged BT0752 was fished by tagged BT0753N.

The purified BT0753N (lane 2) fishes out untagged BT0752 from CFE (lane 3), the flow though has less BT0752 protein left (lane 4). After a series of elution with imidazole concentration from 10 mM, 30 mM, 50 mM, 75 mM, to 100 mM (lanes 5-9), in the final 100 mM elution, the proteins are pure enough for crystallization (lane 9).

The pure complex (BT0752/BT0753N) was set up for sitting-drop crystallization trials on four screens (JCSG, PACT, Classics, and Newcastle, see Appendix B). Unfortunately, no crystals were obtained in any of the conditions tested.

6.3 Discussion

B. thetaiotaomicron possesses 88 PULs to sense the environment variation (Bjursell *et al.*, 2006). In which, ECF-PULs are similar to Fec system of *E. coli*. They employ ECF sigma/anti-sigma factors as sensors, and SusC transducers as polysaccharide transducers to accomplish the sensing and transport process. Compared to the SusC transducers in Non-ECF PULs, ECF-PULs ones have N-extensions, which is about 80 residues after a signal peptide. This extra domain plays the signalling function. Binding of polysaccharide to the SusC transducer causes the N-extension to bind the periplasmic domain of the anti-sigma factor. This signal is passed to the cytoplasmic domain of the anti-sigma through the membrane. The change of cytoplasmic domain will stimulate the sigma factor to join the core RNA polymerase to form a complete RNAP, which will start to transcript the catabolic enzymes required to degrade the initially bound polysaccharide. This pathway enables *B. thetaiotaomicron* to sense the variation of niches in human gut accurately and efficiently, which enhances the competition ability of the bacterium which lives in a very densely populated environment. This kind of ECF sigma factor sensing pathway is common in many organisms that are able to respond to a large number of environmental variations, such as *Bacillus subtilis*, *Mycobacterium tuberculosis*, *Caulobacter crescentus*, *Pseudomonas aeruginosa*, and *Streptomyces coelicolor*. In many cases ECF sigmas outnumber all other types of sigma factors (Table 1.1) (Helmann, 2002).

However, *B. thetaiotaomicron* genome contains more two component systems (79 classic plus 32 hybridized) than ECF sigma/anti-sigma pairs (25 in number) (Xu *et al.*, 2003). The coexistence of TCS and ECF sigma factors signalling mechanisms reflects the high stress level from human gut and multiple response policy of *B. thetaiotaomicron*. TCS may respond to the signalling in human gut at regular level, while ECF mechanisms are only activated when signalling level is lower, i.e., stressful. The high number of TCS and high level of sequence similarity may be sufficient to allow some level of cross-talk, which has been detected in other organism, such as *E. coli* (Siryaporn and Goulian, 2008). This kind of cross-talking capacity may be an economy way for *B. thetaiotaomicron*, as in some cases, a set of signalings from human gut may come together regularly, such as plant cellulose always come with hemicellulose, the signalling from cellulose could initiate hemicellulose hydrolases to transcript. Further research is needed to support this

hypothesis. Data presented here shows that ECF-PULs in *B. thetaiotaomicron* have no cross-talk recognition of anti-sigma factors to sigma factors, or sigma factors to SusC transducers. The recognition only happens between proteins in the same gene cluster. This could explain the sequence similarity variations of sigma/anti-sigma factors, and SusC transducers. The low sequence similarity of both N- and C-terminal of sigma factors, but high similarity in the middle region is because the N-terminal recognize special promoter of RNA polymerase, while the C-terminal recognize matched N-terminal of anti-sigma factor only. But the main function of ECF sigma factor is conserved, so the middle region is highly similar. Meanwhile, the lack of binding of sigma/anti-sigma factors also explained the low similarity of N-terminal of anti-sigma factor (cytoplasmic domain). The lack of binding of anti-sigma C-terminal to N-terminal of SusC explains the low similarity of C-terminal 15 residues of anti-sigma factor and N-terminal 80 residues of SusC transducers. And the high similarity of other fraction of C-terminal of anti-sigma factors addressed the functional conservation, while the high similarity of barrel of SusC transducers indicates the functional conservation as well.

The binding of sigma/anti-sigma factors, anti-sigma factor/SusC N-extension has a stoichiometry of 1:1 and the interaction is an exothermic reaction though the entropy changes and association constants vary considerably (Table 6. 9 and Table 6. 10). It is currently not clear why there is such difference in the affinity and thermodynamic signature of the different protein pairs. It is possible the actual residues involved in the interactions have diverged significantly between proteins in different PULs. However it is also possible the differences seen are due to errors in the fit of the data.

The unknown *E. coli* cytoplasmic protease that cleaves the periplasmic domain of anti-sigma factors can not be inhibited by commercial protease inhibitor, or urea and the Trx tag and GST tag could not stop the fused protein being cleaved either. It is possible that this event is biologically relevant, but it is hard to envisage how cleaving half of the periplasmic domain of the anti-sigma in the inactive (unbound) form would have any useful role in the system. It seems more likely that this is simply an artifact of the expression system as the protein is present in the cytoplasm of *E. coli*, but is not a periplasmic protein in *B. thetaiotaomicron*. However, the highly conserved sequence of the cleavage site and that fact that the same proteolytic even occurs with FecR, suggests that this region

is important for function, even if it is not normally cleaved. Interestingly, expression of only the C-terminal portion of the periplasmic domain of the anti-sigma after the cleavage site produces a protein that is able to bind the cognate SusC transducers' N-extension indicating a distinct two-domain structure of the periplasmic region of the anti-sigma. One domain (the C-terminal) binds the transducers N-extension, while the other (the N-terminal extending from the inner membrane) may be involved in signal transduction across the membrane.

Comparing the K_a of BT0753C Δ /BT0754N to BT4706C Δ /BT4707N, a possible reason to explain the dissociation of BT0753C Δ /BT0754N complex from gel filtration (Table 6. 9) is that the affinity is not high enough. If we employ the complex of BT4706C Δ /BT4707N for crystallization, a higher purity of complex may be achieved. A possible explanation for the crystallization failure of BT0752/BT0753N is the His₆ tag takes a large ratio of BT0753N when compared to the size of this protein. This kind of tag could be very flexible and prevent the complex to form into proper crystal lattice. If the His tag was removed, the chance to crystallize may be higher. NMR could be another potential method to solve the structures of the complexes.

Chapter 7 Final Discussion

Combining all result chapters, a final discussion was drawn to further understand the overall mechanism of carbohydrate sensing and utilisation system of *B. thetaiotaomicron*.

B. thetaiotaomicron is prominent in the normal human gut microflora as a symbiont (Xu *et al.*, 2003). Its genome revealed a large group of proteins involved in the sensing, acquisition and degradation of carbohydrates (Xu *et al.*, 2003). These proteins include 209 SusC/D homologues, 241 carbohydrate-active enzymes, 76 ECF- σ /anti- σ factors, 111 TCS (included HTCS), and 41 related transporters (Magnus *et al.*, 2006; Shipman *et al.*, 2000; Xu *et al.*, 2003).

Glycoside hydrolase and polysaccharide lyase genes are frequently adjacent to SusC and SusD homologues and sensors, such as pairs of ECF-sigma/anti-sigma factors and hybrid two-component system. These linked genes combine into polysaccharide utilisation loci (PULs) (18 ECF-PULs and 70 non-ECF-PULs) to sense, bind, and degrade specific polysaccharides (Bjursell *et al.*, 2006). The PULs expand the carbohydrate utilisation ability from human and establish the dominate status of *B. thetaiotaomicron* in gut. So far, the carbohydrate sensing and utilisation mechanism is not fully understood at molecular level.

An ECF-PUL gene cluster comprises SusC transducer, ECF- σ /anti- σ factors, and at least one glycoside hydrolase or one polysaccharide lyase, in which SusC transducer has an N-terminal tail as a signal peptide (chapter 6). The tail binds to the anti-sigma factor in periplasm which passes the signal to the cytoplasmic domain through the transmembrane domain. Finally, the associated sigma factor is released by anti-sigma factor to form a RNAP to transcribe the polysaccharide enzymes. Compared to non-ECF PULs, the N-tail of SusC transducer is unique, which indicates that in non-ECF PULs, the mechanisms are different. ECF-PULs have no cross recognition of anti-sigma factors to sigma factors, or sigma factors to SusC transducers in the systems tested. If we want to fully understand level of cross-talk, we need to analyse all the ECF-sigma and anti-sigma factors, possibly using a yeast 2-hybrid system. ECF-PULs transfer signals from C-terminal of SusC transducer to N-terminal tail, then to C-terminal of anti-sigma factor (periplasmic domain), followed by the N-terminal of anti-sigma factor interacts with C-terminal domain of ECF sigma

factor (Figure 6. 1) (Braun and Mahren, 2005). All system components signal from C- to N-terminus.

In comparison, the sensory component of hybrid two component systems of non-ECF PULs has lower binding affinity (chapter 3: BT1754peri against fructose, the K_a is only $2.3 \times 10^5 \text{ M}^{-1}$). Classical two component systems are predicted to have cross-talk opportunity according to other two component system studies where the response regulator is free in the cytoplasm (Siraporn and Goulian, 2008). Hybrid two component system is different to two component system as the fusion of the response regulator to the histidine kinase means it is unlikely to have any cross talk in these systems or signal amplification. Two component systems (including HTCS) are N- to C-terminal signalling (Figure 1. 13). The sensor histidine kinase is predicted to be a homodimer according to the crystal structure (Lee *et al.*, 2008), even though BT1754peri presents as a monomer in solution (section 3.2.1.3.1). According to the comparison of ECF-PULs and non-ECF-PULs, the sensing mechanism difference is obvious.

In the operon controlled by HTCS BT1754, there are four GHs which belong to GH32 family, only one of which is an endo-acting enzyme, BT1760 (chapter 4). BT1760 functions a mixture of endo-acting and exo-acting. The low rate of endo-acting in a group of same activity enzymes could be an effective arrangement for carbohydrate digestion. BT1760 is located on the outer membrane and cuts levan into small oligosaccharides which are transferred into the *B. thetaiotaomicron* periplasm. BT1760 could not digest inulin, which may not be a problem *in vivo* because native inulin are $< \sim 20$ DP which may translocate through the barrel of SusC transducer straight away, while levans are generally much longer chains (Figure 4. 6E). BT1759 and BT3082 appear to have similar specificity and both are periplasmic. There is no clue as why the two enzymes play one role, although there may be redundancy. Knockout of one of the two chondroitin sulfate lyases in *B. thetaiotaomicron* does not affect its growth on chondroitin, so redundancy may be a common theme in *B. thetaiotaomicron* (Shaya *et al.*, 2008). BT1765 prefers short oligosaccharides and likely accomplishes the final digestion step of fructans in cytoplasm. Totally this group of GH32s could digest all kinds of fructans including oligosaccharides and sucrose. Their substrate preferences complement each other to increase fructan digestion from long chains to monosaccharides.

The outer membrane SusD homologue BT1762 could bind levan, but not inulin, while BT1760 GH32, which is located on the outer membrane as well, could digest levan, but not inulin. The identical ligand-specificity revealed that these two proteins are cooperative. BT1762 binds polymeric levan tightly, but oligosaccharides only weakly. When BT1760 cleaves the levan chain while the polysaccharide is bound to BT1762, the binding will become much weaker and this may facilitate transport through the SusC homologue BT1763. BT1762 presents identical TPR motif to SusD though, which indicated BT1762 may play a scaffold role to link SusC transducer BT1763, levanase BT1760, and other related proteins on the outer membrane. Although we could see no interaction between BT1762 and BT1760 by native gels, if we could mimic the lipid-layer environment of outer membrane and also express full lengths of both proteins, it may be possible to detect an interaction between them. Levan-binding of BT1762 is probably based on the endo-recognition of the secondary structure rather than the exo-action against terminal fructose units, as SusD could form hydrogen bonds with G4, G5, and G6 of helical maltoheptaose, which is a cycle unit of starch and levan is predicted to have such a helical secondary structure as well, but more extended than starch (Koropatkin *et al.*, 2008). This kind of endo-binding of BT1762 could help the endo-acting of levanase BT1760.

Generally, all the components of *B. thetaiotaomicron* sensing and utilisation of carbohydrates are organized as a team to accomplish the whole process from sensing carbohydrate in the gut environment and stimulating related enzymes to transcript to digesting carbohydrate for the consumption of *B. thetaiotaomicron* and host (Figure 7. 1). SusC/D homologues locate on the outer membrane associated with GH enzyme to mediate polysaccharides to get into periplasm. In ECF-PULs, the N-terminal tail of SusC transducer could interact with C-terminal of anti- σ periplasmic domain to change the N-terminal structure of anti- σ and release the C-terminal associated σ factor which then binds RNAP to initiate the transcription of ECF-PULs. In non-ECF-PULs the activating signal is sensed by the periplasmic domain of dimerized sensor HKs of HTCS to phosphorelay to cognate RR which then binds DNA directly to activate transcription of the catabolic genes.

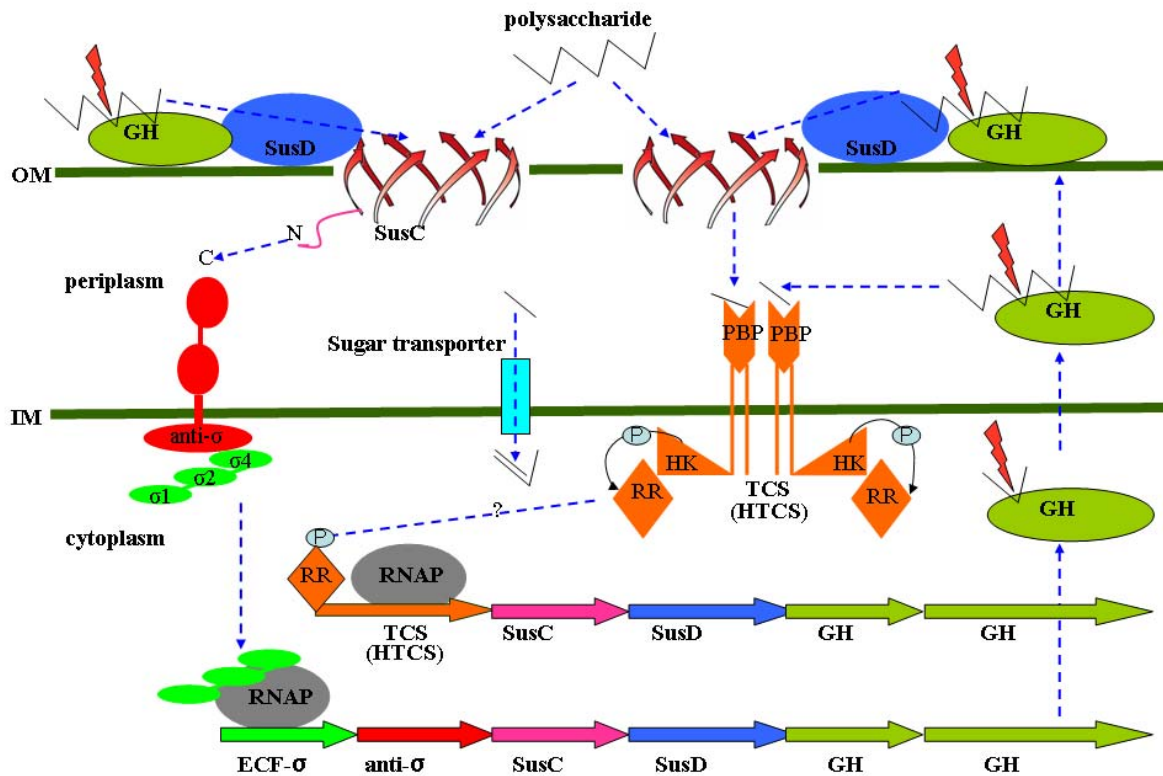


Figure 7. 1 Overall process of *B. thetaiotaomicron* sensing and utilisation of polysaccharides.

Blue dash arrow means the movement of polysaccharides and proteins. SusD homologues (light blue circle) located on the outer membrane (OM: upper green line), associated with SusC homologues (purple barrel) and GH enzyme (grass green) to mediate polysaccharides to get into periplasm. The N-terminal tail of SusC transducer could interact with C-terminal of anti- σ periplasmic domain to change the N-terminal structure of anti- σ and release the C-terminal associated σ factor (σ_4) which then binds RNAP (grey cycle) to initiate the transcription of ECF-PULs. In non-ECF-PULs the activating signal is recognised directly by the periplasmic domain of dimerized sensor HKs of HTCS (orange) to initiate the phosphorelay that results in activated cognate response regulator (RR). The phosphorylated RR output domain then binds DNA directly to activate transcription of the catabolic genes under its control. Whether this occurs by proteolytic release of the RR or the DNA is brought to the membrane and the RR remains attached to the HK domain is currently unknown.

Compared to polysaccharide utilisation mechanisms in other bacteria, which mainly use CBMs to bind plant cell wall polysaccharides and secreted enzymes to degrade the polysaccharides, *B. thetaiotaomicron* catabolic enzymes are different. Instead *B. thetaiotaomicron* uses the Sus complex and more specifically SusD homologues, to bind polysaccharides and pull them into the periplasm for degradation. This adaptation will provide *B. thetaiotaomicron* with a competitive advantage in the very densely populated niche of the human gut as none of its products of degradation will be available for use by its competitors.

Future work

Some plans for future studies in the area of polysaccharide utilisation by *Bacteroides thetaiotaomicron* are as follows:

- 1) What is the mechanism of signal transduction in BT1754? Is it the same in other HTCS?
- 2) How does the RR domain of the HTCS bind the genomic DNA to activate transcription? Is it proteolytically cleaved or does the DNA come to the membrane? Research on ToxR suggested that DNA may come to the membrane to bind RR output domain (Crawford *et al.*, 2003).
- 3) How is the mechanism in other PULs that degrade more complex polysaccharides than fructan?
- 4) What is the levan structure bound to levanase BT1760 and SusD homologue BT1762? Will this help explain the unique activity of BT1760 and understand the role of BT1762 better?
- 5) What is the role of the N- and C-terminal domains in GH32s? Why does BT1760 have no N-terminal domain?
- 6) How many proteins involved in Sus complex and role and structure of each?

B. thetaiotaomicron has a lot of interesting genes with potential function to sense and utilise carbohydrates which could be utilised in industrial applications, such as production of biofuel and biosensors. Further understanding of this organism will also enhance our knowledge of its role in nutrition of human gut.

References

- Adams, P., Fowler, R., Howell, G., Kinsella, N., Skipp, P., Coote, P. and O'Connor, C.D. (1999) Defining protease specificity with proteomics: a protease with a dibasic amino acid recognition motif is regulated by a two-component signal transduction system in Salmonella. *Electrophoresis*, **20**, 2241-2247.
- Ajit, V., Richard Cummings, Jeffrey Esko, Hudson Freeze, Gerald Hart and Marth, J. (1999) *Essentials of Glycobiology*. Cold Spring Harbor Laboratory.
- Alberto, F., Bignon, C., G. S., Henrissat, B. and Czjzek, M. (2004) The three-dimensional structure of invertase (beta-fructosidase) from *Thermotoga maritima* reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J Biol Chem*, **279**, 18903-18910.
- Alberto, F., Jordi, E., Henrissat, B. and Czjzek, M. (2006) Crystal structure of inactivated *Thermotoga maritima* invertase in complex with the trisaccharide substrate raffinose. *Biochem J*, **395**, 457-462.
- Alvarez, M., C. E., Lourenco, R.F., Baldini, R.L., Laub, M.T. and Gomes, S.L. (2007) The ECF sigma factor sigma(T) is involved in osmotic and oxidative stress responses in *Caulobacter crescentus*. *Mol Microbiol*, **66**, 1240-1255.
- Angerer, A., Enz, S., Ochs, M. and Braun, V. (1995) Transcriptional regulation of ferric citrate transport in *Escherichia coli* K-12. Fecl belongs to a new subfamily of sigma 70-type factors that respond to extracytoplasmic stimuli. *Mol Microbiol*, **18**, 163-174.
- Anthony, J.R., Newman, J.D. and Donohue, T.J. (2004) Interactions between the *Rhodobacter sphaeroides* ECF sigma factor, sigma(E), and its anti-sigma factor, ChrR. *J Mol Biol*, **341**, 345-360.
- Appleby, J.L., Parkinson, J.S. and Bourret, R.B. (1996) Signal transduction via the multi-step phosphorelay: not necessarily a road less traveled. *Cell*, **86**, 845-848.
- Aravind, L. and Ponting, C.P. (1997) The GAF domain: an evolutionary link between diverse phototransducing proteins. *Trends Biochem Sci*, **22**, 458-459.
- Aravind, L. and Ponting, C.P. (1999) The cytoplasmic helical linker domain of receptor histidine kinase and methyl-accepting proteins is common to many prokaryotic signalling proteins. *FEMS Microbiol Lett*, **176**, 111-116.
- Arraiz, N., Salazar, L., Lopez, G., Rodriguez, R., Casart, Y. and Takiff, H. (2001) [Characterization of the expression and function of SigM an ECF sigma factor in mycobacteria]. *Acta Cient Venez*, **52 Suppl 1**, 40-41.
- Artymiuk, P.J., Rice, D.W., Mitchell, E.M. and Willett, P. (1990) Structural resemblance between the families of bacterial signal-transduction proteins and of G proteins revealed by graph theoretical techniques. *Protein Eng*, **4**, 39-43.
- Ayers, R.A. and Moffat, K. (2008) Changes in quaternary structure in the signaling mechanisms of PAS domains. *Biochemistry*, **47**, 12078-12086.

- Backhed, F., Ley, R.E., Sonnenburg, J.L., Peterson, D.A. and Gordon, J.I. (2005) Host-bacterial mutualism in the human intestine. *Science*, **307**, 1915-1920.
- Baikalov, I., Schroder, I., Kaczor-Grzeskowiak, M., Cascio, D., Gunsalus, R.P. and Dickerson, R.E. (1998) NarL dimerization? Suggestive evidence from a new crystal form. *Biochemistry*, **37**, 3665-3676.
- Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res*, **28**, 304-305.
- Bantscheff, M., Perraud, A.L., Bock, A., Rippe, K., Weiss, V., Glocker, M. and Gross, R. (2000) Structure-function relationships in the Bvg and Evg two-component phosphorelay systems. *Int J Med Microbiol*, **290**, 317-323.
- Bauer, M., Kube, M., Teeling, H., Richter, M., Lombardot, T., Allers, E., Wurdemann, C.A., Quast, C., Kuhl, H., Knaust, F., Woebken, D., Bischof, K., Mussmann, M., Choudhuri, J.V., Meyer, F., Reinhardt, R., Amann, R.I. and Glockner, F.O. (2006) Whole genome analysis of the marine Bacteroidetes 'Gramella forsetii' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol*, **8**, 2201-2213.
- Bayer, E.A., Belaich, J.P., Shoham, Y. and Lamed, R. (2004) The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol*, **58**, 521-554.
- Beier, D. and Gross, R. (2006) Regulation of bacterial virulence by two-component systems. *Curr Opin Microbiol*, **9**, 143-152.
- Beier, D., Schwarz, B., Fuchs, T.M. and Gross, R. (1995) In vivo characterization of the unorthodox BvgS two-component sensor protein of Bordetella pertussis. *J Mol Biol*, **248**, 596-610.
- Belcheva, A. and Golemi-Kotra, D. (2008) A close-up view of the VraSR two-component system. A mediator of Staphylococcus aureus response to cell wall damage. *J Biol Chem*, **283**, 12354-12364.
- Berg, J., Tymoczko, J.L. and Stryer, L. (2002) *Biochemistry*.
- Berg, R.D. (1996) The indigenous gastrointestinal microflora. *Trends Microbiol*, **4**, 430-435.
- Bernet, M.F., Brassart, D., Neeser, J.R. and Servin, A.L. (1994) Lactobacillus acidophilus LA 1 binds to cultured human intestinal cell lines and inhibits cell attachment and cell invasion by enterovirulent bacteria. *Gut*, **35**, 483-489.
- Bewsey, K.E., Johnson, M.E. and Huff, J.P. (1991) Rapid isolation and purification of DNA from agarose gels: the phenol-freeze-fracture method. *Biotechniques*, **10**, 724-725.
- Bilwes, A.M., Alex, L.A., Crane, B.R. and Simon, M.I. (1999) Structure of CheA, a signal-transducing histidine kinase. *Cell*, **96**, 131-141.
- Bingham, S.A. (1999) High-meat diets and cancer risk. *Proc Nutr Soc*, **58**, 243-248.
- Biville, F., Cwerman, H., Letoffe, S., Rossi, M.S., Drouet, V., Ghigo, J.M. and Wandersman, C. (2004) Haemophore-mediated signalling in Serratia marcescens: a new mode of regulation for an extra cytoplasmic function (ECF)

sigma factor involved in haem acquisition. *Mol Microbiol*, **53**, 1267-1277.

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. *J Biol Chem*, **281**, 36269-36279.

Bons, N., Mestre-Frances, N., Guiraud, I. and Charnay, Y. (1997) Prion immunoreactivity in brain, tonsil, gastrointestinal epithelial cells, and blood and lymph vessels in lemurian zoo primates with spongiform encephalopathy. *C R Acad Sci III*, **320**, 971-979.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J*, **382**, 769-781.

Borgstahl, G.E., Williams, D.R. and Getzoff, E.D. (1995) 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry*, **34**, 6278-6287.

Bourret, R.B., Borkovich, K.A. and Simon, M.I. (1991) Signal transduction pathways involving protein phosphorylation in prokaryotes. *Annu Rev Biochem*, **60**, 401-441.

Brünger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472-475.

Braun-Fahrlander, C., Riedler, J., Herz, U., Eder, W., Waser, M., Grize, L., Maisch, S., Carr, D., Gerlach, F., Bufe, A., Lauener, R.P., Schierl, R., Renz, H., Nowak, D. and von Mutius, E. (2002) Environmental exposure to endotoxin and its relation to asthma in school-age children. *N Engl J Med*, **347**, 869-877.

Braun, M., Endriss, F., Killmann, H. and Braun, V. (2003a) In vivo reconstitution of the FhuA transport protein of *Escherichia coli* K-12. *J Bacteriol*, **185**, 5508-5518.

Braun, V. and Mahren, S. (2005) Transmembrane transcriptional control (surface signalling) of the *Escherichia coli* Fec type. *FEMS Microbiol Rev*, **29**, 673-684.

Braun, V., Mahren, S. and Ogierman, M. (2003b) Regulation of the FecI-type ECF sigma factor by transmembrane signalling. *Curr Opin Microbiol*, **6**, 173-180.

Brinkman, F.S., Schoofs, G., Hancock, R.E. and De Mot, R. (1999) Influence of a putative ECF sigma factor on expression of the major outer membrane protein, OprF, in *Pseudomonas aeruginosa* and *Pseudomonas fluorescens*. *J Bacteriol*, **181**, 4746-4754.

Brown, K.L. and Hughes, K.T. (1995) The role of anti-sigma factors in gene regulation. *Mol Microbiol*, **16**, 397-404.

Browning, D.F., Whitworth, D.E. and Hodgson, D.A. (2003) Light-induced carotenogenesis in *Myxococcus xanthus*: functional characterization of the ECF sigma factor CarQ and antisigma factor CarR. *Mol Microbiol*, **48**, 237-251.

Brutsche, S. and Braun, V. (1997) SigX of *Bacillus subtilis* replaces the ECF sigma factor fecI of *Escherichia coli* and is

inhibited by RsiX. *Mol Gen Genet*, **256**, 416-425.

- Bulow, L. and Mosbach, K. (1982) Ligation of restriction endonuclease-generated DNA fragments using immobilized T4 DNA ligase. *Biochem Biophys Res Commun*, **107**, 458-464.
- Burger, M., Woods, R.G., McCarthy, C. and Beacham, I.R. (2000) Temperature regulation of protease in *Pseudomonas fluorescens* LS107d2 by an ECF sigma factor and a transmembrane activator. *Microbiology*, **146 Pt 12**, 3149-3155.
- Carter, R.A., Worsley, P.S., Sawers, G., Challis, G.L., Dilworth, M.J., Carson, K.C., Lawrence, J.A., Wexler, M., Johnston, A.W. and Yeoman, K.H. (2002) The vbs genes that direct synthesis of the siderophore vicibactin in *Rhizobium leguminosarum*: their expression in other genera requires ECF sigma factor RpoI. *Mol Microbiol*, **44**, 1153-1166.
- Chang, C. and Stewart, R.C. (1998) The two-component system. Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiol*, **117**, 723-731.
- Chen, P.M., Chen, H.C., Ho, C.T., Jung, C.J., Lien, H.T., Chen, J.Y. and Chia, J.S. (2008) The two-component system ScnRK of *Streptococcus mutans* affects hydrogen peroxide resistance and murine macrophage killing. *Microbes Infect*, **10**, 293-301.
- Cheng, Q., Yu, M.C., Reeves, A.R. and Salyers, A.A. (1995) Identification and characterization of a *Bacteroides* gene, csuF, which encodes an outer membrane protein that is essential for growth on chondroitin sulfate. *J Bacteriol*, **177**, 3721-3727.
- Cheung, J. and Hendrickson, W.A. (2008) Crystal Structures of C4-Dicarboxylate Ligand Complexes with Sensor Domains of Histidine Kinases DcuS and DctB. *J Biol Chem*, **283**, 30256-30265.
- Cho, J.C. and Giovannoni, S.J. (2003) *Croceibacter atlanticus* gen. nov., sp. nov., a novel marine bacterium in the family Flavobacteriaceae. *Syst Appl Microbiol*, **26**, 76-83.
- Christie, J.M., Salomon, M., Nozue, K., Wada, M. and Briggs, W.R. (1999) LOV (light, oxygen, or voltage) domains of the blue-light photoreceptor phototropin (nph1): binding sites for the chromophore flavin mononucleotide. *Proc Natl Acad Sci U S A*, **96**, 8779-8783.
- Clark, D.J. (1968) The regulation of DNA replication and cell division in *E. coli* B-r. *Cold Spring Harb Symp Quant Biol*, **33**, 823-838.
- Cohen, S.N., Chang, A.C. and Hsu, L. (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci U S A*, **69**, 2110-2114.
- Conly, J.M., Stein, K., Worobetz, L. and Rutledge-Harding, S. (1994) The contribution of vitamin K2 (menaquinones) produced by the intestinal microflora to human nutritional requirements for vitamin K. *Am J Gastroenterol*, **89**, 915-923.
- Coulton, J.W., Reid, G.K. and Campana, A. (1988) Export of hybrid proteins FhuA'-LacZ and FhuA'-PhoA to the cell

envelope of *Escherichia coli* K-12. *J Bacteriol*, **170**, 2267-2275.

Crawford, J.A., Krukonis, E.S. and DiRita, V.J. (2003) Membrane localization of the ToxR winged-helix domain is required for TcpP-mediated virulence gene activation in *Vibrio cholerae*. *Mol Microbiol*, **47**, 1459-1473.

Crosson, S., Rajagopal, S. and Moffat, K. (2003) The LOV domain family: photoresponsive signaling modules coupled to diverse output domains. *Biochemistry*, **42**, 2-10.

Cuneo, M.J., Beese, L.S. and Hellinga, H.W. (2008) Ligand-induced conformational changes in a thermophilic ribose-binding protein. *BMC Struct Biol*, **8**, 50.

Cuneo, M.J., Tian, Y., Allert, M., Hellinga, H.W. . (2008) The backbone structure of the thermophilic *Thermoanaerobacter tengcongensis* ribose binding protein is essentially identical to its mesophilic *E. coli* homolog. *Bmc StructBiol* **8**.

D'Andrea, L.D. and Regan, L. (2003) TPR proteins: the versatile helix. *Trends Biochem Sci*, **28**, 655-662.

D'Elia, J.N. and Salyers, A.A. (1996) Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *J Bacteriol*, **178**, 7173-7179.

Daguer, J.P., Geissmann, T., Petit-Glatron, M.F. and Chambert, R. (2004) Autogenous modulation of the *Bacillus subtilis* *sacB-levB-yveA* levansucrase operon by the *levB* transcript. *Microbiology*, **150**, 3669-3679.

Dathe, M. and Wieprecht, T. (1999) Structural features of helical antimicrobial peptides: their potential to modulate activity on model membranes and biological cells. *Biochim Biophys Acta*, **1462**, 71-87.

David, P.L.I. and Cynthia, A.H. (1998) Apoplastic Sugars, Fructans, Fructan Exohydrolase, and Invertase in Winter Oat: Responses to Second-Phase Cold Hardening. *PlantPhysiol*, **116**, 403-408.

Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure*, **3**, 853-859.

Davies, G.J., Ducros, V., Lewis, R.J., Borchert, T.V. and Schulein, M. (1997) Oligosaccharide specificity of a family 7 endoglucanase: insertion of potential sugar-binding subsites. *J Biotechnol*, **57**, 91-100.

Dean, C.R. and Poole, K. (1993) Cloning and characterization of the ferric enterobactin receptor gene (*pfeA*) of *Pseudomonas aeruginosa*. *J Bacteriol*, **175**, 317-324.

DeBoy, R.T., Mongodin, E.F., Fouts, D.E., Tailford, L.E., Khouri, H., Emerson, J.B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H.J. and Nelson, K.E. (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J Bacteriol*, **190**, 5455-5463.

Demirci, F. (2008) Applied Thin-Layer Chromatography-Best Practice and Avoidance of Mistakes. 2nd ed. By E. Hahn-Deinstrop. Translated by R. G. Leach. Wiley-VCH, Weinheim. 2007. xvi +330 pp. 16.5 x 24.5 cm. \$190.00 or pound100.00. ISBN 978-3-527-31553-6 (hardcover). ISBN 978-3-527-61025-9 (e-book). *J Nat Prod*, **71**, 1661.

Divne, C., Stahlberg, J., Teeri, T.T. and Jones, T.A. (1998) High-resolution crystal structures reveal how a cellulose

- chain is bound in the 50 Å long tunnel of cellobiohydrolase I from *Trichoderma reesei*. *J Mol Biol*, **275**, 309-325.
- Djordjevic, S., Goudreau, P.N., Xu, Q., Stock, A.M. and West, A.H. (1998) Structural basis for methyltransferase CheB regulation by a phosphorylation-activated domain. *Proc Natl Acad Sci U S A*, **95**, 1381-1386.
- Dona, V., Rodrigue, S., Dainese, E., Palu, G., Gaudreau, L., Manganelli, R. and Provvedi, R. (2008) Evidence of complex transcriptional, translational and posttranslational regulation of the ECF sigma factor σ^E in *Mycobacterium tuberculosis*. *J Bacteriol*.
- Downey, N. (2003) Extraction of DNA from agarose gels. *Methods Mol Biol*, **235**, 137-139.
- Dutta, R., Qin, L. and Inouye, M. (1999) Histidine kinases: diversity of domain organization. *Mol Microbiol*, **34**, 633-640.
- Dwyer, M.A. and Hellinga, H.W. (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol*, **14**, 495-504.
- Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. and Relman, D.A. (2005) Diversity of the human intestinal microbial flora. *Science*, **308**, 1635-1638.
- Elich, T.D. and Chory, J. (1997) Phytochrome: if it looks and smells like a histidine kinase, is it a histidine kinase? *Cell*, **91**, 713-716.
- Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, **60**, 2126-2132.
- Engels, S., Schweitzer, J.E., Ludwig, C., Bott, M. and Schaffer, S. (2004) *clpC* and *clpP1P2* gene expression in *Corynebacterium glutamicum* is controlled by a regulatory network involving the transcriptional regulators ClgR and HspR as well as the ECF sigma factor σ^H . *Mol Microbiol*, **52**, 285-302.
- Enz, S., Brand, H., Orellana, C., Mahren, S. and Braun, V. (2003) Sites of interaction between the FecA and FecR signal transduction proteins of ferric citrate transport in *Escherichia coli* K-12. *J Bacteriol*, **185**, 3745-3752.
- Fehr, M., Frommer, W.B. and Lalonde, S. (2002) Visualization of maltose uptake in living yeast cells by fluorescent nanosensors. *Proc Natl Acad Sci U S A*, **99**, 9846-9851.
- Fehr, M., Lalonde, S., Lager, I., Wolff, M.W. and Frommer, W.B. (2003) In vivo imaging of the dynamics of glucose uptake in the cytosol of COS-7 cells by fluorescent nanosensors. *J Biol Chem*, **278**, 19127-19133.
- Filippou, P.S., Kasemian, L.D., Panagiotidis, C.A. and Kyriakidis, D.A. (2008) Functional characterization of the histidine kinase of the *E. coli* two-component signal transduction system AtoS-AtoC. *Biochim Biophys Acta*.
- Finegold, S.M., Sutter, V.L. and Mathisen, G.E. (1983) Normal indigenous intestinal flora, 3-31. In D.J. Hentges (ed), *Human intestinal microflora in health and disease Academic Press, New York*

- Fisher, H.F. and Singh, N. (1995) Calorimetric methods for interpreting protein-ligand interactions. *Methods Enzymol*, **259**, 194-221.
- Flegal, K.M. and Troiano, R.P. (2000) Changes in the distribution of body mass index of adults and children in the US population. *Int J Obes Relat Metab Disord*, **24**, 807-818.
- Forst, D., Welte, W., Wacker, T. and Diederichs, K. (1998) Structure of the sucrose-specific porin ScrY from *Salmonella typhimurium* and its complex with sucrose. *Nat Struct Biol*, **5**, 37-46.
- Fortman, J.L., Chhabra, S., Mukhopadhyay, A., Chou, H., Lee, T.S., Steen, E. and Keasling, J.D. (2008) Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol*, **26**, 375-381.
- Fukami-Kobayashi, K., Tateno, Y. and Nishikawa, K. (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol*, **286**, 279-290.
- Fulop, V. and Jones, D.T. (1999) Beta propellers: structural rigidity and functional diversity. *Curr Opin Struct Biol*, **9**, 715-721.
- Galperin, M.Y. (2006) Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *J Bacteriol*, **188**, 4169-4182.
- Garcia-Herrero, A. and Vogel, H.J. (2005) Nuclear magnetic resonance solution structure of the periplasmic signalling domain of the TonB-dependent outer membrane transporter FecA from *Escherichia coli*. *Mol Microbiol*, **58**, 1226-1237.
- Garrity, L.F., Schiel, S.L., Merrill, R., Reizer, J., Saier, M.H., Jr. and Ordal, G.W. (1998) Unique regulation of carbohydrate chemotaxis in *Bacillus subtilis* by the phosphoenolpyruvate-dependent phosphotransferase system and the methyl-accepting chemotaxis protein McpC. *J Bacteriol*, **180**, 4475-4480.
- Garrod, L.P. and al., e. (1981) *Antibiotic and Chemotherapy*.
- Goldberg, M.B., Boyko, S.A., Butters, J.R., Stoebner, J.A., Payne, S.M. and Calderwood, S.B. (1992) Characterization of a *Vibrio cholerae* virulence factor homologous to the family of TonB-dependent proteins. *Mol Microbiol*, **6**, 2407-2418.
- Golomb, M. and Chamberlin, M. (1974) Characterization of T7-specific ribonucleic acid polymerase. IV. Resolution of the major in vitro transcripts by gel electrophoresis. *J Biol Chem*, **249**, 2858-2863.
- Gong, W., Hao, B., Mansy, S.S., Gonzalez, G., Gilles-Gonzalez, M.A. and Chan, M.K. (1998) Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction. *Proc Natl Acad Sci U S A*, **95**, 15177-15182.
- Gorham, H.C., McGowan, S.J., Robson, P.R. and Hodgson, D.A. (1996) Light-induced carotenogenesis in *Myxococcus xanthus*: light-dependent membrane sequestration of ECF sigma factor CarQ by anti-sigma factor CarR. *Mol Microbiol*, **19**, 171-186.

- Gottschalk, S., Bygebjerg-Hove, I., Bonde, M., Nielsen, P.K., Nguyen, T.H., Gravesen, A. and Kallipolitis, B.H. (2008) The two-component system CesRK controls the transcriptional induction of cell envelope-related genes in *Listeria monocytogenes* in response to cell wall-acting antibiotics. *J Bacteriol*, **190**, 4772-4776.
- Griffitts, J.S., Carlyon, R.E., Erickson, J.H., Moulton, J.L., Barnett, M.J., Toman, C.J. and Long, S.R. (2008) A *Sinorhizobium meliloti* osmosensory two-component system required for cyclic glucan export and symbiosis. *Mol Microbiol*, **69**, 479-490.
- Groom, M.J., Gray, E.M. and Townsend, P.A. (2008) Biofuels and biodiversity: principles for creating better policies for biofuel production. *Conserv Biol*, **22**, 602-609.
- Gross, C.A., Chan, C., Dombroski, A., Gruber, T., Sharp, M., Tupy, J. and Young, B. (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harb Symp Quant Biol*, **63**, 141-155.
- Guarner, F. and Malagelada, J.R. (2003) Gut flora in health and disease. *Lancet*, **361**, 512-519.
- Guarner, F. and Schaafsma, G.J. (1998) Probiotics. *Int J Food Microbiol*, **39**, 237-238.
- Guisez, Y., Fache, I., Campfield, L.A., Smith, F.J., Farid, A., Plaetinck, G., Van der Heyden, J., Tavernier, J., Fiers, W., Burn, P. and Devos, R. (1998) Efficient secretion of biologically active recombinant OB protein (leptin) in *Escherichia coli*, purification from the periplasm and characterization. *Protein Expr Purif*, **12**, 249-258.
- Gunsekere, I.C., Kahler, C.M., Ryan, C.S., Snyder, L.A., Saunders, N.J., Rood, J.I. and Davies, J.K. (2006) Ecf, an alternative sigma factor from *Neisseria gonorrhoeae*, controls expression of *msrAB*, which encodes methionine sulfoxide reductase. *J Bacteriol*, **188**, 3463-3469.
- Hagemann, T.L. and Kwan, S.P. (1999) ABI sequencing analysis. Manipulation of sequence data from the ABI DNA sequencer. *Mol Biotechnol*, **13**, 137-152.
- Hanlon, D.W. and Ordal, G.W. (1994) Cloning and characterization of genes encoding methyl-accepting chemotaxis proteins in *Bacillus subtilis*. *J Biol Chem*, **269**, 14038-14046.
- Hansen, P. and Lindeberg, G. (1995) Importance of the alpha-amino group in the selective purification of synthetic histidine peptides by immobilised metal ion affinity chromatography. *J Chromatogr A*, **690**, 155-159.
- Hardy, M.R., Townsend, R.R. and Lee, Y.C. (1988) Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection. *Anal Biochem*, **170**, 54-62.
- Heller, K. and Kadner, R.J. (1985) Nucleotide sequence of the gene for the vitamin B12 receptor protein in the outer membrane of *Escherichia coli*. *J Bacteriol*, **161**, 904-908.
- Helmann, J.D. (2002) *The extracytoplasmic function (ECF) sigma factors*.
- Helmann, J.D. and Chamberlin, M.J. (1988) Structure and function of bacterial sigma factors. *Annu Rev Biochem*, **57**, 839-872.

- Henrissat, B. and Bairoch, A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J*, **293 (Pt 3)**, 781-788.
- Henry, J.B. (2001) *Clinical diagnosis and Management by Laboratory Methods* Saunders, Philadelphia, PA,.
- Hettwer, U., Gross, M. and Rudolph, K. (1995) Purification and characterization of an extracellular levansucrase from *Pseudomonas syringae* pv. phaseolicola. *J Bacteriol*, **177**, 2834-2839.
- Hill, M.J. (1991) Bacterial N-nitrosation and gastric carcinogenesis in humans. *Ital J Gastroenterol*, **23**, 17-23.
- Hirschman, S.Z. and Felsenfeld, G. (1966) Determination of DNA composition and concentration by spectral analysis. *J Mol Biol*, **16**, 347-358.
- Ho, Y.S., Burden, L.M. and Hurley, J.H. (2000) Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *Embo J*, **19**, 5288-5299.
- Holm, L. and Park, J. (2000) DaliLite workbench for protein structure comparison. *Bioinformatics*, **16**, 566-567.
- Hooper, L.V. and Gordon, J.I. (2001) Commensal host-bacterial relationships in the gut. *Science*, **292**, 1115-1118.
- Hooper, L.V., Midtvedt, T. and Gordon, J.I. (2002) How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr*, **22**, 283-307.
- Hooper, L.V., Stappenbeck, T.S., Hong, C.V. and Gordon, J.I. (2003) Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol*, **4**, 269-273.
- Hooper, L.V., Wong, M.H., Thelin, A., Hansson, L., Falk, P.G. and Gordon, J.I. (2001) Molecular analysis of commensal host-microbial relationships in the intestine. *Science*, **291**, 881-884.
- Hooper, L.V., Xu, J., Falk, P.G., Midtvedt, T. and Gordon, J.I. (1999) A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A*, **96**, 9833-9838.
- Horie, H., Kanazawa, K., Okada, M., Narushima, S., Itoh, K. and Terada, A. (1999) Effects of intestinal bacteria on the development of colonic neoplasm: an experimental study. *Eur J Cancer Prev*, **8**, 237-245.
- Horsburgh, M.J. and Moir, A. (1999) Sigma M, an ECF RNA polymerase sigma factor of *Bacillus subtilis* 168, is essential for growth and survival in high concentrations of salt. *Mol Microbiol*, **32**, 41-50.
- Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y. and Schomburg, D. (2003) Crystal structure and snapshots along the reaction pathway of a family 51 alpha-L-arabinofuranosidase. *Embo J*, **22**, 4922-4932.
- Huang, X. and Helmann, J.D. (1998) Identification of target promoters for the *Bacillus subtilis* sigma X factor using a consensus-directed search. *J Mol Biol*, **279**, 165-173.
- Hughes, K.T. and Mathee, K. (1998) *The anti-sigma factors*.

- Hughes, R., Cross, A.J., Pollock, J.R. and Bingham, S. (2001) Dose-dependent effect of dietary meat on endogenous colonic N-nitrosation. *Carcinogenesis*, **22**, 199-202.
- Hulko, M., Berndt, F., Gruber, M., Linder, J.U., Truffault, V., Schultz, A., Martin, J., Schultz, J.E., Lupas, A.N. and Coles, M. (2006) The HAMP domain structure implies helix rotation in transmembrane signaling. *Cell*, **126**, 929-940.
- Husebye, E. (2005) The pathogenesis of gastrointestinal bacterial overgrowth. *Chemotherapy*, **51 Suppl 1**, 1-22.
- Inouye, M. (2006) Signaling by transmembrane proteins shifts gears. *Cell*, **126**, 829-831.
- Isolauri, E., Salminen, S. and Ouwehand, A.C. (2004) Microbial-gut interactions in health and disease. Probiotics. *Best Pract Res Clin Gastroenterol*, **18**, 299-313.
- Jaenicke, R. and Rudolph, R. (1986) Refolding and association of oligomeric proteins. *Methods Enzymol*, **131**, 218-250.
- Janausch, I.G., Zientz, E., Tran, Q.H., Kroger, A. and Uden, G. (2002) C4-dicarboxylate carriers and sensors in bacteria. *Biochim Biophys Acta*, **1553**, 39-56.
- Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G. and Gibson, T.J. (1998) Multiple sequence alignment with Clustal X. *Trends Biochem Sci*, **23**, 403-405.
- Joseleau-Petit, D., Kepes, F., Peutat, L., D'Ari, R. and Kepes, A. (1987) DNA replication initiation, doubling of rate of phospholipid synthesis, and cell division in *Escherichia coli*. *J Bacteriol*, **169**, 3701-3706.
- Kang, E.Y., Coleman, R.D., Pownall, H.J., Gotto, A.M., Jr. and Yang, C.Y. (1990) Analysis of the carbohydrate composition of glycoproteins by high-performance liquid chromatography. *J Protein Chem*, **9**, 31-35.
- Karpowich, N.K., Huang, H.H., Smith, P.C. and Hunt, J.F. (2003) Crystal structures of the BtuF periplasmic-binding protein for vitamin B12 suggest a functionally important reduction in protein mobility upon ligand binding. *J Biol Chem*, **278**, 8429-8434.
- Kay, C.W., Mennenga, B., Gorisch, H. and Bittl, R. (2004) Characterisation of the PQQ cofactor radical in quinoprotein ethanol dehydrogenase of *Pseudomonas aeruginosa* by electron paramagnetic resonance spectroscopy. *FEBS Lett*, **564**, 69-72.
- Kelly, S.M., Jess, T.J. and Price, N.C. (2005) How to study proteins by circular dichroism. *Biochim Biophys Acta*, **1751**, 119-139.
- Khachatourians, G.G. and Huzyk, L. (1974) Relationship between nucleoside triphosphate pools and DNA replication in the cell cycle of *Escherichia coli*. *Can J Microbiol*, **20**, 747-750.
- Kim, I., Stiefel, A., Plantor, S., Angerer, A. and Braun, V. (1997) Transcription induction of the ferric citrate transport genes via the N-terminus of the FecA outer membrane protein, the Ton system and the electrochemical potential of the cytoplasmic membrane. *Mol Microbiol*, **23**, 333-344.

- Kim, K., Oh, J., Han, D., Kim, E.E., Lee, B. and Kim, Y. (2006) Crystal structure of PilF: functional implication in the type 4 pilus biogenesis in *Pseudomonas aeruginosa*. *Biochem Biophys Res Commun*, **340**, 1028-1038.
- Kleinschmidt, J.H. (2005) *Folding and Stability of monomeric β -barrel Membrane proteins*, In: *Protein-Lipid Interactions: From Membrane Domains to Cellular Networks*. Ed. Tamm, L.K. Wiley-VCH Weinheim.
- Koebnik, R. (2005) TonB-dependent trans-envelope signalling: the exception or the rule? *Trends Microbiol*, **13**, 343-347.
- Koebnik, R., Locher, K.P. and Van Gelder, P. (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol Microbiol*, **37**, 239-253.
- Kolida, S. and Gibson, G.R. (2007) Prebiotic capacity of inulin-type fructans. *J Nutr*, **137**, 2503S-2506S.
- Koropatkin, N., Martens, E.C., Gordon, J.I. and Smith, T.J. (2009) Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont. *Biochemistry*, **48**, 1532-1542.
- Koropatkin, N.M., Martens, E.C., Gordon, J.I. and Smith, T.J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure*, **16**, 1105-1115.
- Krinos, C.M., Coyne, M.J., Weinacht, K.G., Tzianabos, A.O., Kasper, D.L. and Comstock, L.E. (2001) Extensive surface diversity of a commensal microorganism by multiple DNA inversions. *Nature*, **414**, 555-558.
- Krzeslak, J., Gerritse, G., van Merkerk, R., Cool, R.H. and Quax, W.J. (2008) Lipase expression in *Pseudomonas alcaligenes* is under the control of a two-component regulatory system. *Appl Environ Microbiol*, **74**, 1402-1411.
- Kunkel, T.A. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc Natl Acad Sci U S A*, **82**, 488-492.
- Kyriakidis, D.A., Theodorou, M.C., Filippou, P.S., Kyriakidis, K.D. and Tiligada, E. (2008) Effect of histamine on the signal transduction of the AtoS-AtoC two component system and involvement in poly-(R)-3-hydroxybutyrate biosynthesis in *Escherichia coli*. *Amino Acids*, **35**, 45-52.
- Labayen, I., Forga, L., Gonzalez, A., Lenoir-Wijnkoop, I., Nutr, R. and Martinez, J.A. (2001) Relationship between lactose digestion, gastrointestinal transit time and symptoms in lactose malabsorbers after dairy consumption. *Aliment Pharmacol Ther*, **15**, 543-549.
- Leavitt, S. and Freire, E. (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr Opin Struct Biol*, **11**, 560-566.
- Lee, J., Tomchick, D.R., Brautigam, C.A., Machius, M., Kort, R., Hellingwerf, K.J. and Gardner, K.H. (2008a) Changes at the KinA PAS-A dimerization interface influence histidine kinase function. *Biochemistry*, **47**, 4051-4064.
- Lee, S.W., Jeong, K.S., Han, S.W., Lee, S.E., Phee, B.K., Hahn, T.R. and Ronald, P. (2008b) The *Xanthomonas oryzae* pv. *oryzae* PhoPQ two-component system is required for AvrXA21 activity, hrpG expression, and virulence. *J Bacteriol*, **190**, 2183-2197.

- Lee, Y.C. (1990) High-performance anion-exchange chromatography for carbohydrate analysis. *Anal Biochem*, **189**, 151-162.
- Li, M., Wang, C., Feng, Y., Pan, X., Cheng, G., Wang, J., Ge, J., Zheng, F., Cao, M., Dong, Y., Liu, D., Wang, J., Lin, Y., Du, H., Gao, G.F., Wang, X., Hu, F. and Tang, J. (2008) SalK/SalR, a two-component signal transduction system, is essential for full virulence of highly invasive *Streptococcus suis* serotype 2. *PLoS ONE*, **3**, e2080.
- Lievin, V., Peiffer, I., Hudault, S., Rochat, F., Brassart, D., Neeser, J.R. and Servin, A.L. (2000) Bifidobacterium strains from resident infant human gastrointestinal microflora exert antimicrobial activity. *Gut*, **47**, 646-652.
- Lohmander, L.S. (1986) Analysis by high-performance liquid chromatography of radioactively labeled carbohydrate components of proteoglycans. *Anal Biochem*, **154**, 75-84.
- Loomis, W.F., Shaulsky, G. and Wang, N. (1997) Histidine kinases in signal transduction pathways of eukaryotes. *J Cell Sci*, **110 (Pt 10)**, 1141-1145.
- Luckey, T.D. (1972) Introduction to intestinal microecology. *Am J Clin Nutr* **25**, 1292-1294.
- Machie, W. and Perlin.S.A. (1966) Pyranose-Furanose and Anomeric Equilibria Influence of Solvent and of Partial Methylation *Canadian Journal of Chemistry*, **44**, 2039-2049.
- Maeda, T., Wurgler-Murphy, S.M. and Saito, H. (1994) A two-component system that regulates an osmosensing MAP kinase cascade in yeast. *Nature*, **369**, 242-245.
- Magnus, K.B., Eric, C.M. and Jeffrey, I.G. (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**, 36269-36279.
- Majamaa, H., Isolauri, E., Saxelin, M. and Vesikari, T. (1995) Lactic acid bacteria in the treatment of acute rotavirus gastroenteritis. *J Pediatr Gastroenterol Nutr*, **20**, 333-338.
- Malhotra, A., Severinova, E. and Darst, S.A. (1996) Crystal structure of a sigma 70 subunit fragment from *E. coli* RNA polymerase. *Cell*, **87**, 127-136.
- Manganelli, R., Voskuil, M.I., Schoolnik, G.K. and Smith, I. (2001) The Mycobacterium tuberculosis ECF sigma factor sigmaE: role in global gene expression and survival in macrophages. *Mol Microbiol*, **41**, 423-437.
- Marina, A., Waldburger, C.D. and Hendrickson, W.A. (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. *Embo J*, **24**, 4247-4259.
- 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**, 447-457.
- Marutani, M., Taguchi, F., Ogawa, Y., Hossain, M.M., Inagaki, Y., Toyoda, K., Shiraishi, T. and Ichinose, Y. (2008) Gac two-component system in *Pseudomonas syringae* pv. *tabaci* is required for virulence but not for hypersensitive

- reaction. *Mol Genet Genomics*, **279**, 313-322.
- Marvin, J.S. and Hellinga, H.W. (2001) Manipulation of ligand binding affinity by exploitation of conformational coupling. *Nat Struct Biol*, **8**, 795-798.
- Mascher, T., Helmann, J.D. and Uden, G. (2006) Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiol Mol Biol Rev*, **70**, 910-938.
- Masuda, S., Murakami, K.S., Wang, S., Anders Olson, C., Donigian, J., Leon, F., Darst, S.A. and Campbell, E.A. (2004) Crystal structures of the ADP and ATP bound forms of the Bacillus anti-sigma factor SpoIIAB in complex with the anti-anti-sigma SpoIIAA. *J Mol Biol*, **340**, 941-956.
- Matitashvili, E. and Zavizion, B. (1997) One-tube extraction of DNA or RNA from agarose gel. *Anal Biochem*, **246**, 260-262.
- Matsuda, K. (1957) Kojibiose (2-O-alpha-D-Glucopyranosyl-D-Glucose): Isolation and Structure: Chemical Synthesis. *Nature*, **180**, 985.
- Matthews, E.E., Zoonens, M. and Engelman, D.M. (2006) Dynamic helix interactions in transmembrane signaling. *Cell*, **127**, 447-450.
- Maunsell, B., Adams, C. and O'Gara, F. (2006) Complex regulation of AprA metalloprotease in *Pseudomonas fluorescens* M114: evidence for the involvement of iron, the ECF sigma factor, PbrA and pseudobactin M114 siderophore. *Microbiology*, **152**, 29-42.
- Mazmanian, S.K., Liu, C.H., Tzianabos, A.O. and Kasper, D.L. (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*, **122**, 107-118.
- Meng, G. and Futterer, K. (2003) Structural framework of fructosyl transfer in *Bacillus subtilis* levansucrase. *Nat Struct Biol*, **10**, 935-941.
- Metzger, H. (1992) Transmembrane signaling: the joy of aggregation. *J Immunol*, **149**, 1477-1487.
- Meyers, C.L. and Meyers, D.J. (2008) Thin-layer chromatography. *Curr Protoc Nucleic Acid Chem*, **Appendix**, 3D.
- Miasnikov, A.N. (1997) Characterization of a novel endo-levanase and its gene from *Bacillus* sp. L7. *FEMS Microbiol Lett*, **154**, 23-28.
- Miller, G.L. (1959) The use of dinitrosalicylic acid reagent for the determination of reducing sugar. *Analytical Chemistry*, **31**, 426-428.
- Miller, S.T., Xavier, K.B., Campagna, S.R., Taga, M.E., Semmelhack, M.F., Bassler, B.L. and Hughson, F.M. (2004) *Salmonella typhimurium* recognizes a chemically distinct form of the bacterial quorum-sensing signal AI-2. *Mol Cell*, **15**, 677-687.
- Miyatake, H., Mukai, M., Park, S.Y., Adachi, S., Tamura, K., Nakamura, H., Nakamura, K., Tsuchiya, T., Iizuka, T. and

- Shiro, Y. (2000) Sensory mechanism of oxygen sensor FixL from *Rhizobium meliloti*: crystallographic, mutagenesis and resonance Raman spectroscopic studies. *J Mol Biol*, **301**, 415-431.
- 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. *J Bacteriol*, **185**, 2219-2226.
- Mooney, R.A., Darst, S.A. and Landick, R. (2005) Sigma and RNA polymerase: an on-again, off-again relationship? *Mol Cell*, **20**, 335-345.
- Moore, D.T., Berger, B.W. and DeGrado, W.F. (2008) Protein-protein interactions in the membrane: sequence, structural, and biological motifs. *Structure*, **16**, 991-1001.
- Moore, W.E. and Moore, L.H. (1995) Intestinal floras of populations that have a high risk of colon cancer. *Appl Environ Microbiol*, **61**, 3202-3207.
- Murakami, K.S., Masuda, S., Campbell, E.A., Muzzin, O. and Darst, S.A. (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*, **296**, 1285-1290.
- Nagem, R.A., Rojas, A.L., Golubev, A.M., Korneeva, O.S., Eneyskaya, E.V., Kulminkaya, A.A., Neustroev, K.N. and Polikarpov, I. (2004) Crystal structure of exo-inulinase from *Aspergillus awamori*: the enzyme fold and structural determinants of substrate recognition. *J Mol Biol*, **344**, 471-480.
- Nau, C.D. and Konisky, J. (1989) Evolutionary relationship between the TonB-dependent outer membrane transport proteins: nucleotide and amino acid sequences of the *Escherichia coli* colicin I receptor gene. *J Bacteriol*, **171**, 1041-1047.
- Nes, I.F. and Eijsink, V.G.H. (1999) Regulation of group II peptide bacteriocin synthesis by quorum-sensing mechanisms. In *Cell-cell signaling in bacteria* Winans, G.M.D.a.S.C. (ed.). American Society for Microbiology Press, Washington, D.C., pp. 175-192.
- Newman, J.D., Falkowski, M.J., Schilke, B.A., Anthony, L.C. and Donohue, T.J. (1999) The *Rhodobacter sphaeroides* ECF sigma factor, sigma(E), and the target promoters *cycA* P3 and *rpoE* P1. *J Mol Biol*, **294**, 307-320.
- Nishiura, T., Fujii, S., Kanayama, Y., Nishikawa, A., Tomiyama, Y., Iida, M., Karasuno, T., Nakao, H., Yonezawa, T., Taniguchi, N. and et al. (1990) Carbohydrate analysis of immunoglobulin G myeloma proteins by lectin and high performance liquid chromatography: role of glycosyltransferases in the structures. *Cancer Res*, **50**, 5345-5350.
- Noverr, M.C., Noggle, R.M., Toews, G.B. and Huffnagle, G.B. (2004) Role of antibiotics and fungal microbiota in driving pulmonary allergic responses. *Infect Immun*, **72**, 4996-5003.
- Oberhelman, R.A., Gilman, R.H., Sheen, P., Taylor, D.N., Black, R.E., Cabrera, L., Lescano, A.G., Meza, R. and Madico, G. (1999) A placebo-controlled trial of *Lactobacillus GG* to prevent diarrhea in undernourished Peruvian children. *J Pediatr*, **134**, 15-20.

- Ochs, M., Veitinger, S., Kim, I., Welz, D., Angerer, A. and Braun, V. (1995) Regulation of citrate-dependent iron transport of *Escherichia coli*: *fecR* is required for transcription activation by *FecI*. *Mol Microbiol*, **15**, 119-132.
- Osipiuk, J., Wu, R., Jedrzejczak, R., Joachimiak, A., . (2008) Crystal structure of beta-xylosidase, family 43 glycosyl hydrolase from *Clostridium acetobutylicum*
- Ottemann, K.M., Xiao, W., Shin, Y.K. and Koshland, D.E., Jr. (1999) A piston model for transmembrane signaling of the aspartate receptor. *Science*, **285**, 1751-1754.
- Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode: Macromolecular crystallography Part A. *Methods Enzymol*, **276**, 307-326.
- Oubrie, A., Rozeboom, H.J., Kalk, K.H., Huizinga, E.G. and Dijkstra, B.W. (2002) Crystal structure of quinoxinoprotein alcohol dehydrogenase from *Comamonas testosteroni*: structural basis for substrate oxidation and electron transfer. *J Biol Chem*, **277**, 3727-3732.
- Paget, M.S. and Helmann, J.D. (2003) The sigma70 family of sigma factors. *Genome Biol*, **4**, 203.
- Pappalardo, L., Janausch, I.G., Vijayan, V., Zientz, E., Junker, J., Peti, W., Zweckstetter, M., Unden, G. and Griesinger, C. (2003) The NMR structure of the sensory domain of the membranous two-component fumarate sensor (histidine protein kinase) DcuS of *Escherichia coli*. *J Biol Chem*, **278**, 39185-39188.
- Parente, J.P., Leroy, Y., Montreuil, J. and Fournet, B. (1984) Separation of sialyl-oligosaccharides by high-performance liquid chromatography. Application to analysis of carbohydrate units of acidic oligosaccharides obtained by hydrazinolysis of hen ovomucoid. *J Chromatogr*, **288**, 147-155.
- Pell, G., Szabo, L., Charnock, S.J., Xie, H., Gloster, T.M., Davies, G.J. and Gilbert, H.J. (2004) Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. *J Biol Chem*, **279**, 11777-11788.
- Pereira, C.S., McAuley, J.R., Taga, M.E., Xavier, K.B. and Miller, S.T. (2008) *Sinorhizobium meliloti*, a bacterium lacking the autoinducer-2 (AI-2) synthase, responds to AI-2 supplied by other bacteria. *Mol Microbiol*, **70**, 1223-1235.
- Perraud, A.L., Weiss, V. and Gross, R. (1999) Signalling pathways in two-component phosphorelay systems. *Trends Microbiol*, **7**, 115-120.
- Pilon-Smits, E., Ebskamp, M., Paul, M.J., Jeuken, M., Weisbeek, P.J. and Smeekens, S. (1995) Improved Performance of Transgenic Fructan-Accumulating Tobacco under Drought Stress. *Plant Physiol*, **107**, 125-130.
- Pinhassi, J., Bowman, J.P., Nedashkovskaya, O.I., Lekunberri, I., Gomez-Consarnau, L. and Pedros-Alio, C. (2006) *Leeuwenhoekella blandensis* sp. nov., a genome-sequenced marine member of the family Flavobacteriaceae. *Int J Syst Evol Microbiol*, **56**, 1489-1493.
- Pons, T., Gomez, R., China, G. and Valencia, A. (2003) Beta-propellers: associated functions and their role in human diseases. *Curr Med Chem*, **10**, 505-524.

- Prins, R.A. (1977) Biochemical activities of gut microorganisms, p. 73-183. In R T J Clarke and T Bauchop (ed), *Microbial ecology of the gut Academic Press, New York*.
- Project, C.C. (1994) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.
- Putrins, M., Ilves, H., Kivisaar, M. and Horak, R. (2008) ColRS two-component system prevents lysis of subpopulation of glucose-grown *Pseudomonas putida*. *Environ Microbiol*.
- Quioco, F.A., Phillips, G.N., Jr., Parsons, R.G. and Hogg, R.W. (1974) Letter: Crystallographic data of an L-arabinose-binding protein from *Escherichia coli*. *J Mol Biol*, **86**, 491-493.
- Rajagopal, S. and Moffat, K. (2003) Crystal structure of a photoactive yellow protein from a sensor histidine kinase: conformational variability and signal transduction. *Proc Natl Acad Sci U S A*, **100**, 1649-1654.
- Rastall, R.A. (2004) Bacteria in the gut: friends and foes and how to alter the balance. *J Nutr*, **134**, 2022S-2026S.
- Reeves, A.R., D'Elia, J.N., Frias, J. and Salyers, A.A. (1996) A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol*, **178**, 823-830.
- 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*. *J Bacteriol*, **179**, 643-649.
- Reid, C.J. and Poole, P.S. (1998) Roles of DctA and DctB in signal detection by the dicarboxylic acid transport system of *Rhizobium leguminosarum*. *J Bacteriol*, **180**, 2660-2669.
- Reinelt, S., Hofmann, E., Gerharz, T., Bott, M. and Madden, D.R. (2003) The structure of the periplasmic ligand-binding domain of the sensor kinase CitA reveals the first extracellular PAS domain. *J Biol Chem*, **278**, 39189-39196.
- Ritsema, T. and Smeekens, S. (2003) Fructans: beneficial for plants and humans. *Curr Opin Plant Biol*, **6**, 223-230.
- Roberfroid, M.B. (2005) Introducing inulin-type fructans. *Br J Nutr*, **93 Suppl 1**, S13-25.
- Roberfroid, M.B., Bornet, F., Bouley, C. and Cummings, J.H. (1995) Colonic microflora: nutrition and health. Summary and conclusions of an International Life Sciences Institute (ILSI) [Europe] workshop held in Barcelona, Spain. *Nutr Rev*, **53**, 127-130.
- Robinson, V.L., Buckler, D.R. and Stock, A.M. (2000) A tale of two components: a novel kinase and a regulatory switch. *Nat Struct Biol*, **7**, 626-633.
- Roman-Leshkov, Y., Barrett, C.J., Liu, Z.Y. and Dumesic, J.A. (2007) Production of dimethylfuran for liquid fuels from biomass-derived carbohydrates. *Nature*, **447**, 982-985.
- Rouvinen, J., Bergfors, T., Teeri, T., Knowles, J.K. and Jones, T.A. (1990) Three-dimensional structure of

cellobiohydrolase II from *Trichoderma reesei*. *Science*, **249**, 380-386.

Rowen, D.W. and Deretic, V. (2000) Membrane-to-cytosol redistribution of ECF sigma factor AlgU and conversion to mucoidy in *Pseudomonas aeruginosa* isolates from cystic fibrosis patients. *Mol Microbiol*, **36**, 314-327.

Rozen, R., Bachrach, G., Bronshteyn, M., Gedalia, I. and Steinberg, D. (2001) The role of fructans on dental biofilm formation by *Streptococcus sobrinus*, *Streptococcus mutans*, *Streptococcus gordonii* and *Actinomyces viscosus*. *FEMS Microbiol Lett*, **195**, 205-210.

Rutz, J.M., Liu, J., Lyons, J.A., Goranson, J., Armstrong, S.K., McIntosh, M.A., Feix, J.B. and Klebba, P.E. (1992) Formation of a gated channel by a ligand-specific transport protein in the bacterial outer membrane. *Science*, **258**, 471-475.

Ryu, H.B., Shin, I., Yim, H.S. and Kang, S.O. (2006) YlaC is an extracytoplasmic function (ECF) sigma factor contributing to hydrogen peroxide resistance in *Bacillus subtilis*. *J Microbiol*, **44**, 206-216.

Saier, M.H., Jr. (1993) Introduction: protein phosphorylation and signal transduction in bacteria. *J Cell Biochem*, **51**, 1-6.

Salyers, A.A. and Leedle, J.A.Z. (1983) Carbohydrate metabolism in the human colon, p. 129-146. In D J Hentges (ed), *Human intestinal microflora in health and disease Academic Press, New York*.

Sato, Y., Ito, T., Udaka, N., Kanisawa, M., Noguchi, Y., Cushman, S.W. and Satoh, S. (1996) Immunohistochemical localization of facilitated-diffusion glucose transporters in rat pancreatic islets. *Tissue Cell*, **28**, 637-643.

Savage, D.C. (1977) *Microbial ecology of the gastrointestinal tract*.

Scupham, A.J., Presley, L.L., Wei, B., Bent, E., Griffith, N., McPherson, M., Zhu, F., Oluwadara, O., Rao, N., Braun, J. and Borneman, J. (2006) Abundant and diverse fungal microbiota in the murine intestine. *Appl Environ Microbiol*, **72**, 793-801.

Sebert, M.E., Patel, K.P., Plotnick, M. and Weiser, J.N. (2005) Pneumococcal HtrA protease mediates inhibition of competence by the CiaRH two-component signaling system. *J Bacteriol*, **187**, 3969-3979.

Sevvana, M., Vijayan, V., Zweckstetter, M., Reinelt, S., Madden, D.R., Herbst-Irmer, R., Sheldrick, G.M., Bott, M., Griesinger, C. and Becker, S. (2008) A ligand-induced switch in the periplasmic domain of sensor histidine kinase CitA. *J Mol Biol*, **377**, 512-523.

Shanks, R.M., Meehl, M.A., Brothers, K.M., Martinez, R.M., Donegan, N.P., Graber, M.L., Cheung, A.L. and O'Toole, G.A. (2008) Genetic evidence for an alternative citrate-dependent biofilm formation pathway in *Staphylococcus aureus* that is dependent on fibronectin binding proteins and the GraRS two-component regulatory system. *Infect Immun*, **76**, 2469-2477.

Shimoda, N., Toyoda-Yamamoto, A., Aoki, S. and Machida, Y. (1993) Genetic evidence for an interaction between the VirA sensor protein and the ChvE sugar-binding protein of *Agrobacterium*. *J Biol Chem*, **268**, 26552-26558.

Shipman, J.A., Berleman, J.E. and Salyers, A.A. (2000) Characterization of four outer membrane proteins involved in

- binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol*, **182**, 5365-5372.
- Shipman, J.A., Cho, K.H., Siegel, H.A. and Salyers, A.A. (1999) Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol*, **181**, 7206-7211.
- Silvester, K.R., Englyst, H.N. and Cummings, J.H. (1995) Ileal recovery of starch from whole diets containing resistant starch measured in vitro and fermentation of ileal effluent. *Am J Clin Nutr*, **62**, 403-411.
- Siryaporn, A. and Goulian, M. (2008) Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in *E. coli*. *Mol Microbiol*, **70**, 494-506.
- Skidmore, W.D. and Duggan, E.L. (1966) Simultaneous spectral determination of both base composition and concentration of DNA. *Anal Biochem*, **14**, 223-236.
- Smith, E.A. and Macfarlane, G.T. (1996) Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. *J Appl Bacteriol*, **81**, 288-302.
- Soderback, E., Reyes-Ramirez, F., Eydmann, T., Austin, S., Hill, S. and Dixon, R. (1998) The redox- and fixed nitrogen-responsive regulatory protein NIFL from *Azotobacter vinelandii* comprises discrete flavin and nucleotide-binding domains. *Mol Microbiol*, **28**, 179-192.
- Sonnenburg, E.D., Sonnenburg, J.L., Manchester, J.K., Hansen, E.E., Chiang, H.C. and Gordon, J.I. (2006) A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A*, **103**, 8834-8839.
- Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.H., Westover, B.P., Weatherford, J., Buhler, J.D. and Gordon, J.I. (2005) Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science*, **307**, 1955-1959.
- Sorenson, M.K., Ray, S.S. and Darst, S.A. (2004) Crystal structure of the flagellar sigma/anti-sigma complex sigma(28)/FlgM reveals an intact sigma factor in an inactive conformation. *Mol Cell*, **14**, 127-138.
- Spollen, W.G. and Nelson, C.J. (1994) Response of Fructan to Water Deficit in Growing Leaves of Tall Fescue. *Plant Physiol*, **106**, 329-336.
- Stock, A.M., Mottonen, J.M., Stock, J.B. and Schutt, C.E. (1989) Three-dimensional structure of CheY, the response regulator of bacterial chemotaxis. *Nature*, **337**, 745-749.
- Stock, A.M., Robinson, V.L. and Goudreau, P.N. (2000) Two-component signal transduction. *Annu Rev Biochem*, **69**, 183-215.
- Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol*, **189**, 113-130.
- Sturm, A. and Tang, G.Q. (1999) The sucrose-cleaving enzymes of plants are crucial for development, growth and carbon partitioning. *Trends Plant Sci*, **4**, 401-407.

- Sugimoto, M., Esaki, N., Tanaka, H. and Soda, K. (1989) A simple and efficient method for the oligonucleotide-directed mutagenesis using plasmid DNA template and phosphorothioate-modified nucleotide. *Anal Biochem*, **179**, 309-311.
- Sun, R., Converse, P.J., Ko, C., Tyagi, S., Morrison, N.E. and Bishai, W.R. (2004) Mycobacterium tuberculosis ECF sigma factor sigC is required for lethality in mice and for the conditional expression of a defined gene set. *Mol Microbiol*, **52**, 25-38.
- Szurmant, H., Bu, L., Brooks, C.L., 3rd and Hoch, J.A. (2008) An essential sensor histidine kinase controlled by transmembrane helix interactions with its auxiliary proteins. *Proc Natl Acad Sci U S A*, **105**, 5891-5896.
- Takeo, K. (1984) Affinity electrophoresis: principles and applications. *Electrophoresis*, **5**, 187-195.
- Takeuchi, M., Takasaki, S., Inoue, N. and Kobata, A. (1987) Sensitive method for carbohydrate composition analysis of glycoproteins by high-performance liquid chromatography. *J Chromatogr*, **400**, 207-213.
- Tam, R. and Saier, M.H., Jr. (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev*, **57**, 320-346.
- Taylor, B.L. and Zhulin, I.B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol Mol Biol Rev*, **63**, 479-506.
- Taylor, J.W., Ott, J. and Eckstein, F. (1985) The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. *Nucleic Acids Res*, **13**, 8765-8785.
- Tepljakov, A., Fedorov, E., Gilliland, G.L., Almo, S.C., Burley, S.K., . (2005) Beta-d-xylosidase (selenomethionine) XYND from *Clostridium Acetobutylicum*
- Thomas, J.A. and Valvano, M.A. (1992) tolQ is required for cloacin DF13 susceptibility in *Escherichia coli* expressing the aerobactin/cloacin DF13 receptor IutA. *FEMS Microbiol Lett*, **70**, 107-111.
- Tian, Y., Cuneo, M.J., Changela, A., Hocker, B., Beese, L.S. and Hellinga, H.W. (2007) Structure-based design of robust glucose biosensors using a *Thermotoga maritima* periplasmic glucose-binding protein. *Protein Sci*, **16**, 2240-2250.
- Tie-xin, T. and Hong, W. (2008) An image analysis system for thin-layer chromatography quantification and its validation. *J Chromatogr Sci*, **46**, 560-564.
- Townsend, R.R. and Hardy, M.R. (1991) Analysis of glycoprotein oligosaccharides using high-pH anion exchange chromatography. *Glycobiology*, **1**, 139-147.
- Townsend, R.R., Hardy, M.R., Hindsgaul, O. and Lee, Y.C. (1988) High-performance anion-exchange chromatography of oligosaccharides using pellicular resins and pulsed amperometric detection. *Anal Biochem*, **174**, 459-470.
- Tu, X., Latifi, T., Bougdour, A., Gottesman, S. and Groisman, E.A. (2006) The PhoP/PhoQ two-component system

- stabilizes the alternative sigma factor RpoS in *Salmonella enterica*. *Proc Natl Acad Sci U S A*, **103**, 13503-13508.
- Turnbaugh, P.J., Ley, R.E., Mahowald, M.A., Magrini, V., Mardis, E.R. and Gordon, J.I. (2006) An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature*, **444**, 1027-1031.
- van der Waaij, L.A., Limburg, P.C., Mesander, G. and van der Waaij, D. (1996) In vivo IgA coating of anaerobic bacteria in human faeces. *Gut*, **38**, 348-354.
- VanDamme, E.J. and D.G.Derycke. (1983) Microbial inulinases. Fermentation process, properties and applications. *Adv Appl Microbiol*, **29**, 139-176.
- Vassylyev, D.G., Sekine, S., Laptenko, O., Lee, J., Vassylyeva, M.N., Borukhov, S. and Yokoyama, S. (2002) Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature*, **417**, 712-719.
- Vercellotti, J.R., Salyers, A.A., Bullard, W.S. and Wilkins, D. (1977) Breakdown of mucin and plant polysaccharides in the human colon. *Can J Biochem*, **55**, 1190-1196.
- Verhaest, M., Ende, W.V., Roy, K.L., De Ranter, C.J., Laere, A.V. and Rabijns, A. (2005) X-ray diffraction structure of a plant glycosyl hydrolase family 32 protein: fructan 1-exohydrolase IIa of *Cichorium intybus*. *Plant J*, **41**, 400-411.
- Verhaest, M., Lammens, W., Le Roy, K., De Coninck, B., De Ranter, C.J., Van Laere, A., Van den Ende, W. and Rabijns, A. (2006) X-ray diffraction structure of a cell-wall invertase from *Arabidopsis thaliana*. *Acta Crystallogr D Biol Crystallogr*, **62**, 1555-1563.
- Verhaest, M., Lammens, W., Le Roy, K., De Ranter, C.J., Van Laere, A., Rabijns, A. and Van den Ende, W. (2007a) Insights into the fine architecture of the active site of chicory fructan 1-exohydrolase: 1-kestose as substrate vs sucrose as inhibitor. *New Phytol*, **174**, 90-100.
- Verhaest, M., Le Roy, K., De Ranter, C.J., Van Laere, A., Van den Ende, W. and Rabijns, A. (2007b) Crystal structure of fructan 1-exohydrolase IIa from *Cichorium intybus* in complex with sucrose. *New Phytol*, **174**, 90-100.
- Wack, M. and Blaschek, W. (2006) Determination of the structure and degree of polymerisation of fructans from *Echinacea purpurea* roots. *Carbohydr Res*, **341**, 1147-1153.
- Waksmundzka-Hajnos, M. (2008) Thin-layer chromatography. *J Chromatogr Sci*, **46**, 289-290.
- Walzer, G., Rosenberg, E. and Ron, E.Z. (2008) Identification of outer membrane proteins with emulsifying activity by prediction of beta-barrel regions. *J Microbiol Methods*.
- Watts, K.J., Sommer, K., Fry, S.L., Johnson, M.S. and Taylor, B.L. (2006) Function of the N-terminal cap of the PAS domain in signaling by the aerotaxis receptor Aer. *J Bacteriol*, **188**, 2154-2162.
- Weinrauch, Y., Penchev, R., Dubnau, E., Smith, I. and Dubnau, D. (1990) A *Bacillus subtilis* regulatory gene product for genetic competence and sporulation resembles sensor protein members of the bacterial two-component signal-

- transduction systems. *Genes Dev*, **4**, 860-872.
- West, A.H. and Stock, A.M. (2001) Histidine kinases and response regulator proteins in two-component signaling systems. *Trends Biochem Sci*, **26**, 369-376.
- Whitworth, D.E., Millard, A., Hodgson, D.A. and Hawkins, P.F. (2008) Protein-protein interactions between two-component system transmitter and receiver domains of *Myxococcus xanthus*. *Proteomics*, **8**, 1839-1842.
- Williamson, N.R., Fineran, P.C., Ogawa, W., Woodley, L.R. and Salmond, G.P. (2008) Integrated regulation involving quorum sensing, a two-component system, a GGDEF/EAL domain protein and a post-transcriptional regulator controls swarming and RhlA-dependent surfactant biosynthesis in *Serratia*. *Environ Microbiol*, **10**, 1202-1217.
- Wilson, M.J. and Lamont, I.L. (2000) Characterization of an ECF sigma factor protein from *Pseudomonas aeruginosa*. *Biochem Biophys Res Commun*, **273**, 578-583.
- Wiseman, T., Williston, S., Brandts, J.F. and Lin, L.N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem*, **179**, 131-137.
- Wollowski, I., Rechkemmer, G. and Pool-Zobel, B.L. (2001) Protective role of probiotics and prebiotics in colon cancer. *Am J Clin Nutr*, **73**, 451S-455S.
- Wriedt, K., Angerer, A. and Braun, V. (1995) Transcriptional regulation from the cell surface: conformational changes in the transmembrane protein FecR lead to altered transcription of the ferric citrate transport genes in *Escherichia coli*. *J Bacteriol*, **177**, 3320-3322.
- Wu, J., Ohta, N., Zhao, J.L. and Newton, A. (1999) A novel bacterial tyrosine kinase essential for cell division and differentiation. *Proc Natl Acad Sci U S A*, **96**, 13068-13073.
- Wurgler-Murphy, S.M. and Saito, H. (1997) Two-component signal transducers and MAPK cascades. *Trends Biochem Sci*, **22**, 172-176.
- Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V. and Gordon, J.I. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science*, **299**, 2074-2076.
- Xu, J., Chiang, H.C., Bjursell, M.K. and Gordon, J.I. (2004) Message from a human gut symbiont: sensitivity is a prerequisite for sharing. *Trends Microbiol*, **12**, 21-28.
- Xu, J. and Gordon, J.I. (2003) Inaugural Article: Honor thy symbionts. *Proc Natl Acad Sci U S A*, **100**, 10452-10459.
- Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R.S., Fulton, L.A., Clifton, S.W., Wilson, R.K., Knight, R.D. and Gordon, J.I. (2007) Evolution of Symbiotic Bacteria in the Distal Human Intestine. *PLoS Biol*, **5**, e156.
- Yan, Q., Wu, X.G., Wei, H.L., Wang, H.M. and Zhang, L.Q. (2008) Differential control of the PcoI/PcoR quorum-sensing system in *Pseudomonas fluorescens* 2P24 by sigma factor RpoS and the GacS/GacA two-component

regulatory system. *Microbiol Res.*

- Yeh, K.C., Wu, S.H., Murphy, J.T. and Lagarias, J.C. (1997) A cyanobacterial phytochrome two-component light sensory system. *Science*, **277**, 1505-1508.
- Yeoman, K.H., May, A.G., deLuca, N.G., Stuckey, D.B. and Johnston, A.W. (1999) A putative ECF sigma factor gene, rpol, regulates siderophore production in *Rhizobium leguminosarum*. *Mol Plant Microbe Interact*, **12**, 994-999.
- Yeoman, K.H., Mittelheiser, S., Sawers, G. and Johnston, A.W. (2003) The ECF sigma factor Rpol of *R. leguminosarum* initiates transcription of the vbsGSO and vbsADL siderophore biosynthetic genes in vitro. *FEMS Microbiol Lett*, **223**, 239-244.
- Younes, H., Coudray, C., Bellanger, J., Demigne, C., Rayssiguier, Y. and Remesy, C. (2001) Effects of two fermentable carbohydrates (inulin and resistant starch) and their combination on calcium and magnesium balance in rats. *Br J Nutr*, **86**, 479-485.
- Zhang, W. and Shi, L. (2005) Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. *Microbiology*, **151**, 2159-2173.
- Zhulin, I.B., Taylor, B.L. and Dixon, R. (1997) PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends Biochem Sci*, **22**, 331-333.
- Zoetendal, E.G., Akkermans, A.D. and De Vos, W.M. (1998) Temperature gradient gel electrophoresis analysis of 16S rRNA from human fecal samples reveals stable and host-specific communities of active bacteria. *Appl Environ Microbiol*, **64**, 3854-3859.
- Zoetendal, E.G., Smith, A.H., Sundset, M.A. and Mackie, R.I. (2008) The BaeSR two-component regulatory system mediates resistance to condensed tannins in *Escherichia coli*. *Appl Environ Microbiol*, **74**, 535-539.
- Zoetendal, E.G., Vaughan, E.E. and de Vos, W.M. (2006) A microbial world within us. *Mol Microbiol*, **59**, 1639-1650.
- Zou, J.Y., Flocco, M.M. and Mowbray, S.L. (1993) The 1.7 Å refined X-ray structure of the periplasmic glucose/galactose receptor from *Salmonella typhimurium*. *J Mol Biol*, **233**, 739-752.

Appendices

Appendix A: Chemicals, Media, and Enzymes used in this study

A1 Chemicals

Acrose Organics

- D-(+)-Turanose
- D-(-)-Tagatose
- L-(-)-Sorbose

Amersham-Boehringer Mannheim

- 2'-Deoxyadenosine 5'-triphosphate (dATP)
- 2'-Deoxycytidine 5'-triphosphate (dCTP)
- 2'-Deoxyguanosine 5'-triphosphate (dGTP)
- 2'-Deoxythymidine 5'-triphosphate (dTTP)

Bio-Rad

- Bradford's Reagent

BioGene

- Electrophoresis grade Agarose

British Drug Houses (BDH)

- Acetic acid (Glacial)
- Acrylamide solution (40% (w/v); Electran)
- Boric acid
- Bromophenol Blue
- Citric Acid
- Calcium chloride
- Chloroform
- Dimethylformamide
- Ethanol (industrial grade)
- Hydrochloric acid
- Isopropanol
- Magnesium chloride
- Magnesium sulphate
- Mineral oil
- Methanol
- Polyethelen glycol MW 8000 (PEG-8000)
- Polyethelen glycol MW 550 monomethyl ether (PEG550 mme)
- Potassium dihydrogen orthophosphate
- Sodium acetate
- Sodium chloride

Sodium dihydrogen orthophosphate
Sulphuric acid
Triton X-100
Xylene cyanol FF

Calbiochem

Deoxymannojirimycin
Lactulose

Fisons

46/48% (w/v) NaOH
Sodium acetate trihydrate

Hampton Research

Aqua Sil®

James Burrough (F.A.D.) Ltd

Ethanol

Megazyme

1-Kestose
1,1-Kestotetraose
1,1,1-Kestopentaose

Melford Laboratories

Isopropyl- β -D-thiogalactoside (IPTG)
Urea (ultrapure)
HEPES

Pharmacia

Adenosine 5'-triphosphate (ATP)
Agarose (ultrapure)

Roche

Complete Protease Inhibitor Cocktail Tablets
Complete Protease Inhibitor Cocktail Tablets, EDTA-free

Sigma

Ammonium persulphate
Ampicillin
Bis tris propane
Bovine serum albumin, fraction V (BSA)
N-Butyl-deoxymannojirimycin
Castanospermine
Chondroitin 4-sulfate sodium salt from bovine trachea
Chondroitin 6-sulfate sodium salt from shark cartilage

Coomassie Brilliant Blue G
Copper sulphate
DEAE-Trisacryl
Deoxynojirimycin
Digalacturonic acid
Dimethylsulphoxide
2, 4-Dinitrophenol
3, 5-Dinitrosalicylic acid (DNSA)
Dithiothrietol (DTT)
Ethidium bromide
Ethelene diamine tetra-aceticacid, disodium salt (EDTA)
Ethylene glycol
D-(-) Fructose
Galacturonic acid
Guanosine 5' diphosphoglucose
Gadolinium chloride
D-Glucose
Gluteraldehyde
Glyceraldehyde-3-phosphate
 Glycerol
 D-Glycerate
 Imidazole
 Inulin from chicory
 D-lactete
 Levan from *Erwinia herbicola*
 Levan from *Zymomonas mobilis*
 Lithium acetate
 Lithium chloride
 β -Mercaptoethanol
 2, 4-methane-pentenediol (MPD)
 Mineral oil (light)
 Nicotinamide adenine dinucleotide-reduced
 Polyethylene glycol MW 350
 Phenol
 Potassium acetate
 Potassium chloride
 Potassium thiocyanate
 D-(+)-Raffinose pentahydrate
 Sodium carbonate
 Sodium dodecyl sulphate (SDS)
Sucrose dihydrogen orthophosphate
Sodium fluoride
Sodium sulphate
di- Sodium hydrohen Phosphate
Sucrose (nuclease free)
N, N, N', N'-Tetramethylethylenediamine (TEMED)

Trichloroacetic acid (TCA)
Trigalacturonic acid
Trizma base (Tris)
D-Xylulose
Zinc sulphate

A2 Media

Difco

Bacto®tryptone
Bacto®yeast extract

Oxoid

Bacteriological Agar No.1

A3 Enzymes

NBI Fermentase

DNA restriction endonucleases

Gibco BRL

Bacteriophage T4 DNA ligase
DNA restriction endonucleases

Novagen

KOD HotStart DNA polymerase

Sigma

Lactate dehydrogenase (LDH)
Pyruvate Kinase
Protease K
Trypsin
 α -Chymotrypsin

Stratagene

Pfu Turbo DNA polymerase
DpnI restriction endonuclease

A4 DNA

BT genome

Prepared by labmates. Kept at 4 °C in EB buffer

MWG Biotech

All primers

A5 Kits

Invitrogen

Zero Blunt™ Cloning Kit

Qiagen

Plasmid mini kit

Eppendorf

Perfectprep® Gel cleanup kit

Stratagene

QuikChange™ Site-Directed Mutagenesis kit

Appendix B Crystallization Screens

B1 Newcastle Screen Composition Table

Number	Salt	Buffer	Precipitant
1	0.2M Lithium Sulphate	0.1M Sodium Acetate pH 5.1	50% PEG 400
2		0.1M Sodium Citrate pH 5.5	20% PEG 3000
3		0.2M Diammonium Hydrogen Citrate pH 5.0	20% PEG 3350
4	0.08M Calcium Chloride	0.1M Sodium Acetate pH 4.6	30% MPD
5		0.2M Magnesium Formate pH 5.9	20% PEG 3350
6	0.2M Lithium Sulphate	0.25M Sodium Dihydrogen Phosphate pH 4.2	20% PEG 1000
7	0.25M Sodium Citrate	0.1M CHES pH 9.5	20% PEG 8000
8		0.2M Ammonium Formate pH 6.6	20% PEG 3350
9		0.2M Ammonium Chloride pH 6.3	20% PEG 3350
10		0.2M Potassium Formate pH 7.3	20% PEG 3350
11	0.2M Ammonium Dihydrogen Phosphate	0.1M Tris pH 8.5	50% MPD
12		0.2M Potassium Nitrate pH 6.9	20% PEG 3350
13	0.8M Ammonium Sulphate	0.1M Citric Acid pH 4.0	
14		0.2M Sodium Thiocyanate pH 6.9	20% PEG 3350
15		0.1M Bicine pH 9.0	20% PEG 6000
16	8% Ethylene Glycol	0.1M HEPES pH 7.5	10% PEG 8000 40% MPD
17		0.1M Sodium Cacodylate pH 7.0	5% PEG 8000
18	0.25M Sodium Citrate	0.25M Sodium Dihydrogen Phosphate pH 5.2	5% PEG 1000 40% Ethanol
19		0.1M Sodium Acetate pH 4.6	8% PEG 4000
20	0.2M Magnesium Chloride	0.1M Tris pH 7.0	10% PEG 8000
21		0.1M Citric Acid pH 5.0	20% PEG 6000
22	0.2M Magnesium Chloride	0.1M Sodium Cacodylate pH 6.6	50% PEG 200
23		1.6M Sodium Citrate pH 6.5	
24		0.2M Potassium Citrate pH 8.3	20% PEG 3350
25	0.02M Calcium Chloride	0.1M Sodium Acetate pH 4.6	30% MPD
26	0.2M Sodium Chloride	0.25M Sodium Dihydrogen Phosphate pH 4.2	20% PEG 8000
27	0.25M Sodium Citrate	0.1M Citric Acid pH 4.0	20% PEG 6000
28	1.0M Lithium Chloride	0.2M Ammonium Nitrate pH 6.3	20% PEG 3350
29		0.1M HEPES pH 7.0	10% PEG 6000
30	0.8M Ammonium Dihydrogen Phosphate 0.8M Potassium Dihydrogen Phosphate	0.1M HEPES pH 7.5	
31		0.25M Sodium Dihydrogen Phosphate pH 5.2	40% PEG 300
32	0.25M Sodium Citrate	0.1M Sodium Acetate pH 4.5	10% PEG 3000
33		0.1M Tris pH 8.5	20% Ethanol
34	25% 1-2-Propanediol	0.1M Sodium Potassium Phosphate pH 6.8	10% Glycerol
35	2% Dioxane	0.1M Bicine pH 9.0	10% PEG 20000
36	2.0M Ammonium Sulphate	0.1M Sodium Acetate pH 4.6	
37			10% PEG 1000 10% PEG 8000
38	20% Glycerol		24% PEG 1000
39	0.2M Magnesium Chloride	0.1M HEPES pH 7.5	30% PEG 400

40	0.2M Sodium Chloride	0.1M Sodium Potassium Phosphate pH 7.2	50% PEG 200
41	0.2M Lithium Sulphate	0.1M Sodium Acetate pH 4.5	30% PEG 8000
42	0.2M Magnesium Chloride	0.1M HEPES pH 7.5	70% MPD
43		0.1M Tris pH 8.5	20% PEG 8000
44	0.2M Lithium Sulphate	0.1M Tris pH 8.4	40% PEG 400
45		0.1M Tris pH 8.0	40% MPD
46	0.17M ammonium Sulphate		25.5% PEG 4000
47	0.2M Calcium Acetate	0.1M Sodium Cacodylate pH 7.0	15% Glycerol
48	0.14M Calcium Chloride	0.07M Sodium Acetate pH 4.6	40% PEG 300
49	0.04M Potassium Dihydrogen Phosphate		14% Isopropanol
50	1.0M sodium Citrate	0.1M Sodium Cacodylate pH 6.5	30% Glycerol
51	0.2M Sodium Chloride	0.1M Sodium Cacodylate pH 6.5	16% PEG 8000
52	0.2M Sodium Chloride	0.1M HEPES pH 7.5	20% Glycerol
53	1.26M ammonium Sulphate		
54	0.2M Lithium Sulphate	0.1M Tris pH 8.5	2.0M Ammonium Sulphate
55		0.1M CAPS pH 10.1	10% Isopropanol
56	0.2M Zinc Acetate	0.1M Imidazol pH 8.0	
57	0.2M Zinc Acetate	0.1M Sodium Cacodylate pH 6.5	40% MPD
58	1.0M Diammonium Hydrogen Phosphate	0.1M Sodium Acetate pH 4.5	20% PEG 3000
59	1.6M Magnesium Sulphate	0.1M MES pH 6.5	10% Isopropanol
60	10% PEG 6000	0.1M Bicine pH 9.0	
61	0.16M Calcium Acetate	0.08M Sodium Cacodylate pH 6.5	14.4% PEG 8000
62	0.05M Caesium Chloride	0.1M Imidazol pH 8.0	20% Glycerol
63	3.2M Ammonium Sulphate	0.1M MES pH 6.5	10% PEG 8000
64		0.1M Citric Acid pH 5.0	30% Jeffamine
65		0.1M Tris pH 8.0	
66	0.2M Magnesium Chloride	0.1M HEPES pH 6.5	20% MPD
67		0.1M Tris pH 8.5	20% Jeffamine
68		0.1M Bicine pH 9.0	50% Ethylene Glycol
69	0.2M Ammonium Sulphate	0.1M Sodium Acetate pH 4.6	10% MPD
70	0.2M Ammonium Sulphate	0.1M MES pH 6.5	30% PEG MME 2000
71	0.01M Zinc Sulphate	0.1M MES pH 6.5	30% PEG MME 5000
72	0.01M Nickel Chloride	0.1M Tris pH 8.5	25% PEG MME 550
73	0.1M Sodium Chloride	0.1M Bicine pH 9.0	20% PEG MME 2000
74	0.005M Magnesium Chloride	0.05M HEPES pH 7.0	20% PEG MME 550
75	0.1M Potassium Chloride		
76	0.015M Magnesium Chloride	0.05M Tris pH 7.5	25% PEG MME 550
77	0.2M Lithium Sulphate	0.1M MES pH 6.0	10% PEG MME 550
78	0.2M Sodium Chloride	0.1M Imidazol pH 8.0	20% 1-4-Butandiol
79		0.1M Sodium Acetate pH 4.5	1M Sodium Potassium Tartrate
80	0.2M Lithium Sulphate	0.1M CHES pH 9.5	20% 1-4-Butandiol
81	35% Propanol	0.1M Sodium Cacodylate pH 6.5	1M Sodium Potassium Tartrate
82	3.5M Sodium Formate	0.1M Tris pH 8.5	35% Ethoxyethanol
83		0.8M Succinic Acid pH 7.0	
84		2.1M Malic Acid pH 7.0	
85	0.2M Potassium Chloride	2.4M Sodium Malonate pH 7.0	
		0.05M HEPES pH 7.5	35% Pentaerythritol

86	0.005M Ammonium Sulphate	0.05M Tris pH 6.5	Propoxylate 30% Pentaerythritol Ethoxylate
87	0.2M Potassium Bromide	0.1M HEPES pH 7.5	25% PEG MME 2000
88	0.2M Potassium Bromide	0.1M Tris pH 8.5	8% PEG 20000
	1.0M Potassium Dihydrogen Phosphate	0.1M Sodium Citrate pH 4.6	8% PEG MME 550
89	0.5M Potassium Dihydrogen Phosphate	0.1M HEPES pH 7.0	
90	0.005M Cadmium Chloride	0.1M Tris pH 8.0	20% PEG 4000
91	0.005M Nickel Chloride	0.1M MED pH 6.5	20% PEG 4000
92			10% PEG 8000
93	0.8M Sodium Formate	0.1M Imidazol pH 8.0	10% PEG 1000
94	0.005M Cadmium Sulphate	0.1M Sodium Cacodylate pH 6.5	15% PEG 4000
95	0.005M Cobalt Chloride	0.1M HEPES pH 7.5	20% PEG 600
			2M Ammonium Sulphate
96		0.1M Tris pH 8.0	10% Jeffamine

B2 The Classics Suite™ (Qiagen) Composition Table

Number	Salt	Buffer	Precipitant
1	0.01 M Cobalt chloride	0.1 M Sodium acetate pH 4.6	1.0 M 1,6-Hexanediol
2		0.1 M tri-Sodium citrate pH 5.6	2.5 M 1,6-Hexanediol
3	0.2 M Magnesium chloride	0.1 M TRIS pH 8.5	3.4 M 1,6-Hexanediol
4			5 %(v/v) Isopropanol, 2.0 M Ammonium sulfate
5		0.1 M HEPES sodium salt pH 7.5	10 %(v/v) Isopropanol, 20 %(w/v) PEG 4000
6	0.2 M Calcium chloride	0.1 M Sodium acetate pH 4.6	20 %(v/v) Isopropanol
7		0.1 M tri-Sodium citrate pH 5.6	20 %(v/v) Isopropanol, 20 %(w/v) PEG 4000
8	0.2 M tri-Sodium citrate	0.1 M HEPES sodium salt pH 7.5	20 %(v/v) Isopropanol
9	0.2 M tri-Sodium citrate	0.1 M Sodium cacodylate pH 6.5	30 %(v/v) Isopropanol
10	0.2 M Magnesium chloride	0.1 M HEPES sodium salt pH 7.5	30 %(v/v) Isopropanol
11	0.2 M Ammonium acetate	0.1 M TRIS.HCl pH 8.5	30 %(v/v) Isopropanol
12			10 %(v/v) Ethanol, 1.5 M Sodium chloride
13		0.1 M TRIS pH 8.5	20 %(v/v) Ethanol
14			25 %(v/v) Ethylene glycol
15	0.02 M Calcium chloride	0.1 M Sodium acetate pH 4.6	30 %(v/v) MPD
16	0.2 M Sodium chloride	0.1 M Sodium acetate pH 4.6	30 %(v/v) MPD
17	0.2 M Ammonium acetate	0.1 M tri-Sodium citrate pH 5.6	30 %(v/v) MPD
18	0.2 M Magnesium acetate	0.1 M Sodium cacodylate pH 6.5	30 %(v/v) MPD
19	0.2 M tri-Sodium citrate	0.1 M HEPES sodium salt pH 7.5	30 %(v/v) MPD
20	0.5 M Ammonium sulfate	0.1 M HEPES pH 7.5	30 %(v/v) MPD
21	0.2 M Ammonium phosphate	0.1 M TRIS pH 8.5	50 %(v/v) MPD
22		0.1 M HEPES pH 7.5	70 %(v/v) MPD
23		0.1 M TRIS pH 8.5	25 %(v/v) tert-Butanol
24		0.1 M tri-Sodium citrate pH 5.6	35 %(v/v) tert-Butanol
25			0.4 M Ammonium phosphate
26		0.1 M tri-Sodium citrate pH 5.6	1.0 M Ammonium phosphate
27		0.1 M TRIS.HCl pH 8.5	2.0 M Ammonium phosphate
28		0.1 M HEPES pH 7.5	2.0 M Ammonium formate
29		0.1 M Sodium acetate pH 4.6	2.0 M Ammonium sulfate
30		0.1 M TRIS.HCl pH 8.5	2.0 M Ammonium sulfate
31			2.0 M Ammonium sulfate
32	0.1 M Sodium chloride	0.1 M HEPES pH 7.5	1.6 M Ammonium sulfate
33	0.01 M Cobalt chloride	0.1 M MES pH 6.5	1.8 M Ammonium sulfate
34	0.2 M K/Na tartrate	0.1 M tri-Sodium citrate pH 5.6	2.0 M Ammonium sulfate
35			1.0 M Imidazole pH 7.0
36			0.4 M K/Na tartrate
37		0.1 M HEPES sodium salt pH 7.5	0.8 M K/Na tartrate
38		0.1 M Imidazole pH 6.5	1.0 M Sodium acetate
39	0.05 M Cadmium sulfate	0.1 M HEPES pH 7.5	1.0 M Sodium acetate
40		0.1 M Sodium cacodylate pH 6.5	1.4 M Sodium acetate
41		0.1 M Sodium acetate pH 4.6	2.0 M Sodium chloride
42	0.1 M Sodium phosphate		
43	0.1 M Potassium phosphate	0.1 M MES pH 6.5	2.0 M Sodium chloride
44		0.1 M HEPES pH 7.5	4.3 M Sodium chloride
45		0.1 M HEPES sodium salt pH 7.5	1.4 M tri-Sodium citrate
			1.6 M tri-Sodium citrate pH 6.5

46		0.1 M HEPES sodium salt pH 7.5	0.8 M Sodium phosphate
47		0.1 M Sodium acetate pH 4.6	0.8 M Potassium phosphate
48			2.0 M Sodium formate
49		0.1 M BICINE pH 9.0	4.0 M Sodium formate
50		0.1 M MES pH 6.5	2 % (v/v) Dioxane, 10 % (w/v) PEG 20000
51			10 % (v/v) Dioxane, 1.6 M Ammonium sulfate
52	0.5 M Sodium chloride	0.1 M tri-Sodium citrate pH 5.6	35 % (v/v) Dioxane
53		0.1 M TRIS pH 8.5	2 % (v/v) Ethylene imine polymer
54	0.5 M Sodium chloride 0.01 M Magnesium chloride		12 % (v/v) Glycerol 1.5 M Ammonium sulphate
55	0.01 M Ferric chloride	0.1 M tri-Sodium citrate pH 5.6	0.01 M CTAB
56		0.1 M HEPES pH 7.5	10 % (v/v) Jeffamine M-600
57	0.5 M Ammonium sulfate	0.1 M tri-Sodium citrate pH 5.6	20 % (v/v) Jeffamine M-600
58	0.01 M Nickel chloride	0.1 M TRIS pH 8.5	1.0 M Lithium sulfate
59		0.1 M HEPES sodium salt pH 7.5	1.0 M Lithium sulfate
60		0.1 M BICINE pH 9.0	1.5 M Lithium sulfate
61			2.0 M Magnesium chloride
62		0.1 M MES pH 6.5	0.2 M Magnesium formate
63		0.1 M TRIS.HCl pH 8.5	1.6 M Magnesium sulfate
64		0.1 M HEPES pH 7.5	8 % (w/v) PEG 8000
65	0.5 M Lithium sulfate		10 % (w/v) PEG 8000
66	0.2 M Zinc acetate	0.1 M Sodium cacodylate pH 6.5	15 % (w/v) PEG 8000
67	0.2 M Calcium acetate	0.1 M Sodium cacodylate pH 6.5	18 % (w/v) PEG 8000
68	0.2 M Magnesium acetate	0.1 M Sodium cacodylate pH 6.5	18 % (w/v) PEG 8000
69	0.05 M Potassium phosphate		20 % (w/v) PEG 8000
70	0.2 M Ammonium sulfate	0.1 M Sodium cacodylate pH 6.5	20 % (w/v) PEG 8000
71	0.2 M Sodium acetate	0.1 M Sodium cacodylate pH 6.5	30 % (w/v) PEG 8000
72	0.2 M Ammonium sulfate		30 % (w/v) PEG 8000
73		0.1 M HEPES sodium salt pH 7.5	30 % (w/v) PEG 8000
74	0.2 M Calcium chloride	0.1 M HEPES sodium salt pH 7.5	2 % (v/v) PEG 400, 2.0 M Ammonium sulfate
75	0.1 M Cadmium chloride	0.1 M Sodium acetate pH 4.6	28 % (v/v) PEG 400
76	0.2 M Magnesium chloride	0.1 M HEPES sodium salt pH 7.5	30 % (v/v) PEG 400
77	0.2 M tri-Sodium citrate	0.1 M TRIS.HCl pH 8.5	30 % (v/v) PEG 400
78	0.1 M Sodium chloride	0.1 M BICINE pH 9.0	30 % (v/v) PEG 400
79	0.01 M Zinc sulfate	0.1 M MES pH 6.5	20 % (w/v) PEG 550 MME
80			25 % (w/v) PEG 550 MME
81			10 % (w/v) PEG 1000, 10 % (w/v) PEG 8000
82	0.01 M Nickel chloride	0.1 M TRIS pH 8.5	30 % (w/v) PEG 1500
83	0.2 M Ammonium sulfate	0.1 M Sodium acetate pH 4.6	20 % (w/v) PEG 2000 MME
84		0.1 M Sodium acetate pH 4.6	30 % (w/v) PEG 2000 MME
85	0.2 M Ammonium sulfate	0.1 M Sodium acetate pH 4.6	8 % (w/v) PEG 4000
86	0.2 M Ammonium acetate	0.1 M Sodium acetate pH 4.6	25 % (w/v) PEG 4000
87	0.2 M Ammonium acetate	0.1 M tri-Sodium citrate pH 5.6	30 % (w/v) PEG 4000
88	0.2 M Magnesium chloride	0.1 M TRIS.HCl pH 8.5	30 % (w/v) PEG 4000
89	0.2 M Lithium sulfate	0.1 M TRIS.HCl pH 8.5	30 % (w/v) PEG 4000
90	0.2 M Sodium acetate	0.1 M TRIS.HCl pH 8.5	30 % (w/v) PEG 4000

91	0.2 M Ammonium sulfate		30 %(w/v) PEG 4000 30 %(w/v) PEG 5000
92	0.2 M Ammonium sulfate	0.1 M MES pH 6.5	MME
93		0.1 M HEPES pH 7.5	10 %(w/v) PEG 6000, 5 %(v/v) MPD
94			10 %(w/v) PEG 6000, 2.0 M Sodium chloride
95		0.1 M HEPES pH 7.5	20 %(w/v) PEG 10000, 8 %(v/v) Ethylene glycol
96		0.1 M MES pH 6.5	12 %(w/v) PEG 20000

B3 The JCSG+ Suite™ (Qiagen) Composition Table

Number	Salt	Buffer	Precipitant
1	0.2 M Lithium sulfate	0.1 M Sodium acetate pH 4.5	50 %w/v PEG 400
2		0.1 M tri-Sodium citrate pH 5.5	20 %w/v PEG 3000
3	0.2 M di-Ammonium citrate pH 5.0		20 %w/v PEG 3350
4	0.02 M Calcium chloride	0.1 M Sodium acetate pH 4.6	30 %v/v MPD
5	0.2 M Magnesium formate pH 5.9		20 %w/v PEG 3350
6	0.2 M Lithium sulfate	0.1 M Phosphate-citrate pH 4.2	20 %w/v PEG 1000
7		0.1 M CHES pH 9.5	20 %w/v PEG 8000
8	0.2 M Ammonium formate pH 6.6		20 %w/v PEG 3350
9	0.2 M Ammonium chloride pH 6.3		20 %w/v PEG 3350
10	0.2 M Potassium formate pH 7.3		20 %w/v PEG 3350
11	0.2 M Ammonium phosphate	0.1 M TRIS pH 8.5	50 %v/v MPD
12	0.2 M Potassium nitrate pH 6.9		20 %w/v PEG 3350
13	0.8 M Ammonium sulfate	0.1 M Citric acid pH 4	
14	0.2 M Sodium thiocyanate pH 6.9		20 %w/v PEG 3350
15		0.1 M BICINE pH 9	20 %w/v PEG 6000 10 %w/v PEG 8000, 8 %v/v Ethylene glycol
16		0.1 M HEPES pH 7.5	40 %v/v MPD, 5 %w/v PEG 8000
17		0.1 M Sodium cacodylate pH 6.5	40 %v/v Ethanol, 5 %w/v PEG 1000
18		0.1 M Phosphate-citrate pH 4.2	8 %w/v PEG 4000
19		0.1 M Sodium acetate pH 4.6	10 %w/v PEG 8000
20	0.2 M Magnesium chloride	0.1 M TRIS pH 7	20 %w/v PEG 6000
21		0.1 M Citric acid pH 5	50 %v/v PEG 200
22	0.2 M Magnesium chloride	0.1 M Sodium cacodylate pH 6.5	
23	1.6 M tri-Sodium citrate		
24	0.2 M tri-Potassium citrate pH 8.3		20 %w/v PEG 3350
25	0.2 M Sodium chloride	0.1 M Phosphate-citrate pH 4.2	20 %w/v PEG 8000
26	1 M Lithium chloride	0.1 M Citric acid pH 4	20 %w/v PEG 6000
27	0.2 M Ammonium nitrate pH 6.3		20 %w/v PEG 3350
28		0.1 M HEPES pH 7	10 %w/v PEG 6000 0,8 M Sodium phosphate, 0,8 M Potassium phosphate
29		0.1 M HEPES pH 7.5	40 %v/v PEG 300
30		0.1 M Phosphate-citrate pH 4.2	10 %w/v PEG 3000
31	0.2 M Zinc acetate	0.1 M Sodium acetate pH 4.5	20 %v/v Ethanol 25 %v/v 1,2 propandiol, 10 %v/v Glycerol
32		0.1 M TRIS pH 8.5	10 %w/v PEG 20000, 2 %v/v Dioxane
33		0.1 M Na/K phosphate pH 6.2	
34		0.1 M BICINE pH 9	
35	2 M Ammonium sulfate	0.1 M Sodium acetate pH 4.6	10 %w/v PEG 1000, 10% w/v PEG 8000 24 %w/v PEG 1500, 20% w/v glycerol
36			
37			
38	0.2 M Magnesium chloride	0.1 M HEPES pH 7.5	30 %v/v PEG 400
39	0.2 M Sodium chloride	0.1 M Na/K phosphate pH 6.2	50 %v/v PEG 200
40	0.2 M Lithium sulfate	0.1 M Sodium acetate pH 4.5	30 %w/v PEG 8000
41		0.1 M HEPES pH 7.5	70 %v/v MPD
42	0.2 M Magnesium chloride	0.1 M TRIS pH 8.5	20 %w/v PEG 8000
43	0.2 M Lithium sulfate	0.1 M TRIS pH 8.5	40 %v/v PEG 400
44		0.1 M TRIS pH 8	40 %v/v MPD 25,5 %w/v PEG 4000, 15 %v/v Glycerol
45	0.17 M Ammonium sulfate		

46	0.2 M Calcium acetate	0.1 M Sodium cacodylate pH 6.5	40 %v/v PEG 300 14 %v/v Isopropanol, 30 %v/v Glycerol 16 %w/v PEG 8000, 20 %v/v Glycerol
47	0.14 M Calcium chloride	0.07 M Sodium acetate pH 4.6	
48	0.04 M Potassium phosphate		
49	1 M tri-Sodium citrate	0.1 M Sodium cacodylate pH 6.5	
50	0.2 M Sodium chloride	0.1 M Sodium cacodylate pH 6.5	2 M Ammonium sulfate
51	0.2 M Sodium chloride	0.1 M HEPES pH 7.5	10 %v/v Isopropanol
52	0.2 M Lithium sulfate	0.1 M TRIS pH 8.5	1,26 M Ammonium sulfate
53		0.1 M CAPS pH 10.5	40 %v/v MPD
54	0.2 M Zinc acetate	0.1 M Imidazole pH 8	20 %w/v PEG 3000
55	0.2 M Zinc acetate	0.1 M Sodium cacodylate pH 6.5	10 %v/v Isopropanol
56	1 M di-Ammonium phosphate	0.1 M Sodium acetate pH 4.5	
57	1.6 M Magnesium sulfate	0.1 M MES pH 6.5	
58		0.1 M BICINE pH 9	10 %w/v PEG 6000 14,4 %w/v PEG 8000, 20 %v/v Glycerol
59	0.16 M Calcium acetate	0.08 M Sodium cacodylate pH 6.5	
60		0.1 M Imidazole pH 8	10 %w/v PEG 8000
61	0.05 M Cesium chloride	0.1 M MES pH 6.5	30 %w/v Jeffamine M-600
62	3.15 M Ammonium sulfate	0.1 M Citric acid pH 5	
63		0.1 M TRIS pH 8	20 %v/v MPD
64		0.1 M HEPES pH 6.5	20 %w/v Jeffamine M-600
65	0.2 M Magnesium chloride	0.1 M TRIS pH 8.5	50 %v/v Ethylene glycol
66		0.1 M BICINE pH 9	10 %v/v MPD
67	0.8 M Succinic acid pH 7.0		
68	2.1 M DL-Malic acid pH 7.0		
69	2.4 M Sodium malonate pH 7.0		0,5 %v/v Jeffamine ED-2001 pH 7.0
70	1.1 M Sodium malonate pH 7.0	0.1 M HEPES pH 7	
71	1 M Succinic acid pH 7.0	0.1 M HEPES pH 7	1 %w/v PEG MME 2000 30 %v/v Jeffamine M-600 pH 7.0
72		0.1 M HEPES pH 7	30 %v/v Jeffamine ED-2001 pH 7.0
73		0.1 M HEPES pH 7	22 %w/v Polyacrylic acid 5100 sodium salt 20 %w/v Polyvinylpyrrolidone K15
74	0.02 M Magnesium chloride	0.1 M HEPES pH 7.5	20 %w/v PEG MME 2000
75	0.1 M Cobalt chloride	0.1 M TRIS pH 8.5	
76	0.2 M Trimethylamine N-oxide 0.005 M Cobalt chloride 0.005 M Cadmium chloride 0.005 M Magnesium chloride	0.1 M TRIS pH 8.5	
77	0.005 M Nickel chloride	0.1 M HEPES pH 7.5	12 %w/v PEG 3350
78	0.2 M Sodium malonate pH 7.0		20 %w/v PEG 3350
79	0.1 M Succinic acid pH 7.0		15 %w/v PEG 3350
80	0.15 M DL-Malic acid pH 7.0		20 %w/v PEG 3350
81	0.1 M Potassium thiocyanate		30 %w/v PEG MME 2000
82	0.15 M Potassium bromide		30 %w/v PEG MME 2000
83	2 M Ammonium sulfate	0.1 M bis-TRIS pH 5.5	
84	3 M Sodium chloride	0.1 M bis-TRIS pH 5.5	
85	0.3 M Magnesium formate	0.1 M bis-TRIS pH 5.5	
86	1 M Ammonium sulfate	0.1 M bis-TRIS pH 5.5	1 %w/v PEG 3350
87		0.1 M bis-TRIS pH 5.5	25 %w/v PEG 3350
88	0.2 M Calcium chloride	0.1 M bis-TRIS pH 5.5	45 %v/v MPD
89	0.2 M Ammonium acetate	0.1 M bis-TRIS pH 5.5	45 %v/v MPD
90	0.1 M Ammonium acetate	0.1 M bis-TRIS pH 5.5	17 %w/v PEG 10,000
91	0.2 M Ammonium sulfate	0.1 M bis-TRIS pH 5.5	25 %w/v PEG 3350
92	0.2 M Sodium chloride	0.1 M bis-TRIS pH 5.5	25 %w/v PEG 3350
93	0.2 M Lithium sulfate	0.1 M bis-TRIS pH 5.5	25 %w/v PEG 3350

94	0.2 M Ammonium acetate	0.1 M bis-TRIS pH 5.5	25 %w/v PEG 3350
95	0.2 M Magnesium chloride	0.1 M bis-TRIS pH 5.5	25 %w/v PEG 3350
96	0.2 M Ammonium acetate	0.1 M HEPES pH 7.5	45 %v/v MPD

B4 The PACT Suite™ (Qiagen) Composition Table

Number	Salt	Buffer	Precipitant
1		0.1M SPG buffer pH 4	25% w/v PEG 1500
2		0.1M SPG buffer pH 5	25% w/v PEG 1500
3		0.1M SPG buffer pH 6	25% w/v PEG 1500
4		0.1M SPG buffer pH 7	25% w/v PEG 1500
5		0.1M SPG buffer pH 8	25% w/v PEG 1500
6		0.1M SPG buffer pH 9	25% w/v PEG 1500
7	0.2M Sodium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
8	0.2M Ammonium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
9	0.2M Lithium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
10	0.2M Magnesium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
11	0.2M Calcium chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
12	0.01M Zinc chloride	0.1M Sodium acetate pH 5	20% w/v PEG 6000
13		0.1M MIB buffer pH 4	25% w/v PEG 1500
14		0.1M MIB buffer pH 5	25% w/v PEG 1500
15		0.1M MIB buffer pH 6	25% w/v PEG 1500
16		0.1M MIB buffer pH 7	25% w/v PEG 1500
17		0.1M MIB buffer pH 8	25% w/v PEG 1500
18		0.1M MIB buffer pH 9	25% w/v PEG 1500
19	0.2M Sodium chloride	0.1M MES pH 6	20% w/v PEG 6000
20	0.2M Ammonium chloride	0.1M MES pH 6	20% w/v PEG 6000
21	0.2M Lithium chloride	0.1M MES pH 6	20% w/v PEG 6000
22	0.2M Magnesium chloride	0.1M MES pH 6	20% w/v PEG 6000
23	0.2M Calcium chloride	0.1M MES pH 6	20% w/v PEG 6000
24	0.01M Zinc chloride	0.1M MES pH 6	20% w/v PEG 6000
25		0.1M PCB buffer pH 4	25% w/v PEG 1500
26		0.1M PCB buffer pH 5	25% w/v PEG 1500
27		0.1M PCB buffer pH 6	25% w/v PEG 1500
28		0.1M PCB buffer pH 7	25% w/v PEG 1500
29		0.1M PCB buffer pH 8	25% w/v PEG 1500
30		0.1M PCB buffer pH 9	25% w/v PEG 1500
31	0.2M Sodium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
32	0.2M Ammonium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
33	0.2M Lithium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
34	0.2M Magnesium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
35	0.2M Calcium chloride	0.1M Hepes pH 7	20% w/v PEG 6000
36	0.01M Zinc chloride	0.1M Hepes pH 7	20% w/v PEG 6000
37		0.1M MMT buffer pH 4	25% w/v PEG 1500
38		0.1M MMT buffer pH 5	25% w/v PEG 1500
39		0.1M MMT buffer pH 6	25% w/v PEG 1500
40		0.1M MMT buffer pH 7	25% w/v PEG 1500
41		0.1M MMT buffer pH 8	25% w/v PEG 1500
42		0.1M MMT buffer pH 9	25% w/v PEG 1500
43	0.2M Sodium chloride	0.1M Tris pH 8	20% w/v PEG 6000
44	0.2M Ammonium chloride	0.1M Tris pH 8	20% w/v PEG 6000
45	0.2M Lithium chloride	0.1M Tris pH 8	20% w/v PEG 6000
46	0.2M Magnesium chloride	0.1M Tris pH 8	20% w/v PEG 6000
47	0.2M Calcium chloride	0.1M Tris pH 8	20% w/v PEG 6000
48	0.01M Zinc chloride	0.1M Tris pH 8	20% w/v PEG 6000
49	0.2M Sodium fluoride		20% w/v PEG 3350
50	0.2M Sodium bromide		20% w/v PEG 3350
51	0.2M Sodium iodide		20% w/v PEG 3350
52	0.2M Potassium thiocyanate		20% w/v PEG 3350

53	0.2M Sodium nitrate		20% w/v PEG 3350
54	0.2M Sodium formate		20% w/v PEG 3350
55	0.2M Sodium acetate		20% w/v PEG 3350
56	0.2M Sodium sulphate		20% w/v PEG 3350
57	0.2M Potassium/sodium tartrate		20% w/v PEG 3350
58	0.2M Sodium/potassium phosphate		20% w/v PEG 3350
59	0.2M Sodium citrate		20% w/v PEG 3350
60	0.2M Sodium malonate		20% w/v PEG 3350
61	0.2M Sodium fluoride	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
62	0.2M Sodium bromide	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
63	0.2M Sodium iodide	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
64	0.2M Potassium thiocyanate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
65	0.2M Sodium nitrate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
66	0.2M Sodium formate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
67	0.2M Sodium acetate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
68	0.2M Sodium sulphate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
69	0.2M Potassium/sodium tartrate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
70	0.2M Sodium/potassium phosphate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
71	0.2M Sodium citrate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
72	0.2M Sodium malonate	0.1M Bis Tris propane pH 6.5	20% w/v PEG 3350
73	0.2M Sodium fluoride	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
74	0.2M Sodium bromide	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
75	0.2M Sodium iodide	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
76	0.2M Potassium thiocyanate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
77	0.2M Sodium nitrate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
78	0.2M Sodium formate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
79	0.2M Sodium acetate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
80	0.2M Sodium sulphate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
81	0.2M Potassium/sodium tartarte	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
82	0.2M Sodium/potassium phosphate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
83	0.2M Sodium citrate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
84	0.2M Sodium malonate	0.1M Bis Tris propane pH 7.5	20% w/v PEG 3350
85	0.2M Sodium fluoride	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
86	0.2M Sodium bromide	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
87	0.2M Sodium iodide	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
88	0.2M Potassium thiocyanate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
89	0.2M Sodium nitrate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
90	0.2M Sodium formate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
91	0.2M Sodium acetate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
92	0.2M Sodium sulphate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
93	0.2M Potassium/sodium tartrate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
94	0.2M Sodium/potassium phosphate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
95	0.2M Sodium citrate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350
96	0.2M Sodium malonate	0.1M Bis Tris propane pH 8.5	20% w/v PEG 3350

B5 The AmSO4 Suite™ (Qiagen) Composition Table

Number	Salt	Buffer	Precipitant 1
1			2.2 M Ammonium sulfate
2	0.2 M Ammonium acetate		2.2 M Ammonium sulfate
3	0.2 M Ammonium chloride		2.2 M Ammonium sulfate
4	0.2 M Ammonium phosphate		2.2 M Ammonium sulfate
5	0.2 M Ammonium fluoride		2.2 M Ammonium sulfate
6	0.2 M Ammonium formate		2.2 M Ammonium sulfate
7	0.2 M di-Ammonium citrate		2.2 M Ammonium sulfate
8	0.2 M di-Ammonium phosphate		2.2 M Ammonium sulfate
9	0.2 M Ammonium iodide		2.2 M Ammonium sulfate
10	0.2 M Ammonium nitrate		2.2 M Ammonium sulfate
11	0.2 M di-Ammonium tartrate		2.2 M Ammonium sulfate
12	0.2 M Cadmium chloride		2.2 M Ammonium sulfate
13	0.2 M Cadmium sulfate		2.2 M Ammonium sulfate
14	0.2 M Cesium chloride		2.2 M Ammonium sulfate
15	0.2 M Cesium sulfate		2.2 M Ammonium sulfate
16	0.2 M Ammonium bromide		2.2 M Ammonium sulfate
17	0.2 M Lithium acetate		2.2 M Ammonium sulfate
18	0.2 M Lithium chloride		2.2 M Ammonium sulfate
19	0.2 M tri-Lithium citrate		2.2 M Ammonium sulfate
20	0.2 M Lithium nitrate		2.2 M Ammonium sulfate
21	0.2 M Lithium sulfate		2.2 M Ammonium sulfate
22	0.2 M Potassium acetate		2.2 M Ammonium sulfate
23	0.2 M Potassium bromide		2.2 M Ammonium sulfate
24	0.2 M Potassium chloride		2.2 M Ammonium sulfate
25	0.2 M tri-Potassium citrate		2.2 M Ammonium sulfate
26	0.2 M Potassium phosphate		2.2 M Ammonium sulfate
27	0.2 M Potassium fluoride		2.2 M Ammonium sulfate
28	0.2 M Potassium formate		2.2 M Ammonium sulfate
29	0.2 M di-Potassium phosphate		2.2 M Ammonium sulfate
30	0.2 M Potassium iodide		2.2 M Ammonium sulfate
31	0.2 M Potassium nitrate		2.2 M Ammonium sulfate
32	0.2 M K/Na tartrate		2.2 M Ammonium sulfate
33	0.2 M Potassium sulfate		2.2 M Ammonium sulfate
34	0.2 M Potassium thiocyanate		2.2 M Ammonium sulfate
35	0.2 M Sodium acetate		2.2 M Ammonium sulfate
36	0.2 M Sodium bromide		2.2 M Ammonium sulfate
37	0.2 M Sodium chloride		2.2 M Ammonium sulfate
38	0.2 M tri-Sodium citrate		2.2 M Ammonium sulfate
39	0.2 M Sodium phosphate		2.2 M Ammonium sulfate
40	0.2 M Sodium fluoride		2.2 M Ammonium sulfate
41	0.2 M Sodium formate		2.2 M Ammonium sulfate
42	0.2 M di-Sodium phosphate		2.2 M Ammonium sulfate
43	0.2 M Sodium iodide		2.2 M Ammonium sulfate
44	0.2 M Sodium malonate		2.2 M Ammonium sulfate
45	0.2 M Sodium nitrate		2.2 M Ammonium sulfate
46	0.2 M Sodium sulfate		2.2 M Ammonium sulfate

47	0.2 M di-Sodium tartate		2.2 M Ammonium sulfate
48	0.2 M Sodium thiocyanate		2.2 M Ammonium sulfate
49		0.1 M Citric acid pH 4.0	0.8 M Ammonium sulfate
50		0.1 M Citric acid pH 5.0	0.8 M Ammonium sulfate
51		0.1 M MES pH 6.0	0.8 M Ammonium sulfate
52		0.1 M HEPES pH 7.0	0.8 M Ammonium sulfate
53		0.1 M TRIS pH 8.0	0.8 M Ammonium sulfate
54		0.1 M BICINE pH 9.0	0.8 M Ammonium sulfate
55		0.1 M Citric acid pH 4.0	1.6 M Ammonium sulfate
56		0.1 M Citric acid pH 5.0	1.6 M Ammonium sulfate
57		0.1 M MES pH 6.0	1.6 M Ammonium sulfate
58		0.1 M HEPES pH 7.0	1.6 M Ammonium sulfate
59		0.1 M TRIS pH 8.0	1.6 M Ammonium sulfate
60		0.1 M BICINE pH 9.0	1.6 M Ammonium sulfate
61		0.1 M Citric acid pH 4.0	2.4 M Ammonium sulfate
62		0.1 M Citric acid pH 5.0	2.4 M Ammonium sulfate
63		0.1 M MES pH 6.0	2.4 M Ammonium sulfate
64		0.1 M HEPES pH 7.0	2.4 M Ammonium sulfate
65		0.1 M TRIS pH 8.0	2.4 M Ammonium sulfate
66		0.1 M BICINE pH 9.0	2.4 M Ammonium sulfate
67		0.1 M Citric acid pH 4.0	3.2 M Ammonium sulfate
68		0.1 M Citric acid pH 5.0	3.2 M Ammonium sulfate
69		0.1 M MES pH 6.0	3.2 M Ammonium sulfate
70		0.1 M HEPES pH 7.0	3.2 M Ammonium sulfate
71		0.1 M TRIS pH 8.0	3.2 M Ammonium sulfate
72		0.1 M BICINE pH 9.0	3.2 M Ammonium sulfate
73	0.1 M tri-Sodium citrate		0.5 M Ammonium sulfate
74			1.0 M Lithium Sulfate
75		0.1 M Sodium acetate pH 4.6	1.0 M Ammonium sulfate
76		0.1 M HEPES sodium salt pH 7.5	1.0 M Ammonium sulfate
77		0.1 M TRIS.HCl pH 8.5	2 % (w/v) PEG 400
78	0.05 M tri-Sodium citrate		1.0 M Ammonium sulfate
79		0.1 M TRIS.HCl pH 8.5	1.2 M Ammonium sulfate
80	0.5 M Lithium chloride		3 % (w/v) Isopropanol
81	1.0 M Lithium sulfate		15 % (w/v) Glycerol
82	0.2 M Sodium chloride	0.1 M HEPES sodium salt pH 7.5	1.6 M Ammonium sulfate
83		0.1 M HEPES sodium salt pH 7.5	1.6 M Ammonium sulfate
84		0.1 M MES sodium salt pH 6.5	2 % (w/v) PEG 1000
85	2.0 M Sodium chloride		1.8 M Ammonium sulfate
86		0.1 M Sodium acetate pH 4.6	2.0 M Ammonium sulfate
87		0.1 M MES sodium salt pH 6.5	2.0 M Ammonium sulfate
88		0.1 M TRIS.HCl pH 8.5	5 % (w/v) PEG 400
89			2.0 M Ammonium sulfate
90			2.2 M Ammonium sulfate
			2.2 M Ammonium sulfate
			20 % (w/v) Glycerol

91	0.1 M tri-Sodium citrate		2.4 M Ammonium sulfate
92			3.0 M Ammonium sulfate
93			1 %(w/v) MPD
			3.0 M Ammonium sulfate
			10 %(w/v) Glycerol
94		0.1 M HEPES sodium salt pH 7.5	3.5 M Ammonium sulfate
95		0.1 M MES sodium salt pH 6.5	3.5 M Ammonium sulfate
96			1 %(w/v) MPD
			3.5 M Ammonium sulfate

Appendix C: Vectors used in this Study

C1 pET-21a

pET-21a-d(+) Vectors

TB036 12/98

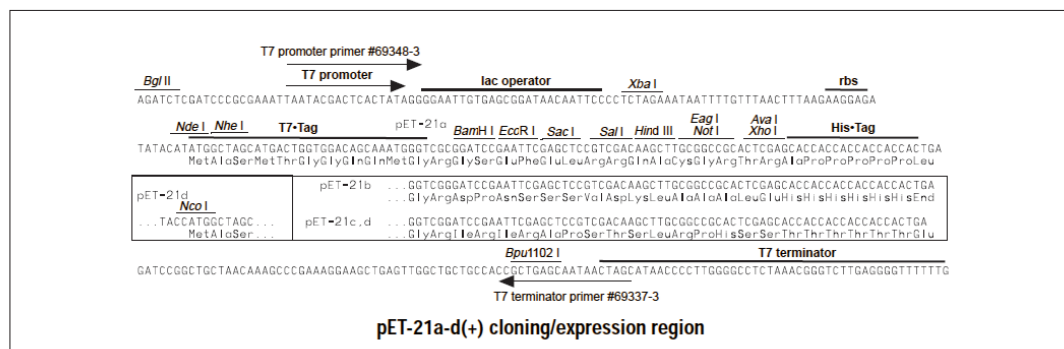
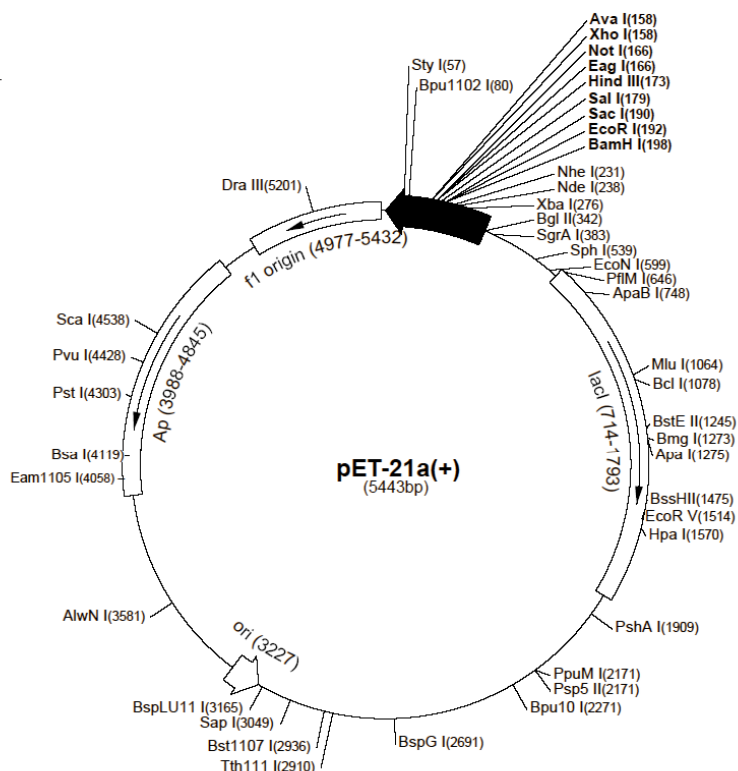
	Cat. No.
pET-21a DNA	69740-3
pET-21b DNA	69741-3
pET-21c DNA	69742-3
pET-21d DNA	69743-3

The pET-21a-d(+) vectors carry an N-terminal T7•Tag[®] sequence plus an optional C-terminal His•Tag[®] sequence. These vectors differ from pET-24a-d(+) only by their selectable marker (ampicillin vs. kanamycin resistance). Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

pET-21a(+) sequence landmarks

T7 promoter	311-327
T7 transcription start	310
T7•Tag coding sequence	207-239
Multiple cloning sites (<i>Bam</i> H I - <i>Xho</i> I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
<i>lacI</i> coding sequence	714-1793
pBR322 origin	3227
<i>bla</i> coding sequence	3988-4845
f1 origin	4977-5432

The maps for pET-21b(+), pET-21c(+) and pET-21d(+) are the same as pET-21a(+) (shown) with the following exceptions: pET-21b(+) is a 5442bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-21c(+) is a 5441bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198. pET-21d(+) is a 5440bp plasmid; the *Bam*H I site is in the same reading frame as in pET-21c(+). An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 238 of pET-21c(+). As a result, *Nco* I cuts pET21d(+) at 234, and *Nhe* I cuts at 229. For the rest of the sites, subtract 3bp from each site beyond position 239 in pET-21a(+). *Nde* I does not cut pET-21d(+). Note also that *Sty* I is not unique in pET-21d(+).



C3 pET-28b

pET-28a-c(+) Vectors

TB074 12/98

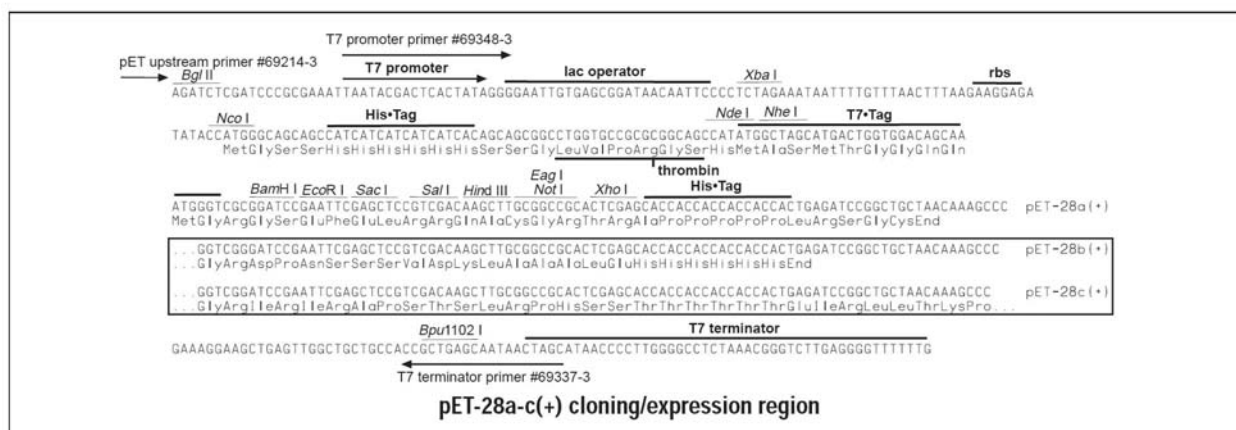
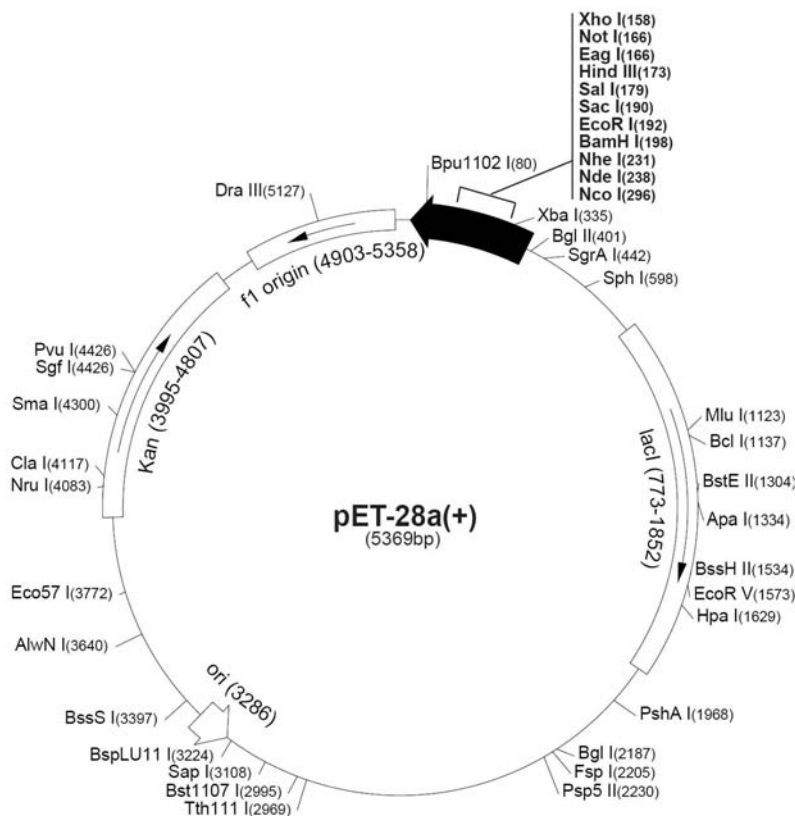
	Cat. No.
pET-28a DNA	69864-3
pET-28b DNA	69865-3
pET-28c DNA	69866-3

The pET-28a-c(+) vectors carry an N-terminal His⁶ Tag[®]/thrombin/T7[®] Tag[®] configuration plus an optional C-terminal His⁶ Tag sequence. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

pET-28a(+) sequence landmarks

T7 promoter	370-386
T7 transcription start	369
His ⁶ Tag coding sequence	270-287
T7 [®] Tag coding sequence	207-239
Multiple cloning sites (<i>Bam</i> H I - <i>Xho</i> I)	158-203
His ⁶ Tag coding sequence	140-157
T7 terminator	26-72
<i>lacI</i> coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.



C4 pET-32b

pET-32a-c(+) Vectors

TB122 12/98

	Cat. No.
pET-32a DNA	69015-3
pET-32b DNA	69016-3
pET-32c DNA	69017-3

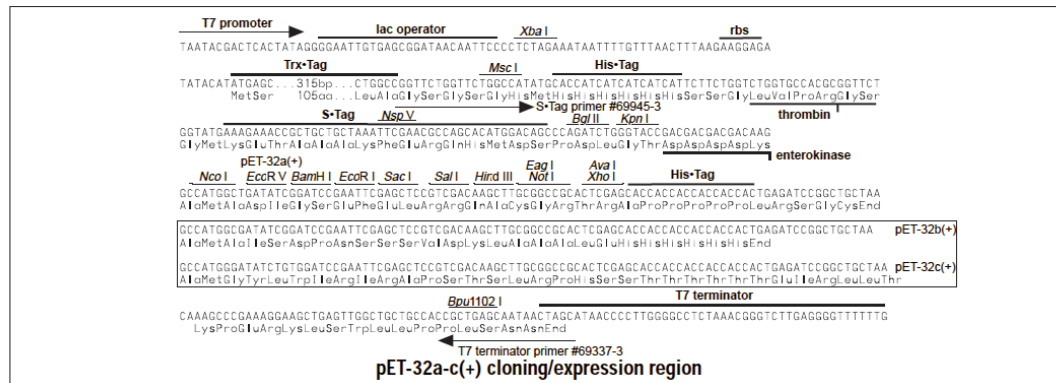
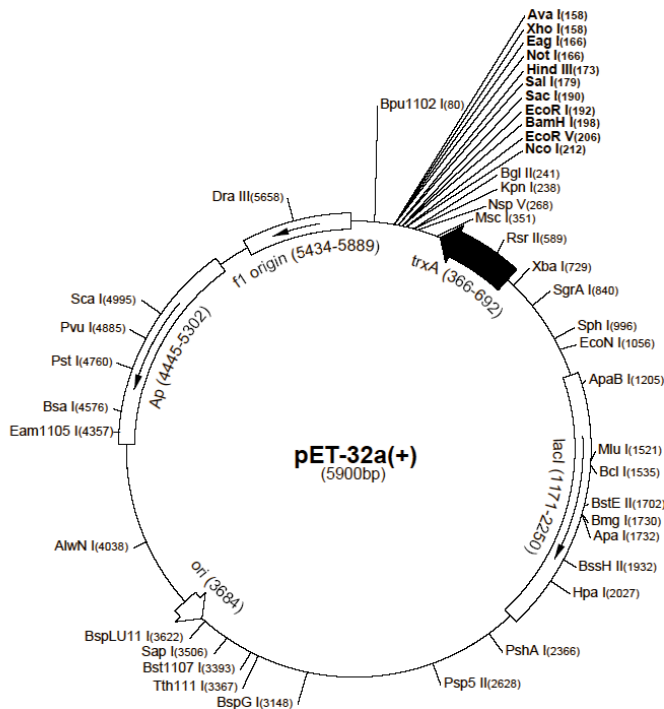
The pET-32 series is designed for cloning and high-level expression of peptide sequences fused with the 109aa Trx•Tag™ thioredoxin protein (1). Cloning sites are available for producing fusion proteins also containing cleavable His•Tag® and S•Tag™ sequences for detection and purification. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single-stranded DNA that corresponds to the coding strand. Therefore, single-stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

1. LaVallie, E.R., DiBlasto, E.A., Kovacic, S., Grant, K.L., Schendel, P.F. and McCoy, J.M. (1993) *BioTechnology* 11, 187-193.

pET-32a(+) sequence landmarks

T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites (<i>Nco</i> I - <i>Xho</i> I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
<i>lac</i> I coding sequence	1171-2250
pBR322 origin	3684
<i>bla</i> coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*HI at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*HI at 198 except for *Eco*R V, which cuts at 209.



C5 pET-39b

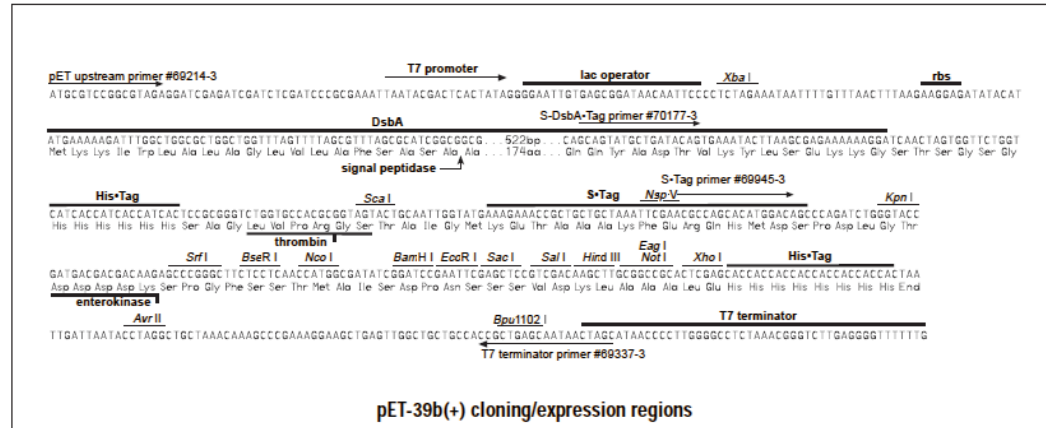
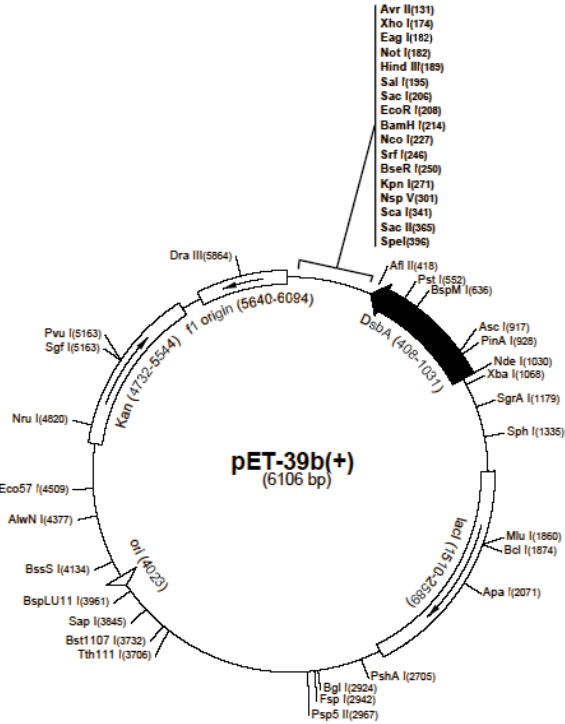
pET-39b(+) Vector

TB230 12/9

The pET-39b(+) vector (Cat. No. 70090-3) is designed for expression of DsbA fusion proteins. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The fl origin is oriented so that infection with helper phage will produce virions containing single stranded DNA that corresponds to the coding strand. Therefore, single stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3).

pET-39b(+) sequence landmarks

T7 promoter	1103-1119
T7 transcription start	1102
DsbA•Tag™ coding seq.	408-1031
His•Tag® coding sequence	369-386
S•Tag™ coding sequence	282-326
Multiple cloning sites (SrfI - Xho I)	174-250
His•Tag coding sequence	150-173
T7 terminator	26-72
lacI coding sequence	1510-2589
pBR322 origin	4023
Kan coding sequence	4732-5544
fl origin	5640-6094



C6 pCR®-Blunt

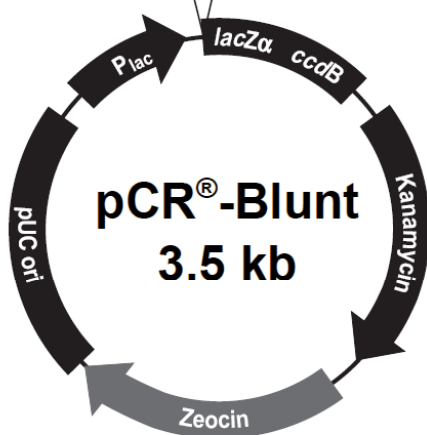
M13 Reverse priming site Mlu I
 201 CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTAT TTAGGTGACG CGTTAGAATA
 GTGTGTCCTT TGTCGATACT GTACTAATG CGGTTGATA AATCCACTGC GCAATCTTAT

Nsi I Hind III Kpn I Sac I BamH I Spe I
 CTC AAGCTAT GCATCAAGCT TGGTACCGAG CTCGGATCCA CTAGTAACGG CCGCCAGTGT
 GAGTTCGATA CGTAGTTCGA ACCATGGCTC GAGCCTAGGT GATCATTGCC GCGGTCACA

EcoR I EcoR I Pst I EcoR V
 GCTGGAATTC AGG **Blunt PCR Product** CCTGAATTCT GCAGATA
 CGACCTTAAG TCC GGACTTAAGA CGTCTAT

Not I Xho I Nsi I Xba I Apa I T7 promoter/priming site
 TCCATCACAC TGGCGGCCGC TCGAGCATGC ATCTAGAGGG CCAATTCCG CCTATAGTGA
 AGGTAGTGTG ACCGCCGGCG AGCTCGTACG TAGATCTCCC GGGTTAAGCG GGATATCACT

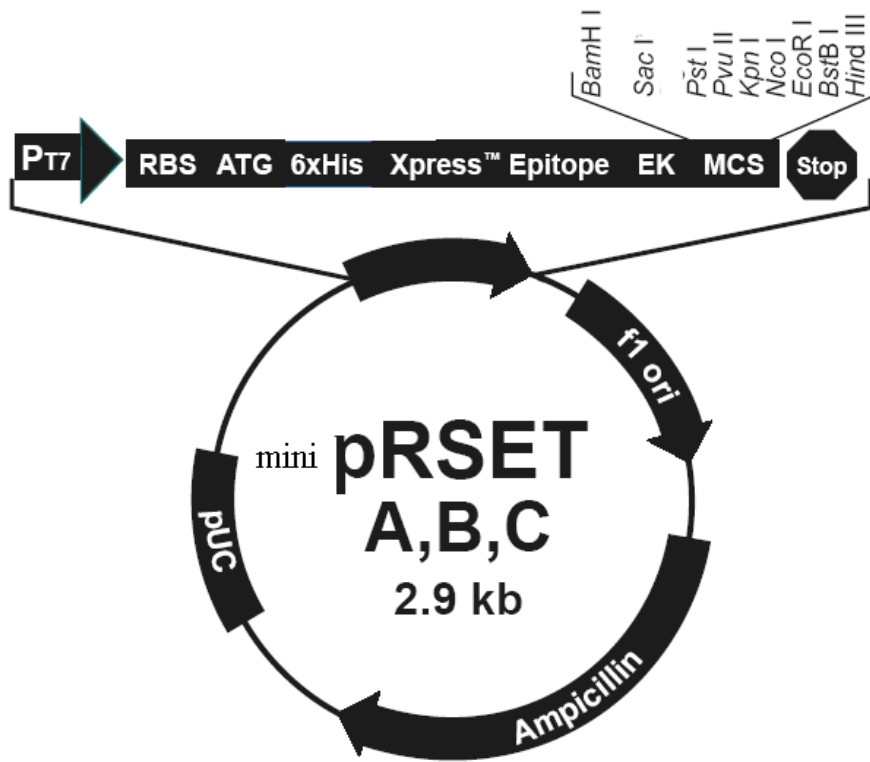
 M13 Forward (-20) priming site
 GTCGTATTAC AATTCACCTGG CCGTCGTTTT ACAACGTCGT GACTGGGAAA ACCCTGGCGT 470
 CAGCATAATG TTAAGTGACC GGCAGCAAAA TGTTGCAGCA CTGACCCTTT TGGGACCGCA



Comments for pCR®-Blunt 3512 nucleotides

Lac promoter/operator region: bases 95-216
 M13 Reverse priming site: bases 205-221
 LacZ-alpha ORF: bases 217-570
 T7 promoter priming site: bases 400-419
 M13 Forward (-20) priming site: bases 427-442
 Fusion joint: bases 571-579
 ccdB lethal gene ORF: bases 580-882
 Kanamycin resistance ORF: bases 1231-2025
 Zeocin resistance ORF: bases 2231-2605
 pUC origin: bases 2673-3386

C7 minipRSET-A



C8 minipRSET-Trx

```

      10      20      30      40      50      60      70      80
      *      *      *      *      *      *      *      *
GATCTCGATC CCGCGAAATT AATACGACTC ACTATAGGGA GACCACAACG GTTCCCTCT AGAAATAAAT TTGTTTAACT

      90      100      110      120      130      140      150
      *      *      *      *      *      *      *
TTAAGAAGGA GATATACAT ATG CAT CAT CAT CAT CAT CAT AGC GAT AAA ATT ATT CAC CTG ACT GAC GAC
      M H H H H H H S D K I I H L T D D>
      _a_a_a_SIX HISTIDINES AND THIOREDUCTASE TAG_a_a_a_>

      160      170      180      190      200      210
      *      *      *      *      *      *
AGT TTT GAC ACG GAT GTA CTC AAA GCG GAC GGG GCG ATC CTC GTC GAT TTC TGG GCA GAG TGG TGC
      S F D T D V L K A D G A I L V D F W A E W C>
      _a_a_a_a_a_a_SIX HISTIDINES AND THIOREDUCTASE TAG_a_a_a_a_a_a_>

      220      230      240      250      260      270      280
      *      *      *      *      *      *      *
GGT CCG TGC AAA ATG ATC GCC CCG AIT CTG GAT GAA ATC GCT GAC GAA TAT CAG GGC AAA CTG ACC
      G P C K M I A P I L D E I A D E Y Q G K L T>
      _a_a_a_a_a_a_SIX HISTIDINES AND THIOREDUCTASE TAG_a_a_a_a_a_a_>

      290      300      310      320      330      340
      *      *      *      *      *      *
GTT GCA AAA CTG AAC ATC GAT CAA AAC CCT GGC ACT GCG CCG AAA TAT GGC ATC CGT GGT ATC CCG
      V A K L N I D Q N P G T A P K Y G I R G I P>
      _a_a_a_a_a_a_SIX HISTIDINES AND THIOREDUCTASE TAG_a_a_a_a_a_a_>

350      360      370      380      390      400      410
      *      *      *      *      *      *      *
ACT CTG CTG TTC AAA AAC GGT GAA GTG GCG GCA ACC AAA GTG GGT GCA CTG TCT AAA GGT CAG
      T L L L F K N G E V A A T K V G A L S K G Q>
      _a_a_a_a_a_a_SIX HISTIDINES AND THIOREDUCTASE TAG_a_a_a_a_a_a_>

      420      430      440      450      460      470      480
      *      *      *      *      *      *      *
TTG AAA GAG TTC CTC GAC GCT AAC CTG GCC GGT TCT GGT TCT GGT GAT GAC GAT GAC AAG CTG GTT
      L K E F L D A N L A G S G S G D D D D K>
      _a_a_a_a_a_a_SIX HISTIDINES AND THIOREDUCTASE TAG_a_a_a_a_a_a_>
      L V>
      _b_>

      >Nsp_V
      |
      >BstB_I
      |
      >Asp718      >Csp45_I
      |            |
      >Sst_I      >Acc65_I      >EcoR_I      >Hind_III
      |            |            |            |
      >Sac_I      >Pvu_II      >Kpn_I      >Sfu_I
      |            |            |            |
      >BamH_I      >Ecl136_II      >Pst_I      >Nco_I      >Asu_II
      |            |            |            |            |
      490      500      510      520      530      540      550
      *      *      *      *      *      *      *
CCG CGT GGA TCC GAGCTCGA CATCTGCAGC TGGTACCATG GAATTCGAAG CTTGATCCGG CTGCTAACAA
      P R G S>
      _THROMBIN_>
      _c_c_10_MULTIPLE CLONING SITE_c_40_c_>
  
```

C9 pGEX-6p-1

pGEX-6P-1 (27-4597-01)

PreScission™ Protease

Leu	Glu	Val	Leu	Phe	Gln	Gly	Pro	Leu	Gly	Ser	Pro	Glu	Phe	Pro	Gly	Arg	Leu	Glu	Arg	Pro	His
CTG	GAA	GTT	CTG	TTC	CAG	GGG	CCC	CTG	GGA	TCC	CCG	GAA	TTC	CCG	GGT	CGA	CTC	GAG	CGG	CCG	CAT
									BamH I		EcoR I		Sma I		Sal I		Xho I		Not I		

