

CLONING OF A NOVEL *BACILLUS PUMILUS* CELLOBIOSE-UTILISING
SYSTEM: FUNCTIONAL EXPRESSION IN *ESCHERICHIA COLI*

BY
RONÉL VAN ROOYEN

Thesis presented in partial fulfilment of the requirements for the
degree of Master of Science at the University of Stellenbosch



December 2002

Supervisor: Prof. W.H. van Zyl
Co-supervisor: Dr. D.C. la Grange

DECLARATION

I, the undersigned, hereby declare that the work contained in this thesis is my own original work and that I have not previously in its entirety or in part submitted it at any university for a degree.

Ronél van Rooyen

SUMMARY

Cellulose, a β -1,4-linked polymer of glucose, is the most abundant renewable carbon source on earth. It is well established that efficient degradation of cellulose requires the synergistic action of three categories of enzymes: endoglucanases (EG), cellobiohydrolases (CBH) and β -glucosidases. β -Glucosidases are a heterogeneous group of enzymes that display broad substrate specificity with respect to hydrolysis of cellobiose and different aryl- and alkyl- β -D-glucosides. They not only catalyse the final step in the saccharification of cellulose, but also stimulate the extent of cellulose hydrolysis by relieving the cellobiose mediated inhibition of EG and CBH. The ability to utilize cellobiose is widespread among gram-negative, gram-positive, and *Archaea* bacterial genera. Cellobiose phosphoenolpyruvate-dependent phosphotransferase systems (PTS) have been reported in various bacteria, including *Bacillus* species.

In this study, we have used a cellobiose chromophore analog, *p*-nitrophenyl- β -D-glucopyranoside (pNPG), to screen a *Bacillus pumilus* genomic library for cellobiose utilization genes that are functionally expressed in *Escherichia coli*. Cloning and sequencing of the most active clone with subsequent sequence analysis allowed the identification of four adjacent open reading frames. An operon of four genes (*celBACH*), encoding a cellobiose phosphotransferase system (PTS): enzyme II (encoded by *celB*, *celA* and *celC*) and a 6-phospho- β -glucosidase (encoded by *celH*) was derived from the sequence data. The amino acid sequence of the *celH* gene displayed good homology with β -glucosidases from *Bacillus halodurans* (74.2%), *B. subtilis* (72.7%) and *Listeria monocytogenes* (62.2%). As implied by sequence alignments, the *celH* gene product belongs to family 1 of the glycosyl hydrolases, which employ a retaining mechanism of enzymatic bond hydrolysis.

In vivo PTS activity assays concluded that the optimal temperature and pH at which the recombinant *E. coli* strain hydrolysed pNPG were pH 7.5 and 45°C, respectively. Unfortunately, at 45°C the CelBACH-associated activity of the recombinant strain was only stable for 20 minutes. It was also shown that the enzyme complex is very sensitive to

glucose. Since active growing cells metabolise glucose very rapidly this feature is not a significant problem.

Constitutive expression of the *B. pumilus celBACH* genes in *E. coli* enabled the host to efficiently metabolise cellobiose as a carbon source. However, cellobiose utilization was only achievable in the presence of 0.01% glucose. This phenomenon could be explained by the critical role of phosphoenolpyruvate (PEP) as the phosphate donor in PTS-mediated transport. Glucose supplementation induced the glycolytic pathway and subsequently the availability of PEP. Furthermore, it could be concluded that the general PTS components (enzyme I and HPr) of *E. coli* must have complemented the CelBACH system from *B. pumilus* to allow functionality of the *celBACH* operon in the recombinant *E. coli* host.

OPSOMMING

Sellulose ('n polimeer van β -1,4-gekoppelde glukose) is die volopste bron van hernubare koostof in die natuur. Effektiewe afbraak van sellulose word deur die sinnergistiese werking van drie ensiemklasse bewerkstellig: endoglukanases (EG), sellobiohidrolases (CBH) en β -glukosidases. β -Glukosidases behoort tot 'n heterogene groep ensieme met 'n wye substraatspesifisiteit m.b.t. sellobiose en verskeie ariel- and alkiel- β -D-glukosidiese verbindings. Alhoewel hierdie ensieme primêr as kataliste vir die omskakeling van sellulose afbraak-produkte funksioneer, stimuleer hulle ook die mate waartoe sellulose hidroliese plaasvind deur eindproduk-inhibisie van EG en CBH op te hef. Sellobiose word algemeen deur verskeie genera van die gram-negatiewe, gram-positiewe en *Archae* bakterieë gemetaboliseer. Die sellobiose-spesifieke fosfoenolpirovaat-fosfotransportsisteam (PTS) is reeds in verskeie bakterië, insluitende die *Bacillus* spesies, beskryf.

In hierdie studie word die sifting van 'n *Bacillus pumilus* genoombiblioteek m.b.v. 'n chromofoor analoog van sellobiose, *p*-nitrofeniel- β -D-glukopiranosied (pNPG), vir die teenwoordigheid van gene wat moontlike sellobiose-benutting in *Escherichia coli* kan bewerkstellig, beskryf. Die DNA-volgorde van die mees aktiewe kloon is bepaal en daaropvolgende analiese van die DNA-volgorde het vier aangrensende ooplesrame geïdentifiseer. 'n Operon (*celBACH*), bestaande uit vier gene, wat onderskeidelik vir die ensiem II (gekodeer deur *celB*, *celA* en *celC*) en fosfo- β -glukosidase (gekodeer deur *celH*) van die sellobiose-spesifieke PTS van *B. pumilus* kodeer, is vanaf die DNA-volgorde afgelei. Die aminosuuropeenvolging van die *celH*-geen het goeie homologie met β -glukosidases van *Bacillus halodurans* (74.2%), *B. subtilis* (72.7%) en *Listeria monocytogenes* (62.2%) getoon. Belyning van die DNA-volgordes het aangedui dat die *celH* geenproduk saam met die familie 1 glikosielhidrolases gegroepeer kan word. Hierdie familie gebruik 'n hidrolitiese meganisme waartydens die stoigiometriesse posisie van die anomeriese koolstof behou word.

PTS-aktiwiteit van die rekombinante *E. coli* ras, wat die *celBACH* gene uitdruk, is *in vivo* bepaal. Die optimale temperatuur en pH waarby die rekombinante ras pNPG hidroliseer, is onderskeidelik pH 7.5 en 45°C. Alhoewel die ensiemkompleks baie sensitief is vir glukose, is dit nie 'n wesenlike probleem nie, omdat aktief groeiende *E. coli* selle glukose teen 'n baie vinnige tempo benut.

Die *celBACH* operon het onder beheer van 'n konstitiewe promotor in *E. coli* die rekombinante gasheer in staat gestel om sellobiose as 'n koolstofbron te benut. Die benutting van sellobiose word egter aan die teenwoordigheid van 'n lae konsentrasie glukose (0.01%) gekoppel. Hierdie verskynsel dui op die kritiese rol van fosfoenolpirovaat (PEP) as die fosfaatdonor gedurende PTS-gebaseerde transport. Glukose speel waarskynlik 'n rol in die indusering van glikoliese, en sodoende die produksie van PEP as tussenproduk. Verder kan afgelei word dat die algemene PTS komponente (ensiem I en HPr) van *E. coli* die *B. pumilis* CelBACH-sisteem komplementeer en derhalwe funksionering van die *celBACH* operon in *E. coli* toelaat.

BIOGRAPHICAL SKETCH

Ronél van Rooyen was born in Lichtenburg, South Africa, on 2 January 1976. She attended the Lichtenburg Primary School and matriculated at the Lichtenburg High School, in 1994. Ronél enrolled at the University of Stellenbosch in 1995 and obtained a B.Sc.Agric degree in Biochemistry and Microbiology 1998.

I dedicate this thesis to my parents

ACKNOWLEDGMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions for their invaluable contributions to the successful completion of this study:

Prof. W.H. van Zyl, Department of Microbiology, University of Stellenbosch, who acted as supervisor, for accepting me as a student and always believing in me, and for his invaluable guidance, enthusiasm, encouragement and insight throughout this project;

My colleagues in the laboratory, particularly **Danie la Grange** (co-supervisor), **Annette Plüddemann** and **Riaan den Haan** for their support and encouragement and all their advice;

The staff at the Department of Microbiology, for their assistance;

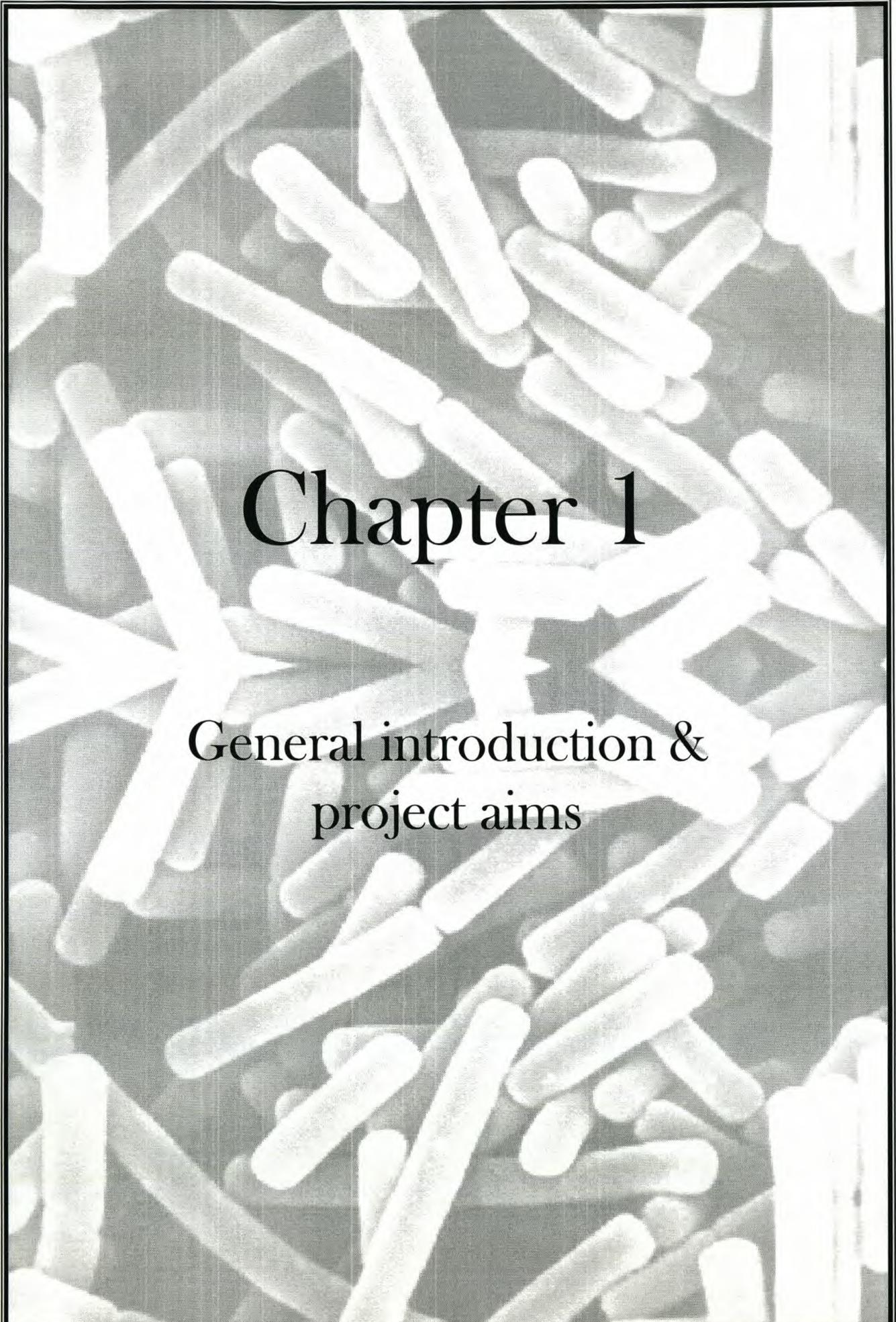
My parents and family for all their love and support and their endless belief in my abilities;

The National Research Foundation (NRF) for financial assistance;

Our Almighty God, for this opportunity and the privilege to be a scientist.

1	INTRODUCTION.....	1
1.1	AIMS OF THIS STUDY.....	2
1.2	REFERENCES.....	2
2	CELLULOSE DEGRADATION.....	4
2.1	THE STRUCTURE AND DISTRIBUTION OF CELLULOSE.....	4
2.2	CELLULOLYTIC ENZYMES.....	6
2.2.1	<i>Cellobiohydrolases (CBH)</i>	7
2.2.2	<i>Endoglucanases (EG)</i>	8
2.2.3	<i>β-Glucosidases</i>	8
2.2.3.1	Classification of Glycosyl Hydrolases.....	9
2.2.3.2	Structures and mechanisms of Glycosyl Hydrolases.....	12
2.2.3.3	Properties of Glycosyl Hydrolases.....	14
2.3	INTERACTIONS AND ASSOCIATIONS OF CELLULOLYTIC ENZYMES.....	17
2.3.1	<i>Complexed Systems</i>	17
2.3.2	<i>Non-complexed Systems</i>	19
2.4	BACTERIAL SUGAR TRANSPORT AND UTILISATION SYSTEMS.....	22
2.4.1.1	Components of the PTS.....	22
2.4.1.2	PTS operons.....	25
2.4.2	<i>The role of the PTS in energy and signal transduction</i>	26
2.4.3	<i>PTS Genes from Bacillus spp.</i>	29
2.5	INDUSTRIAL IMPORTANCE OF β -GLUCOSIDASE ENZYMES.....	33
2.5.1	<i>Bioethanol Production</i>	33
2.5.2	<i>Flavour and Wine-making Industries</i>	33
2.5.3	<i>Metabolic Engineering of Bacteria for the Conversion of cellulose degradation products to ethanol</i>	35
2.6	REFERENCES.....	37
3	CLONING OF A NOVEL <i>BACILLUS PUMILUS</i> CELLOBIOSE UTILISING SYSTEM: FUNCTIONAL EXPRESSION IN <i>ESCHERICHIA COLI</i>.....	45
3.1	ABSTRACT.....	45
3.2	INTRODUCTION.....	45
3.3	MATERIALS AND METHODS.....	47
3.3.1	<i>Bacterial Strains and plasmids</i>	47
3.3.2	<i>Media for bacterial growth</i>	48
3.3.3	<i>Cloning of the <i>B. pumilus</i> celBACH operon</i>	48

3.3.4	<i>Polymerase chain reaction</i>	48
3.3.5	<i>Plasmid constructions</i>	49
3.3.6	<i>DNA sequencing and sequence analysis</i>	50
3.3.7	<i>In vivo assay of PTS activity</i>	50
3.4	RESULTS.....	51
3.4.1	<i>Cloning of the cel operon</i>	51
3.4.2	<i>DNA sequence analysis</i>	51
3.4.3	<i>Identification of cloned cel genes by homology</i>	53
3.4.4	<i>Characterisation of the CelBACH complex</i>	54
3.5	FUNCTIONAL ANALYSIS OF GENES IN THE <i>B. PUMILUS CEL</i> OPERON IN <i>E. COLI</i>	55
3.6	DISCUSSION.....	56
3.7	ACKNOWLEDGEMENTS	60
3.8	REFERENCES.....	60
4	GENERAL DISCUSSION AND CONCLUSIONS	63
4.1	IDENTIFICATION AND ISOLATION OF GENES FOR A NOVEL <i>B. PUMILUS</i> CELLOBIOSE-DEPENDENT PHOSPHOTRANSFERASE SYSTEM (PTS)	63
4.2	GENE TRANSCRIPTION IN <i>ESCHERICIAE COLI</i>	64
4.3	ASSESSMENT OF THE RECOMBINANT <i>E. COLI</i> STRAIN FOR THE ABILITY TO UTILIZE CELLOBIOSE AS A CARBON SOURCE.....	64
4.4	CONCLUSIONS	65
4.5	FUTURE RESEARCH.....	66
4.6	REFERENCES.....	67



Chapter 1

General introduction &
project aims

1 INTRODUCTION

Cellulose is the most abundant biopolymer on earth, and its microbial degradation is a key process in carbon cycling. An estimated rate of cellulose synthesis is approximately 4×10^9 tons per year (Coughlan, 1990). Aerobic biodegradation of cellulose is performed predominantly by fungi, whereas, in anaerobic environments, bacteria are the main cellulose degraders (Tomme *et al.*, 1995). The chemical structure of cellulose is simple, consisting of D-glucose residues linked by β -1,4-glucosidic bonds to form a linear polymer. Although highly crystalline, the structure of cellulose is not uniform and contains both highly crystalline and less ordered amorphous regions (Hon, 1994). Usually, native cellulose displays about 70% crystallinity (Gardner and Blackwell, 1974).

Cellulolytic organisms typically produce endoglucanase, cellobiohydrolase, and β -glucosidase enzymes that interact synergistically to degrade cellulose (Béguin, 1990; Clarke, 1997; Henrissat, 1994; Teeri, 1997; Tomme *et al.*, 1995; Wood, 1991). Endoglucanase and exocellobiohydrolase act cooperatively on cellulose to produce cellobiose (Väljamäe *et al.*, 1999, 1998), which is then cleaved by β -glucosidase to glucose. Both endoglucanase and cellobiohydrolase activities are inhibited by cellobiose and the latter has to be removed by enzymatic activity (Wright *et al.*, 1992; Kadam and Demain, 1989, Coughlan, 1985; Woodward and Wiseman, 1982). This disaccharide is among the most abundant soluble substrates for microbial growth in nature and is assimilated in preference to glucose by some rumen microorganisms (Thurston *et al.*, 1993; Helaszek and White, 1991).

The widespread ability among gram-positive and gram-negative, as well as *Archaea*, to metabolize cellobiose is generally accepted (Coughlan and Mayer, 1992). Bacterial cellobiase (β -glucosidase) activity is typically cell associated, although, multiple pathways for cellobiose utilization can exist within a single microorganism (El Hassouni, 1992, Hall and Xu, 1992; Wu and Saier, 1990). The three commonly known pathways are: (i) extracellular hydrolysis of cellulose by β -glucosidase and subsequent transport of the glucose residues via ATP-dependent active transport; (ii) transport of cellobiose via the cellobiose-specific phosphotransferase system (PTS) and subsequent intracellular hydrolysis by phospho- β -glucosidase; and (iii) direct uptake of cellobiose via an active

ATP-dependent transport system and phosphorylytic cleavage of the molecule by cellobiose phosphorylase.

1.1 Aims of this study

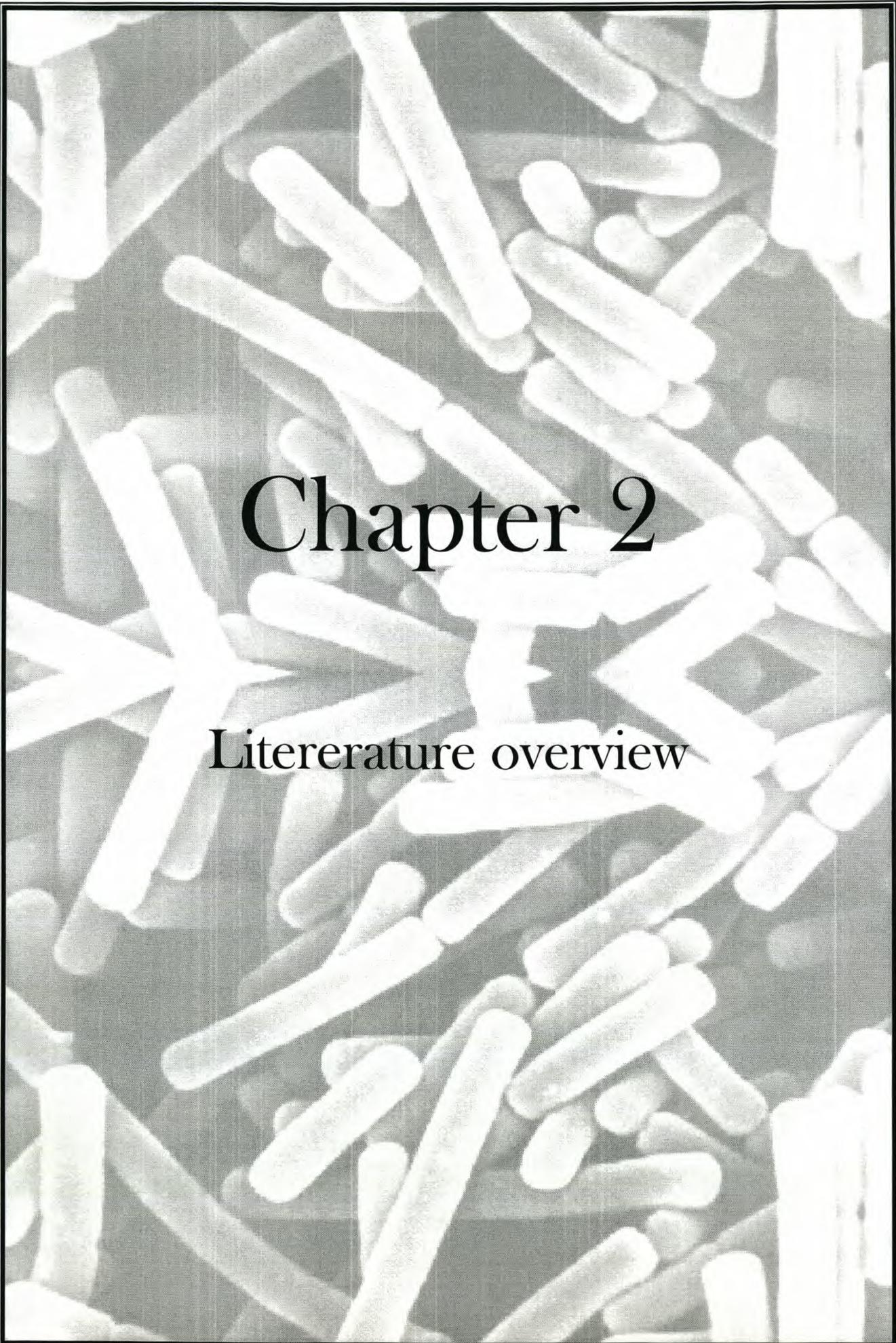
The aim of this study was to explore the cellobiose-utilising system of the gram-positive bacterium, *B. pumilus*. The specific aims of the present study were as follows:

- ◆ identification of a novel *Bacillus pumilus* cellobiose-utilising system;
- ◆ isolation of the *B. pumilus* gene(s) encoding the protein(s) required for cellobiose metabolism and functional expression in *Escherichia coli*;
- ◆ *in vivo* characterization of the cellobiose-associated enzyme activities in recombinant *E. coli*;
- ◆ assessment of the recombinant *E. coli* strain for the ability to utilize cellobiose as a carbon source.

1.2 References

- Béguin, P. 1990. Molecular biology of cellulose degradation. *Annu.Rev. Microbiol* **44**:219–248.
- Clarke, A.J. 1997. Biodegradation of cellulose: enzymology and biotechnology. Lancaster, PA: Technomic.
- Coughlan, M. P. 1985. The properties of fungal and bacterial cellulases with comment on their production and application. *Biotechnol. Genet. Eng. Rev.* **3**:39–109.
- Coughlan, M. P., and F. Mayer. 1992. The cellulose-decomposing bacteria and their enzyme systems, p. 460–516. In A. Balows, H. G. Trüper, M. Dworkin, W. Harder, and K. Schleifer (ed.), *The prokaryotes*, 2nd ed. A handbook on the biology of bacteria: ecophysiology, isolation, identification, applications, vol. I. Springer-Verlag, Berlin, Germany.
- Coughlan, M.P. 1990. Cellulose degradation by fungi. p 1-36. In W.M. Fogarty and C.T. Kelly (ed.), *Microbial enzymes and biotechnology*. Elsevier Applied Science, London.
- El Hassouni, M., B. Henrissat, M. Chippaux, and F. Barras. 1992. Nucleotide sequences of the *arb* genes, which control β -glucoside utilization in *Erwinia chrysanthemi*: comparison with the *Escherichia coli* *bgl* operon and evidence for a new β -glycohydrolase family including enzymes from eubacteria, archeobacteria, and humans. *J. Bacteriol.* **174**:765–777.
- Gardner, H.K., and J. Blackwell. 1974. The structure of native cellulose. *Biopolymers* **13**:1975-2001.
- Hall, B. G., and L. Xu. 1992. Nucleotide sequence, function, activation, and evolution of the cryptic *asc* operon of *Escherichia coli* K12. *Mol. Biol. Evol.* **9**:688–706.

- Helaszek, C. T., and B. A. White.** 1991. Cellobiose uptake and metabolism by *Ruminococcus flavefaciens*. *Appl. Environ. Microbiol.* **57**:64–68.
- Henrissat, B.** 1994. Cellulases and their interaction with cellulose. *Cellulose* **1**:169–196.
- Hon, S.-N.D.** 1994. Cellulose: a random walk along its historical path. *Cellulose* **1**:1-25.
- Kadam, S. K., and A. L. Demain.** 1989. Addition of cloned β -glucosidase enhances the degradation of crystalline cellulose by *Clostridium thermocellum* cellulase complex. *Biochem. Biophys. Res. Commun.* **161**:706–711.
- Teeri, T.T.** 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends Biotechnol.* **15**:160–167.
- Thurston, B., K. A. Dawson, and H. J. Strobel.** 1993. Cellobiose versus glucose utilization by the ruminal bacterium *Ruminococcus albus*. *Appl. Environ. Microbiol.* **59**:2631–2637.
- Tomme, P., R.A.J. Warren, and N.R. Gilkes.** 1995. Cellulose hydrolysis by bacteria and fungi. *Adv. Microbial Physiol.* **37**:1–81.
- Väljamäe, P., V.V. Sild, A. Nutt, G. Pettersson, and G. Johansson.** 1999. Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. *Eur. J. Biochem.* **266**:327-334.
- Väljamäe, P; Sild, V; Pettersson, G; Johansson, G.** 1998. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. *Eur.J.Biochem.* **253**:469-475.
- Wood TM.** 1991. Fungal cellulases. p 491–533, In: Haigler CH, Weimer PJ (ed.). *Biosynthesis and biodegradation of cellulose*. New York: Marcel Dekker.
- Woodward, J., and A. Wiseman.** 1982. Fungal and other β -glucosidases: their properties and applications. *Enzyme Microb. Technol.* **4**:73–79.
- Wright, R.M., M.D. Yablonsky, Z.P. Shalita, A.K. Goyal, and D. E. Eveleigh.** 1992. Cloning, characterization, and nucleotide sequence of a gene encoding *Microbispora bispora* BglB, a thermostable β -glucosidase expressed in *Escherichia coli*. *Appl. Environ. Microbiol.* **57**:3455–3465.
- Wu, L.-F., and M. H. Saier, Jr.** 1990. Nucleotide sequence of the *fruA* gene, encoding the fructose permease of *Rhodobacter capsulatus* phosphotransferase system, and analyses of the deduced protein sequence. *J. Bacteriol.* **172**:7167–7178.



Chapter 2

Litererature overview

2 CELLULOSE DEGRADATION

Plants synthesize about 4×10^9 tons of cellulose annually (Cox *et al.*, 1999, Coughlan, 1990). Most of this material does not accumulate because fungi and bacteria efficiently degrade plant biomass as a source of energy and carbon, ultimately recycling carbon dioxide into the ecosystem. Previously, fundamental studies on cellulolytic microorganisms were aimed at the prevention of microbial attack on plants and plant material of importance, whereas in the last two decades, more attention is focussed on their biotechnological application for the production of fermentable sugars. In order to understand the mechanisms of cellulose degradation by cellulases, it is important to realise that although chemically simple, the properties of cellulose are more complex than simply a homopolymer of β -1,4-linked glucose units would indicate. The structure and function of cellulolytic systems are determined at both the genetic level of cellulose production and protein-substrate level of synergistic cooperation.

2.1 The structure and distribution of cellulose

Cellulose fulfills a vital structural role in plants and other living systems. The cellulose content in plant cells is relatively constant across all species and represents 40-50% of cell wall substances (Kubicek *et al.*, 1993; Coughlan, 1990). Each cellulose molecule is an unbranched polymer of 1,000 to 1 million D-glucose residues, linked together with β -1,4-glycosidic bonds (Fig. 1). The degree of polymerisation, i.e. the number of glucose units included in a cellulose chain, is generally in the range of 7 500 to 15 000 for plant cellulose. Cellulose fibers are embedded in an amorphous matrix of other structural biopolymers (mainly hemicelluloses and lignin) with the exception of a few instances (e.g. cotton balls), where it is present in an almost pure state. Although these matrix interactions differ with plant cell type and with maturity, they are a major structural feature limiting the rate and degree of utilisation of intact, crude biomass materials (Lynd *et al.*, 2002).

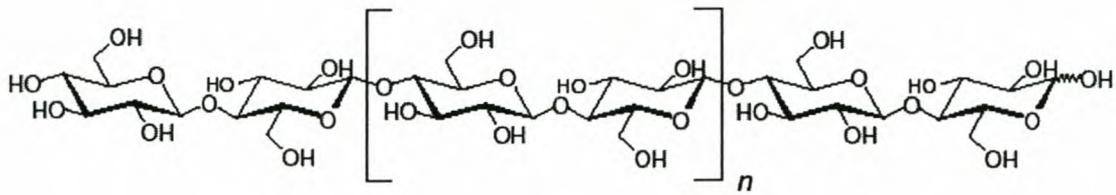


FIGURE 1. Schematic illustration of the structure of a cellulose chain (n = repeat cellobiose unit).

A relatively unique and important characteristic of cellulose is its crystalline structure. In nature, cellulose is synthesised as individual molecules which go through a process of self-assembly at the site of biosynthesis (Saxena and Brown, 2000). In natural cellulose, the cellulose molecules have a parallel orientation with all reducing chain ends to the one side. About thirty individual molecules are assembled into larger units known as elementary fibrils (proto-fibrils), which are consecutively packed into larger rod-like units called micro-fibrils. These micro-fibrils are associated through hydrogen and van der Waals bonds, forming a very rigid macromolecular structure (Fig. 2).

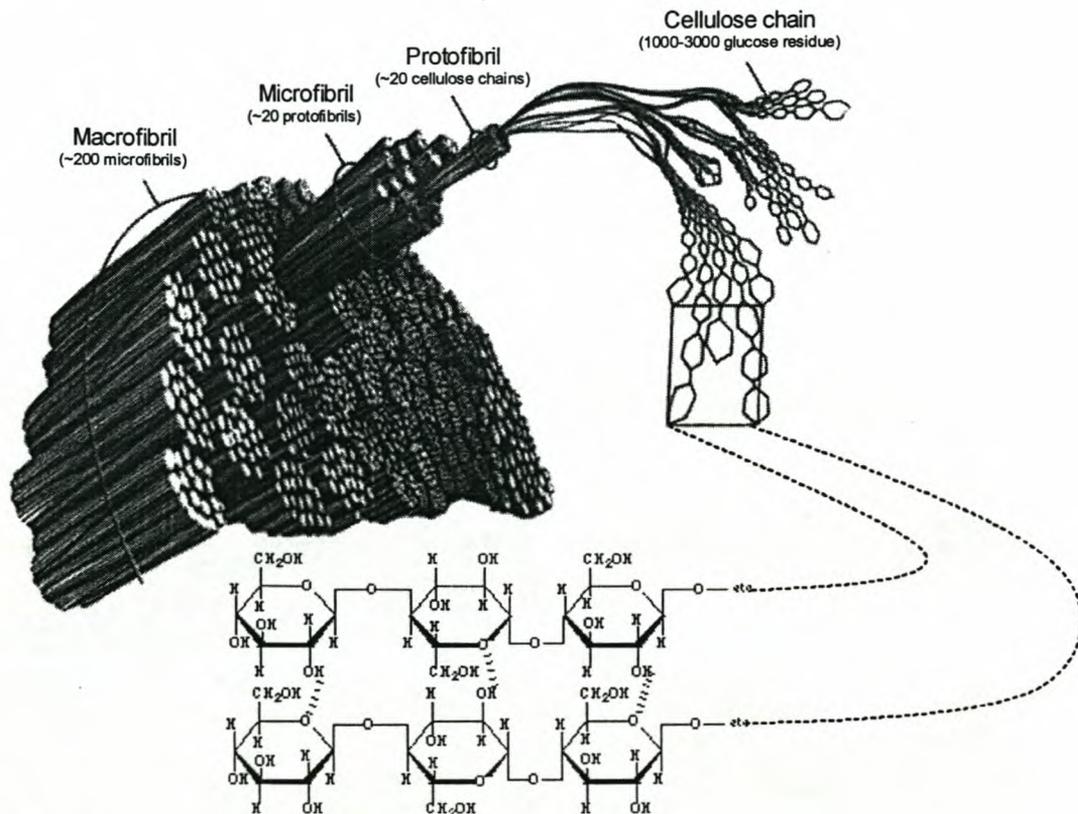


FIGURE 2. The macromolecular structure of cellulose.

In cellulose I (the most abundant allomorph of cellulose in nature), the neighbouring microfibrils are staggered to ensure the highest degree of hydrogen bonding and result in the formation of highly ordered crystalline regions (Bayer *et al.*, 1998). Although highly crystalline, the structure of cellulose is not uniform. A significant amount of physical and chemical evidence indicates that native cellulose contains both highly crystalline and less ordered amorphous (or paracrystalline) regions (Hon, 1994). An important feature of the crystalline structure is its ability to prevent the penetration of water or enzymes, and therefore explains its insolubility. On the other hand, amorphous cellulose, allows the penetration of endoglucanases, another subgroup of cellulases, that catalyses the hydrolysis of internal bonds. The natural consequence of this difference in the crystalline structure is that the rate of hydrolysis for amorphous cellulose is much faster than for crystalline cellulose. In addition to crystallinity, the protective biopolymers surrounding the cellulose in plants, also limit the diffusion of enzymes into the reaction sites and therefore play an important role in determining the rate of hydrolysis (Woodward *et al.*, 1988). This illustrates only some of the complexities faced by microorganisms that hydrolyse cellulose.

2.2 Cellulolytic Enzymes

Cellulases are a group of enzymes secreted by a wide variety of fungal and bacterial species and randomly attack and hydrolyse the β -1,4-glycosidic bonds of cellulose to produce cello-oligosaccharides. Depending on their site of action on cellulose, they have traditionally been divided into two groups: i) cellobiohydrolases (or exoglucanases) and ii) endoglucanases (Bayer *et al.*, 1998). However, efficient degradation of cellulose requires the synergistic action of at least three distinct enzymes: Cellobiohydrolases (exoglucanases; EC 3.2.1.91); endoglucanases (EC 3.2.1.4), and β -glucosidases (EC 3.2.1.74) (Clarke, 1997; Teeri, 1997; Tomme *et al.*, 1995). Fig. 3 illustrates the concerted action of these enzymes.

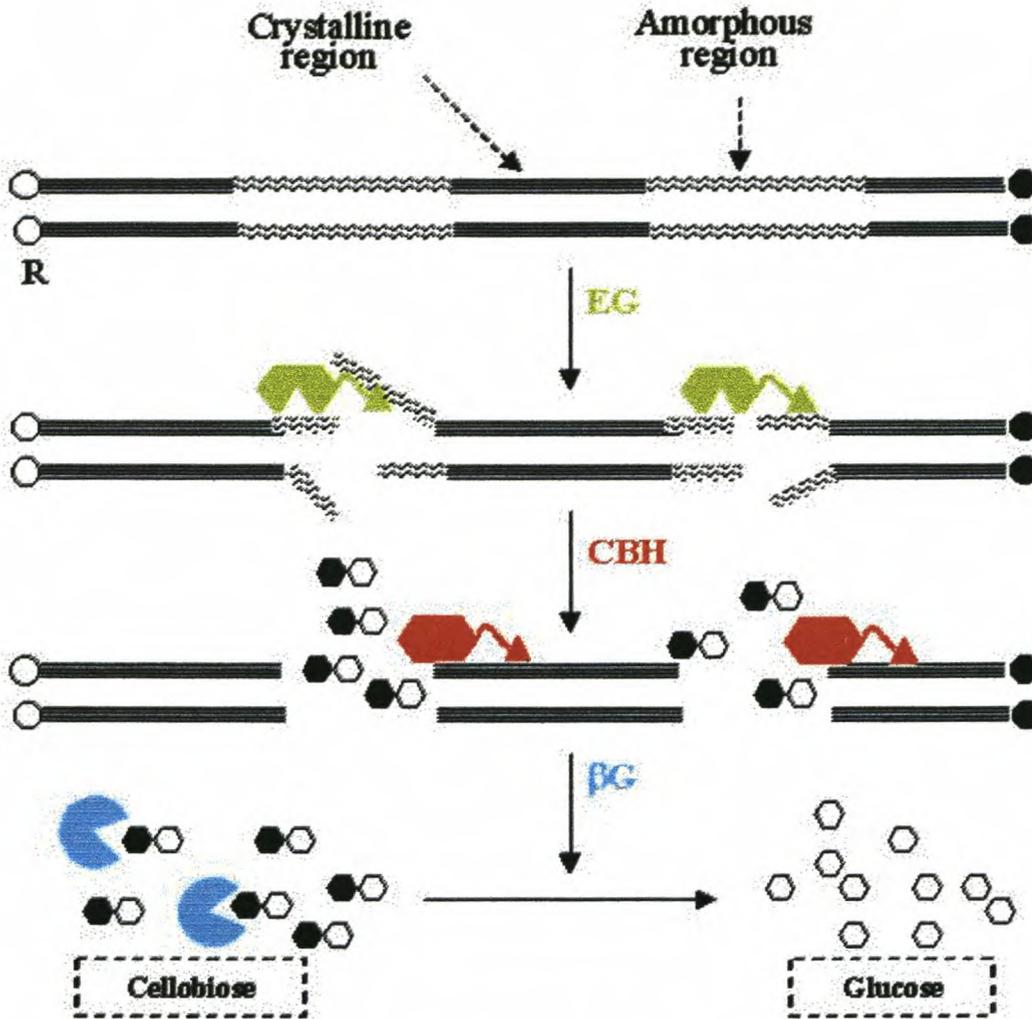


FIGURE 3. A simplified model of the hydrolysis of cellulose by the combined action of three types of enzymatic activities: Endoglucanases (green), acting at random in amorphous regions and producing new ends; cellobiohydrolases (red), that liberate cellobiose/cellodextrins from the crystalline regions in a processive manner; and β -glucosidases (blue), which release D-glucose units from soluble cellodextrins.

2.2.1 Cellobiohydrolases (CBH)

Cellobiohydrolases release cellobiose units predominantly from the chain end and degrade preferentially crystalline cellulose in a processive manner. Cellobiohydrolases appear to be the main enzyme responsible for the degradation of crystalline cellulose (Teeri, 1997). Also, the mode of digestion of crystalline substrates is related to the specific three-dimensional structure of cellobiohydrolases. Several structural studies revealed that the catalytic site of these enzymes is covered by long loops, resulting in a tunnel morphology (Varrot *et al.*, 1999; Divne *et al.*, 1998). The loops can undergo large movements, leading to the closing or opening of the tunnel roof. When the roof is open, an endo type of attack

of the polymeric substrate becomes possible. Once entrapped inside the catalytic tunnel, a cellulose chain is threaded through the tunnel and sequentially hydrolysed, one cellobiose unit at a time. This mechanism which is specific to cellobiohydrolases, relies on one of the fundamental characteristics of these enzymes. Once initiated at the reducing or non-reducing end of a cellulose chain, it moves along processivity (Tomme *et al.*, 1996).

2.2.2 Endoglucanases (EG)

There are two sites in the enzymes (CBH and EG) that mediate binding: The active site of the catalytic domain and the separately folded and functionally independent CBM (cellulose binding modules). The catalytic domain and CBM are connected via a linker (commonly proline-threonine-serine- or PTS-boxes) that acts as a flexible arm which allows the catalytic domain (limited) movement while the CBM is fixed to the cellulose surface (Schwarz, 2001). Endoglucanases randomly attack bonds in amorphous regions on the microfibril surface of cellulose to generate new chain ends for attack by exo-acting cellobiohydrolases. In turn, the erosion of the microfibril surface by cellobiohydrolases exposes further amorphous regions for endoglucanase attack. This is the so-called “endo-exo synergy” described in numerous studies (Srisodsuk *et al.*, 1997; Teeri, 1997; Ooshima *et al.*, 1991). In addition to the endo-exo synergism, a so-called exo-exo synergism between the CBHI and CBHII from *T. reesei* was also observed (Medve *et al.*, 1998). This phenomenon is discussed in more detail under the heading “Non-complexed systems”.

2.2.3 β -Glucosidases

β -Glucosidases are defined as enzymes that hydrolyse compounds containing β -glucosidic linkages (bond between a sugar and an alcohol) by splitting of the terminal, non-reducing β -D-glucose residues and releasing β -D-glucose (Sternberg *et al.*, 1977). Prior to discussing the numerous properties of the different types of β -glucosidases, it is important to begin with a general introduction to the classification of enzymes and enzyme nomenclature with specific focus on the subclass of glycosyl hydrolases of which β -glucosidases are a member.

2.2.3.1 Classification of Glycosyl Hydrolases

During the late 1950's, the International Union of Biochemistry and Molecular Biology (IUBMB) (formerly the International Union of Biochemistry) set up an International Commission on Enzymes to bring order into the general nomenclature of enzymes, or into that of particular groups of enzymes during a period when the number of known enzymes was increasing rapidly. Their classification system was based on grouping enzymes according to the type of reaction catalysed. Since the publication of the first edition of Enzyme Nomenclature (1961), the Enzyme List has increased steadily from 712 to 3196 entries in the latest publication (sixth edition of Enzyme Nomenclature, 1992). This system for classification of enzymes also served as a basis for assigning code numbers to them, generally known as the EC number of a specific enzyme. The code numbers, prefixed by EC, contain four digits separated by points, with the following meaning: (i) The first digit shows to which of the six main divisions (classes) the enzyme belongs, as illustrated in Fig. 4; (ii) the second digit indicates the subclass; (iii) the third digit gives the sub-subclass; (iv) The fourth digit is the serial number of the enzyme in its sub-subclass.

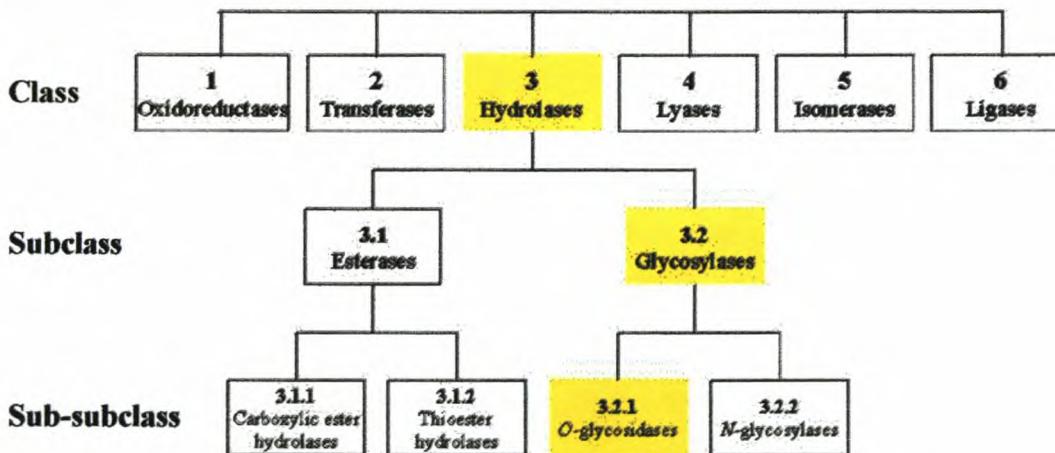


FIGURE 4. Schematic representation of the enzyme classification system

More accurately for the hydrolases (class 3), the second digit in the EC number indicates the nature of the bond hydrolysed; EC 3.1 is the *esterases*; EC 3.2 the *glycosylases*, etc. Whereas, the third digit specifies the nature of the substrate, e.g. *O*- or *S*-glycosyl compounds (EC 3.2.1 and EC 3.2.3) or *N*-glycosyl compounds (EC 3.2.2).

For the purpose of this review, only the *O*-glycosyl hydrolases (EC 3.2.1.) will be discussed. *O*-Glycosyl hydrolases (or β -glycosidases) are defined as enzymes that can

hydrolyse β -1,4-glycosidic bonds between two or more sugars or between a sugar and non-sugar moiety. Over 2000 β -glycosidases (which include cellulases, cellobiohydrolases and β -glucosidases) from a diversity of microorganisms have been characterised. At first, the genes and proteins were named randomly as they were identified, and given the appropriate designations according to the IUBMB Enzyme Nomenclature. As previously mentioned, the IUBMB Enzyme Nomenclature of glycosyl hydrolases are based on their substrate specificity and seldom on their molecular mechanism. Therefore, this classification system did not reflect the structural features of the enzymes. More recently, a classification system based on amino acid sequence similarities and catalytic domains was established (Henrissat and Bairoch, 1996; Henrissat *et al.*, 1995; Henrissat and Bairoch, 1993; Henrissat, 1991; Henrissat *et al.*, 1989). According to this classification system β -glycosidases have been assigned to over 77 different families. In Table 1 only the families that include the cellulolytic and xylanolytic hydrolases (Clarke, 1997; Henrissat and Bairoch, 1996) are listed. Since there is a direct relationship between the amino acid sequence and the folding of an enzyme, such a classification is expected to: (i) reflect the structural features of these enzymes more accurately than substrate specificity alone; (ii) help to expose the evolutionary relationships between these enzymes; and (iii) provide an accessible tool to obtain mechanistic information from the protein sequence data (Henrissat and Bairoch, 1993; Henrissat, 1991).

TABLE 1. Classification of Glycosyl Hydrolases (SWISS-PROT entries; <http://expasy.org>).

Family (Clan)	Members	EC Number	Taxonomic range
1 (GH-A)	β -Glucosidases	3.2.1.21	Eukaryota
	6-Phospho- β -galactosidases	3.2.1.85	Bacteria
	6-Phospho- β -glucosidases	3.2.1.86	Archaea
	Lactase-phlorizin hydrolases	3.2.1.108	
	Myrosinases	3.2.3.1	
3	β -Glucosidases	3.2.1.21	Eukaryota Bacteria
5 (GH-A)	Endoglucanases	3.2.1.4	Eukaryota
	β -Mannanases	3.2.1.78	Bacteria
	Exo-1,3-glycanases	3.2.1.58	
3	β -Glucosidases	3.2.1.21	Eukaryota Bacteria
5 (GH-A)	Endoglucanases	3.2.1.4	Eukaryota
	β -Mannanases	3.2.1.78	Bacteria
	Exo-1,3-glycanases	3.2.1.58	
6	Endoglucanases	3.2.1.4	Eukaryota
	Cellobiohydrolases	3.2.1.91	Bacteria
7 (GH-B)	Endoglucanases	3.2.1.4	Eukaryota
	Cellobiohydrolases	3.2.1.91	

TABLE 1. Classification of Glycosyl Hydrolases (continued).

Family (Clan)	Members	EC Number	Taxonomic range
7 (GH-B)	Endoglucanases	3.2.1.4	Eukaryota
	Cellobiohydrolases	3.2.1.91	
8	Endoglucanases	3.2.1.4	Bacteria
9	Endoglucanases	3.2.1.4	Eukaryota
			Bacteria
10 (GH-A)	Xylanases	3.2.1.8	Eukaryota
			Bacteria
11 (GH-C)	Xylanases	3.2.1.8	Eukaryota
			Bacteria
12 (GH-C)	Endoglucanases	3.2.1.4	Eukaryota
			Bacteria
16 (GH-B)	β -Glucanases	3.2.1.73	Bacteria
17 (GH-A)	Glucan endo-1,3- β -glucosidases	3.2.1.39	Eukaryota (Plants & fungi)
	Lichenases	3.2.1.73	
26 (GH-A)	Mannan endo-1,4- β -mannosidases	3.2.1.78	Eukaryota
			Bacteria
39 (GH-A)	β -Xylosidases	3.2.1.37	Eukaryota (Mammalian)
	α -l-Iduronidases	3.2.1.76	Bacteria
43 (GH-F)	Bifunctional	3.2.1.37/	Eukaryota
	β -xylosidases/ α -l-arabinofuranosidases	3.2.1.55	Bacteria
44	Endoglucanases	3.2.1.4	Bacteria
45	Endoglucanases	3.2.1.4	Eukaryota
			Bacteria
48	Endoglucanases	3.2.1.4	Bacteria
	Cellobiohydrolases	3.2.1.91	
52	β -Xylosidases	3.2.1.37	Bacteria
55	Exo-1,3-glucanases	3.2.1.58	Eukaryota
	Endo-1,3-glucanases	3.2.1.39	
61	Endoglucanases	3.2.1.4	Eukaryota
64	Endo-1,3-glucanases	3.2.1.39	Bacteria

Given that the catalytic domains of certain families are better conserved than their peptide sequences, some of the families can be grouped in "clans" (GH families) as indicated in Table 2. Because some glycosyl hydrolases are multi-functional enzymes, it is also possible that a specific enzyme complex could contain catalytic domains that belong to different GH families.

TABLE 2. The GH families (clans) (<http://ca.expasy.org>).

Name	Families belonging to the clan
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 51, 53
GH-B	7, 16
GH-C	11, 12
GH-D	27, 36
GH-E	33, 34
GH-F	43, 62
GH-G	37, 63
GH-H	13, 70

2.2.3.2 Structures and mechanisms of Glycosyl Hydrolases

The abundance of information provided by the latest structure determinations of many different glycosyl hydrolases confirms that the substrate specificity and the mode of action of these enzymes are regulated by detailed three-dimensional structures rather than by their overall folds. Enzymatic hydrolysis of the glycosidic bond proceeds via general catalysis that involves two critical residues: a proton donor and a nucleophile/base (Sinnott, 1990). The hydrolysis occurs via two major mechanisms that result in either an overall retention, or an inversion, of anomeric configuration (Fig. 5). In both the retaining and the inverting mechanisms, the position of the proton donor is identical, suggesting that it is within hydrogen-bonding distance of the glycosidic oxygen. In retaining enzymes, the nucleophilic catalytic base is in close vicinity of the sugar anomeric carbon. This base, however, is more distant in inverting enzymes, which have to accommodate a water molecule between the base and the sugar. This difference results in an average distance between the two catalytic residues of $\sim 5.5 \text{ \AA}$ in retaining enzymes as opposed to $\sim 10 \text{ \AA}$ in inverting enzymes (McCarter and Withers, 1994).

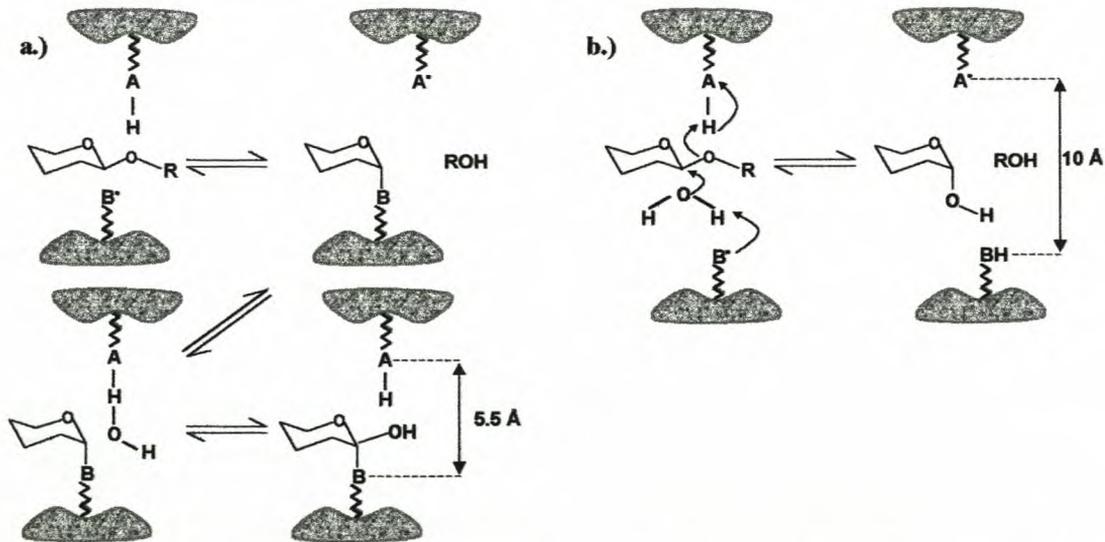


FIGURE 5. The two major mechanisms of enzymatic glycosidic bond hydrolysis (Davies and Henrissat, 1995) (a) The retaining mechanism, in which the glycosidic oxygen is protonated by the acid catalyst (AH) and nucleophilic assistance to aglycon departure is provided by the base B^- . A water molecule hydrolyses the resulting glycosyl enzyme and this second nucleophilic substitution at the anomeric carbon generates a product with the same stereochemistry as the substrate. (b) The inverting mechanism, in which protonation of the glycosidic oxygen and aglycon removal are accompanied by a simultaneous attack of a water molecule that is activated by the base residue (B^-). This single nucleophilic substitution yields a product with opposite stereochemistry to the substrate.

Table 3 lists several of the significant glycosyl hydrolase families for which at least one three-dimensional (3-D) structure has been determined, together with the mechanism of glycosidic bond hydrolysis. Many glycosyl hydrolases have a modular structure consisting of a catalytic domain and one or more non-catalytic domains, some of which are concerned with substrate binding, but most of which have unknown functions. Fig. 6 illustrates the main folds found in the catalytic domains of different glycosyl hydrolases.

TABLE 3. Structures and mechanisms in various families of glycosyl hydrolases (Davies and Henrissat, 1995).

Family	Enzyme	EC number	Organism	Mechanism
1	β -Glucosidase	3.2.1.21	<i>Trifolium repens</i>	Retaining
5	Endoglucanase A	3.2.1.4	<i>Clostridium cellulolyticum</i>	Retaining
6	Cellobiohydrolase II	3.2.1.91	<i>Trichoderma reesei</i>	Inverting
	Endoglucanase	3.2.1.4	<i>Thermonospora fusca</i>	Inverting
7	Cellobiohydrolase I	3.2.1.91	<i>Trichoderma reesei</i>	Retaining
	Endoglucanase I	3.2.1.4	<i>Humicola insolens</i>	Retaining
9	Endoglucanase D	3.2.1.4	<i>Clostridium thermocellum</i>	Inverting
10	Xylanase A	3.2.1.8	<i>Streptomyces lividans</i>	Retaining
11	Xylanase	3.2.1.8	<i>Bacillus circulans</i>	Retaining
45	Endoglucanase V	3.2.1.4	<i>Humicola insolens</i>	Inverting

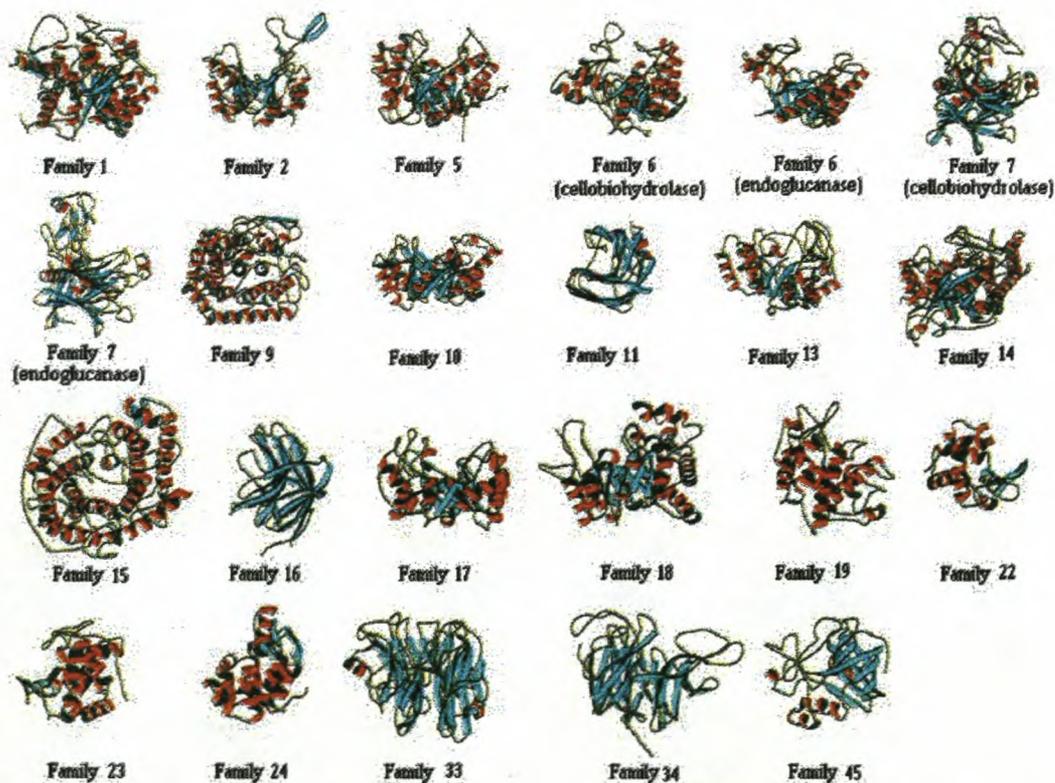


FIGURE 6. Ribbon representation of the main fold of the catalytic domain in various glycosyl hydrolase families (Davies and Henrissat, 1995).

2.2.3.3 Properties of Glycosyl Hydrolases

Even though various protein folds are represented in the 22 families for which the three-dimensional structure is known, the overall topologies of the active sites fall into three general classes (irrespective of whether the enzyme is inverting or retaining). The three types of active sites found in glycosyl hydrolases are (i) the pocket, (ii) the cleft and (iii) the tunnel as illustrated in Fig. 7 (Henrissat *et al.*, 1995).

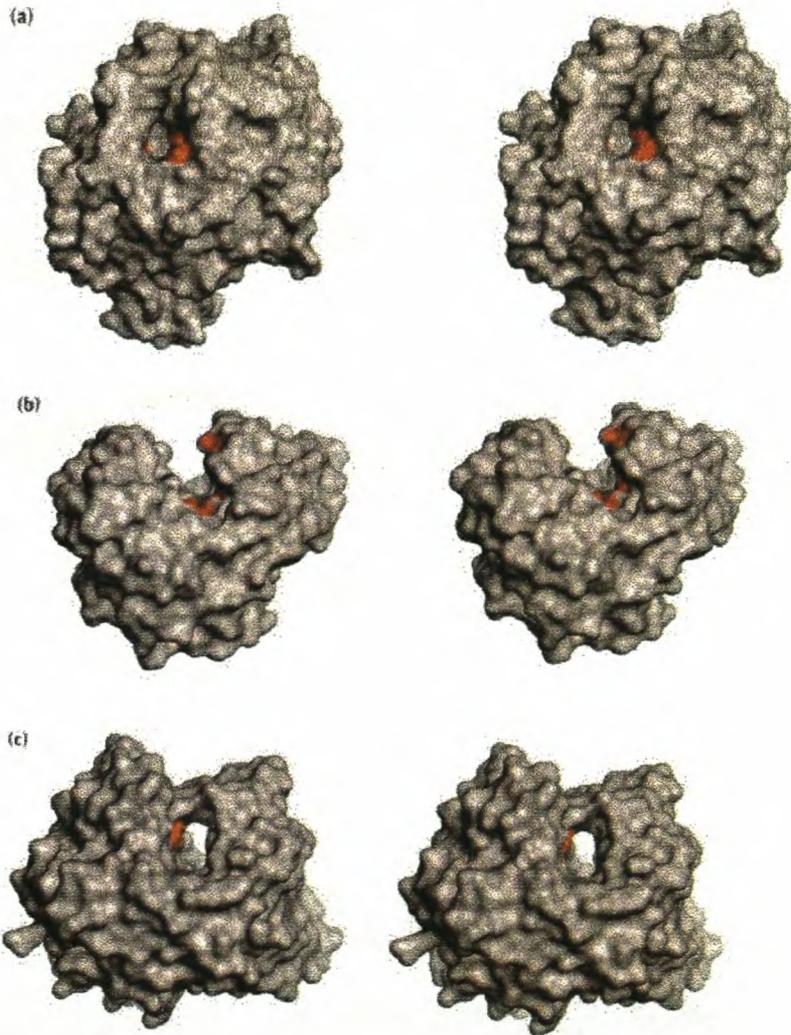


FIGURE 7. The three types of active sites found in glycosyl hydrolases. (a) The pocket topology of the active site found in β -glucosidases, β -galactosidases and glucoamylases. The cleft (b) and tunnel (c) topologies of the active site found in xylanases and cellobiohydrolases, respectively. The proposed catalytic residues are shaded in red (Davies and Henrissat, 1995).

The pocket (or crater) topology is most favourable for the recognition of a non-reducing end of a sugar and is found in enzymes such as β -glucosidases, β -galactosidases, glucoamylases, etc. (Fig. 7a) (Jenkins *et al.*, 1995). This is particularly important in the

case of β -glucosidases, which are defined as enzymes that hydrolyse compounds containing β -glucosidic linkages by splitting of the terminal, non-reducing β -D-glucose residues thereby releasing β -D-glucose. Such enzymes are tailored to accommodate substrates having a large number of "available" non-reducing chain ends at the surface, which implies that they are not very efficient for fibrous substrates such as native cellulose, which has almost no free chain ends. The remaining two active site topologies, the cleft (or groove) and tunnel, are respectively found in (i) xylanases, β -1,3-glucanases, β -1,4-glucanases, β -1,3-1,4-glucanases, etc. (Fig. 7b) and (ii) cellobiohydrolases (Fig. 7c).

Initially, the results obtained from 3-D structure analysis of the first glycosyl hydrolases, identified the two catalytic amino acids as aspartate and glutamate residues (Matthews and Remington, 1974; Blake *et al.*, 1965). In most glycosyl hydrolases studied since, only aspartate and/or glutamate residues have been found to perform catalysis. Additional data now also suggest that other residues may sometimes be involved in glycosidic bond cleavage. Within the group of β -glucosidase enzymes (predominantly members of Family 1 & 3) aspartate, glutamate and histidine are the amino acids identified to be involved in enzymatic hydrolysis using an acid catalytic mechanism (Grabnitz *et al.*, 1991).

Since this study mainly focuses on the β -glucosidase, phospho- β -glucosidase and phospho- β -galactosidase enzymes, it is important to elaborate on some of the interesting features concerning members of family 1. To date, the crystal structure of twelve enzymes has been solved (http://afmb.cnrs-mrs.fr/CAZY/GH_1.html). The overall structure of this family is highly conserved. All enzymes have the TIM-barrel fold with insertions at the carboxy ends of the β -strands. Two acidic amino acids at the carboxy ends of the fourth and seventh strands of the $(\beta\alpha)_8$ -barrel act as general acid/base and nucleophile/leaving group, respectively. Two of the enzymes with characteristics that are worth discussing, are: *Bacillus polymyxa* β -glucosidase (BglA) (Sanz-Aparicio *et al.*, 1998a, 1998b; Lopez-Camacho *et al.*, 1996) and *Lactococcus lactis* phospho- β -galactosidase (PBGAL) (Wiesman *et al.*, 1997). Ribbon images of both enzymes are presented in Fig. 8.

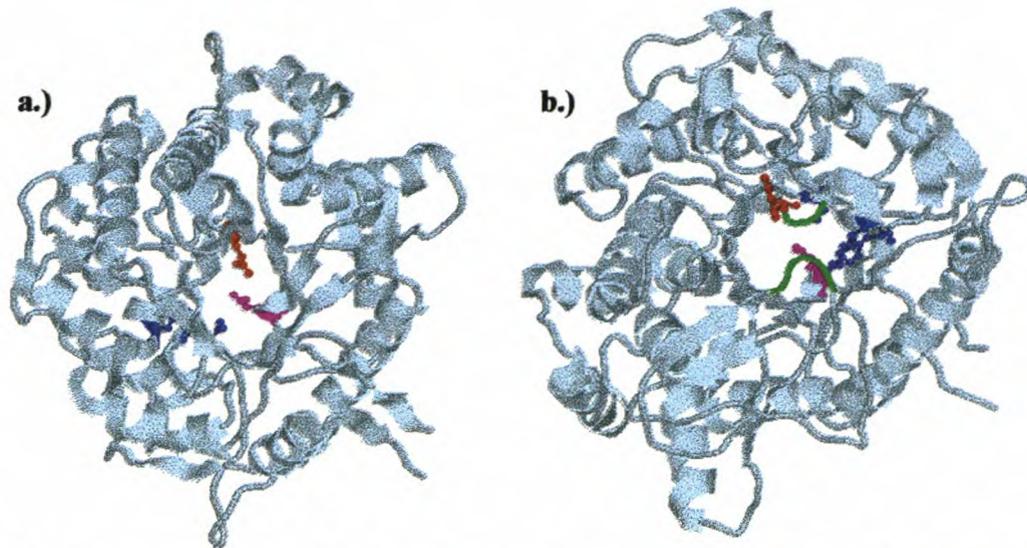


FIGURE 8. Ribbon representations of the three-dimensional structure of a) *B. polymyxa* β -glucosidase and b) *L. lactis* phospho- β -glucosidase. In both enzymes the active center is in a cavity in the center of the barrel and a long channel connects it with the solvent. The major differences occur only at the residues near the channel entrance. The acid/base catalyst is indicated in red and the nucleophile in pink. The residues in blue determine glucosidase/galactosidase activity and phosphate binding in enzymes of *B. polymyxa* and *L. lactis*, respectively. The mobile loops of the phospho- β -galactosidase are indicated in green.

The crystal structure of *B. polymyxa* BglA revealed significant structural differences in the loops surrounding the active center cavity. These differences cause a wide and extended cavity in BglA, which make it possible to accommodate substrates longer than cellobiose (its natural substrate). In addition, the presence of a third sub-site is connected with the *transglycosylating* activity associated with this enzyme. Residues that were identified to be involved in the recognition of the substrate are: Gln20, His121, Tyr296, Glu405 and Trp406. Interactions between the Gln20 and Glu405 are considered to play a role in the ability of most members of the family for displaying both glucosidase and galactosidase activity (Sanz-Aparicio *et al.*, 1998a, 1998b).

Schulte and Hengstenberg (2000) succeeded in identifying several structurally important residues for the specificity of the *L. lactis* phospho- β -glucosidase. According to their results, the difference between β -glucosidases and phospho- β -galactosidases is not caused by the difference in the phosphate-binding loop, but rather in the loops near the channel entrance. Phospho- β -galactosidases have two mobile loops that close after substrate binding as indicated in Fig. 8 (Wiesman *et al.*, 1997). It has also been shown that the amino acid at position 429 (Trp in phospho- β -galactosidases and Ala in phospho- β -glucosidases) is important for the distinction between glucosides and galactosides (Schulte and Hengstenberg, 2000).

2.3 Interactions and Associations of Cellulolytic Enzymes

As a result of the taxonomic and ecological diversity of cellulolytic microorganisms it is not unexpected that the ways in which cellulase systems are organised appear equally diverse. Hazlewood and Gilbert (1993) divided all cellulose-degrading systems into two broad categories, namely, complexed and non-complexed. Complexed (or aggregating) systems are generally characteristic of anaerobic microorganisms, including bacteria and fungi that colonise anaerobic environments in the rumen and hindgut of herbivores, composting biomass and sewage (Kajikawa and Masaki, 1999). In some anaerobic bacteria, particularly *Clostridium* spp., the complexed enzymes are contained in distinct high-molecular-weight protein complexes called cellulosomes. By contrast, non-complexed systems (also described as “non-aggregating” or “free enzyme” systems), are representative of aerobic fungi and bacteria and consist of several soluble cellulases and related polysaccharide depolymerases which are secreted into the culture medium. However, while this distinction between cellulase systems of aerobes and anaerobes is very common, it does not apply to all systems. For instance, some cellulolytic anaerobes, e.g. *Bacillus* spp., secrete non-complexed systems (Lo *et al.*, 1988), whereas some cellulases in aerobic bacteria may be cell-bound (Schlochtermeier *et al.*, 1992).

2.3.1 Complexed Systems

In the early 1980's, it was discovered that cellulose degrading anaerobic bacteria, e.g. clostridia and ruminococci, organize their cellulolytic enzymes into a tightly associated, extracellular cellulase complexes called cellulosomes (Shoham *et al.*, 1999). An important aspect of the cellulosome model proposed by Mayer *et al.* (1988) is that the enzyme complex hydrolyses crystalline cellulose with extraordinary efficiency, not because the enzyme components themselves are exceptionally active, but because the activities of the particular enzymes are highly coordinated. The cellulosome complex contains many different types of glycosyl hydrolases, including cellulases, hemicellulases and even carbohydrate esterases. In addition to their catalytic domain, all these enzymes were found to contain a second “dockerin” domain, which is characterized by a 22-residue repeat sequence. An early advance in this field was the cloning and sequencing of a large multi-domain cellulosomal subunit, later called “scaffoldin” (Shoseyov *et al.*, 1992). This subunit contained a single cellulose binding module (CBM) and up to nine similar

repeating “cohesion” domains. Scaffoldin was found to be responsible for both cellulose binding (via its CBM) and for assembly of the enzymes into the cellulosome complex through the intersubunit cohesion-dockerin interaction (Béguin and Lemaire, 1996). Interestingly, the CBM has been discovered to have broad binding specificity for different sites on crystalline cellulose (Schwarz, 2001). In *C. thermocellum*, scaffoldin attaches to the host cell via a distinct set of surface proteins and this anchoring function is facilitated by a second type of cohesion-dockerin interaction (Lemaire *et al.*, 1998; Fujino *et al.*, 1993). The key difference between cellulosomal and non-cellulosomal enzymes is that the latter lack a dockerin domain. Cellulosomal enzymes essentially bear a dockerin domain and usually rely on the scaffoldin CBM for effective binding to the crystalline substrate. A schematic view of the cellulosome and its interaction with cellulose and the cell surface is presented in Fig. 9. The catalytic activity of the cellulosome is dependent on the presence of lipids (Bolobova *et al.*, 1994). It has also been shown that Ca^{++} ions stimulate the activity of cellulosomes, but not that of individual, soluble components. This phenomenon can be explained by the crucial role Ca^{++} ions play in the correct folding of the dockerin, which in turn guarantees close binding of the catalytic components to the scaffoldin (Schwarz, 2001).

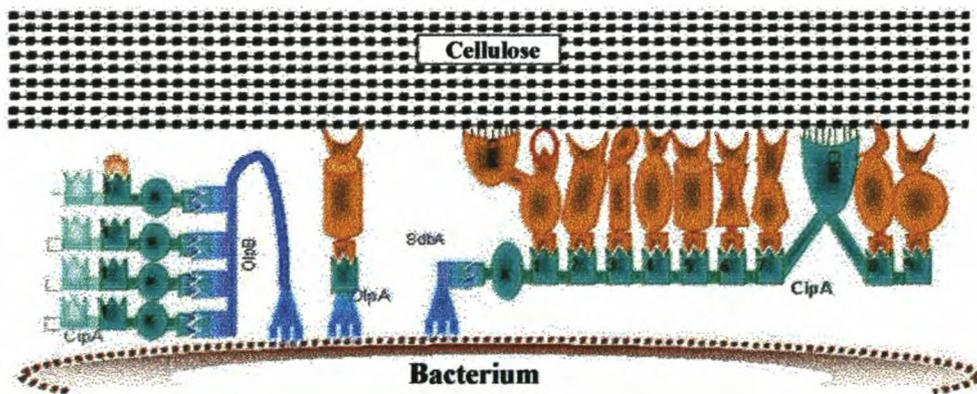


FIGURE 9. The model for the organisation of the cellulosome produced by *C. thermocellum*. The scaffoldin (CipA) protein is indicated in green, catalytical components in orange, cell-wall binding components in blue, the bacterial cell in brown and the cellulose microfibril in black. The cellulose binding domains/modules are designated CBM. The S-layer is dotted. The cohesin modules are numbered 1 to 9. The binding of the enzymes to specific positions (cohesins) is hypothetical, as is the linear orientation of the scaffoldin (Schwarz, 2001).

The enzymes associated with the cellulosome include classical cellulases, both endo- and exo-acting β -glucanases, which attack the cellulose chain internally or at one of the ends,

respectively. It appears that enzymes that cleave the cellulose chain sequentially are the most effective in their action on crystalline substrates. Interestingly, the cellulosomal enzymes also include xylanases, mannanases, lichenases, and even chitinase. It is concluded that the abundance of non-cellulolytic enzymes in the cellulosome are involved in the removal or detachment of plant cell wall polymers – hemicellulose and lignin – that are usually found in close contact with cellulose. The composition of the enzyme complex depends on induction, therefore it varies between different carbon sources (Bhat *et al.*, 1994).

A recent study showed that the *C. thermocellum* cellulosome is extremely efficient at solubilizing cellulosic substrates of the highest known crystalline content (Boisset *et al.*, 1999). It suggests that the specific activity of the cellulosome for such substrates is higher than that of free enzyme systems. According to Schwarz (2001), the arrangement of enzymes into a multi-enzyme cellulosome complex has four major advantages over non-complexed enzyme systems (e.g. that of *T. reesei*):

1. Synergism is optimized by the correct ratio together with the orientation of the components both with respect to each other and to the substrate.
2. Non-productive adsorption is prevented by the optimal spacing of components working together in synergistic mode.
3. Competitiveness in binding is avoided by binding the entire complex to a single site through a strong binding domain with low specificity.
4. A cessation in hydrolysis on depletion of one structural type of cellulose at the site of adsorption is avoided by the presence of other enzymes with different specificities.

2.3.2 Non-complexed Systems

The degradation of complex carbohydrates, such as cellulose, by heterotrophic microorganisms typically involves multiple glycosyl hydrolases whose concerted actions are needed to support growth on polysaccharidic substrates (Beguin and Aubert, 1994; Walker *et al.*, 1992). Sometimes these enzymes are assembled in large subcellular units, such as the cellulose-degrading cellulosomes produced by certain clostridia (Schwarz, 2001; Stalbrand *et al.*, 1998; Gal *et al.*, 1997). In other cases, the cellulases are not complex-associated and are produced as individual enzymes such as those found in some fungi, including *Aspergillus niger* (Ashadi *et al.*, 1996), several *Trichoderma* species (Nidetzky *et al.*, 1994), and *Thermomonospora fusca* (Irwin *et al.*, 1993). As the

knowledge about polysaccharide-degrading enzyme systems expanded, it has become evident that the interaction among cellulases is complicated but essential for efficient degradation of both crystalline and amorphous cellulose. Most of the early work on enzymatic degradation of cellulose involved the isolation of cellulases from the supernatants of fungal cultures. Several independent studies demonstrated that mixtures of isolated components interacted synergistically, i.e. their combined activity on cellulose was greater than the sum of their individual activities. This effect is generally observed with crystalline substrates, but not soluble derivatives such as carboxymethyl-cellulose (CMC) (Tomme *et al.*, 1995).

Numerous cellulose degradation studies have been done to explain the nature of the endoglucanase (endo)/exoglucanase (exo), endo/endo, and exo/exo cellulase interactions (Ashadi *et al.*, 1996; Irwin *et al.*, 1993; Nidetzky *et al.*, 1994; Philippidis *et al.*, 1993; Poulsen and Petersen, 1992; Stalbrand *et al.*, 1998). Competitive adsorption, ratio and concentration of enzymes have been considered and the need for a loose enzyme-enzyme complex has been postulated to be a necessary for an attack on crystalline cellulose (Klyosov, 1990; Woodward *et al.*, 1988; Wood and McCrae, 1978). Currently, the accepted model is that there is a sequential mechanism for synergistic action whereby endoglucanase initiates the attack on cellulose by forming new chain ends, which then serve as attack points for processive hydrolysis by the end-acting cellobiohydrolases (Väljamäe *et al.*, 1999; Tomme *et al.*, 1995). This however is a simplistic view of this general model.

In *T. reesei*, the synergistic action between CBHI and EGI display a 1.7- to 1.8-fold increase in degradation rate with highly crystalline cellulose substrates such as bacterial microcrystalline cellulose (BMCC). A simple production of new chain ends is improbable to account for this synergism. Therefore these observations indicate a more interactive mechanism for endo-exo synergism based on simultaneous action of these two enzymes. Usually, this phenomenon would be explained by the formation of a partial complex between enzymes on the cellulose surface, which acts differently from the separate components acting independently. However, no proof for the presence of such loose *in situ* complexes exists. A more recent model offers a mechanistic explanation for interactive synergism based on the role of cellulose changes during hydrolysis (Väljamäe *et al.*, 1999).

Evidence indicated a strong lag of BMCC hydrolysis by CBHI caused by enzyme-generated alteration of the cellulose surface (Väljamäe *et al.*, 1998). Repetitive hydrolytic actions (the result of a processive enzyme) causes erosion of the cellulose surface in a manner that randomly left isolated chains that possibly form obstructions for the hydrolysis of chains in the underneath layer (Fig. 10). The central function of EGI in this new model lies in its ability to attack these isolated chains, thus assisting subsequent processive action of CBHI. EGI thus supports CBHI as a coexisting “scavenging” agent that efficiently removes the isolated chains that remain on the “eroded” crystalline surface (Fig. 11). In the same context, the erosion effect by CBHI makes crystalline cellulose more susceptible to the EGI attack and therefore the continuous interaction between the “new-end-producing” and the “scavenging” mechanisms adds to the full synergistic effect (Väljamäe *et al.*, 1999).

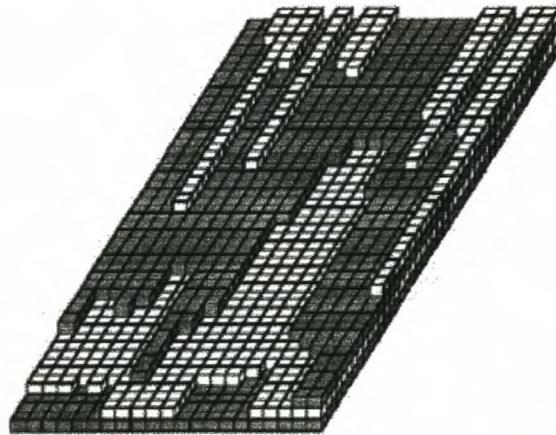


FIGURE 10. The erosion of the cellulose surface by a processive enzyme from *T. reesei* CBHI (Väljamäe *et al.*, 1999).

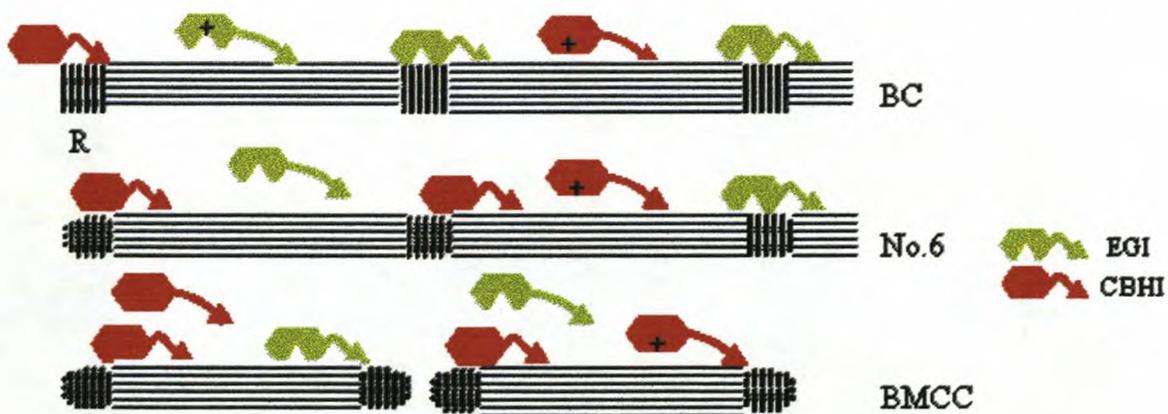


FIGURE 11. Schematic representation of cellulose hydrolysis by CBHI and EGI. Only a longitudinal section of the cellulose microfibril is shown. R = reducing end. Crystalline parts are represented by parallel lines and dotted regions are amorphous. Catalytic domains of non-productively bound enzymes are marked with a cross (Väljamäe *et al.*, 1999).

2.4 Bacterial Sugar Transport and Utilisation Systems

Bacteria have developed different mechanisms of transport to allow the movement of molecules across the cytoplasmic membrane. The most common transport mechanisms can be divided into two major groups. The first group is passive transport, which include both passive and facilitated diffusion. The second group is active transport, that entails mechanisms which either do not modify the sugars (chemiosmotic coupling or direct phosphate bond energy coupling that needs the participation of a binding protein) or modify the substrates during the process (group translocation) (Fouet *et al.*, 1989) (Fig. 12).

The best-known group translocation system is the phosphoenolpyruvate-dependant phosphotransferase system (PTS), a process in which a molecule is transported into the cell while being phosphorylated.

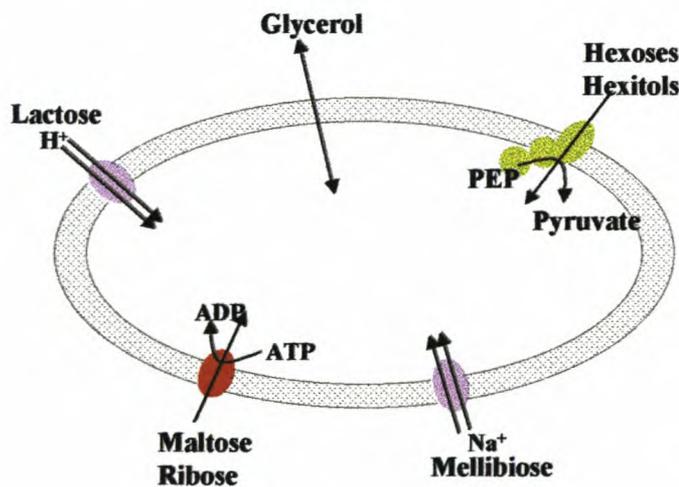


FIGURE 12. The various transport mechanisms used by bacteria. Passive transport includes passive (\square) and facilitated diffusion (purple). Active transport is divided in ATP-dependent transport (red) and group translocation (green) (Siebold *et al.*, 2001).

2.4.1.1 Components of the PTS

PTS's consist of two cytoplasmic proteins, enzyme I (EI) and histine-containing protein (HPr), and a variable amount of sugar-specific transport complexes; enzymes II (EII^{sugar}) (Fig. 13). The EI and HPr components of the PTS serve to transfer the phosphate moiety derived from PEP to the sugar-specific EII transporters. The amino acid sequences of EI and HPr are strongly conserved in all bacteria (Siebold *et al.*, 2001). EI (encoded by *ptsI*) and HPr (encoded by *ptsH*) are commonly referred to as the general PTS proteins and are

synthesised constitutively by the bacterial cell, while EII transporters are produced after induction with the corresponding sugar (Hengstenberg *et al.*, 1993). PTS transporters consist of three functional units, IIA, IIB and IIC, which present themselves either as protein subunits in a complex (e.g. IIA^{Glc}·IICB^{Glc}) or as domains of a single polypeptide (e.g. IICBA^{Glc}). IIA and IIB sequentially transfer phosphoryl groups from HPr to the sugars being transported across the membrane. IIC constitutes the sugar-binding site. The phosphorylation of the sugar molecule by IIC while being transported generate intracellular sugar-phosphate. EI, HPr and IIA are phosphorylated at a His residue, while IIB domains are phosphorylated at either Cys or His residues (depending on the specific transporter) (Lengeler *et al.*, 1994). The number and structure of PTS transporters vary between species. They can be grouped into six structurally very different families based on the homology of their deduced amino acid sequences (Table 4) (Kotrba *et al.*, 2001).

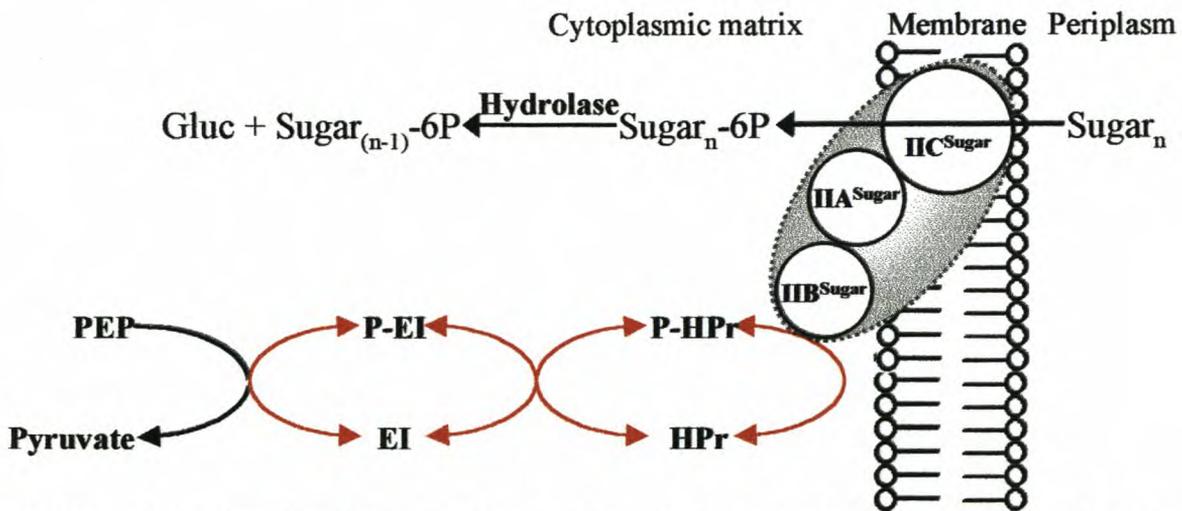


FIGURE 13. The PTS-mediated glow of a phosphoryl group to the transported sugar and organization of the PTS components. Phosphoryl groups are sequentially transferred from PEP to the general PTS components, EI and HPr, and subsequently to the sugar-specific transporter IIBAC. The commonly three-component EI complex consists of two hydrophylic domains (IIA and IIB) and a sugar-selective transmembrane domain (IIC). The red arrows highlight the reversibility of the phosphotransfer reaction.

TABLE 4. Families of the PTS transporters according to Transport Commission (TC) classification (Kotrba *et al.*, 2001)

TC no.	PTS family	Features
4.A.1.	Glucose-glucoside (Glc) family	Transport of glucose, <i>N</i> -acetylglucosamine, various α - and β -glucosides; some lack specific IIA and function with IIA ^{Glc}
4.A.2.	Fructose-mannitol (Fru) family	Transport fructose and mannitol; several putative permeases of unknown specificity
4.A.3.	Lactose- <i>N,N</i> -diacetylchitobiose- β -glucoside (Lac) family	Transport of lactose, <i>N,N</i> -diacetylchitobiose, lichenan and aryl- β -glucosides

TABLE 4. Families of the PTS transporters (continued)

TC no.	PTS family	Features
4.A.4.	Glucitol (Gut) family	Permeases found in both Gram-positive and -negative bacteria, so far only glucitol-specific
4.A.5.	Galactitol (Gat) family	Only one member known at present (<i>E. coli</i>)
4.A.6.	Mannose-fructose-sorbose (Man) family	Members usually show broad sugar specificity; transport of mannose, glucosamine, sorbose, galactoseamine, N-acetyl-galactosamine, fructose; permease complex possess a transmembrane IID domain

It is important to stress that the coupled translocation-phosphorylation of PTS sugars requires a single PEP, which is equivalent to one ATP formed in the pyruvate kinase reaction. The phosphorylated sugar (e.g. glucose-6-P) released in the cytoplasm is the first metabolic intermediate so that uptake and catabolism are tightly coupled. The utilisation of the sugar taken up by a non-PTS active transporter requires (in addition to energy for transport) an extra ATP for phosphorylation of the molecule in the substrate kinase reaction. In rumen bacteria and clostridia it has been found that additional ATP, beyond that obtained from catabolism of monosaccharides, is potentially available as a result of the action of cellobiose phosphorylase (CbP) and cellodextrin phosphorylase (CdP) (Lynd *et al.*, 2002; Kajikawa and Masaki, 1999). CbP catalyses a Pi-mediated (ATP-independent) phosphorylysis reaction:



The initial metabolism of soluble cellulose degradation products involves either hydrolytic or phosphorolytic cleavage (Schimz *et al.*, 1983). The simultaneous presence of extracellular β -glucosidase, intracellular phospho- β -glucosidase and intracellular CbP suggests that cellobiose metabolism can occur via several processes as indicated in Fig. 14. In general, the metabolic pathways which use phosphorolytic cleavage conserve more energy and is of some importance because of the potential for ATP production. Since bacterial energy sources in the rumen are often restricted and the yield of ATP is usually low in anaerobic microorganisms, efficient utilisation of substrate is significant for the growth of ruminal bacteria. Thus, it is not surprising that phosphorolytic cleavage mechanisms have been found in several ruminal bacteria (Kajikawa and Masaki, 1999; Wells *et al.*, 1995; Thurston *et al.*, 1993; Helaszek and White, 1991; Ayers, 1958). The presence of both phosphorylytic and hydrolytic routes for cellobiose metabolism in these

cellulolytic organisms suggests that it could be of selective importance. Regulating the relative carbon flow via the two different routes may provide a means to adjust the rate of ATP supply in response to environmental factors (e.g. availability of substrate or nutrients) (Lynd *et al.*, 2002). The CbP of *Cellvibrio gilvus* has also been demonstrated to permit the *in vitro* synthesis of numerous unusual di- and trisaccharides (Percy *et al.*, 1998). This may be of certain importance in the production of both intracellular glycogen and extracellular polysaccharides (Guedon *et al.*, 2000a, 200b).

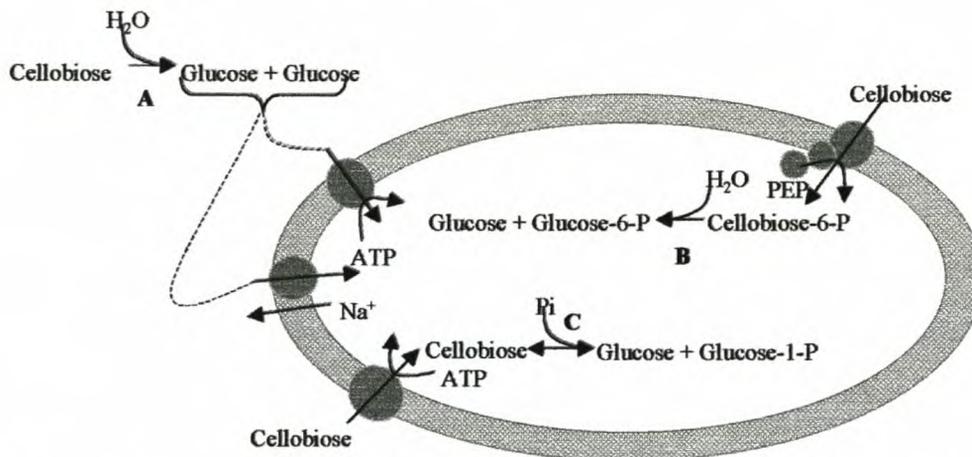


FIGURE 14. Different routes by which cellobiose is transported and metabolized: 1) Extracellular hydrolysis of cellulose by β -glucosidase (A) and subsequent transport of the glucose residues via ATP-dependent active transport or Na^+ antiport. 2) Transport of cellobiose via the cellobiose-specific phosphotransferase system (PTS) and subsequent hydrolysis by phospho- β -glucosidase (B). 3) Direct uptake of cellobiose via an active ATP-dependent transport system and phosphorylytic cleavage of the molecule by cellobiose phosphorylase (C). Note that this reaction is ATP-independent and fully reversible.

2.4.1.2 PTS operons

The PTS is the first step in the catabolism of sugars. However, to enter the central metabolic pathways (Embden-Meyerhof and Entner-Doudoroff) the sugars, other than glucose and fructose need to be converted to glucose- and fructose-6-phosphates or other intermediates of the individual pathways. Genes encoding the enzymes necessary for these conversions (hydrolysis, oxidation, isomerisation or epimerisation) are often located within the same operon as the sugar-specific transporters. A good example is the *bgIGFB* operon of *E. coli* which contain genes necessary for regulated transport and utilisation of β -glucosides such as arbutin and salicin (Postma *et al.*, 1993) (Fig. 15). Alternatively, the essential genes may be transcribed from different loci but respond to the same specific regulation. The expression of catabolic PTS operons is regulated by transcriptional

repressors or activators and anti-terminators (Reizer *et al.*, 1999; Postma *et al.*, 1993). This aspect will be discussed in more detail under “The role of the PTS in energy and signal transduction”.

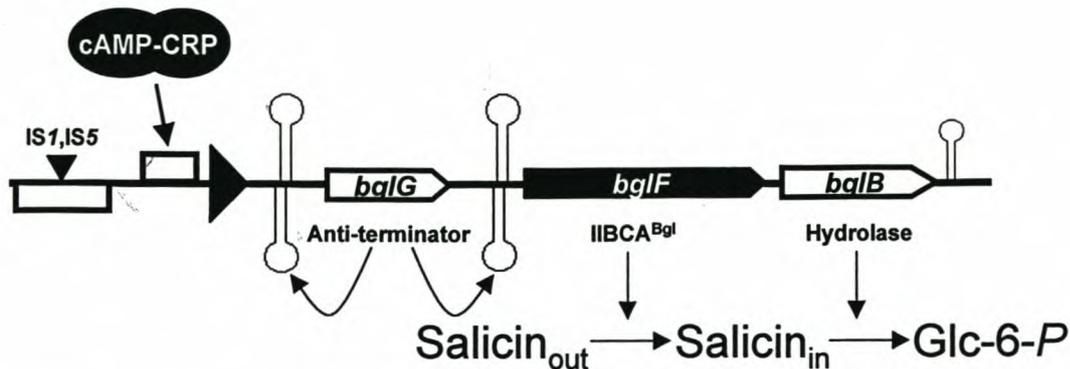


FIGURE 15. The β -glucoside (e.g. salicin) specific *bgl* operon from *E. coli*. The specific enzymatic reactions responsible for the conversion (and subsequent integration) of salicin into the cell catabolism are shown at the bottom. The ribonucleic terminator is indicated as a hairpin and the ribonucleic terminator/anti-terminator (RAT) element as two overlapping hairpins. CAP: CAMP receptor protein (Kotrba *et al.*, 2001).

2.4.2 The role of the PTS in energy and signal transduction

In addition to their function in transport, some PTS proteins play an important role in intracellular signal transduction (Fig. 16). These subunits regulate their target proteins either allosterically or by phosphorylation. Their regulatory activity depends on the degree of phosphorylation of EII (ratio of non-phosphorylated/phosphorylated units), which varies with the ratio of sugar-dependent dephosphorylation and PEP-dependent rephosphorylation (Saier *et al.*, 1995). Proteins of the PTS are involved in carbon catabolite repression (CCR). CCR can be explained as the phenomenon where the expression of genes and/or operons involved in the transport and catabolism of sugars are expressed only if (i) the corresponding sugar/inducer is present in the medium and (ii) the preferred carbon sources (more rapidly metabolized sugar) are absent (Reizer *et al.*, 1999; Saier and Reizer, 1994; Steinmetz *et al.*, 1989;). Additionally, synthesis of secondary metabolites (e.g. antibiotics), as well as developmental pathways (e.g. synthesis of extracellular enzymes and spore formation) is subject to CCR (Krüger *et al.*, 1996).

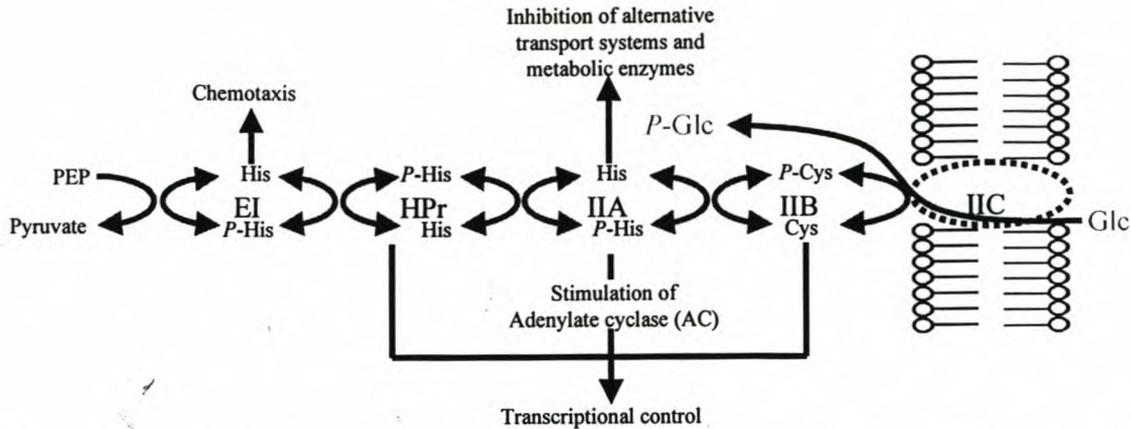


FIGURE 16. A simplified schematic representation of the key components of the PTS that are involved in various regulatory interactions in the cell (Siebold *et al.*, 2001).

The identification of the PTS in *E. coli* provided the understanding of the mechanism by which exogenous glucose causes CCR (Amster-Choder *et al.*, 1989). The indirect mechanism by which *E. coli* achieves CCR, involves the positive regulator CAP (catabolite activator protein) that, when associated with cyclic AMP (cAMP), activates transcription of catabolite repression-sensitive operons. The CAP-cAMP complex binds to a specific site upstream of the -35 region of these promoters thereby activating transcription. The glucose-specific enzyme IIA (IIA^{Glc}) of the PTS appears to be the central regulatory protein, since it controls both the intracellular level of cAMP and several non-PTS permeases (Krüger *et al.*, 1996; Schnetz and Rak, 1990). For example, the presence of glucose in the external medium causes the accumulation of unphosphorylated EI^{Glc}, which inhibits the activity of permeases for various non-PTS sugars such as lactose. On the contrary, the absence of glucose leads to an increase in phosphorylated EI^{Glc}, which activates adenylate cyclase and causes an increase in cAMP-levels and subsequent activation of cAMP-CRP dependent promoters. Thus, the absence of glucose transported by the system leads to accumulation of phosphorylated intermediates, whereas the presence and transport of glucose drain the system of high-energy phosphates and results in the accumulation of unphosphorylated enzymes.

Carbon catabolite repression in *Bacillus subtilis* applies different mechanisms, as very low concentrations of cAMP are present and a negative regulatory mechanism triggers this type of regulation. Previous attempts to identify *cis*-acting elements mediating CCR of several genes led to the establishment of a consensus sequence for a catabolite responsive element (CRE) (Hueck *et al.*, 1994; Weickert and Chambliss, 1990.). Mutations in CRE sequences

(found upstream of various catabolic operons), result in loss of glucose repression. *B. subtilis* mutants, relieved from glucose repression, possessed the *crsA* mutation (located in the gene for the σ^A -subunit of RNA polymerase) and as a result caused the developmental pathways (e.g. sporulation) to be insensitive to glucose repression (Sun and Takahashi, 1984). Mutations in *ptsH* (encoding the HPr protein of the PTS), and in *ccpA* (encoding the putative *trans*-acting repressor of CCR), revealed pleiotropic effects (single gene that influenced a variety of evident features) (Deutscher *et al.*, 1994). In contrast to HPr of *E. coli*, the *B. subtilis* protein possesses a second phosphorylation site at serine 46 which can be phosphorylated by an ATP-dependent protein kinase (Deutscher *et al.*, 1986). A loss of this phosphorylation (due to the *ptsHI* mutation, Ser46Ala) causes the expression of several catabolic genes to become glucose resistant (Deutscher *et al.*, 1994, Krüger *et al.*, 1993.). The CcpA protein belongs to the LacI-GalR family of transcriptional regulators (Grundy *et al.*, 1993; Weickert and Adhya, 1992). Recently, interaction between HPr-Ser-P and CcpA has been confirmed, signifying that this complex binds to CREs mediating CCR (Deutscher *et al.*, 1995; Hueck and Hillen, 1995).

The direct involvement of the PTS in the regulation of promoter activity is mediated via regulatory proteins. Recent studies identified a class of positive regulators that is a target for specific induction and/or CCR. All of these positive regulators include a conserved domain called PRD (PTS regulatory domain), which is the target for both positive and negative control (Stülke *et al.*, 1998). Proteins containing a PRD are found in both Gram-negative and Gram-positive bacteria. Most of them are positive transcriptional regulators (either activators or anti-terminators) controlling the synthesis of enzymes that are either involved in the catabolism sugars transported by the PTS or create PTS substrates. The activator and anti-terminator proteins that contain PRDs are often subject to dual control by the PTS: (i) in the absence of the inducer of the controlled operon, they are phosphorylated and thus inactivated; (ii) their activity may also depend on an additional HPr-dependent phosphorylation, which occurs only in the absence of a repressing carbon source, thus providing a means for hierarchical expression of genes required for the utilisation of secondary carbon sources (Stülke *et al.*, 1998).

Interestingly, anti-terminators containing a PRD are generally involved in the regulation of systems controlling the catabolism of β -glucoside sugars. Five anti-terminators regulating

bgl operons have been sequenced and an additional two incomplete sequences are available (Rutberg, 1997). Anti-terminator proteins control the transcription of genes and operons at mRNA level by binding to ribonucleic anti-terminators (RAT) sequences located at the 5' end. Upon binding, the anti-terminator is assumed to stabilise the RAT secondary structure, preventing the formation of a hairpin structure acting as a terminator and therefore allowing transcription of the target gene (Aymerich and Steinmetz, 1992). For example, in *E. coli* the anti-terminator protein BglG can be phosphorylated by the β -glucoside-specific EII^{Bgl} protein, thereby controlling BglG binding to the RAT sequence, subsequently preventing termination of the *bgl* operon (Houman *et al.*, 1990) (Fig. 13). However, the BglG equivalent in *Bacillus* (named LicT) is both negatively regulated by EII^{Bgl} and positively by EI and HPr *in vivo* and *in vitro* (Deutscher *et al.*, 1997; Schnetz *et al.*, 1996).

All these mechanisms of catabolite repression mediated by PTS proteins, suggest the existence of an extensive repertoire of pathways by which the activity of individual promoters are either directly or indirectly subjected to external carbon sources. The discovery of new PTS-associated sequences in the bacterial genomes indicates that PTSs may be involved in additional and still unknown functions. For example, EI^{Ntr} from *Pseudomonas aeruginosa* has been shown to play a role in its pathogenicity (Tan *et al.*, 1999). In addition to EI^{Ntr}, other PTS proteins have also been associated with bacterial virulence. Several lines of evidence suggest that the general PTS components (EI and HPr) may play a significant role in the ability of bacteria to migrate towards PTS sugars. In both *E. coli* (Lux *et al.*, 1999) and *B. subtilis* (Garrity *et al.*, 1998) the EI/EI-P ratio modulates the activity of chemotactic sensor kinase (CheA) and subsequently direct chemotaxis towards PTS-sugar attractants.

2.4.3 PTS Genes from *Bacillus* spp.

Soil bacteria, including many members of the family *Bacillaceae*, are efficient in utilising a wide variety of carbohydrates, which include cellulose, β -1,3-glucan (laminarin) and β -1,3-1,4-glucan (lichenan). In 1999, the complete set of genes encoding PTS and PTS-associated proteins in the model bacterium *B. subtilis* was presented by Reizer *et al.* (Table 5) (1999). Initially, enzymes with distinct linkage specificities degrade the polymers

extracellularly. The generated β -glucosides are taken up via an inducible, binding protein-dependent phosphotransport system.

TABLE 5. Genes encoding PTS and PTS-associated proteins in *B. subtilis* (Reizer *et al.*, 1999).

Gene	Encoded protein/ substrate	Domain structure	No. of residues
General energy-coupling proteins			
<i>ptsH</i>	HPr	H	87
<i>ptsI</i>	EI	I	570
Glucose permease (Glc) family			
<i>gamP</i> (<i>ybfS</i>)	Glucosamine	IICBA	631
<i>nagP</i> (<i>yflF</i>)	N-Acetylglucosamine	IICB	452
<i>malP</i> (<i>glvC</i>)	Maltose	IICB	527
<i>ptsG</i>	Glucose	IICBA	699
<i>ypqE</i>	Unknown	IIA	168
<i>yyzE</i>	Unknown	IIA	76
Sucrose permease (Scr) family			
<i>ybbF</i>	Unknown β -glucoside	IIBC	446
<i>treP</i>	Trehalose	IIBC	470
<i>sacP</i>	Sucrose	IIBC	460
<i>sacX</i>	Sucrose	IIBC	459
<i>bglP</i> (<i>sytA</i>)	Aryl- β -glucosides	IIBCA	609
Lactose permease (Lac) family			
<i>ywbA</i>	Unknown	IIC	444
<i>licA</i> (<i>celC</i>)	Oligo- β -glucoside	IIA	110
<i>licB</i> (<i>celA</i>)	Oligo- β -glucoside	IIB	102
<i>licC</i> (<i>celB</i>)	Oligo- β -glucoside	IIC	452
<i>ydhM</i>	Oligo- β -mannoside	IIB	103
<i>ydhN</i>	Oligo- β -mannoside	IIA	110
<i>ydhO</i>	Oligo- β -mannoside	IIC	442
Mannose permease (Man) family			
<i>LevD</i>	Fructose	IIA	146
<i>levE</i>	Fructose	IIB	163
<i>levF</i>	Fructose	IIC	269
Fructose/Mannitol permease (Fru) family			
<i>mtlA</i>	Mannitol	IICBA	610
<i>fruA</i>	Fructose	IIABC	635
<i>manP</i> (<i>yjdD</i>)	Mannitol	IIBCA	589
<i>levG</i>	Fructose	IID	275

TABLE 5. Genes encoding PTS and PTS-associated proteins in *B. subtilis* (continued).

Gene	Encoded protein/ substrate	Domain structure	No. of residues
Metabolic enzymes			
<i>glpK</i>	Glycerol kinase	-	496
<i>pykK</i>	Pyruvate kinase	-	585
Protein kinases/phosphatases			
<i>prkA</i>	Serine protein kinase	-	631
<i>ptsK</i>	HPr(Ser) kinase/phosphatase	-	310
Transcriptional regulators			
Transcriptional anti-terminators	Regulated operon		
<i>glcT (ykwA)</i>	ptsGHI		285
<i>sacT</i>	sacPA		276
<i>sacY (sacS)</i>	sacB		280
<i>licT</i>	bglPH; licS (bglS)		277
Transcriptional activators			
<i>mtlR (ydaA)</i>	mtlAD		694
<i>manR (yjdC)</i>	yjdDEF		648
<i>levR</i>	levDEFG; sacC		938
<i>licR (celR)</i>	licBCAH (celABCD)		641

In *B. subtilis*, several genes and operons involved in β -glucoside utilization have been identified. β -Glucosides, such as lichenan and its hydrolysis products, can be considered alternative carbon sources that are utilized under limited glucose conditions. Two of the operons that have been studied extensively are the *bglPH* operon (Krüger and Hecker, 1995; Le Coq *et al.*, 1995) and the *licBCAH* operon (Tobisch *et al.*, 1997).

The *bglPH* operon encodes a PTS enzyme IIBCA protein and a phospho- β -glucosidase (BglH). This particular β -glucoside utilization system is induced by salicin and arbutin. Induction is mediated via the LicT anti-terminator that binds to a RAT-terminator structure located downstream of the *bgl* promoter (Krüger and Hecker, 1995). This protein belongs to the BglG-SacY family of bacterial antitermination proteins (Krüger *et al.*, 1996). The activity of LicT is negatively regulated by BglP but requires the proteins of the PTS (Le Coq *et al.*, 1995.). Furthermore, BglP synthesis is subject to CCR. A *cis*-acting site similar to CREs has been discovered, and CcpA and HPr are involved in this regulation (Krüger and Hecker, 1995). The CcpA protein is involved in catabolite repression of *B. subtilis*.

Previously, a model has suggested that CcpA and HPr act together, possibly by protein-protein interaction, to bind DNA and repress transcription (Deutscher *et al.*, 1994). Since then, it has also been established that the action of CcpA interferes with transcription initiation and offers evidence for another, HPr-dependent mechanism for CCR mediated via LicT (Fig. 17)(Krüger *et al.*, 1996).

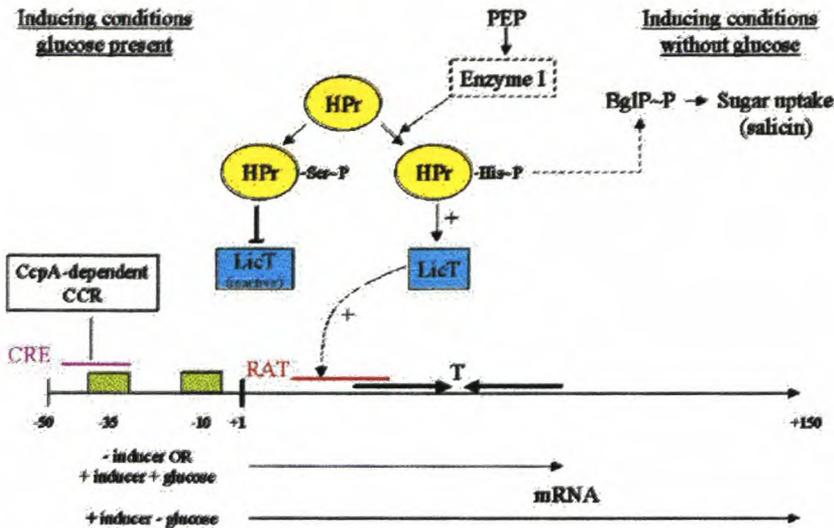


FIGURE 17. Model of complex regulation of the *bgIPH* operon of *B. subtilis*. The regulation of CCR by CcpA via CRE, as well as the novel type of regulation involving HPr, LicT, and the target, RAT, is indicated. HPr can be phosphorylated by two different enzymes (ATP-dependent kinase and EI) which direct HPr-P in different regulatory pathways. As a result, HPr plays a critical role. Arrows and perpendicular bars indicate positive and negative regulation, respectively.

In 1997, Tobisch *et al.* described the *licBCAH* operon and its role in the uptake and subsequent hydrolysis of oligomeric β -glucosides, which are produced by extracellular enzymes on substrates such as lichenan or barley glucan. The *lic*-PTS is encoded within the *licBCAH* operon that also includes a 6-phospho- β -glucosidase encoding gene, *licH*. The *licBCAH* operon shares high similarities to the *cel* operons from *E. coli* (Parker and Hall, 1990) and *B. stearothermophilus* (Lai and Ingram, 1993). The *lic* operon is transcribed from a σ^A -dependent promoter and is inducible by lichenan, lichenan hydrolysate, and cellobiose. Induction of the operon requires a DNA sequence with dyad symmetry located directly upstream of the *licBCAH* promoter. Expression of the *lic* operon is positively controlled by the LicR regulator protein, which contains two potential helix-turn-helix motifs, two PTS regulation domains (PRDs), and a domain similar to PTS enzyme IIA (EIIA) (Stülke *et al.*, 1998). The activity of LicR is stimulated by modification (most likely phosphorylation) of both PRD-I and PRD-II by the general PTS components and is negatively regulated by modification (probably phosphorylation) of its EIIA domain

by the specific EII^{Lic} in the absence of oligomeric β -glucosides. Furthermore, the *lic* operon is subject to carbon catabolite repression (CCR). CCR takes place via a CcpA-dependent mechanism and a CcpA-independent mechanism in which the general PTS enzyme HPr is involved (Tobish *et al.*, 1999).

2.5 Industrial Importance of β -Glucosidase Enzymes

2.5.1 Bioethanol Production

β -Glucosidase enzymes are widely used in various biotechnological processes, including the production of fuel ethanol from cellulosic residues. Bioethanol has become one of the most promising alternatives for fossil fuels since it has many beneficial properties as a fuel: High octane values and reduction of air-pollution and smog-forming compounds including uncombusted hydrocarbons, carbon monoxide (CO), nitrogen oxides (NO_x) and exhaust emission of reactive aldehydes (acetaldehyde, formaldehyde, etc.) (Tengborg, 2000). Agricultural biomass represents an abundant renewable resource that can be used for the production of bioethanol, provided that the cellulosic component is converted to fermentable sugars. A consortium of enzymes (endoglucanase, exoglucanase and β -glucosidase) is involved in the enzymatic saccharification of cellulose. The enzyme β -glucosidase that converts cellobiose and soluble cellodextrins to glucose, has been indicated to be the major rate-limiting step in the saccharification process. The accumulation of glucose represses the activity of β -glucosidase, subsequently resulting in a build-up of cellobiose and additional glucose. These, in turn, cause the overall rate of saccharification to decrease with time. One strategy used to overcome the accumulation of glucose has been to simultaneously saccharify and ferment (SSF) the substrate. In this process, an ethanol-producing organism is combined with the enzymatic saccharification step in order to constantly convert the fermentable sugars to ethanol, thereby relieving end-product inhibition (Stenberg *et al.*, 2000).

2.5.2 Flavour and Wine-making Industries

In the flavour industry, β -glucosidases are also key enzymes in the enzymatic release of aromatic compounds from glucosidic precursors present in fruits and fermenting products. Natural occurring flavour compounds (monoterpene alcohols) accumulate in fruits as

flavourless precursors linked to mono- or diglycosides and require enzymatic or acidic hydrolysis for the liberation of their fragrances. It has been well established that the release of certain monoterpenols present in grape berries (e.g. linalol, geraniol, nerol, citronelol, α -terpineol and linalol oxide), contribute significantly to the aroma of the wine (Wirth *et al.*, 2001). Unlike acidic hydrolysis, enzymatic hydrolysis is highly efficient and does not result in modifications of the aromatic character. Grape and yeast glucosidases exhibit limited activity on monoterpenyl-glucosides during the process of winemaking, and the addition of exogenous β -glucosidases during or following fermentation are required to effectively release the aromatic compounds to enhance wine flavour (Riou *et al.*, 1998). The most favourable enzyme properties required for β -glucosidases suitable for the wine industry are stability at a low pH (pH 2.5 to 3.8) and resistance to inhibition by glucose (10 to 20%) and ethanol (10 to 15%) (Saha and Botharst, 1996). The same principle is used in the Japanese brewing industry during the making of sweet potato *shochu*. In this process the key enzyme is the β -glucosidase of *Aspergillus kawachii*, which releases several free monoterpene alcohols that contribute to the sweet potato *shochu* flavour (Iwashita *et al.*, 1998).

The musts from red grapes are mainly composed of pigments called anthocyanins. These phenolic molecules are made up of a flavylum ion glycosylated with a β -glycosidic bond. Cleavage of this bond releases the corresponding anthocyanidin, which results in a decolourisation of the juice. Future contributions of β -glucosidase (commonly termed anthocyanase) to the wine industry will be the prevention of sediment formation during storage of the bottles and, even of greater importance, the recovery of free-run juice from red grape varieties that can be used for the production of white wines with lower red colour (Sánchez-Torres *et al.*, 1998).

Finally, β -glucosidases are also used to improve the organoleptic properties of citrus fruit juices. The bitterness of citrus juices is in part due to a glucosidic compound, naringin (4',5,7-trihydroxyflavanone-7-rhamnoglucoside) whose hydrolysis requires, in succession, the action of an α -rhamnosidase and a β -glucosidase (Zheng and Shetty, 2000; Riou *et al.*, 1998).

2.5.3 Metabolic Engineering of Bacteria for the Conversion of cellulose degradation products to ethanol

Metabolic engineering has been defined as “improving product formation or cellular properties through the modification of a specific biochemical reaction(s) or the introduction of new one(s) with recombinant DNA technology (Baily, 1991). Currently, the most successful attempts to generate a recombinant organism able to ferment glucose and cellulosic oligosaccharides involve two bacterial hosts: *Klebsiella oxytoca* and *E. coli* KO11.

K. oxytoca naturally transports and utilises cellobiose and cellotriose, the soluble intermediates from cellulose hydrolysis and therefore only required a redirection of the central metabolism to ethanol (Ingram *et al.*, 1999). The homo-ethanol producer that came under investigation was the bacterium *Zymomonas mobilis*. It contains a very efficient two-enzyme pathway for the conversion of pyruvate (and NADH) to ethanol and CO₂ (and NAD⁺), which consists of pyruvate decarboxylase and alcohol dehydrogenase. Subsequently the genes for *Z. mobilis* pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adhB*) were cloned and placed under the control of a single promoter, creating the PET (production of ethanol) operon (Ingram *et al.*, 1987). Transformation of the plasmid containing the artificial PET operon and further integration into the chromosome resulted in a *K. oxytoca* strain P2, able to produce ethanol from a variety of monomeric sugars and disaccharides (cellobiose, xylobiose, sucrose), trisaccharides (cellotriose, xylotri-ose, raffinose) and tetrasaccharides (stachiose) (Fig. 18). The combined effect of the high PET overexpression and the low apparent K_m value of the pyruvate decarboxylase enzyme for pyruvate (i.e. high affinity for pyruvate) effectively divert carbon flow to ethanol even in the presence of the native fermentation enzymes (e.g. lactate dehydrogenase [LDH]) (Zaldivar *et al.*, 2001; Aristidou and Penttila, 2000). In cellulose-containing medium, *K. oxytoca* P2 together with commercial cellulases (EG and CBH) produced ethanol with yields that exceeded 70% of the theoretical (Doran and Ingram, 1993). Further enhancement of *K. oxytoca* P2 cellulolytic potential was obtained by chromosomal integration of the *celZ* gene, encoding a *C. thermocellum* EG. Recently, efficient secretion (more than 50% of the EG) to the surrounding medium was achieved through expression of the *out* genes from *Erwinia chrysantemii* (Zhou and Ingram, 1999).

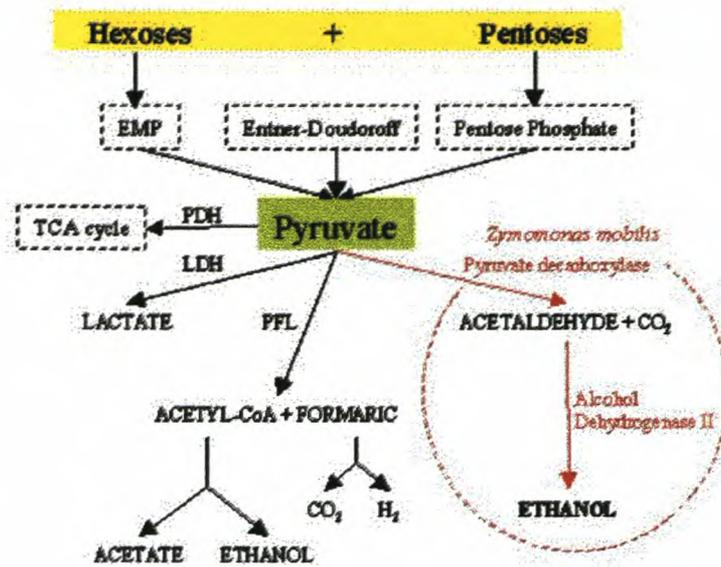


FIGURE 18. Fermentation pathway in recombinant *K. oxytoca* P2 (and *E. coli* KO11) which contains *Z. mobilis* genes for the conversion of pyruvate to ethanol. Abbreviations: EMP, Embden-Meyerhof pathway; PDH, pyruvate decarboxylase; LDH, lactate dehydrogenase; PFL, pyruvate formate-lyase.

E. coli KO11 was constructed to produce ethanol from acid hydrolysates of hemicellulose (pentoses and hexoses) by an identical approach described previously for *K. oxytoca*. Unfortunately, *E. coli* lacks the cellobiose-specific PTS for the transport and subsequent metabolism of cellobiose. KO11 was further engineered for the fermentation of cellulose by adding the *K. oxytoca casAB* genes encoding Enzyme II^{cellobiose} and phospho- β -glucosidase. Although the two *K. oxytoca* genes were well expressed in cloning hosts such as DH5 α , both were expressed poorly in *E. coli* KO11 (Lai *et al.*, 1997). Subsequently, screening for spontaneous mutants resulted in the isolation of clones that exhibited more than 15-fold higher specific activities for cellobiose metabolism. Interestingly, the mutations of these mutants existed in the plasmid and all contained similar internal deletions which eliminated the *casAB* promoter and operator regions. KO11 harbouring mutant plasmids rapidly fermented cellobiose to ethanol and the yield was more than 90% of the theoretical yield. Two of these strains were used with commercial cellulases to ferment mixed-waste office paper to ethanol (Moniruzzaman *et al.*, 1997).

With the rapid advances of genetic tools and genome sequencing we can expect to see more organisms added to the list of “potential biocatalysts” that will be genetically engineered for the production of fuel ethanol from cellulosic material. In the bacteria category there

will probably be more focus on less common species, for example the extremophiles and photosynthetic organisms.

2.6 References

- Amster-Choder, O., F. Houtman, and A. Wright. 1989. Protein phosphorylation regulates transcription of the β -glucosidase utilization operon in *E. coli*. *Cell* **58**:847-855.
- Aristidou, A., and M. Penttila. 2000. Metabolic engineering applications to renewable resource utilization. *Curr. Opin. Biotechnol.* **11**:187-98.
- Ashadi R.W., K. Shimokawa, K. Ogawa. 1996. The mechanism of enzymatic cellulose degradation (I). Purification and some properties of cellulolytic enzymes from *Aspergillus niger* UC. *J. Gen. Appl. Microbiol.* **42**:93-102.
- Ayers, W. A. 1958. Phosphorylation of cellobiose and glucose by *Ruminococcus flavefaciens*. *J. Bacteriol.* **76**:515-517.
- Aymerich, S. and M. Steinmetz. 1992. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the BglG/SacY family. *Proc. Natl. Acad. Sci. USA* **89**:10410-10414.
- Bailey, J.E. 1991. Towards a science of metabolic engineering. *Science* **252**:1668-1674.
- Bayer, E. A., H. Chanzy, R. Lamed, and Y. Shoham. 1998. Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* **8**:557.
- Béguin, P., and J. Aubert. 1994. The biological degradation of cellulose. *FEMS Microbiol. Rev.* **13**:25-58.
- Béguin, P., and M. Lemaire. 1996. The cellulosome: an extracellular, multiprotein complex specialized in cellulose degradation. *Crit. Rev. Biochem. Mol. Biol.* **31**:201-236.
- Bhat, S., P.W. Goodenough, M.K. Bhat, and E. Owen. 1994. Isolation of four major subunits from *Clostridium thermocellum* cellulosome and their synergism in the hydrolysis of crystalline cellulose. *Int. J. Biol. Macromol.* **16**:335-42.
- Blake, C.C.F., D.F. Koenig, F.A. Mair, A.C.T. North, D.C. Phillips, and V.R. Sarma. 1965. Structure of hen egg white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **206**:757-763.
- Boisset, C., H. Chanzy, B. Henrissat, R. Lamed, Y. Shoham, and E.A. Bayer. 1999. Digestion of crystalline cellulose substrate by the *Clostridium thermocellum* cellulosome: structural and morphological aspects. *Biochem. J.* **340**:829-835.
- Bolobova, A.V., A.V. Zhukov, and A.A. Klyosov. 1994. Lipids and fatty acids in cellulosomes of *Clostridium thermocellum*. *Appl. Environ. Microbiol.* **42**:128-133.
- Clarke, A.J. 1997. Families of β -glycosidases: classification of cellulolytic and xylanolytic enzymes, p191-199. *Biodegradation of cellulose: Enzymology and biotechnology*. Technomic Publishing Company, Inc., Basel.
- Coughlan, M.P. 1990. Cellulose degradation by fungi, p1-36. in W.M. Fogarty and C.T. Kelly (ed.), *Microbial enzymes and biotechnology*. Elsevier Applied Science, London.
- Cox, P.M., R.A. Betts, C.D. Jones, S.A. Spall and I.J. Totterdell. 2000. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature* **408**:184-187.

- Davies, G.J. and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* 3:853-859.
- Deutscher, J., B. Pevec, K. Beyreuther, H.H. Kiltz, and W. Hengstenberg. 1986. Streptococcal phosphoenolpyruvate-sugar phosphotransferase system: amino acid sequence and site of ATP-dependent phosphorylation of HPr. *Biochemistry* 25:6543-6551.
- Deutscher, J., C. Fischer, V. Charrier, A. Galinier, C. Lindner, E. Darbon, and V. Dossonnet. 1997. Regulation of carbon metabolism in gram-positive bacteria by protein phosphorylation. *Folia Microbiologica* 42:171-178.
- Deutscher, J., E. Kster, U. Bergstedt, V. Charrier, and W. Hillen. 1995. Protein kinase-dependent HPr/CcpA interaction links glycolytic activity to carbon catabolite repression in gram-positive bacteria. *Mol. Microbiol.* 15:1049-1053.
- Deutscher, J., J. Reizer, C. Fischer, A. Galinier, M.H. Saier, Jr., and M. Steinmetz. 1994. Loss of protein kinase-catalyzed phosphorylation of HPr, a phosphocarryer protein of the phosphotransfer system, by mutation of the *ptsH* gene confers catabolite repression resistance to several catabolic genes of *Bacillus subtilis*. *J. Bacteriol.* 176:3336-3344.
- Divne, C., J. Ståhlberg, T.T. Teeri, and T.A. Jones. 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.
- Doran, J.B., and L.O. Ingram. 1993. Fermentation of crystalline cellulose to ethanol by *Klebsiella oxytoca* containing chromosomally integrated *Zymomonas mobilis* genes. *Biotechnol. Prog.* 9:533-38.
- Driskill, L.E., M.W. Bauer, R.M. Kelly. 1999 Synergistic Interactions Among β -laminarinase, β -1,4-Glucanase, and β -Glucosidase from the hyperthermophilic archaeon *Pyrococcus furiosus* during hydrolysis of β -1,4-, β -1,3-, and mixed-linked polysaccharides. *Biotechnol. Bioeng.* 66:51-60.
- Fouet, A., A. Arnaud, A. Klier, and G. Rapoport. 1989. Genetics of the phosphotransferase system of *Bacillus subtilis*. *FEMS Microbiol. Rev.* 63:175-182.
- Fujino, T; Béguin, P; Aubert, J P. 1993. Organization of a *Clostridium thermocellum* gene cluster encoding the cellulosomal scaffolding protein CipA and a protein possibly involved in attachment of the cellulosome to the cell surface. *J. Bacteriol.* 175:1891-1899
- Gal, L., S. Pages, C. Gaudin, A. Belaich, C. Reverbel-Leroy, C. Tardif, J. Belaich. 1997. Characterization of the cellulolytic complex (cellulosome) produced by *Clostridium cellulolyticum*. *Appl. Environ. Microbiol* 63:903-909.
- Garrity, L.F., S.L. Schiela, R. Merrill, J. Reizer, M.H. Saier, Jr, and G.W. Ordal. 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.
- Grabnitz, F., M. Seiss, K.P. Rucknagel, W.L. Staudenbauer. 1991. Structure of the β -glucosidase gene *bglA* of *Clostridium thermocellum*. Sequence analysis reveals a superfamily of cellulases and beta-glycosidases including human lactase/phlorizin hydrolase. *Eur. J. Biochem.* 200:301-309.
- Grundy, F.J., D.A. Waters, S.H. Allen, and T.M. Henkin. 1993. Regulation of the *Bacillus subtilis* acetate kinase gene by CcpA. *J. Bacteriol.* 175:7348-55.
- Guedon, E., M. Desvaux, and H. Petitdemange. 2000a. Kinetic analysis of *Clostridium cellulolyticum* carbohydrate metabolism: importance of glucose-1-phosphate and glucose-6-phosphate branch points for distribution of carbon fluxes inside and outside cells as revealed by steady-state continuous culture. *J. Bacteriol.* 182:2010-2017.

- Guedon, E., S. Payot, M. Desvaux, and H. Petitdemange. 2000b. Relationships between cellobiose catabolism, enzyme levels, and metabolic intermediates in *Clostridium cellulolyticum* grown in a synthetic medium. *Biotechnol. Bioeng.* **67**:327-335.
- Hazlewood, G.P., and H.J. Gilbert. 1993. Xylan and cellulose utilization by the clostridia. *Biotechnol.* **25**:311-41.
- Helaszek, C. T., and B. A. White. 1991. Cellobiose uptake and metabolism by *Ruminococcus flavefaciens*. *Appl. Environ. Microbiol.* **57**:64-68.
- Hengstenberg W., D. Kohlbrecher, E. Witt, R. Kruse, I. Christiansen, D. Peters, R. Pogge von Strandmann, P. Stadler, B. Koch, H.R. Kalbitzer. 1993. Structure and function of proteins of the phosphotransferase system and of 6-phospho-beta-glycosidases in gram-positive bacteria. *FEMS Microbiol. Rev.* **12**:149-63.
- Henrissat B., I. Callebaut, S. Fabrega, P. Lehn, J.P. Mornon, and G. Davies. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc. Natl. Acad. Sci. USA.* **92**:7090-7094.
- Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarity. *Biochim. J.* **280**: 309-316.
- Henrissat, B. and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **293**:781-788.
- Henrissat, B. and A. Bairoch. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**:695-696.
- Henrissat, B. and A. Romeu. 1995. Families, superfamilies and subfamilies of glycosyl hydrolases. *Biochem. J.* **311**:350-351.
- Henrissat, B., M. Claeyssens, P. Tomme, L. Lemesle, and J.P. Mornon. 1989. Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81**:83-95.
- Hon, S. 1994. Cellulose: a random walk along its historical path. *Cellulose* **1**:1-25.
- Houman, F., M. R. Diaz-Torres, and A. Wright. 1990. Transcriptional antitermination in the *bgl* operon of *E. coli* is modulated by a specific RNA binding protein. *Cell* **62**:1153-1163.
- Hueck, C.J., and W. Hillen. 1995. Catabolite repression in *Bacillus subtilis*: A global regulatory mechanism for the Gram-positive bacteria? *Mol. Microbiol.* **15**:395-401.
- Hueck, C.J., W. Hillen, and M.H. Saier. 1994. Analysis of a *cis*-active sequence mediating catabolite repression in gram-positive bacteria. *Res. Microbiol.* **145**:503-518.
- Ingram, L.O., H.C. Aldrich, A.C.C. Borges, T.B. Causey, A. Martinez, F. Morales, A. Saleh, S.A. Underwood, L.P. Yomano, S.W. York, J. Zaldivar and S. Zhou. 1999. Enteric bacterial catalysts for fuel ethanol production. *Biotech. Bioeng.* **58**:204-214.
- Ingram, L.O., T. Conway, D.P. Clark, G.W. Sewell and J.F. Preston. 1987. Genetic engineering of ethanol production in *Escherichia coli*. *Appl. Environ. Microbiol.* **53**:2420-2425.
- Irwin, D.C., M. Spezio, L.P. Walker, and D.B. Wilson. 1993. Activity studies of eight purified cellulases: specificity, synergism, and binding domain effects. *Biotechnol. Bioeng.* **42**:1002-1013.
- Iwashita, K., K. Todoroki, H. Kimura, H. Shimoi, and K. Ito. 1998. Purification and characterization of extracellular and cell wall-bound β -glucosidases from *Aspergillus kawachii*. *Biosci. Biotech. Biochem.* **62**:1938-1946.

- Jenkins, J. I. Lo Leggio, G. Harris, and R. Pickersgill. 1995. β -Glucosidase, β -galactosidase, family A cellulases, family F xylanases and two barley glycanases for a superfamily with 8-fold/architecture and with two conserved glutamates near the carboxy-terminal ends of strands four and seven. *FEBS. Lett.* **362**:281-285.
- Kajikawa, H. and S. Masaki. 1999. Cellobiose transport by mixed ruminal bacteria from a cow. *Appl. Environ. Microbiol.* **65**:2565-2569.
- Klyosov, A.A. 1990. Trends in biochemistry and enzymology of cellulose degradation. *Biochemistry.* **29**:10577-10585.
- Kotrba, P., M. Inui, and H. Yukawa. 2001. Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *J. Biosci. Bioeng.* **92**:502-517.
- Krüger, S. and M. Hecker. 1995. Regulation of the putative *bglPH* operon for aryl- β -glucoside utilization in *Bacillus subtilis*. *J. Bacteriol.* **177**:5590-5597.
- Krüger, S., S. Gertz, and M. Hecker. 1996. Transcriptional analysis of *bglPH* expression in *Bacillus subtilis*: Evidence for two distinct pathways mediating carbon catabolite repression. *J. Bacteriol.* **178**:2637-2644.
- Kubicek, C.P., R. Messner, F. Gruber, R.L. Mach and E.M. Kubicek-Pranz. 1993. The *Trichoderma* cellulase regulatory puzzle: From the interior life of a secretory fungus. *Enzyme. Microb. Technol.* **15**:92-0-99.
- Lai, X. and L. O. Ingram. 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *J. Bacteriol.* **175**:6441-6450.
- Lai, X., F. C. Davis, R. B. Hespell, and L. O. Ingram. 1997. Cloning of cellobiose phosphoenolpyruvate-dependent phosphotransferase genes: Functional expression in recombinant *Escherichia coli* and identification of a putative binding region for disaccharides. *Appl. Environ. Microbiol.* **63**:335-363.
- Le Coq, D., C. Lindner, S. Krüger, M. Steinmetz, and J. Stülke. 1995. New β -glucoside (*bgl*) genes in *Bacillus subtilis*: The *bglP* gene product has both transport and regulatory functions similar to those of BglF, its *Escherichia coli* homolog. *J. Bacteriol.* **177**:1527-1535.
- Lemaire, M., I. Miras, P. Gounon, and P. Béguin. 1998. Identification of a region responsible for binding to the cell wall within the S-layer protein of *Clostridium thermocellum*. *Microbiology* **144**:211-217.
- Lengeler, J.W., K. Jahreis, and U.F. Whemeier. 1994. Enzymes II of the phosphoenolpyruvate-dependent phosphotransferase systems: Their structure and function in carbohydrate transport. *Biochim. Biophys. Acta.* **1188**:1-28.
- Lo, A.C., R.M. MacKay, V.L. Seligy, and G.E. Willick. 1988. *Bacillus subtilis* β -1,4-endoglucanase products from intact and truncated genes are secreted into the extracellular medium by *Escherichia coli*, *Appl. Environ. Microbiol.* **54**:2287-2292
- Lopez-Camacho, C., J. Salgado, J. L. Lequerica, A. Madarro, E. Ballestar, L. Franco, and J. Polaina. 1996. Amino acid substitutions enhancing thermostability of *Bacillus polymyxa* β -glucosidase A. *Biochem. J* **314**:833-838.
- Lux, R., V.R. Munasinghe, F. Castellano, J.W. Lengeler, J.E. Corrie, and S. Khan. 1999. Elucidation of a PTS-carbohydrate chemotactic signal pathway in *Escherichia coli* using a time-resolved behavioral assay. *Mol. Biol. Cell* **10**:1133-1146.
- Lynd, L.R., P.J. Weimer, W.H. van Zyl, and I.S. Pretorius. 2002. Microbial cellulose utilization: Fundamentals & Biotechnology. *Microbiol. Mol. Biol. Rev.* (in press)

- Matthews, B.W. and S.J. Remington. 1974. The three-dimensional structure of lysozyme from bacteriophage T4. *Proc. Natl. Acad. Sci. USA.* **71**:4178-4182.
- Mayer, F. 1988. Cellulolysis: ultrastructural aspects of bacterial systems. *Electron Microsc. Rev.* **1**:69-85.
- McCarter, J.D. and S.G. Withers. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **4**:885-892.
- Medve, J., J. Karlsson, J. Lee, and F. Tjerneld. 1998. Hydrolysis of microcrystalline cellulose by cellobiohydrolase I and endoglucanase II from *Trichoderma reesei*: Adsorption, sugar production and synergism of the enzymes. *Biotechnol. Bioeng.* **59** :621-634.
- Moniruzzaman, M., X. Lai, S. W. York, and L. O. Ingram. 1997. Isolation and molecular characterisation of high-performance cellobiose-fermenting spontaneous mutants of ethanologenic *Escherichia coli* KO11 containing the *Klebsiella oxytoca casAB* operon. *Appl. Environ. Microbiol.* **63**:4633-4637.
- Nidetzky, B., W. Steiner, M. Hayn, and M. Claeysens. 1994. Cellulose hydrolysis by the cellulases from *Trichoderma reesei*: a new model for synergistic interaction. *Biochem. J.* **298**:705-710.
- Ooshima, H., M. Kurakake, J. Kato, and Y. Harano Y. 1991. Enzymatic activity of cellulase adsorbed on cellulose and its change during hydrolysis. *Appl. Biochem. Biotechnol.* **31**:3253-3266.
- Parker, L. L., and B. G. Hall. 1990. Characterization and nucleotide sequence of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**:455-471.
- Percy, A., H. Ono. And K. Hayashi. 1998. Acceptor specificity of cellobiose phosphorylase from *Cellvibrio gilvus*: synthesis of three branched trisaccharides. *Carbohydr. Res.* **308**:423-429.
- Philippidis, G.P., T.K. Smith, and C.E. Wyman. 1993. Study of the enzymatic hydrolysis of cellulose for production of fuel ethanol by the simultaneous saccharification and fermentation process. *Biotechnol. Bioeng.* **41**:846-853.
- Postma, P.W., J.W. Lengeler, and G.R. Jacobson. 1993. Phosphoenolpyruvate:carbohydrate phosphotransfer system of bacteria. *Microbiol. Rev.* **57**:543-594.
- Poulsen, O.M., and L.W. Petersen. 1992. Degradation of microcrystalline cellulose: synergism between different endoglucanases of *Cellulomonas* sp. ATCC 21399. *Biotechnol. Bioeng.* **39**:121-123.
- Reizer, J., S. Bachem, A. Reizer, M. Arnaud, M.H. Saier, and J. Stülke. 1999. Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. *Microbiology* **145**:3419-3429.
- Riou, C., J. M. Salmon, M. J. Vallier, Z. Gunata, and P. Barre. 1998. Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β -glucosidase from *Aspergillus oryzae*. *Appl. Environ. Microbiol.* **64**:3607-3614.
- Rutberg, B. 1997. Antitermination of transcription of catabolic operons. *Mol. Microbiol.* **23**:413-421.
- Saha, B.C., and R.J. Bothast. 1996. Production, purification, and characterization of a highly glucose-tolerant novel beta-glucosidase from *Candida peltata*, *Appl. Environ. Microbiol.* **62**:3165-3170
- Saier M.H., Jr, S. Chauvaux, J. Deutscher, J. Reizer, and J.J. Ye. 1995. Protein phosphorylation and regulation of carbon metabolism in gram-negative versus gram-positive bacteria. *Trends Biochem. Sci.* **20**:267-71.
- Saier, M. H. and J. Reizer. 1994. The bacterial phosphotransferase system: New frontiers 30 years later. *Mol. Microbiol.* **13**:755-764.

- Sánchez-Torres, P., L. González-Candelas, and D. Ramón. 1998. Heterologous expression of a *Candida molischiana* anthocyanin- β -glucosidase in a wine yeast strain. *J. Agric. Food Chem.* **46**:354-360.
- Sanz-Aparicio, J., J. A. Hermoso, M. Martínez-Ripoll, B. Gonzalez, C. Lopez-Camacho, and J. Polaina. 1998a. Structural basis of increased resistance to thermal denaturation induced by single amino acid substitution in the sequence of β -glucosidase A from *Bacillus polymyxa*. *Proteins* **33**:567-576.
- Sanz-Aparicio, J., J. A. Hermoso, M. Martínez-Ripoll, J. L. Lequerica, and J. Polaina. 1998b. Crystal structure of β -glucosidase A from *Bacillus polymyxa*: Insights into the catalytic activity in family I glycosyl hydrolases. *J. Mol. Biol.* **275**:491-502.
- Saxena, I.M., and R.M. Brown, Jr. 2000. Cellulose synthases and related enzymes, *Curr. Opin. Plant Biol.* **3**:523-531.
- Schimz, K. L., B. Broll, and B. John. 1983. Cellobiose phosphorylase (EC 2.4.1.20) of *Cellulomonas* occurrence, induction, and its role in cellobiose metabolism. *Arch. Microbiol.* **135**:241-249.
- Schimz, K.L., B. Broll, and B. John. 1983. Cellobiose phosphorylase (EC 2.4.1.20) of *Cellulomonas*: occurrence, induction, and its role in cellobiose metabolism. *Arch. Microbiol.* **135**:241-249.
- Schlichtermeier, A; S. Walter, J. Schröder, M. Moorman, and H. Schrempf. 1992. The gene encoding the cellulase (Avicelase) Cell from *Streptomyces reticuli* and analysis of protein domains. *Mol. Microbiol.* **6**:3611-3621.
- Schnetz, K. and B. Rak. 1990. β -Glucoside permease represses the *bgl* operon of *Escherichia coli* by phosphorylation of the antiterminator protein and also interacts with glucose-specific enzyme III, the key element in catabolite control. *Proc. Natl. Acad. Sci. USA* **87**:5074-5078.
- Schnetz, K., J. Stülke, S. Gertz, S. Krüger, M. Krieg, M. Hecker, and B. Rak. 1996. LicT, a *Bacillus subtilis* transcriptional antiterminator protein of the BglG family. *J. Bacteriol.* **178**:1971-1979.
- Schulte, D., and W. Hengstenberg. 2000. Engineering the active center of the 6-phospho-beta-galactosidase from *Lactococcus lactis*, *Prot. Eng.* **13**:515-518.
- Schwarz, W.H. 2001. The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* **56**:634-649.
- Shoham, Y., R. Lamed, and E.A. Bayer. 1999. The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides. *Trends Microbiol.* **7**:275-281.
- Shoseyov, O., M. Takagi, M.A. Goldstein, and R.H. Doi. 1992. Primary sequence analysis of *Clostridium cellulovorans* cellulose binding protein A. *Proc. Natl. Acad. Sci. USA.* **89**:3483-3487.
- Siebold, C., K. Flükiger, R. Beutler, and B. Erni. 2001, Carbohydrate transporters of the bacterial phosphoenolpyruvate: sugar phosphotransferase system (PTS). *FEBS Lett.* **504**:104-111.
- Sinnott, M.L. 1990. Catalytic mechanisms of enzymic glycosyl transfer. *Chem. Rev.* **90**:1171-1202.
- Srisodsuk, M., J. Lehtio, M. Linder, E. Margolles-Clark, T. Reinikainen, and T.T. Teeri. 1997. *Trichoderma reesei* cellobiohydrolase I with an endoglucanase cellulose-binding domain: action on bacterial microcrystalline cellulose. *J. Biotechnol.* **57**:49-57.
- Stalbrand, H., S.D. Mansfield, J.N. Saddler, D.G. Kilburn, R.A.J. Warren, and N.R. Gilkes. 1998. Analysis of molecular size distribution of cellulose molecules during hydrolysis of cellulose by recombinant *Cellulomonas fimi* β -1,4-glucanases. *Appl. Environ. Microbiol.* **64**:2374-2379.

- Steinmetz, M., D. Le Coq, and S. Aymerich. 1989. Induction of saccharolytic enzymes by sucrose in *Bacillus subtilis*: evidence for two partially interchangeable regulatory pathways. *J. Bacteriol.* **171**:1519-1523.
- Stenberg, K., M. Bollók, G. Zacchi, and M. Galbe. 2000. Effect of substrate and cellulase concentration on simultaneous saccharification and fermentation of steam-pretreated softwood for ethanol production. *Biotechnol. Bioeng.* **68**:204-210.
- Sternberg, D., P. Vijayakumar, and E. T. Reese. 1977. β -Gucosidase: Microbial production and effect on enzymatic hydrolysis of cellulose. *Can. J. Microbiol.* **23**:139-147.
- Stülke, J., M. Arnaud, G. Rapoport, and I. Martin-Verstraete. 1998. PRD - a protein domain involved in PTS-dependent induction and carbon catabolite repression of catabolic operons in bacteria. *Mol. Microbiol.* **28**:865-874.
- Sun, D.X., and I. Takahashi. 1984. A catabolite-resistance mutation is localized in the *rpo* operon of *Bacillus subtilis*, *Can. J. Microbiol.* **30**: 423-429.
- Tan, M.W., L.G. Rahme, J.A. Sternberg, R.G. Tompkins, and F.M. Ausubel. 1999. *Pseudomonas aeruginosa* killing of *Caenorhabditis elegans* used to identify *P. aeruginosa* virulence factors, *Proc. Natl. Acad. Sci. USA* **96**:2408-2413
- Teeri, T.T. 1997. Crystalline cellulose degradation: new insight into the function of cellobiohydrolases. *Trends. Biotechnol.* **15**:160-167.
- Tengborg, C. 2000. Bioethanol Production: Pretreatment and Enzymatic Hydrolysis of Softwood. Ph.D. thesis. Department of Chemical Engineering1, Lund University, Sweden.
- Thurston, B., K. A. Dawson, and H. J. Strobel. 1993. Cellobiose versus glucose utilization by the ruminal bacterium *Ruminococcus albus*. *Appl. Environ. Microbiol.* **59**:2631-2637.
- Tobisch, S., J. Stülke, and M. Hecker. 1999. Regulation of the *lic* operon of *Bacillus subtilis* and characterization of potential phosphorylation sites of the LicR regulator protein by site-directed mutagenesis. *J. Bacteriol.* **181**:4995-5003.
- Tobisch, S., P. Glaser, S. Kruger, and M. Hecker. 1997. Identification and characterization of a new beta-glucoside utilization system in *Bacillus subtilis*. *J. Bacteriol.* **179**:496-506.
- Tomme, P., E. Kwan, N. R. Gilkes, D. G. Kilburn, and A. J. Warren. 1996. Characterization of CenC, an enzyme from *Cellulomonas fimi* with both endo- and exoglucanase activities. *J. Bacteriol.* **178**:4216-4223.
- Tomme, P., R. A. J. Warren, and N. R. Gilkes. 1995. Cellulose Hydrolysis by Bacteria and Fungi. *Adv. Microb. Physiol.* **37**:1-81.
- Väljamäe, P., V.V. Sild, A. Nutt, G. Pettersson, and G. Johansson. 1999 Acid hydrolysis of bacterial cellulose reveals different modes of synergistic action between cellobiohydrolase I and endoglucanase I. *Eur. J. Biochem.* **266**:327-334.
- Väljamäe, P; Sild, V; Pettersson, G; Johansson, G. 1998. The initial kinetics of hydrolysis by cellobiohydrolases I and II is consistent with a cellulose surface-erosion model. *Eur. J. Biochem.* **253**:469-475.
- Varrot, A., S. Hastrup, and G. J. Davies. 1999. Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from *Humicola insolens*, at 1.92 Å resolution. *Biochem. J* **337**:297-304.

- Walker, L.P., D.B. Wilson, D.C. Irwin, C. McQuire, and M. Price. 1992. Fragmentation of cellulose by the major *Thermomonospora fusca* cellulases, *Trichoderma reesei* CBHI, and their mixtures. *Biotechnol. Bioeng.* **40**:1019-1026.
- Weickert M.J., and S. Adhya. 1992. A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* **267**:15869-15874.
- Wells, J. E., J. B. Russell, Y. Shi, and P. J. Weimer. 1995. Cellodextrin efflux by the cellulolytic ruminal bacterium *Fibrobacter succinogenes* and its potential role in the growth of nonadherent bacteria. *Appl. Environ. Microbiol.* **61**:1757-1762.
- Wickert, M.J. and G.H. Chambliss. 1990. Site-directed mutagenesis of a catabolite repression operator sequence in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA* **87**:6238-6242.
- Wiesmann C, Hengstenberg W, Schulz GE. 1997. Crystal structures and mechanism of 6-phospho-beta-galactosidase from *Lactococcus lactis*. *J. Mol. Biol.* **269**:851-60.
- Wirth, J., W. Guo, R. Baume, and Z. Gunata. 2001. Volatile compounds released by enzymatic hydrolysis of glycoconjugates of leaves and grape berries from *Vitis vinifera* Muscat of Alexandria and Shiraz cultivars. *J. Agric. Food Chem.* **49**:2917-2923.
- Wood, T.M., and S.I. McCrae. 1978. The cellulase of *Trichoderma koningii*. Purification and properties of some endoglucanase components with special reference to their action on cellulose when acting alone and in synergism with the cellobiohydrolase. *Biochem. J.* **171**:61-72.
- Woodward, J., M. Lima, and N.E. Lee. 1988. The role of cellulase concentration in determining the degree of synergism in the hydrolysis of microcrystalline cellulose. *Biochem. J.* **255**:895-899.
- Zaldivar, J., J. Nielsen, and L. Olsson. 2001. Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. *Appl. Microbiol. Biotechnol.* **56**:17-34.
- Zheng, Z. and K. Shetty. 2000. Solid-state bioconversion of phenolics from cranberry pomace and role of *Lentinus edodes* β -glucosidase. *J. Agric. Food Chem.* **48**:895-900.
- Zhou, S., and L.O. Ingram. 1999. Engineering endoglucanase-secreting strains of ethanologenic *Klebsiella oxytoca* P2. *J Ind. Microbiol.* **2**:607

Chapter 3

Cloning of a novel *Bacillus pumilus*
cellobiose utilising system: Functional
expression in *Escherichia coli*

Submitted to:
Journal of Bacteriology

3 CLONING OF A NOVEL *BACILLUS PUMILUS* CELLOBIOSE UTILISING SYSTEM: FUNCTIONAL EXPRESSION IN *ESCHERICHIA COLI*

RONÉL VAN ROOYEN, DANIËL C. LA GRANGE and WILLEM H. VAN ZYL

Department of Microbiology, University of Stellenbosch, Stellenbosch 7600, South Africa

3.1 Abstract

A genomic DNA library of the bacterium *Bacillus pumilus* PLS in *Escherichia coli* was screened for the production of β -glucosidase activity on the chromophore, *p*-nitrophenyl- β -D-glucopyranoside (pNPG) and positive isolates were retained. Plasmid DNA was retrieved from one of the isolates and DNA sequence analysis of the insert revealed the presence of a novel catabolic system in *B. pumilus* involved in the utilisation of β -glucoside compounds. It consists of an operon of four genes (*celBACH*), encoding a phosphotransferase system (PTS): enzyme II (encoded by *celB*, *celA* and *celC*), and a 6-phospho- β -glucosidase (encoded by *celH*). A 3.4-Kb fragment, containing the complete operon, was amplified from a 5-kb genomic DNA fragment with the aid of the polymerase chain reaction (PCR) technique and inserted into plasmid pGEM-T-*easy*[®], generating plasmid pBPU-CelBACH. The presence of pBPU-CelBACH permitted growth of the recombinant *Escherichia coli* on cellobiose as carbon source. The optimal temperature and pH at which the recombinant *E. coli* strain hydrolysed pNPG were 45°C and pH 7.5, respectively. The CelBACH-associated activity in *E. coli* was stable at 45°C for 20 minutes, but activity decreased significantly at higher temperature.

3.2 Introduction

Cellulose, a linear polymer of β -1,4-linked glucose units, is the major carbohydrate synthesized by plants (Yan *et al.*, 1998). In nature cellulose is degraded by the synergistic action of at least three distinct enzymes: cellobiohydrolases (exoglucanases; EC 3.2.1.91); endoglucanases (EC 3.2.1.4), and β -glucosidases. β -Glucosidases are a heterogenous group of enzymes that display a broad substrate specificity with respect to hydrolysis of

cellobiose and different aryl- and alkyl- β -D-glucosides and occur widely in animals, plants, fungi and bacteria. Bacteria have developed several mechanisms for molecules (e.g. cellobiose) to be transported across the cytoplasmic membrane. The three known routes by which cellobiose is transported and metabolized are summarised in Fig. 1.

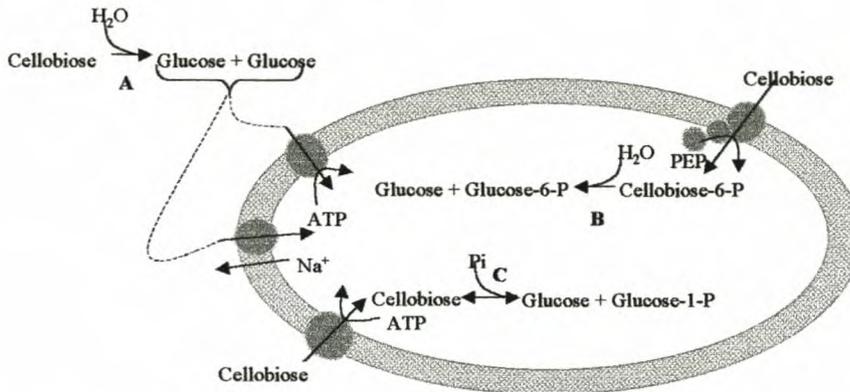


FIGURE 1. A cartoon depicting the different routes by which cellobiose is transported and metabolized: 1.) Extracellular hydrolysis of cellulose by β -glucosidase (A) and subsequent transport of the glucose residues via ATP-dependent active transport or Na^+ antiport. 2.) Transport of cellobiose via the cellobiose-specific phosphotransferase system (PTS) and subsequent hydrolysis by phospho- β -glucosidase (B). 3.) Direct uptake of cellobiose via an active ATP-dependent transport system and phosphorylytic cleavage of the molecule by cellobiose phosphorylase (C). Note that this reaction is ATP-independent and fully reversible.

The presence of several transport systems for the same molecule - in this case cellobiose - suggests an important role for regulation at this level. Up to now, the phosphorylytic pathway has only been described in clostridia and rumen bacteria (Kajikawa and Masaki, 1999; Tanaka *et al.*, 1995). Although the relative importance of these alternatives is not well understood in cellulolytic bacteria, the potential for ATP production via the phosphorylytic cleavage pathway could be advantageous for anaerobic cellulose utilisation (Lynd *et al.*, 2002). Cellobiose-degrading enzymes are of particular interest primarily because of their involvement in the biological saccharification of cellulosic material. They not only catalyse the final step in the saccharification of cellulose, but also stimulate the extent of cellulose hydrolysis by relieving the cellobiose mediated inhibition of exo- and endoglucanases. The industrial applications of β -glucosidases extend from enzymatic release of aromatic compounds from glucosidic precursors present in fruits and fermenting products (Iwashita *et al.*, 1998; Riou *et al.*, 1998) to the production of fuel ethanol from cellulosic residues (Himmel *et al.*, 1999; Leclerc *et al.*, 1986).

Soil bacteria, including members of the family *Bacillaceae*, are efficient in utilising a wide variety of carbohydrates, which include amorphous cellulose, β -1,3-glucan (laminarin) and β -1,3-1,4-glucan (lichenan) (Reizer *et al.*, 1999). Initially, enzymes with distinct linkage specificities degrade the polymers extracellularly. The generated β -glucosides are taken up via an inducible, cellobiose-binding protein-dependent phosphotransport system (PTS). In the model bacterium *B. subtilis*, the complete set of genes encoding PTS and PTS-associated proteins is known and several genes and operons involved in β -glucoside utilisation have been identified. β -Glucosides, such as lichenan and its hydrolysis products, can be considered alternative carbon sources that are utilized under limited glucose conditions. Two of the operons of *B. subtilis* that have been studied extensively are the *bglPH* operon (Krüger and Hecker, 1995; Le Coq *et al.*, 1995) and the *licBCAH* operon (Tobisch *et al.*, 1997). Both the *bglPH* and *licBCAH* encode a PTS enzyme II and a phospho- β -glucosidase.

The degradation of cellulosic biomass represents a significant part of the carbon cycle within the biosphere. Therefore the treatment of cellulose by cellulolytic enzymes has been of ongoing interest to biotechnologists. In this paper we describe the cloning of the *B. pumilus* cellobiose utilisation system (*celBACH*) and its functional expression in *E. coli*.

3.3 Materials and methods

3.3.1 Bacterial Strains and plasmids

The bacterial strains and plasmids used in this study are listed in Table 1.

TABLE 1. Strains and plasmids used in this study.

Strain or plasmid	Genetic characteristic(s)	Source
<i>B. pumilus</i> PLS	Wild type strain	(La Grange <i>et al.</i> , 1997)
<i>E. coli</i>		
XL1-Blue	<i>endA1 supE44 thi-1 recA1 gyrA96 relA1 lac[F' proAB lacI^qZΔM15 Tn10 (tet)]</i>	Stratagene
Plasmids		
YEp6 Δ BamHI	<i>bla LEU2 (BamHI deleted)</i>	(Van Zyl <i>et al.</i> , 1989)
pGEM-T-easy [®]	<i>bla</i>	Promega
pBluescript KS(+)	<i>bla</i>	Stratagene

3.3.2 Media for bacterial growth

E. coli XL1 Blue was cultured at 37°C in Luria Bertani (LB) medium supplemented with ampicillin (100 µg/ml) plasmid selection (Sambrook *et al.*, 1989). 5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside (XGal) and isopropyl-β-D-thiogalactopyranoside (IPTG) were added when required to a final concentration of 20 mg/ml and 200 mg/ml, respectively. Recombinant strains of *E. coli* were evaluated for β-glucoside utilisation using M9 minimal media containing 0.2% carbon source (cellobiose, arbutin, or salicin) without casamino acids (Ausubel *et al.*, 1994). In the modified M9 minimal media the casamino acids were substituted with a vitamin solution (0.005 µg/l biotin; 0.1 µg/l calcium pantothenate; 2.5 µg/l *myo*-inositol; and 0.1 µg/l thiamine-HCl) and trace elements (0.45 µg/l ZnSO₄·7H₂O; 0.1 µg/l MnCl₂·2H₂O; 0.03 µg/l CoCl₂·6H₂O; 0.03 µg/l CuSO₄·5H₂O; 0.04 µg/l Na₂MoO₄·2H₂O; 0.45 µg/l CaCl₂·2H₂O; 0.3 µg/l FeSO₄·7H₂O; 0.1 µg/l H₃BO₃; and 0.01 µg/l KI).

3.3.3 Cloning of the *B. pumilus* *celBACH* operon

Procedures used for the preparation of *E. coli* plasmids, the assembly of recombinant DNA, and the transformation of *E. coli* have been described previously (Sambrook *et al.*, 1989). Digestions with restriction enzymes were carried out as recommended by the manufacturers. Transformants of the *B. pumilus* genomic library constructed in YEp62Δ*Bam*HI (La Grange *et al.*, 1997), were screened for β-glucosidase activity on LB plates containing 1 mM *p*-nitrophenyl-β-D-glucopyranoside (pNPG) and ampicillin (100 µg/ml).

3.3.4 Polymerase chain reaction

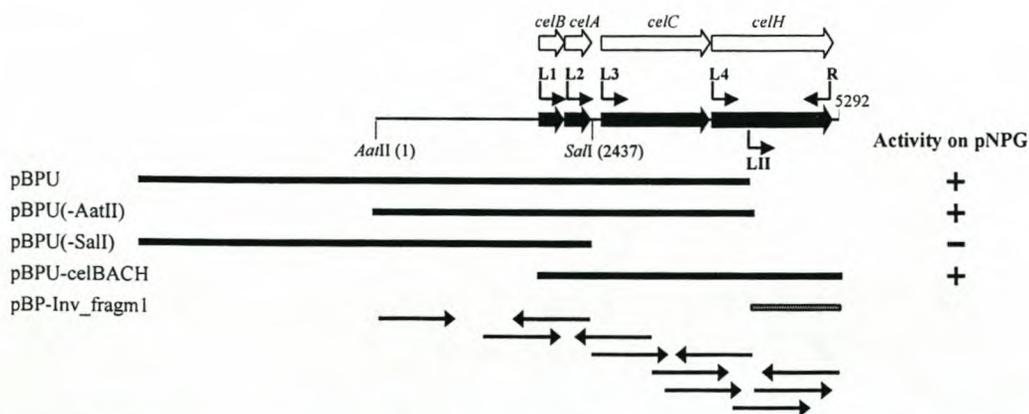
PCR products were amplified from either plasmid DNA (15 ng) or *B. pumilus* genomic DNA (200 ng) with the aid of sequence-specific primers. The reaction (50 µl) mixture contained the following components: 10x reaction buffer, 500 µM of each of the nucleotide triphosphates, 0.25 µM of each primer, DNA template, and 2.5 U EXPAND polymerase (Roche). Table 2 summarises of the different PCR primers used in this study.

TABLE 2. Summary of the PCR primers used. The underlined sequences represent the *EcoRI* and *XhoI* sites in the left and right primers, respectively. The ATG-initiation- and TTA-stop codons are indicated in bold.

Primer	Primer sequence	Amplified fragment (bp)
BP-CEL-L1	5'-TCGCGAGAATTCATGAAACGTAATTTATTAGCATG-3'	3409-bp
BP-CEL-R	5'-GACTCTCGAGTTAAAAACCGTTATTCTTTGAA-3'	
BP-CEL-L2	5'-TCGCGAGAATTCATGGAAAAGAAAACACTCGA-3'	3094-bp
BP-CEL-R		
BP-CEL-L3	5'-TCGCGAGAATTCATGCTTGCGTTCCCG-3'	2649-bp
BP-CEL-R		
BP-CEL-L4	5'-TCGCGAGAATTCATGGAAACTAACAAGCAATCA-3'	1400-bp
BP-CEL-R		
BP-CEL-LII	5'-TTGCGGCCGCGCCGATGCTTTCTTTAAC-3'	~1000-bp
YEp62-L	5'-AATGACAGCTCGGTCGAC-3'	
BP-CEL-LII		~1000-bp
YEp62-R	5'-CGACGTTGTAAAACGACG-3'	

3.3.5 Plasmid constructions

A 6.8-kb genomic DNA insert was amplified from one of the *B. pumilus* clones that were able to produce a yellow zone on LB-plates containing pNPG. PCR primers YEp62-L and YEp62-R, annealing on either side of the cloning site of YEp62 Δ BamHI were used to subclone the fragment into pGEM-T-*easy*[®]. The resulting construct, pBPU, was used to create a number of sub-clones for sequencing (Fig. 2). Sequence analysis revealed that ~1000 bp was absent at the 3' end of the last ORF (*celH*). In spite of this, the recombinant *E. coli* expressing pBPU was able to produce a yellow zone on LB/pNPG.

**FIGURE 2.** Sequencing strategy, and subclones constructed for sequencing. The position of the annealing PCR primers (summarised in Table 2) are indicated.

To obtain the remaining ~1000 bp, inverse PCR, with 2 sets of primers (BP-CEL-LII/YEp62-L and BP-CEL-LII/YEp62-R), were used to isolate fragments in the range of 1-kb to 3-kb from the *B. pumilus* genomic library. Fragments of approximately 1.5-kb were isolated and cloned into pGEM-T-*easy*[®] for DNA sequencing. Sequence analysis of

these clones provided the 3'-end of the *celH* gene. Assembly of the sequence data provided the complete sequence of the *celBACH* operon that was used for the design of primers, BP-CEL-L1 and BP-CEL-R. These primers were used to amplify a 3409-kb DNA fragment from *B. pumilus* chromosomal DNA. The PCR product was cloned into pGEM-T-*easy*[®] pre-digested with *EcoRI* and *XhoI* to yield pBPU-CelBACH. Plasmid pBPU-CelACH, containing the *celACH* genes, was amplified from genomic DNA of *B. pumilus* with primers BP-CEL-L2 and BP-CEL-R, *celCH* with primers BP-CEL-L3 and BP-CEL-R, and *celH* with primers BP-CEL-L4 and BP-CEL-R (Table 2). All of these PCR fragments were cloned into pGEM-T-*easy*[®] and transformed to *E. coli* and tested for β -glucosidase activity.

3.3.6 DNA sequencing and sequence analysis

Eleven subclones were constructed for sequencing of *celBACH*. The *celBACH* nucleotide sequence was determined by amplifying DNA fragments with the Big Dye Terminator Cycle sequencing Reader reaction with Amplitaq DNA polymerase F5 (Applied Biosystems kit) using fluorescently labeled nucleotides, and reaction mixtures were subjected to electrophoresis on an Applied Biosystems automatic DNA sequencer (model ABI Prism[™] 377). Sequence data was analyzed by using the PC/GENE software package (IntelliGenetics, Inc. Mountain View, California) and the DNA sequence was deposited at Genbank (accession number AY124778). Sequence analysis revealed four open reading frames (ORF's) organized in an operon, which were compared with sequence information on the NCBI database using the BLAST program (Altschul *et al.*, 1990).

3.3.7 *In vivo* assay of PTS activity

All four genes on the cloned *B. pumilus* DNA fragment were necessary to allow *E. coli* to hydrolyse pNPG. Recombinant *E. coli* strains expressing the *B. pumilus* genes were cultivated over night in Luria Bertani (LB) medium supplemented with ampicillin. Cells (2 ml samples) were harvested and resuspended in physiological salt solution (0.9% NaCl). The suspension of intact cells were incubated with 5 mM pNPG substrate in 100 mM McIlvaine's citric acid-phosphate buffer (pH 7.5) (McIlvaine, 1921) for one hour at 37°C to determine β -glucosidase activity. An equal volume of 1 M sodium bicarbonate was added to stop the reaction and the liberated *p*-nitrophenyl was measured spectrophotometrically at 400 nm.

The optimal pH for hydrolysis of pNPG by the recombinant *E. coli* producing the CelBACH complex was determined over the pH range of 2.2-8.0. Similarly, a temperature range of 10 to 60°C was employed. The temperature stability of the recombinant *E. coli* strain hydrolyzing pNPG was assayed after pre-incubating the cells in the absence of the substrate for different time intervals at 37, 45 and 50°C, respectively. Glucose sensitivity of the CelBACH system produced in *E. coli* was determined by performing the assays in the presence of 0.01-0.1 mM glucose. All the activity assays were incubated at 37°C for one hour as described above.

3.4 Results

3.4.1 Cloning of the *cel* operon

A genomic library of *B. pumilus* in *E. coli* XL1 Blue was screened for activity on pNPG. Five isolates that hydrolysed pNPG were identified from approximately 2000 colonies tested. One isolate, containing plasmid pBPU, exhibited the strongest activity and was selected for further study. A number of subclones of pBPU were prepared, but only pBPU1(-AatII) exhibited β -glucosidase activity. A restriction enzyme map of plasmid pBPU is shown in Fig. 2. However, further deletion of 2209 bp from plasmid pBPU(-AatII) resulted in the loss of activity. Testing subclones localised the minimal coding region for the activity on pNPG to a 3800 bp DNA fragment.

3.4.2 DNA sequence analysis

The entire 4476 bp DNA fragment in plasmid pBPU(-AatII) was sequenced in both directions (Fig. 2). Three complete open reading frames (ORF's) and one incomplete ORF were found. All genes are transcribed in the same direction. Inverse-PCR was used to obtain the complete fourth ORF. The complete sequence of all four genes is presented in Fig. 3. The coding region for the first gene utilises a TTG-codon to initiate transcription. This was changed to an ATG-codon prior to PCR. Previous studies concluded that the preferred initiation codon in *B. subtilis* (in the order of frequency) is ATG>TTG>GTG (Vellanoweth and Rabinowitz, 1992). The high AT-base pair content of the promoter region was found to be a predominant feature of *Bacillus* species (Doi, 1982). It is interesting to note that the potential ribosome binding sequences of the second and fourth

ORF overlapped the stop codons from the preceding genes and may provide translational coupling. A similar phenomenon was observed in the *B. stearothermophilus* cellobiose-specific PTS (Lai and Ingram, 1993) and the *B. subtilis* sucrose-specific PTS (Fouet *et al.*, 1987).

```

1   CAAAGAGCGTAAATAACGCTCTTTTTATATTAATACCAGACCAATAAATTTAATGTCTAGACATTTTAAAATATAAATAAATA
      uUUCuCC          celB          -35          -10
91  TATATTATTGGAAGCGCTTTTATTACCACCAATGGGGTGAAGAAATGAAACGATTTTATTAGCATGCAGTTCGGGAATGTCTACGAGT
      RB          M K R I L L A C S S G M S T S
181 TTACTAGTGACAAAGATGAAAGCACATGCCGATTCAATCGGGGACGAAGCAGAAATTTGGGCAGTTGGTCAAGATCAAGCAAAAAAGAA
      L L V T K M K A H A D S I G D E A E I W A V G Q D Q A K K E
271 ATGGCGAATGCAGATGTTGTGTTAATCGGCCCGCAATGAGCTTTTTAAAAGGAGACCTTCAAAAAGAAGCAAAAATACGGGATTGAG
      M A N A D V V L I G P Q M S F L K G D L Q K E A T K Y G I E
      UuuCCUcc
361 GTAGAGGTCATTGATATGCAGGCATACGGGCTTGCCGATGGACAAAAAGCGTATGAACAGGCACTTACATTGATGGGAGAATCTTAAGG
      V E V I D M Q A Y G L A D G Q K A Y E Q A L T L M G E S *
      RB
celA
451 TGGAAAAGAAAACACTCGAAGGGCTCACGGAGGAACAAGTGAGCTTTCATTGATCTTACACAGCGGGAATGCACGCAGCAAATGCTTC
      M E K K T L E G L T E E Q V S F Q L I L H S G N A R S K L L
541 AGGCTCTAAAACAATATCGAGAAGGCTTAGAAGAAGAAGCTTTTGATCTTATGAAAGAAGCGGAAGAAGATTTACGGTTGCACATGACA
      Q A L K Q Y R E G L E E E A F D L M K E A E E D L R L A H D
631 TCCATTTTCAACTCGTGCAGAAAGAGGCGAGGAGAAAGGGCATCCTTTTCTTACTGCTCATGCACGCTGAAGATCACTTAATGTGCA
      I H F Q L V Q K E A G G E R A S F S L L L M H A E D H L M S
      uUUCuCC          celC
721 CCATTACCATCAAGGAGCTTGTGGGAGAGCTCCTGCCGATATTCCAATCGCTTAAACAATAAAGGGTGTAAACAAGGGGAAACCAATG
      T I T I K E L V G E L L P I F Q S L K Q *
      RB          M
810 TTTGATAAAAATAAGTGCCATTTTGGTCCGATTGCTGGGAGATTGAACAACAACCGCTATTTAGCGGTTTACGTGATGCCTTTTATGCTT
      F D K I S A I L V P I A G R L N N N R Y L G V L R D A F M L
901 GCGTTCGCGCTCACCATCTTTGGTTCATTATGGTTGTGCTGATGAATTTGCCATTTTATAGATAAAATCATGAGTAAAACCGTACTGGAA
      A F P L T I F G S I M V V L M N L P F L D K I M S K T V L E
991 GCGTCCAGTCTGCAGTGAATATCGCACCGAGTGCACCATCAGCATTATGAGTGTGTTTGTGATTGCGGACTCGCTACTATTTATCA
      G V Q S A L N I A P S A T I S I M S V F V V F G I G Y Y L S
1081 AAAAGCTATGATGTAGAAGCTGTCTTTGGCGGCGTATTGCTTTAGCCTCTTCTCGTCTTAACTCCGTTTTTACTAGAACAAAGAGGC
      K S Y D V E A V F G G V I A L A S F L L L T P F L L E Q E G
1171 GGTGCAACGATTGCAGCGCTTATCCCAGTTGATCGTTTAGTGCAAAAGGAATGTCTCCTAGGAATGATCACAGGCTTTTTGTCTGCTGAA
      G A T I A G V I P V D R L G A K G M F L G M I T G F L S A E
1261 ATTTATCGTTATTTTGCCAGAAGAAATTTGTGATCAATATGCCGCAAGGGTTCGCCAGCTGTTCCAAGTCTTTTGTGCTGACTGATC
      I Y R Y F V Q K K F V I N M P Q G V P P A V S K S F A A L I
1351 CCGGCTACTTTGACACTTACAACCTTTTGGTATTAATATCATCATTACACAAGGCTTCAAAACAAATATGCATGAATTTGATTATCAT
      P A T L T L T T F L V I N I I I T Q G F K T N M H E F D Y H
1441 GCCATTCAGGCGCGCTTGTGCGTCTTGAAGCGGTATCATTCCGACCGTATCGCTGTTTTCTTATACCAATCCTTGGTCCTTTGGC
      A I Q A P L V G L G S G I I P T V I A V F F I P N P W S F G
1531 CTTCATGGGCAAAATATTATTAACCTCTGCATTGGATCCGATTTGGAATACATTGTCCATACAAAACCTAGAAAGCTATACAAAAACAGGT
      L H G P N I I N S A L D P I W N T L S I Q N L E S Y T K T G
1621 GAGGTTCTAATATCATCAGCAAGCAGTTTATTGAAATTTATACGGTTGGAATGGTGGAAACCGGTATGACGCTTGTGTTGTTTACG
      E V P N I I S K Q F I E I Y T V G M G G T G M T L A V V F T
1711 ATTTAATTTTCTGAAAAGTAAGCAATTAACAAGTAGCCAAAGCTGGGCTTAGGACCAGGATTTTAAAGTCAATGAACCGATTATC
      I L I F L K S K Q L K Q V A K L G L G P G L F N V N E P I I
1801 TTCGGTTTACCGATTGTCATGAATCCGCTTATTCTCATTCCGTGGATTTTGGCGCAATGGTCATTACATGTATTACTTATTTGCGATG
      F G L P I V M N P L I L I P W I L A P M V I T C I T Y F A M
1891 GCATCAGGCATTGTACCGCCGCCGACAGCGTGAATATCCCATTGACAGTACCGATCTTTATTAGCGGAATGATGGCAACAAATCACTA
      A S G G I V P P P T G V N I P W T V P I F I S G M M A T N S L
1981 GCGGGCTGCTTCCAGCTGTTAACTGATGATGTTGTGCTTGGTCCGTTCTGAAATTTATTGACCGTATGAATGTGAAA
      A G G L L Q L F N L M I V F V I W F P F L K F I D R M N V K

```

uUCCUCc celH

2071 AATGAAAAATACCGCCACCAGAAAAACGAAGGCGAAATATCAAAGGCGAAGGAGACTCAATACATGTCGATGGAAACTAACAGCA
 N E K I P P P E K T K A A N I K G E G D S I H V D G N *

RB M E T N K Q

2161 ATCATATACACTCCCAGAAGGTTTCTGGTGGGATCTTCTGCTTCTGCGACGAAACAGAGGGAGCGTCCCGGGCGATGGAAAGGGACC
 S Y T L P E G F W W G S S A S A T Q T E G S V P G D G K G P

2251 GAACATTTGGGATCATTGGTTTGAACAAGAGCCGACTCGTTTTTTTGTGGCGTAGGCGCGGACGTGACATCTCAATTTTATCGAAAAATA
 N I W D H W F E Q E P T R F F D G V G P D V T S Q F Y R K Y

2341 TAAAGAGGATATCCGTCTCATGAAAGAAATGGCCACAATTCATTTCCGCTGTGATTCATGGTCACGCCTCTTACCTGAAGGAACAGG
 K E D I R L M K E I G H N S F R L S I S W S R L L P E G T G

2431 TGCCGTCAATGAAGAAGCGGTGTCTTTTACAATGATGTGATCAATGAACTGATTCACATGATATCGAGCCATTTGTGAACCTATACCA
 A V N E E A V S F Y N D V I N E L I A H D I E P F V N L Y H

2521 TTTCGATATGCCAATGGCCTTCCAGGAGAGGGCGGCTGGTGAACAGACAGTGGTGCAGACTATGTCAGCTATCGCGCATATTATGCTT
 F D M P M A L Q E K G G W V N R Q V V D D Y V S Y A A L C F

2611 CCAATATTTGGTGACCGGGTAAAAATGGTTTACCCATAACGAACCGATTGTGCCTGTAGAAGGGCGGATATATGTATAGCTTTCATTA
 Q L F G D R V K K W F T H N E P I V P V E G G Y M Y S F H Y

2701 TCCAAACGAAGTCGATTTCCAAAGGCTGTACAGGTCGGTTTTTATGAAATTTTATCAAATGCGAAGGCGATTGCGCGCTTATAAAAGCT
 P N E V D F Q K A V Q V G F H E I L S N A K A I A A Y K K L

2791 GCAGCAAGCGTAAGATTGGCATCATCTAAACCTGACTCCTTCGTACCCGGAAGCCAGCATCCGCGCGATGTAGAAGCAGCAGAGAT
 Q Q G G K I G I I L N L T P S Y P R S Q H P R D V E A A E M

2881 GGCCGATGCTTTCTTAAACCGTCTTCTTAGATCCTGCGGTGAAGGGGCATTTCCCGAAAAGCTAGTAGAAGTCTTAAAGAAAGAGG
 A D A F F N R S F L D P A V K G H F P E K L V E V L K K E G

2971 CTTTATGCCAAGTATGGAAGAAGGGACCTTGGCTCATTGAGAGAATACGATTGATCTTCTTGGCATCAACTACTATCAGCCAAGAG
 F M P S M E E G D L E L I R E N T I D L L G I N Y Y Q P R R

3061 AGTGAAGCGAAAGAGCATATGCCGCACCTCAAGCACCATTATGCGCAACCATTATTTGATTCTTTGAGATGCCGGGAAGAAAAAT
 V K A K E H M P H P Q A P F M P N H Y F D S F E M P G R K M

3151 GAATGTGTATCGAGGCTGGGAAATCTATGAAAGGGCATTATGACATATGAAAAAGTCCAAACAGACTATGATAATATTGAATGTTT
 N V Y R G W E I Y E K G I Y D I L K K V Q T D Y D N I E C F

3241 TATTTCTGAAAATGGCATGGGTGTGGAAGGAGAAGAGCGGTTTAAGGATGAAGAAGGAATGATTCATGACGATTACCGCATTGACTTTAT
 I S E N G M G V E G E E R F K D E E G M I H D D Y R I D F I

3331 TTCTGAGCATCTCAAATGGGTTTCATCGAGCTATTCAAGAAGGAAGCAATGTGAAAGGCTATCATCTATGGACCTTTATGGACAACCTGGTC
 S E H L K W V H R A I Q E G S N V K G Y H L W T F M D N W S

3421 CTGGTCGAATGCATATAAAAAATCGTTATGGATTGTCTCAGTCGATCTTCAAAAAGACGGGAAACGAACGATTAAGAAAAGCGGATACTG
 W S N A Y K N R Y G F V S V D L Q K D G K R T I K K S G Y W

3511 GTTTCATCCGTTTCAAAGAATAACGGTTTTTAATCAAATACTTGCTCTTTCCTTGAGAA
 F Q S V S K N N G F *

FIGURE 3. Nucleotide sequence of the 3541-kb DNA fragment from *B. pumilus* that contains the complete *cel* operon. The deduced amino acid sequences for the *cel* operon are listed below the second nucleotide of the corresponding codon. A putative promoter region for the *cel* operon is underlined, with the -35 and -10 regions labeled. The sequence of the 3' terminus of the 16S rRNA (from *B. subtilis*) is shown above the potential ribosomal binding region (underlined and labeled RB) from the *cel* genes. Genes within the *cel* operon are highlighted at their start codons. Stop codons are indicated by asterisks.

3.4.3 Identification of cloned *cel* genes by homology

Because pNPG is considered an analog of cellobiose, the sequenced genes from *B. pumilus* were designated the *cel* operon. The genes (*celB*, *celA*, *celC* and *celH*) in this operon only conferred activity on pNPG when expressed together. Individual gene expression and expression in different combinations did not show activity. NCBI BLAST searches identified the different genes according to highest similarity (Table 3). Our particular interest was the production of the phospho- β -glucosidase (encoded by the *celH*). The translated protein sequence of the *B. pumilus celH* displayed good homology with the

B. subtilis, *B. halodurans*, and *Listeria monocytogenes* β -glucosidases. Based on this homology, CelH should group into family 1 of the glycosyl hydrolases (Henrissat and Bairoch, 1996; Henrissat and Davies, 1995; Henrissat and Bairoch, 1993).

TABLE 2. The results obtained with NCBI BLAST searches with the cloned *B. pumilus celBACH* genes.

Gene	Organism	Highest homology	Enzyme	% Identity
<i>celB</i>	<i>Bacillus halodurans</i>	Cellobiose-specific enzyme IIB, PTS system		79.6
	<i>Bacillus subtilis</i>	Cellobiose PTS enzyme II celA (ydhM)		76.7
	<i>Bacillus stearothermophilus</i>	Cellobiose phosphotransferase system celA		47.6
<i>celA</i>	<i>Bacillus halodurans</i>	Cellobiose-specific enzyme IIA, PTS system		61.8
	<i>Bacillus subtilis</i>	Cellobiose phosphotransferase system enzyme II	Cellobiose phosphotransferase system celC	60.0
<i>celC</i>	<i>Bacillus subtilis</i>	Cellobiose phosphotransferase system enzyme II	Cellobiose phosphotransferase system celB	65.6
	<i>Bacillus halodurans</i>	Cellobiose-specific enzyme IIC, PTS system		63.8
	<i>Listeria innocua</i>	Cellobiose-specific enzyme IIC, PTS system		52.0
<i>celH</i>	<i>Bacillus halodurans</i>	β -Glucosidase		74.2
	<i>Bacillus subtilis</i>	β -Glucosidase		72.7
	<i>Listeria monocytogenes</i>	β -Glucosidase		62.2
	<i>Clostridium thermocellum</i>	β -Glucosidase		40.0

3.4.4 Characterisation of the CelBACH complex

The optimal pH and temperature at which the recombinant *E. coli* expressing the *celBACH* operon hydrolysed pNPG was 7.5 (Fig. 4A) and 45°C (Fig. 4B), respectively. Unfortunately, the CelBACH-associated activity in *E. coli* decreased after 20 minutes at the optimum temperature (Fig. 4C). Thermosensitivity assays concluded that the ability of the recombinant strain to hydrolyse pNPG remains constant at 37°C, but at a higher temperature the activity decreases significantly. Inhibition studies showed that the enzyme complex was very sensitive to glucose inhibition (Fig. 4D).

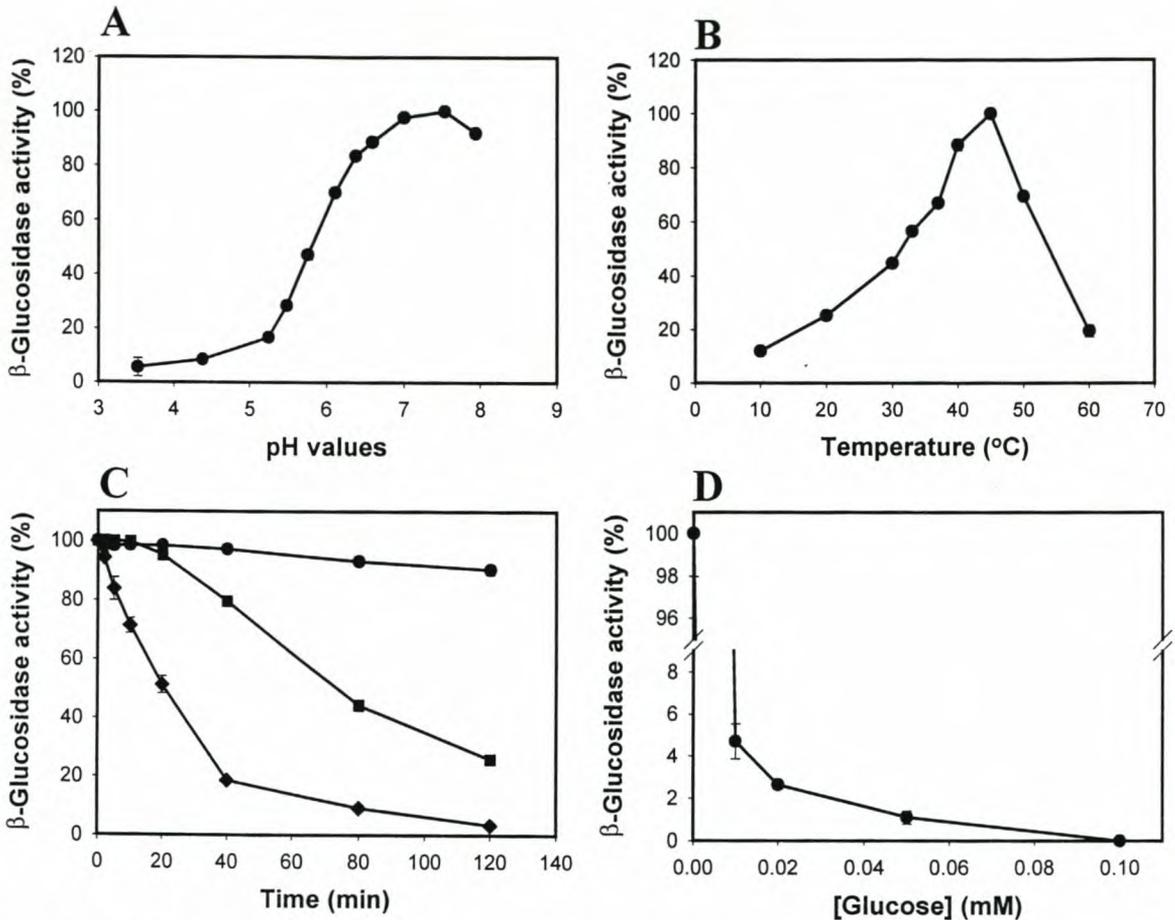


FIGURE 4. The effect of (A) pH and (B) temperature on the ability of *E. coli* expressing the *celBACH* operon to hydrolyse pNPG. The highest activity was measured at pH 7.5 and 45 $^{\circ}$ C, respectively. (C) The temperature stability of the CelBACH-associated activity of the recombinant strain at 37 $^{\circ}$ C (●), 45 $^{\circ}$ C (■) and 50 $^{\circ}$ C (◆) was determined by pre-incubating the cells at these temperatures in the absence of the substrate for 0, 5, 10, 20, 40, 80 and 120 min before determining the β -glucosidase activity on pNPG. The β -glucosidase activity prior to the pre-incubations (time 0 min) was taken as 100%. (D) Glucose sensitivity of the cellobiose-specific PTS system (CelBACH) of *B. pumilus* expressed in *E. coli* was determined by measuring the β -glucosidase activity on pNPG in the presence of different glucose concentrations (0.01, 0.02, 0.05, 0.1 mM). The β -glucosidase activity at 0 mM was taken as 100%.

3.5 Functional analysis of genes in the *B. pumilus cel* operon in *E. coli*

The *B. pumilus cel* genes were constitutively expressed in *E. coli* XL1Blue (pBPU-CelBACH). Recombinant cells exhibited strong pNPG activity when incubated at 37 $^{\circ}$ C. Previous studies demonstrated that the general components (enzyme I and HPr) of the *E. coli* PTS, are capable of complementing sugar-specific components from gram-positive bacteria (De Vos *et al.*, 1990; Fouet *et al.*, 1987). Based on the high degree of homology of these systems across bacteria it is assumed that they are interchangeable. Functional expression of the CelBACH complex enabled the recombinant *E. coli* strain to grow on cellobiose as carbon source (Fig. 5). The addition of 0.01% glucose was, however,

essential to stimulate cellobiose utilization. Results also showed that the CelBACH does not sustain any growth on the aromatic β -glucosides, arbutin and salicin (data not shown).

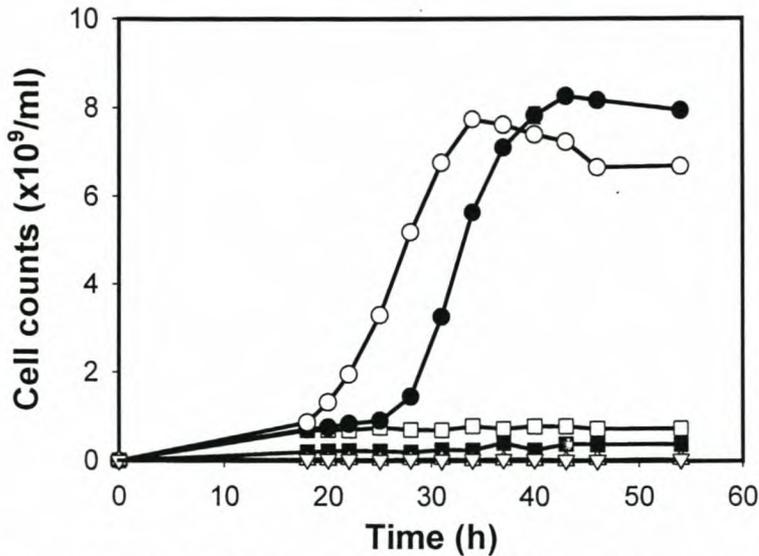


FIGURE 5. Growth of *E. coli* XL1-Blue expressing the *celBACH* operon on M9-minimal media with 0.01% glucose (□), with 0.21% glucose (○), with 0.21% cellobiose (■) and with 0.2% cellobiose plus 0.01% glucose (●). The *E. coli* strain with the pGEM-T-easy vector was inoculated into M9-minimal media containing 0.21% cellobiose (△) and 0.02% cellobiose plus 0.01% glucose (▽), respectively.

3.6 Discussion

In this paper, the cloning and functional expression of a novel *B. pumilus* cellobiose PTS operon in *E. coli* are reported. A *B. pumilus* genomic library in *E. coli* XL1-Blue was screened for activity on pNPG and positive isolates were identified. Sequencing data from the one isolate with the highest β -glucosidase activity revealed the presence of four ORF's encoding the cellobiose-specific genes of the PEP-dependent phosphotransferase system (PTS). The genes (*celB*, *celA*, *celC* and *celH*) in this operon only conferred activity on pNPG when expressed together.

Previous papers have described the organization of the bacterial PTS in significant detail (Saier and Reizer, 1992). The system comprised two general energy-coupling proteins, enzyme I and HPr, as well as a sugar-specific permease, universally termed the enzyme II complex. Though the enzyme II complex may be composed of one, two, three or four individual polypeptide chains, each complex contains at least three functional domains. The functions of the respective domains include that of a hydrophobic trans-membrane domain which binds and transports the sugar, a closely linked hydrophilic domain that

possesses the first phosphorylation site and a second hydrophilic domain containing an additional phosphorylation site. *celC* presumably encodes the membrane-spanning polypeptide based on its high hydrophobic character (Fig. 6), which forms the trans-membrane channel (Fig. 7). The *celC* product has a high predicted pI (Table 3). The *celA* product is more hydrophilic but contains a hydrophobic tail, which may serve as a membrane anchor (Fig. 6). Since *celC* and *celA* encode domains that form a single polypeptide in many organisms (Saier and Reizer, 1992), these two gene products can be assumed to interact closely in the enzyme II complex. The *celB* product is mostly hydrophilic, although a hydrophobic surface appears to be present near the amino terminus (Fig. 6). This enzyme has a low predicted pI (Table 3), which may allow the formation of ionic interactions with both the *celC*- and *celA*-encoded proteins (Fig. 7). The *celH* product is extremely hydrophilic and is proposed to encode an intracellular enzyme for the cleavage of cellobiose-6-phosphate to glucose-6P and glucose.

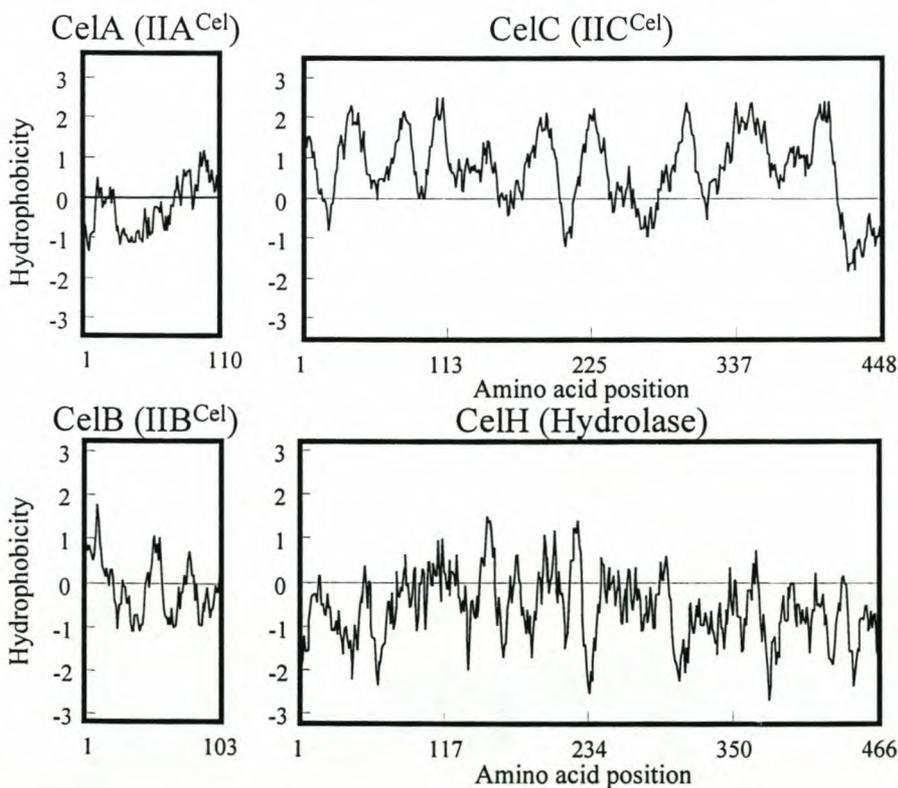


FIGURE 6. Hydrophobicity profiles of the four proteins encoded by the *B. pumilus celBACH* operon. Amino acid position and hydrophobicity are represented by the X- and Y-axis, respectively.

TABLE 3. Predicted properties of the *cel* operon from *B. pumilus*

Gene	PTS component	No. of nucleotides	%GC	Total no. of amino acids	Predicted Mr	Predicted Charge	Predicted pI
<i>celB</i>	IIB ^{Cel}	312	42	103	11 443	-3.06	4.74
<i>celA</i>	IIA ^{Cel}	333	42	110	14 688	-4.59	5.41
<i>celC</i>	IIC ^{Cel}	1347	45	448	43 051	-5.44	9.64
<i>celH</i>	P- β -glucosidase	1401	45	466	54 332	-10.6	5.73

As indicated by the sequence alignment, the *celH* gene product belongs to family 1 of the glycosyl hydrolases which utilise a retaining mechanism of enzymatic glycosidic bond hydrolysis (Davies and Henrissat, 1995). Other members of this family include the *B. polymyxa* β -glucosidase A, *Lactococcus lactis* 6-phospho- β -galactosidase and *Clostridium thermocellum* β -glucosidase A. In the recombinant *E. coli* expressing pNPG activity, enzyme I (and presumably HPr) must have been supplied by the host and complemented the PTS genes (*celBACH*) from *B. pumilus*. Other examples of functional complementation with heterologous PTSs from gram-positive bacteria in recombinant *E. coli* have been reported in various studies (Lai and Ingram, 1993; De Vos *et al.*, 1990; Fouet *et al.*, 1987).

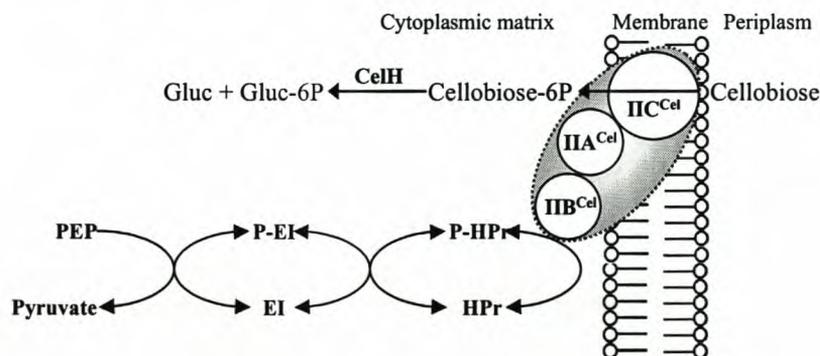


FIGURE 7. Schematic representation of the proposed *in vivo* function of the various gene products encoded by the *B. pumilus celBACH* operon (based on the cellobiose specific PTS of *B. stearothermophilus*) (Lai and Ingram, 1993). The high-energy phosphate is transferred from phosphoenolpyruvate (PEP) to enzyme II (CelBAC) with the aid of enzyme I (EI) and histidine-containing protein (HPr). Therefore, as a cellobiose molecule is transported across the membrane by enzyme II, it is phosphorylated. The cellobiose-6P is subsequently hydrolysed by the phospho- β -glucosidase (CelH). Interestingly, enzyme II transports only specific sugars and varies with PTS, whereas enzyme I and HPr are common to all PTSs.

In vivo assays concluded that the optimal temperature and pH for *E. coli* expressing the *celBACH* operon to hydrolyse pNPG are 7.5 and 45°C, respectively (Fig. 4A and B). Unfortunately, the CelBACH-associated activity in *E. coli* decreased after 20 minutes at

45°C (Fig. 4C). Although the recombinant strain remained stable at 37°C with regards to pNPG hydrolysis, its activity decreases significantly at 50°C. It was also shown that the enzyme complex was extremely sensitive to glucose (Fig. 4D). This phenomenon does not present a significant problem, since active growing *E. coli* cells metabolise glucose very rapidly.

The medium developed for growth studies consisted of M9-salts, MgSO₄, vitamins, trace elements and ampicillin and was not able to sustain any growth in the absence of a carbon source. Results also indicated that recombinant *E. coli* could not be sustained on cellobiose as sole carbon source. Glucose was added to a final concentration of 0.01% to allow product flux through the glycolysis, thus producing PEP as intermediate. Fig. 5 established that growth could be sustained on cellobiose in the presence of 0.01% glucose. The *E. coli* strain containing the pGEM-T-easy vector (used as reference strain) did not allow growth on 0.21% cellobiose or 0.2% cellobiose plus 0.01% glucose. We therefore concluded that the cryptic *cel* genes in *E. coli* reported previously (Thompson *et al.*, 1999; Parker and Hall, 1990; Krickler and Hall, 1987) are not responsible for the growth observed on cellobiose.

The recombinant strain expressing the *cel* operon reached cell densities of up to $\sim 82 \times 10^8$ cells/ml, which compared favorably with that on glucose ($\sim 76 \times 10^8$ cells/ml). The percentage glucose equivalent of the (0.2% cellobiose + 0.01% glucose)-medium composition was calculated as 0.221% glucose. Thus, the additional 0.011% glucose may account for the higher cell density on the (cellobiose + glucose)-medium. The calculated growth rate (μ_{\max}) of the recombinant *E. coli* during log-phase growth was approximately 0.19 h^{-1} and 0.20 h^{-1} for glucose and cellobiose, respectively. According to these values the doubling time for recombinant *E. coli* XL1-blue growing in minimal medium was about 300 minutes on glucose and 316 minutes on cellobiose, respectively. The modified M9-minimal medium sustained no growth without the addition of a carbon source. We thus postulate that the availability of PEP was the rate-limiting factor for sustained growth of *E. coli* XL1-Blue expressing the *CelBACH* operon and that PEP was provided by the addition of 0.01% glucose.

In this study we have constructed a recombinant *E. coli* strain expressing the *celBACH* operon of *B. pumilus* genes that encodes the gene products necessary for the PTS-mediated

uptake and subsequent hydrolysis of cellobiose. Functional expression enabled the host to efficiently metabolise cellobiose as a sole carbon source.

3.7 Acknowledgements

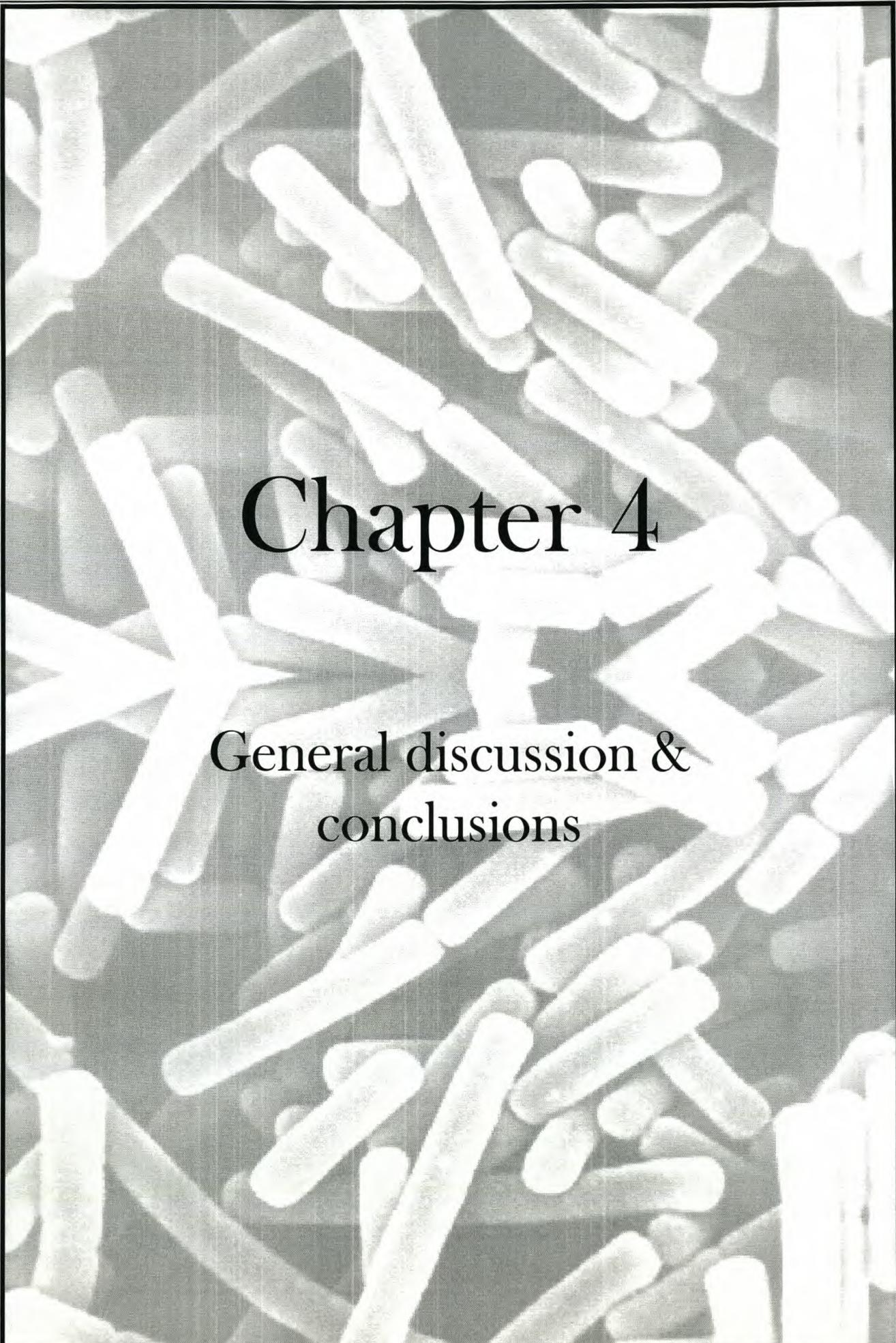
This work was funded by a grant from the Foundation for Research Development.

3.8 References

- Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-10.
- Ausubel, F.M., R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith and K. Struhl. 1994. *Current protocols in molecular biology*. Volume 1. John Wiley & Sons, Inc., Canada.
- Davies, G. J. and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**:853-859.
- De Vos, W.M., I. Boerrigter, R.J. van Rooyen, B. Reiche, and W. Hengstenberg. 1990. Characterization of the lactose-specific enzymes of the phosphotransferase system in *Lactococcus lactis*. *J. Biol. Chem.* **265**:22554-22560.
- Doi, R.H. 1982. RNA polymerase of *Bacillus subtilis*. pp 98-101, *In: The molecular biology of the Bacilli*. D.A. Dubnau (ed) Academic Press, New York.
- Fouet, A., M. Arnaud, A. Klier, and G. Rapoport. 1987. *Bacillus subtilis* sucrose-specific enzyme II of the phosphotransferase system: expression in *Escherichia coli* and homology to enzymes II from enteric bacteria. *Proc. Natl. Acad. Sci.* **84**:8773-8777.
- Henrissat B., and A. Bairoch. 1993. New families in the classification of glycosyl hydrolases based on amino-acid sequence similarities. *Biochem. J.* **293**:781-788.
- Henrissat B., and A. Bairoch. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**:695-696.
- Davies G., and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**:853-859.
- Himmel M.E., M.F. Ruth, and C.E. Wyman. 1999. Cellulase for commodity products from cellulosic biomass. *Curr. Opin. Biotechnol.* **10**:358-64.
- Iwashita, K., K. Todoroki, H. Kimura, H. Shimoi, and K. Ito. 1998. Purification and characterization of extracellular and cell wall-bound β -glucosidases from *Aspergillus kawachii*. *Biosci. Biotech. Biochem.* **62**:1938-1946.
- Kajikawa, H. and S. Masaki. 1999. Cellobiose transport by mixed ruminal bacteria from a cow. *Appl. Environ. Microbiol.* **65**:2565-2569.
- Kricker, M., and B.G. Hall. 1987. Biochemical genetics of the cryptic gene system for cellobiose utilization in *Escherichia coli* K12. *Genetics* **115**:419-429.

- Kruger, S., and M. Hecker. 1995. Regulation of the putative *bglPH* operon for aryl- β -glucoside utilization in *Bacillus subtilis*. *J. Bacteriol.* **177**:5590-5597.
- La Grange, D.C., I.S. Pretorius, and W.H. van Zyl. 1997. Cloning of the *Bacillus pumilus* β -xylosidase gene (*xynB*) and its expression in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **47**:262-266.
- Lai, X. and L. O. Ingram. 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *J. Bacteriol.* **175**:6441-6450.
- Leclerc, M., P. Chemardin, A. Arnaud, R. Ratomahenina, P. Galzy, C. Gerbaud, and A. Raynal. 1986. Biosynthesis regulation of the β -glucosidase produced by a yeast strain transformed by genetic engineering. *Arch. Microbiol.* **146**:115-7.
- Le Coq, D., C. Lindner, S. Krüger, M. Steinmetz, and J. Stülke. 1995. New β -glucoside (*bgl*) genes in *Bacillus subtilis*: The *bglP* gene product has both transport and regulatory functions similar to those of BglF, its *Escherichia coli* homolog. *J. Bacteriol.* **177**:1527-1535.
- McIlvaine, T.C. 1921. A buffer solution for colorimetric comparison. *J. Biol. Chem.* **49**:183-186.
- Leclerc, M., P. Chemardin, A. Arnaud, R. Ratomahenina, P. Galzy, C. Gerbaud, and A. Raynal. 1986. Biosynthesis regulation of the β -glucosidase produced by a yeast strain transformed by genetic engineering. *Arch. Microbiol.* **146**:115-117.
- Parker, L.L., and B.G. Hall. 1990. Mechanisms of activation of the cryptic *cel* operon of *Escherichia coli* K12. *Genetics* **124**:473-482.
- Reizer, J., S. Bachem, A. Reizer, M. Arnaud, M. H. Saier, and J. Stülke. 1999. Novel phosphotransferase system genes revealed by genome analysis - the complete complement of PTS proteins encoded within the genome of *Bacillus subtilis*. *Microbiology* **145** :3419-3429.
- Reizer, J., S.L. Sutrina, L.F. Wu, J. Deutscher, P. Reddy and M.H. Saier. 1992. Functional interactions between proteins of the phosphoenolpyruvate:sugar phosphotransferase systems of *Bacillus subtilis* and *Escherichia coli*. *J. Biol. Chem.* **267**:9158-9169.
- Riou, C., J. M. Salmon, M. J. Vallier, Z. Gunata, and P. Barre. 1998. Purification, characterization, and substrate specificity of a novel highly glucose-tolerant β -glucosidase from *Aspergillus oryzae*. *Appl. Environ. Microbiol.* **64**:3607-3614.
- Saier, M. H. and J. Reizer. 1992. Proposed uniform nomenclature for the proteins and protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.* **174**:1433-1438.
- Sambrook J, E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning : A laboratory manual*. 2nd edn. Cold Spring Harbor Laboratory Press, New York.
- Tanaka, K., T. Kawaguchi, Y. Imada, T. Ooi and M. Arai. 1995. Purification and properties of cellobiose phosphorylase from *Clostridium thermocellum*. *J. Ferment. Bioeng.* **79**:212-216.
- Tobisch, S., P. Glaser, S. Krüger, and M. Hecker. 1997. Identification and characterization of a new β -glucoside utilization system in *Bacillus subtilis*. *J. Bacteriol.* **179**:496-506.
- Thompson, J., S.B. Ruvinov, D.I. Freedberg, and B.G. Hall. 1999. Cellobiose-6-phosphate hydrolase (CelF) of *Escherichia coli*: characterization and assignment to the unusual family 4 of glycosylhydrolases. *J. Bacteriol.* **181**:7339-7345.

- Van Zyl, W.H., N. Wills, and J.R. Broach.** 1989. A general screen for mutants of *Saccharomyces cerevisiae* deficient in tRNA biosynthesis. *Genetics* **123**:55- 68.
- Vellanoweth, R.L., and J.C. Rabinowitz.** 1992. The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli* *in vivo*. *Mol. Microbiol.* **6**:1105-1114.
- Yan, T., Y. Lin, and C. Lin.** 1998. Purification and characterization of an extracellular β -glucosidase II with high hydrolysis and transglucosylation activities from *Aspergillus niger*. *J. Agric. Food Chem.* **46**:431-437.



Chapter 4

General discussion &
conclusions

4 GENERAL DISCUSSION AND CONCLUSIONS

The gram-positive soil bacterium *Bacillus subtilis*, together with other members of the *Bacillaceae* family, is able to use a wide diversity of carbohydrates, including amorphous cellulose, β -1,3-glucan and β -1,3-1,4-glucan as sole carbon source (Reizer *et al.*, 1999). For the utilization of insoluble carbon sources, the organism secretes hydrolytic enzymes which degrade the polymers extracellularly. The subsequent soluble oligosaccharides (e.g. cellobiose and cellotriose) are transported by specific transport systems and introduced into the central metabolism (Tobisch *et al.*, 1999). The preferred carbon source is glucose, and expression of genes and operons whose gene products are involved in utilization of alternative carbon sources is strongly regulated, i.e., inducible by the specific substrate and repressed by preferred carbon sources (carbon catabolite repression [CCR]) (Hueck and Hillen, 1995; Steinmetz, 1993).

The bacterial phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS) is dynamically involved in the regulation of gene expression (Postma *et al.*, 1993). Transport and coupled phosphorylation of its substrates comprise the fundamental function of the PTS. Other roles of the PTS include involvement in chemotaxis and in regulation of other metabolic processes (Postma *et al.*, 1993, Saier, 1989). The PTS is composed of the general proteins enzyme I (EI) and HPr and of the sugar-specific enzyme II (EII). The EII complexes represent the sugar-specific permeases, which consist of three or four subunits either fused in single multidomain proteins or built up of separate polypeptides (Saier and Reizer, 1992). The proteins of the PTS transfer a phosphoryl group from phosphoenolpyruvate (PEP) (the phosphoryl donor) to the sugar to be transported (Postma *et al.*, 1993).

4.1 Identification and isolation of genes for a novel *B. pumilus* cellobiose-dependent phosphotransferase system (PTS)

In this study, the cloning and functional expression of a novel *B. pumilus* cellobiose PTS operon in *E. coli* are reported. A *B. pumilus* genomic library in *E. coli* was screened for activity on pNPG and positive isolates were identified. Sequencing data from the one isolate with the highest β -glucosidase activity revealed the presence of four ORF's encoding the cellobiose-specific genes of the PEP-dependent phosphotransferase system

(PTS). As inferred by sequence alignments, the *celH* gene product belongs to family 1 of the glycosyl hydrolases which utilise a retaining mechanism of enzymatic glycosidic bond hydrolysis (Davies and Henrissat, 1995). Other members of this family include the *B. polymyxa* β -glucosidase A, *Lactococcus lactis* 6-phospho- β -galactosidase and *Clostridium thermocellum* β -glucosidase A.

4.2 Gene transcription in *Escherichiae coli*

The genes (*celB*, *celA*, *celC* and *celH*) comprise an operon and only conferred activity on pNPG when expressed jointly. In the recombinant *E. coli* expressing pNPG activity, enzyme I (and presumably HPr) must have been supplied by the host and complemented the PTS genes (*celBACH*) from *B. pumilus*. Other cases of functional complementation with heterologous PTSs from gram-positive bacteria in recombinant *E. coli* have been reported in various studies (Lai and Ingram, 1993; De Vos *et al.*, 1990; Fouet *et al.*, 1987).

4.3 Assessment of the recombinant *E. coli* strain for the ability to utilize cellobiose as a carbon source.

Results indicated that recombinant *E. coli* expressing the *celBACH* operon, could not be sustained on cellobiose as sole carbon source. However, with the addition of 0.01% glucose, the recombinant strain expressing the *cel* operon reached cell densities of $\sim 240 \times 10^7$ cells/ml on 0.2% cellobiose-containing medium. These results compared favourably with that of the positive control on glucose ($\sim 225 \times 10^7$ cells/ml). We proposed that the availability of PEP was the rate-limiting factor for sustained growth of *E. coli* expressing the *CelBACH* operon. Glucose allowed product flux through glycolysis, thus producing PEP as intermediate for subsequent PEP-dependent transport and metabolism of the cellobiose (Fig. 1).

- ◆ *In vivo* PTS activity assays determined that the recombinant strain hydrolysed pNPG at a pH and temperature optimum of pH 7.5 and 45°C, respectively. The CelBACH-associated activity in *E. coli* remained stable at 37°C, but decreased significantly after 20 minutes at 45°C. Unfortunately, inhibition studies showed that the CelBACH enzyme complex was very sensitive to glucose.
- ◆ The recombinant *E. coli* strain expressing the *B. pumilus celBACH* operon was able to grow on cellobiose as carbon source in the presence of 0.01% glucose. The glucose is speculated to be of importance in providing PEP (via activation of the glycolysis pathway).

4.5 Future research

Potential applications of the current work include genetic engineering of bacterial strains that naturally produce biocommodities, for example ethanol and lactic acid, to utilise cellobiose. Saccharification of cellulose (either chemically or enzymatically) produces cellobiose as the major end product and is therefore considered an abundant and renewable energy source. Efficient removal during SSF (simultaneous saccharification and fermentation) is also of great importance in relieving cellobiose-mediated inhibition of both endoglucanase and cellobiohydrolase activities.

Introduction of the *B. pumilus* cellobiose-specific PTS can efficiently catalyse the bioconversion of cellobiose to glucose, which can be used as a carbon source for the production of valuable compounds. *Zymomonas mobilis*, with ethanol productivity superior to *Saccharomyces cerevisiae*, is a potential target for genetic engineering. Also, L-lactic acid-producing bacteria such as *Lactobacillus fermentum* and *Leuconostoc lactis* can be of great biotechnological value as recombinant cellobiose fermenters. Lactic acid is an important raw material for the chemical synthesis of biodegradable plastics (polylactate). The PTS-mediated cellobiose utilisation could provide the recombinant organism with a selective advantage with regards to intracellular glucose formation. It reduces the possibility of contamination with non-cellobiose utilising organisms since glucose is not present in the external medium.

4.6 References

- Hueck, C. J., and W. Hillen. 1995. Catabolite repression in *Bacillus subtilis*: a global regulatory mechanism for Gram-positive bacteria? *Mol. Microbiol.* **15**:395–401.
- Steinmetz, M. 1993. Carbohydrate catabolism: pathways, enzymes, genetic regulation, and evolution, p. 157–170. In A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria: biochemistry, physiology, and molecular genetics. American Society for Microbiology, Washington, D.C.
- Tobisch, S., J. Stulke and M. Hecker. 1999. Regulation of the *lic* operon of *Bacillus subtilis* and characterization of potential phosphorylation sites of the LicR regulator protein by site-directed mutagenesis. *J. Bacteriol.* **181**:4995–5003.
- Postma, P.W., J.W. Lengeler, and G.R. Jacobson. 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. *Microbiol. Rev.* **57**:543–594.
- Saier, M. H. 1989. Protein phosphorylation and allosteric control of inducer exclusion and catabolite repression by the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *Microbiol. Rev.* **53**:109–120.
- Saier, M.H., and J. Reizer. 1992. Proposed uniform nomenclature for the proteins and the protein domains of the bacterial phosphoenolpyruvate:sugar phosphotransferase system. *J. Bacteriol.* **174**:1433–1438.
- Davies, G. J. and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* **3**:853–859.
- De Vos, W.M., I. Boerrigter, R.J. van Rooyen, B. Reiche, and W. Hengstenberg. 1990. Characterization of the lactose-specific enzymes of the phosphotransferase system in *Lactococcus lactis*. *J. Biol. Chem.* **265**:22554–22560.
- Fouet, A., M. Arnaud, A. Klier, and G. Rapoport. 1987. *Bacillus subtilis* sucrose-specific enzyme II of the phosphotransferase system: expression in *Escherichia coli* and homology to enzymes II from enteric bacteria. *Proc. Natl. Acad. Sci.* **84**:8773–8777.
- Lai, X. and L. O. Ingram. 1993. Cloning and sequencing of a cellobiose phosphotransferase system operon from *Bacillus stearothermophilus* XL-65-6 and functional expression in *Escherichia coli*. *J. Bacteriol.* **175**:6441–6450.