Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 2994320 C 2023/05/16

(11)(21) 2 994 320

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(22) Date de dépôt/Filing Date: 2006/12/15

(41) Mise à la disp. pub./Open to Public Insp.: 2007/06/28

(45) Date de délivrance/Issue Date: 2023/05/16

(62) Demande originale/Original Application: 2 833 029

(30) Priorités/Priorities: 2005/12/22 (US60/753,258); 2005/12/22 (FI20151318)

(51) Cl.Int./Int.Cl. *C12N 15/56* (2006.01), *C12N 1/15* (2006.01), *C12N 1/21* (2006.01), *C12N 9/24* (2006.01), *C12N 9/42* (2006.01), *C12P 19/14* (2006.01)

(72) Inventeurs/Inventors:
KALLIO, JARNO, FI;
VOUTILAINEN, SANNI, FI;
VIIKARI, LIISA, FI;
HOOMAN, SATU, FI;
HALONEN, TEEMU, FI;
SIIKA-AHO, MATTI, FI;
...

(73) Propriétaire/Owner:

(54) Titre : TRAITEMENT DE MATERIEL CELLULOSIQUE ET ENZYMES POUVANT ETRE EMPLOYEES DANS CE TRAITEMENT

(54) Title: TREATMENT OF CELLULOSIC MATERIAL AND ENZYMES USEFUL THEREIN

(57) Abrégé/Abstract:
The present invention relates to the production of sugar hydrolysates from cellulosic material. The method may be used e.g. for producing fermentable sugars for the production of bioethanol from lignocellulosic material. Cellulolytic enzymes and their production by recombinant technology is described, as well as uses of the enzymes and enzyme preparations.

CA 2994320 C 2023/05/16
(11)(21) 2 994 320
(13) C

(72) Inventeurs(suite)/Inventors(continued): VEHMAANPERA, JARI, FI; ALAPURANEN, MARIKA, FI; PURANEN, TERHI, FI
(73) Propriétaires(suite)/Owners(continued):ROAL OY, FI
(74) Agent: NORTON ROSE FULBRIGHT CANADA LLP/S.E.N.C.R.L., S.R.L.

(57) Abstract

The present invention relates to the production of sugar hydrolysates from cellulosic material. The method may be used e.g. for producing fermentable sugars for the production of bioethanol from lignocellulosic material. Cellulolytic enzymes and their production by recombinant technology is described, as well as uses of the enzymes and enzyme preparations.

Treatment of cellulosic material and enzymes useful therein

Field of the Invention

The present invention relates to the production of sugar hydrolysates from cellulosic material. More precisely the invention relates to production of fermentable sugars from lignocellulosic material by enzymatic conversion. The fermentable sugars are useful e.g. in the production of bioethanol, or for other purposes. In particular the invention is directed to a method for treating cellulosic material with cellobiohydrolase, endoglucanase, beta-glucosidase, and optionally xylanase, and to enzyme preparations and the uses thereof. The invention is further directed to novel cellulolytic polypeptides, polynucleotides encoding them, and to vectors and host cells containing the polynucleotides. Still further the invention is directed to uses of the polypeptides and to a method of preparing them.

Background of the Invention

Sugar hydrolysates can be used for microbial production of a variety of fine chemicals or biopolymers, such as organic acids e.g. lactic acid, or ethanol or other alcohols e.g. n-butanol, 1,3-propanediol, or polyhydroxyalkanoates (PHAs). The sugar hydrolysates may also serve as raw material for other non-microbial processes, e.g., for enrichment, isolation and purification of high value sugars or various polymerization processes. One of the major uses of the sugar hydrolysates is in the production of biofuels. The production of bioethanol and/or other chemicals may take place in an integrated process in a biorefinery (Wyman 2001).

Limited resources of fossil fuels, and increasing amounts of CO_2 released from them and causing the greenhouse phenomenon have raised a need for using biomass as a renewable and clean source of energy. One promising, alternative technology is the production of biofuels i.e. ethanol from cellulosic materials. In the transportation sector biofuels are for the time being the only option, which could reduce the CO_2 emissions by an order of magnitude. The ethanol can be used in existing vehicles and distribution systems and thus it does not require expensive infrastructure investments. Sugars derived from lignocellulosic renewable raw materials can also be used as raw materials for a variety of chemical products that can replace oil-based chemicals.

Most of the carbohydrates in plants are in the form of lignocellulose, which essentially consists of cellulose, hemicellulose, pectin and lignin. In a lignocellulose-to-ethanol process the lignocellulosic material is first pretreated either chemically or physically to make the cellulose fraction more accessible to hydrolysis. The cellulose fraction is then hydrolysed to obtain sugars that can be fermented by yeast into ethanol. Lignin is obtained as a main co-product that may be used as a solid fuel.

Bioethanol production costs are high and the energy output is low, and there is continuous research for making the process more economical. Enzymatic hydrolysis is considered the most promising technology for converting cellulosic biomass into fermentable sugars. However, enzymatic hydrolysis is used only to a limited amount at industrial scale, and especially when using strongly lignified material such as wood or agricultural waste the technology is not satisfactory. The cost of the enzymatic step is one of the major economical factors of the process. Efforts have been made to improve the efficiency of the enzymatic hydrolysis of the cellulosic material (Badger 2002).

US 2002/019 2774 A1 describes a continuous process for converting solid lignocellulosic biomass into combustible fuel products. After pretreatment by wet oxidation or steam explosion the biomass is partially separated into cellulose, hemicellulose and lignin, and is then subjected to partial hydrolysis using one or more carbohydrase enzymes (EC 3.2). Celluclast™, a commercial product by Novo Nordisk A/S containing cellulase and xylanase activities is given as an example.

US 2004/000 5674 A1 describes novel enzyme mixtures that can be used directly on lignocellulose substrate, whereby toxic waste products formed during pretreatment processes may be avoided, and energy may be saved. The synergistic enzyme mixture contains a cellulase and an auxiliary enzyme such as cellulase, xylanase, ligninase, amylase, protease, lipidase or glucuronidase, or any combination thereof. Cellulase in considered to include endoglucanase (EG), beta-glucosidase (BG) and cellobiohydrolase (CBH). The examples illustrate the use of a mixture of *Trichoderma* xylanase and cellulase preparations.

Kurabi *et al.* (2005) have investigated enzymatic hydrolysis of steam-exploded and ethanol organosolv-pretreated Douglas-fir by novel and commercial fungal cellulases. They tested two commercial *Trichoderma reesei* cellulase preparations, and two novel preparations produced by mutant strains

of *Trichoderma* sp. and *Penicillium* sp. The *Trichoderma* sp. preparation showed significantly better performance than the other preparations. The better performance was believed to be at least partly due to a significantly higher beta-glucosidase activity, which relieves product inhibition of cellobiohydrolase and endoglucanase.

US 2004/005 3373 A1 pertains a method of converting cellulose to glucose by treating a pretreated lignocellulosic substrate with an enzyme mixture comprising cellulase and a modified cellobiohydrolase I (CBHI). The CBHI has been modified by inactivating its cellulose binding domain (CBD). Advantages of CBHI modification are e.g. better recovery and higher hydrolysis rate with high substrate concentration. The cellulase is selected from the group consisting of EG, CBH and BG. The CBHI is preferably obtained from *Trichoderma*.

US 2005/016 4355 A1 describes a method for degrading lignocellulosic material with one or more cellulolytic enzymes in the presence of at least one surfactant. Additional enzymes such as hemicellulases, esterase, peroxidase, protease, laccase or mixture thereof may also be used. The presence of surfactant increases the degradation of lignocellulosic material compared to the absence of surfactant. The cellulolytic enzymes may be any enzyme involved in the degradation of lignocellulose including CBH, EG, and BG.

There is a huge number of publications disclosing various cellulases and hemicellulases.

Cellobiohydrolases (CBHs) are disclosed e.g. in WO 03/000 941, which relates to CBHI enzymes obtained from various fungi. No physiological properties of the enzymes are provided, nor any examples of their uses. Hong *et al.* (2003b) characterizes CBHI of *Thermoascus aurantiacus* produced in yeast. Applications of the enzyme are not described. Tuohy *et al.* (2002) describe three forms of cellobiohydrolases from *Talaromyces emersonii*.

Endoglucanases of the cel5 family (EGs fam 5) are described e.g. in WO 03/062 409, which relates to compositions comprising at least two thermostable enzymes for use in feed applications. Hong *et al.* (2003a) describe production of thermostable endo-β-1,4-glucanase from *T. aurantiacus* in yeast. No applications are explained. WO 01/70998 relates to β-glucanases from *Talaromyces*. They also describe β-glucanases from *Talaromyces emersonii*. Food, feed, beverage, brewing, and detergent applications are discussed. Lignocellulose hydrolysis is not mentioned. WO 98/06 858 describes beta-1,4-

endoglucanase from *Aspergillus niger* and discusses feed and food applications of the enzyme. WO 97/13853 describes methods for screening DNA fragments encoding enzymes in cDNA libraries. The cDNA library is of yeast or fungal origin, preferably from *Aspergillus*. The enzyme is preferably a cellulase. Van Petegem *et al.* (2002) describe the 3D-structure of an endoglucanase of the cel5 family from *Thermoascus aurantiacus*. Parry *et al.* (2002) describe the mode of action of an endoglucanase of the cel5 family from *Thermoascus aurantiacus*.

Endoglucanases of the cel7 family (EGs fam 7) are disclosed e.g. in US 5,912,157, which pertains *Myceliphthora* endoglucanase and its homologues and applications thereof in detergent, textile, and pulp. US 6,071,735 describes cellulases exhibiting high endoglucanase activity in alkaline conditions. Uses as detergent, in pulp and paper, and textile applications are discussed. Bioethanol is not mentioned. US 5,763,254 discloses enzymes degrading cellulose/hemicellulose and having conserved amino acid residues in CBD.

Endoglucanases of the cel45 family (EGs fam 45) are described e.g. in US 6,001,639, which relates to enzymes having endoglucanase activity and having two conserved amino acid sequences. Uses in textile, detergent, and pulp and paper applications are generally discussed and treating of lignocellulosic material is mentioned but no examples are given. WO 2004/053039 is directed to detergent applications of endoglucanases. US 5,958,082 discloses the use of endoglucanase, especially from *Thielavia terrestris* in textile application. EP 0495258 relates to detergent compositions containing *Humicola* cellulase. US 5,948,672 describes a cellulase preparation containing endoglucanase, especially from *Humicola* and its use in textile and pulp applications. Lignocellulose hydrolysis is not mentioned.

A small amount of beta-glucosidase (BG) enhances hydrolysis of biomass to glucose by hydrolyzing cellobiose produced by cellobiohydrolases. Cellobiose conversion to glucose is usually the major rate-limiting step. Beta-glucosidases are disclosed e.g. in US 2005/021 4920, which relates to BG from *Aspergillus fumigatus*. The enzyme has been produced in *Aspergillus oryzae* and *Trichoderma reesei*. Use of the enzyme in degradation of biomass or detergent applications is generally discussed but not exemplified. WO02/095 014 describes an *Aspergillus oryzae* enzyme having cellobiase activity. Use in the production of ethanol from biomass is generally discussed but

not exemplified. WO2005/074656 discloses polypeptides having cellulolytic enhancing activity derived e.g. from *T. aurantiacus*; *A. fumigatus*; *T. terrestris* and *T. aurantiacus*. WO02/26979 discloses enzymatic processing of plant material. US 6,022,725 describes cloning and amplification of the beta-glucosidase gene of *Trichoderma reesei*, and US 6,103,464 describes a method for detecting DNA encoding a beta-glucosidase from a filamentous fungus. No application examples are given.

Xylanases are described e.g. in FR2786784, which relates to a heat-stable xylanase, useful e.g. in treating animal feed and in bread making. The enzyme is derived from a thermophilic fungus, particularly of the genus *Thermoascus.*

US 6,197,564 describes enzymes having xylanase activity, and obtained from *Aspergillus aculeatus*. Their application in baking is exemplified. WO 02/24926 relates to *Talaromyces* xylanases. Feed and baking examples are given. WO01/42433 discloses thermostable xylanase from *Talaromyces emersonii* for use in food and feed applications.

The best-investigated and most widely applied cellulolytic enzymes of fungal origin have been derived from *Trichoderma reesei* (the anamorph of *Hypocrea jecorina*). Consequently also most of the commercially available fungal cellulases are derived from *Trichoderma reesei*. However, the majority of cellulases from less known fungi have not been applied in processes of practical importance such as in degrading cellulosic material, including lignocellulose.

There is a continuous need for new methods of degrading cellulosic substrates, in particular lignocellulosic substrates, and for new enzymes and enzyme mixtures, which enhance the efficiency of the degradation. There is also a need for processes and enzymes, which work at high temperatures, thus enabling the use of high biomass consistency and leading to high sugar and ethanol concentrations. This approach may lead to significant saving in energy and investments costs. The high temperature also decreases the risk of contamination during hydrolysis. The present invention aims to meet at least part of these needs.

Brief Description of the Invention

It has now surprisingly been found that cellulolytic enzymes, and especially cellobiohydrolases obtainable from *Thermoascus aurantiacus*, *Acremonium thermophilum*, or *Chaetomium thermophilum* are particularly use-

ful in hydrolyzing cellulosic material. In addition to cellobiohydrolases these fungi also have endoglucanases, beta-glucosidases and xylanases that are very suitable for degrading cellulosic material. The enzymes are kinetically very effective over a broad range of temperatures, and although they have high activity at high temperatures, they are also very efficient at standard hydrolysis temperatures. This makes them extremely well suited for varying cellulosic substrate hydrolysis processes carried out both at conventional temperatures and at elevated temperatures.

The present invention provides a method for treating cellulosic material with cellobiohydrolase, endoglucanase and beta-glucosidase, whereby said cellobiohydrolase comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 2, 4, 6 or 8, or to an enzymatically active fragment thereof.

The invention further provides an enzyme preparation comprising cellobiohydrolase, endoglucanase and beta-glucosidase, wherein said cellobiohydrolase comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 2, 4, 6 or 8, or to an enzymatically active fragment thereof.

The use of said enzyme preparation for degrading cellulosic material is also provided, as well as the use of said method in a process for preparing ethanol from cellulosic material.

The invention is also directed to a polypeptide comprising a fragment having cellulolytic activity and being selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence having at least 66% identity to SEQ ID NO:4, 79% identity to SEQ ID NO:6, 78% identity to SEQ ID NO:12, 68% identity to SEQ ID NO:14, 72% identity to SEQ ID NO:16, 68% identity to SEQ ID NO:20, 74% identity to SEQ ID NO:22 or 24, or 78% identity to SEQ ID NO:26;

b) a variant of a) comprising a fragment having cellulolytic activity; and

c) a fragment of a) or b) having cellulolytic activity.

One further object of the invention is an isolated polynucleotide selected from the group consisting of:

a) a nucleotide sequence of SEQ ID NO: 3, 5, 11, 13, 15, 19, 21, 23 or 25, or a sequence encoding a polypeptide of claim 35;

b) a complementary strand of a)

c) a fragment of a) or b) comprising at least 20 nucleotides; and

d) a sequence that is degenerate as a result of the genetic code to any one of the sequences as defined in a), b) or c).

The invention still further provides a vector, which comprises said polynucleotide as a heterologous sequence, and a host cell comprising said vector. *Escherichia coli* strains having accession number DSM 16728, DSM 16729, DSM 17324, DSM 17323, DSM 17729, DSM 16726, DSM 16725, DSM 17325 or DSM 17667 are also included in the invention.

Other objects of the invention are enzyme preparations comprising at least one of the novel polypeptides, and the use of said polypeptide or enzyme preparation in fuel, textile, detergent, pulp and paper, food, feed or beverage industry.

Further provided is a method for preparing a polypeptide comprising a fragment having cellulolytic activity and being selected from the group consisting of:

a) a polypeptide comprising an amino acid sequence having at least 66% identity to SEQ ID NO:4, 79% identity to SEQ ID NO:6, 78% identity to SEQ ID NO:12, 68% identity to SEQ ID NO:14, 72% identity to SEQ ID NO:16, 68% identity to SEQ ID NO:20, 74% identity to SEQ ID NO:22 or 24, or 78% identity to SEQ ID NO:26;

b) a variant of a) comprising a fragment having cellulolytic activity; and

c) a fragment of a) or b) having cellulolytic activity,

said method comprising transforming a host cell with a vector encoding said polypeptide, and culturing said host cell under conditions enabling expression of said polypeptide, and optionally recovering and purifying the polypeptide produced.

Still further provided is a method of treating cellulosic material with a spent culture medium of at least one microorganism capable of producing a polypeptide as defined above, wherein the method comprises reacting the cellulosic material with the spent culture medium to obtain hydrolysed cellulosic material.

Specific embodiments of the invention are set forth in the dependent claims.

Other objects, details and advantages of the present invention will become apparent from the following drawings, detailed description and examples.

Brief Description of the Drawings

Figure 1. Temperature dependencies of the cellulase and beta-glucosidase activities in the supernatants of the tested six fungal strains. The incubation time in the assay was 60 min at the given temperature, the assay pH was 5.0 (MUL-activity) or 4.8 (CMCase or BGU). Activity obtained at 60°C is set as the relative activity of 100%. **A)** *Thermoascus aurantiacus* ALKO4239, **B)** *Thermoascus aurantiacus* ALKO4242, **C)** *Acremonium thermophilum* ALKO4245, **D)** *Talaromyces thermophilus* ALKO4246, **E)** *Chaetomium thermophilum* ALKO4261, **F)** *Chaetomium thermophilum* ALKO4265.

Figure 2. Schematic picture of the expression cassettes used in the transformation of *Trichoderma reesei* protoplasts for producing the recombinant fungal proteins. The recombinant genes were under the control of *T. reesei cbh*1 (*cel*7A) promoter (*cbh*1 prom) and the termination of the transcription was ensured by using *T. reesei cbh*1 terminator sequence (*cbh*1 term). The *amd*S gene was included as a transformation marker.

Figure 3. A) pH optima of the recombinant CBH/Cel7 protein preparations from *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 determined on 4-methylumbelliferyl-β-D-lactoside (MUL) at 50°C, 10 min. The results are given as mean (±SD) of three separate measurements. **B)** Thermal stability of recombinant CBH/Cel7 protein preparations from *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 determined on 4-methylumbelliferyl-β-D-lactoside (MUL) at the optimum pH for 60 min. The results are given as mean (±SD) of three separate measurements. Both reactions contained BSA (100 µg/ml) as a stabilizer.

Figure 4. Crystalline cellulose (Avicel) hydrolysis by the purified recombinant cellobiohydrolases at 45°C. Substrate concentration 1% (w/v), pH 5.0, enzyme concentration 1.4 µM. **A)** Cellobiohydrolases harboring a CBD, **B)** cellobiohydrolases (core) without a CBD.

Date Reçue/Date Received 2022-03-03

Figure 5. Crystalline cellulose (Avicel) hydrolysis by the purified recombinant cellobiohydrolases at 70°C. Substrate concentration 1% (w/v), pH 5.0, enzyme concentration 1.4 µM. **A)** Cellobiohydrolases harboring a CBD, **B)** cellobiohydrolases (core) without a CBD.

It is provided a polypeptide having cellulolytic activity and comprising an amino acid sequence having at least 80% identity over the full length to SEQ ID NO: 22.

It is further provided an isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of: the nucleotide sequence of SEQ ID NO: 21; a sequence encoding the polypeptide as provided ; and the full length complement of the nucleotide sequence recited.

It is also provided method for preparing a polypeptide having cellulolytic activity and comprising an amino acid sequence having at least 80% identity over the full length to SEQ ID NO: 22, the method comprising transforming a host cell with a vector encoding the polypeptide, and culturing the host cell under conditions enabling expression of the polypeptide, and recovering and purifying the polypeptide produced.

Date Reçue/Date Received 2022-03-03

Figure 6. A) The pH dependency of the heterologously produced *Acremonium* EG_40/Cel45A, EG_40_like/Cel45B and *Thermoascus* EG_28/Cel5A activity was determined with CMC substrate in a 10 min reaction at 50°C. **B)** Temperature optimum of the *Acremonium* EG_40/Cel45A, EG_40_like/Cel45B and *Thermoascus* EG_28/Cel5A was determined at pH 5.5, 4.8, and 6.0, respectively. The reaction containing CMC as substrate was performed for 60 min, except for EG_28/Cel5A for 10 min. BSA (100 µg/ml) was added as a stabilizer.

Figure 7. A) The pH dependency of the heterologously produced *Acremonium* BG_101/Cel3A, *Chaetomium* BG_76/Cel3A, and *Thermoascus* BG_81/Cel3A activity was determined with 4-nitrophenyl-β-D-glucopyranoside substrate in a 10 min reaction at 50°C. **B)** Temperature optimum of the *Acremonium* βG_101/Cel3A, *Chaetomium* βG_76/Cel3A, and *Thermoascus* βG_81/Cel3A was determined at pH 4.5, 5.5, and 4.5, respectively. The reaction containing 4-nitrophenyl-β-D-glucopyranosid as substrate was performed for 60 min, BSA (100 µg/ml) was added as a stabilizer.

Figure 8. A) The pH dependency of the heterologously produced *Thermoascus* XYN_30/Xyn10A xylanase activity was determined with birch xylan substrate in a 10 min reaction at 50°C. **B)** Temperature optimum of XYN_30/Xyn10A was determined at pH 5.3 in a 60 min reaction, BSA (100 µg/ml) was added as a stabilizer.

Figure 9. Hydrolysis of washed steam exploded spruce fibre (10 mg/ml) with a mixture of thermophilic enzymes (MIXTURE 1) and *T. reesei* enzymes at 55 and 60°C. Enzyme dosage is given by FPU/g dry matter of substrate, FPU assayed at 50°C, pH 5. Hydrolysis was carried out for 72 h at pH 5, with mixing. The results are given as mean (±SD) of three separate measurements.

Figure 10. Hydrolysis of steam exploded corn stover (10 mg/ml) with a mixture of thermophilic enzymes (MIXTURE 2) and *T. reesei* enzymes at 45, 55 and 57.5°C. Enzyme dosage was for "MIXTURE 2" 5 FPU/g dry matter of substrate and for *T. reesei* enzymes 5 FPU/g dry matter Celluclast supplemented with 100 nkat/g dry matter Novozym 188 (filter paper activity was assayed at 50°C, pH 5). Hydrolysis was carried out for 72 h at pH 5, with mixing. The results are given as mean (±SD) of three separate measurements.

Date Reçue/Date Received 2022-03-03

The substrate contained soluble reducing sugars (ca 0.7 mg/ml). This background sugar content was subtracted from the reducing sugars formed during the hydrolysis.

Figure 11. Hydrolysis of steam exploded corn stover (10 mg/ml) with a mixture of thermophilic enzymes containing a new thermophilic xylanase from *Thermoascus aurantiacus* (MIXTURE 3) and *T. reesei* enzymes at 45, 55 and 60°C. Enzyme dosage was for "MIXTURE 3" 5 FPU/g dry matter of substrate and for *T. reesei* enzymes 5 FPU/g dry matter Celluclast supplemented with 100 nkat/g dry matter Novozym 188 (filter paper activity was assayed at 50°C, pH 5). Hydrolysis was carried out for 72 h at pH 5, with mixing. The results are given as mean (±SD) of three separate measurements. The substrate contained soluble reducing sugars (ca 0.7 mg/ml). This background sugar content was subtracted from the reducing sugars formed during the hydrolysis.

Figure 12. Hydrolysis of steam exploded spruce fibre (10 mg/ml) with a mixture of thermophilic enzymes containing a new thermophilic xylanase XYN_30/Xyn10A from *Thermoascus aurantiacus* (MIXTURE 3) and *T. reesei* enzymes at 45, 55 and 60°C. Enzyme dosage for "MIXTURE 3" was 5 FPU/g dry matter of substrate and for *T. reesei* enzymes 5 FPU/g dry matter Celluclast supplemented with 100 nkat/g dry matter Novozym 188 (filter paper activity was assayed at 50°C, pH 5). Hydrolysis was carried out for 72 h at pH 5, with mixing. The results are given as mean (±SD) of three separate measurements.

Figure 13. The effect of glucose on activity of different β-glucosidase preparations. The standard assay using *p*-nitrophenyl-β-D-glucopyranoside as substrate was carried out in the presence of glucose in the assay mixture. The activity is presented as percentage of the activity obtained without glucose.

Figure 14. FPU activities of the enzyme mixtures at temperatures from 50°C to 70°C, presented as a percentage of the activity under the standard conditions (50°C, 1 h).

Figure 15. The relative cellulase activity of two different *T. reesei* strains grown in media containing untreated Nutriose (N0) or BG_81/Cel3A pretreated Nutriose (NBG81) as a carbon source.

Detailed Description of the Invention

Cellulose is the major structural component of higher plants. It provides plant cells with high tensile strength helping them to resist mechanical

stress and osmotic pressure. Cellulose is a β-1,4-glucan composed of linear chains of glucose residues joined by β-1,4-glycosidic linkages. Cellobiose is the smallest repeating unit of cellulose. In cell walls cellulose is packed in variously oriented sheets, which are embedded in a matrix of hemicellulose and lignin. Hemicellulose is a heterogeneous group of carbohydrate polymers containing mainly different glucans, xylans and mannans. Hemicellulose consists of a linear backbone with β-1,4-linked residues substituted with short side chains usually containing acetyl, glucuronyl, arabinosyl and galactosyl. Hemicellulose can be chemically cross-linked to lignin. Lignin is a complex cross-linked polymer of variously substituted p-hydroxyphenylpropane units that provides strength to the cell wall to withstand mechanical stress, and it also protects cellulose from enzymatic hydrolysis.

Lignocellulose is a combination of cellulose and hemicellulose and polymers of phenol propanol units and lignin. It is physically hard, dense, and inaccessible and the most abundant biochemical material in the biosphere. Lignocellulose containing materials are for example: hardwood and softwood chips, wood pulp, sawdust and forestry and wood industrial waste; agricultural biomass as cereal straws, sugar beet pulp, corn stover and cobs, sugar cane bagasse, stems, leaves, hulls, husks, and the like; waste products as municipal solid waste, newspaper and waste office paper, milling waste of e.g. grains; dedicated energy crops (e.g., willow, poplar, swithcgrass or reed canarygrass, and the like). Preferred examples are corn stover, switchgrass, cereal straw, sugarcane bagasse and wood derived materials.

"Cellulosic material" as used herein, relates to any material comprising cellulose, hemicellulose and/or lignocellulose as a significant component. "Lignocellulosic material" means any material comprising lignocellulose. Such materials are e.g. plant materials such as wood including softwood and hardwood, herbaceous crops, agricultural residues, pulp and paper residues, waste paper, wastes of food and feed industry etc. Textile fibres such as cotton, fibres derived from cotton, linen, hemp, jute and man made cellulosic fibres as modal, viscose, lyocel are specific examples of cellulosic materials.

Cellulosic material is degraded in nature by a number of various organisms including bacteria and fungi. Cellulose is typically degraded by different cellulases acting sequentially or simultaneously. The biological conversion of cellulose to glucose generally requires three types of hydrolytic enzymes: (1) Endoglucanases which cut internal beta-1,4-glucosidic bonds; (2) Exocellobio-

hydrolases that cut the dissaccharide cellobiose from the end of the cellulose polymer chain; (3) Beta-1,4-glucosidases which hydrolyze the cellobiose and other short cello-oligosaccharides to glucose. In other words the three major groups of cellulases are cellobiohydrolases (CBH), endoglucanases (EG) and beta-glucosidases (BG).

Degradation of more complex cellulose containing substrates requires a broad range of various enzymes. For example lignocellulose is degraded by hemicellulases, like xylanases and mannanases. Hemicellulase is an enzyme hydrolysing hemicellulose.

"Cellulolytic enzymes" are enzymes having "cellulolytic activity", which means that they are capable of hydrolysing cellulosic substrates or derivatives thereof into smaller saccharides. Cellulolytic enzymes thus include both cellulases and hemicellulases. Cellulases as used herein include cellobiohydrolase, endoglucanase and beta-glucosidase.

T. reesei has a well known and effective cellulase system containing two CBH's, two major and several minor EG's and BG's. *T. reesei* CBHI (Cel7A) cuts sugar from the reducing end of the cellulose chain, has a C-terminal cellulose binding domain (CBD) and may constitute up to 60% of the total secreted protein. *T. reesei* CBHII (Cel6A) cuts sugar from the non-reducing end of the cellulose chain, has an N-terminal cellulose binding domain and may constitute up to 20% of the total secreted protein. Endoglucanases EGI (Cel7B), and EGV (Cel45A) have a CBD in their C-terminus, EGII (Cel5A) has an N-terminal CBD and EGIII (Cel12A) does not have a cellulose binding domain at all. CBHI, CBHII, EGI and EGII are so called "major cellulases" of *Trichoderma* comprising together 80–90% of total secreted proteins. It is known to a man skilled in the art that an enzyme may be active on several substrates and enzymatic activities can be measured using different substrates, methods and conditions. Identifying different cellulolytic activities is discussed for example in van Tilbeurgh *et al.* 1988.

In addition to a catalytic domain/core expressing cellulolytic activity cellulolytic enzymes may comprise one or more cellulose binding domains (CBDs), also named as carbohydrate binding domains/modules (CBD/CBM), which can be located either at the N- or C-terminus of the catalytic domain. CBDs have carbohydrate-binding activity and they mediate the binding of the cellulase to crystalline cellulose but have little or no effect on cellulase hydro-

lytic activity of the enzyme on soluble substrates. These two domains are typically connected via a flexible and highly glycosylated linker region.

"Cellobiohydrolase" or "CBH" as used herein refers to enzymes that cleave cellulose from the end of the glucose chain and produce mainly cellobiose. They are also called 1,4-beta-D-glucan cellobiohydrolases or cellulose 1,4-beta-cellobiosidases. They hydrolyze the 1,4-beta-D-glucosidic linkages from the reducing or non-reducing ends of a polymer containing said linkages, such as cellulose, whereby cellobiose is released. Two different CBHs have been isolated from *Trichoderma reesei*, CBHI and CBHII. They have a modular structure consisting of a catalytic domain linked to a cellulose-binding domain (CBD). There are also cellobiohydrolases in nature that lack CBD.

"Endoglucanase" or "EG" refers to enzymes that cut internal glycosidic bonds of the cellulose chain. They are classified as EC 3.2.1.4. They are 1,4-beta-D-glucan 4-glucanohydrolases and catalyze endohydrolysis of 1,4-beta-D-glycosidic linkages in polymers of glucose such as cellulose and derivatives thereof. Some naturally occurring endoglucanases have a cellulose binding domain, while others do not. Some endoglucanases have also xylanase activity (Bailey *et al.*, 1993).

"Beta-glucosidase" or "BG" or "βG" refers to enzymes that degrade small soluble oligosaccharides including cellobiose to glucose. They are classified as EC 3.2.1.21. They are beta-D-glucoside glucohydrolases, which typically catalyze the hydrolysis of terminal non-reducing beta-D-glucose residues. These enzymes recognize oligosaccharides of glucose. Typical substrates are cellobiose and cellotriose. Cellobiose is an inhibitor of cellobiohydrolases, wherefore the degradation of cellobiose is important to overcome end-product inhibition of cellobiohydrolases.

Xylanases are enzymes that are capable of recognizing and hydrolyzing hemicellulose. They include both exohydrolytic and endohydrolytic enzymes. Typically they have endo-1,4-beta-xylanase (EC 3.2.1.8) or beta-D-xylosidase (EC 3.2.1.37) activity that breaks down hemicellulose to xylose. "Xylanase" or "Xyn" in connection with the present invention refers especially to an enzyme classified as EC 3.2.1.8 hydrolyzing xylose polymers of lignocellulosic substrate or purified xylan.

In addition to this cellulases can be classified to various glycosyl hydrolase families according their primary sequence, supported by analysis of the three dimensional structure of some members of the family (Henrissat

1991, Henrissat and Bairoch 1993, 1996). Some glycosyl hydrolases are multifunctional enzymes that contain catalytic domains that belong to different glycosylhydrolase families. Family 3 consists of beta-glucosidases (EC 3.2.1.21) such as Ta BG_81, At BG_101 and Ct BG_76 described herein. Family 5 (formerly known as celA) consists mainly of endoglucanases (EC 3.2.1.4) such as Ta EG_28 described herein. Family 7 (formerly cellulase family celC) contains endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91) such as Ct EG_54, Ta CBH, At CBH_A, At CBH_C and Ct CBH described herein. Family 10 (formerly celF) consists mainly of xylanases (EC 3.2.1.8) such as Ta XYN_30 and At XYN_60 described herein. Family 45 (formerly celK) contains endoglucanases (EC 3.2.1.4) such as At EG_40 and At EG_40_like described herein.

Cellulolytic enzymes useful for hydrolyzing cellulosic material are obtainable from *Thermoascus aurantiacus*, *Acremonium thermophilum*, or *Chaetomium thermophilum*. "Obtainable from" means that they can be obtained from said species, but it does not exclude the possibility of obtaining them from other sources. In other words they may originate from any organism including plants. Preferably they originate from microorganisms e.g. bacteria or fungi. The bacteria may be for example from a genus selected from *Bacillus*, *Azospirillum* and *Streptomyces*. More preferably the enzyme originates from fungi (including filamentous fungi and yeasts), for example from a genus selected from the group consisting of *Thermoascus, Acremonium, Chaetomium, Achaetomium, Thielavia, Aspergillus, Botrytis, Chrysosporium, Collybia, Fomes, Fusarium, Humicola, Hypocrea, Lentinus, Melanocarpus, Myceliophthora, Myriococcum, Neurospora, Penicillium, Phanerochaete, Phlebia, Pleurotus, Podospora, Polyporus, Rhizoctonia, Scytalidium*, *Pycnoporus, Trametes* and *Trichoderma*.

According to a preferred embodiment of the invention the enzymes are obtainable from *Thermoascus aurantiacus* strain ALKO4242 deposited as CBS 116239, strain ALKO4245 deposited as CBS 116240 presently classified as *Acremonium thermophilium*, or *Chaetomium thermophilum* strain ALKO4265 deposited as CBS 730.95.

The cellobiohydrolase preferably comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 2, 4, 6 or 8, or an enzymatically active fragment thereof.

Cellobio-hydrolase	Gene	Obtainable from	CBD	nucleic acid SEQ ID NO:	amino acid SEQ ID NO:
Ta CBH	Ta *cel7A*	*T. aurantiacus*	-	1	2
At CBH_A	At cel7B	*A. thermophilum*	-	3	4
At CBH_C	At *cel7A*	*A. thermophilum*	+	5	6
Ct CBH	Ct *cel7A*	*C. thermophilum*	+	7	8

These CBHs have an advantageous cellulose inhibition constant compared to that of *Trichoderma reesei* CBH, and they show improved hydrolysis results when testing various cellulosic substrates. SEQ ID NO: 2 and 4 do not comprise a CBD. Particularly enhanced hydrolysis results may be obtained when a cellulose binding domain (CBD) is attached to a CBH that has no CBD of its own. The CBD may be obtained e.g. from a *Trichoderma* or *Chaetomium* species, and it is preferably attached to the CBH via a linker. The resulting fusion protein containing a CBH core region attached to a CBD via a linker may comprise an amino acid sequence having at least 80 % identity to SEQ ID NO: 28 or 30. Polynucleotides comprising a sequence of SEQ ID NO: 27 or 29 encode such fusion proteins.

The endoglucanase may comprise an amino acid sequence having at least 80% identity to SEQ ID NO: 10, 12, 14 or 16, or an enzymatically active fragment thereof. These endoglucanases have good thermostability.

Endo-glucanase	Gene	Obtainable from	CBD	nucl. acid SEQ ID NO:	amino acid SEQ ID NO:
Ta EG_28	Ta *cel5A*	*T. aurantiacus*	-	9	10
At EG_40	At *cel45A*	*A. thermophilum*	+	11	12
At EG40_like	At *cel45B*	*A. thermophilum*	-	13	14
Ct EG_54	Ct *cel7B*	*C. thermophilum*	+	15	16

The beta-glucosidase may comprise an amino acid sequence having at least 80% identity to SEQ ID NO: 22, 24 or 26, or an enzymatically active fragment thereof. These beta-glucosidases have good resistance to glucose inhibition, which is advantageous to avoid end product inhibition during enzymatic hydrolysis of cellulosic material. The beta-glucosidases may also be used in preparing sophorose, a cellulase inducer used in cultivation of *T. reesei*.

Beta-glucosidase	Gene	Obtainable from	nucleic acid SEQ ID NO:	amino acid SEQ ID NO:
Ta BG_81	Ta *cel*3A	*T. aurantiacus*	21	22
At BG_101	At *cel*3A	*A. thermophilum*	23	24
Ct BG_76	Ct *cel*3A	*C. thermophilum*	25	26

The xylanase may comprise an amino acid sequence having at least 80% identity to SEQ ID NO: 18 or 20, or an enzymatically active fragment thereof.

Xylanase	Gene	Obtainable from	CBD	nucleic acid SEQ ID NO:	amino acid SEQ ID NO:
Xyn_30	Ta *xyn*10A	*T. aurantiacus*	+	17	18
Xyn_60	At *xyn*10A	*A. thermophilum*	-	19	20

By the term "identity" is here meant the global identity between two amino acid sequences compared to each other from the first amino acid encoded by the corresponding gene to the last amino acid. The identity of the full-length sequences is measured by using Needleman-Wunsch global alignment program at EMBOSS (European Molecular Biology Open Software Suite; Rice *et al.*, 2000) program package, version 3.0.0, with the following parameters: EMBLOSUM62, Gap penalty 10.0, Extend penalty 0.5. The algorithm is described in Needleman and Wunsch (1970). The man skilled in the art is aware of the fact that results using Needleman-Wunsch algorithm are comparable only when aligning corresponding domains of the sequence. Consequently comparison of e.g. cellulase sequences including CBD or signal sequences with sequences lacking those elements cannot be done.

According to one embodiment of the invention, a cellulolytic polypeptide is used that has at least 80, 85, 90, 95 or 99% identity to SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24 or 26 or at least to its enzymatically active fragment.

By the term "enzymatically active fragment" is meant any fragment of a defined sequence that has cellulolytic activity. In other words an enzymatically active fragment may be the mature protein part of the defined sequence,

or it may be only an fragment of the mature protein part, provided that it still has cellobiohydrolase, endoglucanase, beta-glucosidase or xylanase activity.

The cellulolytic enzymes are preferably recombinant enzymes, which may be produced in a generally known manner. A polynucleotide fragment comprising the enzyme gene is isolated, the gene is inserted under a strong promoter in an expression vector, the vector is transferred into suitable host cells and the host cells are cultivated under conditions provoking production of the enzyme. Methods for protein production by recombinant technology in different host systems are well known in the art (Sambrook *et al.*, 1989; Coen, 2001; Gellissen, 2005). Preferably the enzymes are produced as extracellular enzymes that are secreted into the culture medium, from which they can easily be recovered and isolated. The spent culture medium of the production host can be used as such, or the host cells may be removed therefrom, and/or it may be concentrated, filtrated or fractionated. It may also be dried.

Isolated polypeptide in the present context may simply mean that the cells and cell debris have been removed from the culture medium containing the polypeptide. Conveniently the polypeptides are isolated e.g. by adding anionic and/or cationic polymers to the spent culture medium to enhance precipitation of cells, cell debris and some enzymes that have unwanted side activities. The medium is then filtrated using an inorganic filtering agent and a filter to remove the precipitants formed. After this the filtrate is further processed using a semi-permeable membrane to remove excess of salts, sugars and metabolic products.

According to one embodiment of the invention, the heterologous polynucleotide comprises a gene similar to that included in a microorganism having accession number DSM 16723, DSM 16728, DSM 16729, DSM 16727, DSM 17326, DSM 17324, DSM 17323, DSM 17729, DSM 16724, DSM 16726, DSM 16725, DSM 17325 or DSM 17667.

The production host can be any organism capable of expressing the cellulolytic enzyme. Preferably the host is a microbial cell, more preferably a fungus. Most preferably the host is a filamentous fungus. Preferably the recombinant host is modified to express and secrete cellulolytic enzymes as its main activity or one of its main activities. This can be done by deleting major homologous secreted genes e.g. the four major cellulases of *Trichoderma* and by targeting heterologous genes to a locus that has been modified to ensure high expression and production levels. Preferred hosts for producing the cellu-

lolytic enzymes are in particular strains from the genus *Trichoderma* or *Aspergillus*.

The enzymes needed for the hydrolysis of the cellulosic material according to the invention may be added in an enzymatically effective amount either simultaneously e.g. in the form of an enzyme mixture, or sequentially, or as a part of the simultaneous saccharification and fermentation (SSF). Any combination of the cellobiohydrolases comprising an amino acid sequence having at least 80% identity to SEQ ID NO: 2, 4, 6 or 8 or to an enzymatically active fragment thereof may be used together with any combination of endoglucanases and beta-glucosidases. If the cellulosic material comprises hemicellulose, hemicellulases, preferably xylanases are additionally used for the degradation. The endoglucanases, beta-glucosidases and xylanases may be selected from those described herein, but are not limited to them. They can for example also be commercially available enzyme preparations. In addition to cellulases and optional hemicellulases one or more other enzymes may be used, for example proteases, amylases, laccases, lipases, pectinases, esterases and/or peroxidases. Another enzyme treatment may be carried out before, during or after the cellulase treatment.

The term "enzyme preparation" denotes to a composition comprising at least one of the desired enzymes. The preparation may contain the enzymes in at least partially purified and isolated form. It may even essentially consist of the desired enzyme or enzymes. Alternatively the preparation may be a spent culture medium or filtrate containing one or more cellulolytic enzymes. In addition to the cellulolytic activity, the preparation may contain additives, such as mediators, stabilizers, buffers, preservatives, surfactants and/or culture medium components. Preferred additives are such, which are commonly used in enzyme preparations intended for a particular application. The enzyme preparation may be in the form of liquid, powder or granulate. Preferably the enzyme preparation is spent culture medium. "Spent culture medium" refers to the culture medium of the host comprising the produced enzymes. Preferably the host cells are separated from the said medium after the production.

According to one embodiment of the invention the enzyme preparation comprises a mixture of CBH, EG and BG, optionally together with xylanase and/or other enzymes. The CBH comprises an amino acid sequence having at least 80% identity to SEQ ID NO: 2, 4, 6 or 8 or to an enzymatically active fragment thereof, and it may be obtained from *Thermoascus aurantia-*

cus, *Acremonium thermophilum*, or *Chaetomium thermophilum*, whereas EG, BG and xylanase may be of any origin including from said organisms. Other enzymes that might be present in the preparation are e.g. proteases, amylases, laccases, lipases, pectinases, esterases and/or peroxidases.

Different enzyme mixtures and combinations may be used to suit different process conditions. For example if the degradation process is to be carried out at a high temperature, thermostable enzymes are chosen. A combination of a CBH of family 7 with an endoglucanase of family 45, optionally in combination with a BG of family 3 and/or a xylanase of family 10 had excellent hydrolysis performance both at 45°C, and at elevated temperatures.

Cellulolytic enzymes of *Trichoderma reesei* are conventionally used at temperatures in the range of about 40–50°C in the hydrolysis, and at 30–40°C in SSF. CBH, EG, BG and Xyn obtainable from *Thermoascus aurantiacus*, *Acremonium thermophilum*, or *Chaetomium thermophilum* are efficient at these temperatures too, but in addition most of them also function extremely well at temperatures between 50°C and 75°C, or even up to 80°C and 85°C, such as between 55°C and 70°C, e.g. between 60°C and 65°C. For short incubation times enzyme mixtures are functional up to even 85°C, for complete hydrolysis lower temperatures are normally used.

The method for treating cellulosic material with CBH, EG, BG and Xyn is especially suitable for producing fermentable sugars from lignocellulosic material. The fermentable sugars may then be fermented by yeast into ethanol, and used as fuel. They can also be used as intermediates or raw materials for the production of various chemicals or building blocks for the processes of chemical industry, e.g. in so called biorefinery. The lignocellulosic material may be pretreated before the enzymatic hydrolysis to disrupt the fiber structure of cellulosic substrates and make the cellulose fraction more accessible to the cellulolytic enzymes. Current pretreatments include mechanical, chemical or thermal processes and combinations thereof. The material may for example be pretreated by steam explosion or acid hydrolysis.

A number of novel cellulolytic polypeptides were found in *Thermoascus aurantiacus*, *Acremonium thermophilum*, and *Chaetomium thermophilum.* The novel polypeptides may comprise a fragment having cellulolytic activity and be selected from the group consisting of a polypeptide comprising an amino acid sequence having at least 66%, preferably 70% or 75%, identity to SEQ ID NO: 4, 79% identity to SEQ ID NO: 6, 78% identity to SEQ ID NO: 12,

68%, preferably 70% or 75%, identity to SEQ ID NO: 14, 72%, preferably 75%, identity to SEQ ID NO: 16, 68%, preferably 70% or 75%, identity to SEQ ID NO: 20, 74% identity to SEQ ID NO: 22 or 24, or 78% identity to SEQ ID NO: 26.

The novel polypeptides may also be variants of said polypeptides. A "variant" may be a polypeptide that occurs naturally e.g. as an allelic variant within the same strain, species or genus, or it may have been generated by mutagenesis. It may comprise amino acid substitutions, deletions or insertions, but it still functions in a substantially similar manner to the enzymes defined above i.e. it comprises a fragment having cellulolytic activity.

The cellulolytic polypeptides are usually produced in the cell as immature polypeptides comprising a signal sequence that is cleaved off during secretion of the protein. They may also be further processed during secretion both at the N-terminal and/or C-terminal end to give a mature, enzymatically active protein. A polypeptide "comprising a fragment having cellulolytic activity" thus means that the polypeptide may be either in immature or mature form, preferably it is in mature form, i.e. the processing has taken place.

The novel polypeptides may further be a "fragment of the polypeptides or variants" mentioned above. The fragment may be the mature form of the proteins mentioned above, or it may be only an enzymatically active part of the mature protein. According to one embodiment of the invention, the polypeptide has an amino acid sequence having at least 80, 85, 90, 95, or 99% identity to SEQ ID NO: 4, 6, 12, 14, 16, 20, 22, 24 or 26, or to a cellulolytically active fragment thereof. It may also be a variant thereof, or a fragment thereof having cellobiohydrolase, endoglucanase, xylanase, or beta-glucosidase activity. According to another embodiment of the invention, the polypeptide consists essentially of a cellulolytically active fragment of a sequence of SEQ ID NO: 4, 6, 12, 14, 16, 20, 22, 24 or 26.

The novel polynucleotides may comprise a nucleotide sequence of SEQ ID NO: 3, 5, 11, 13, 15, 19, 21, 23 or 25, or a sequence encoding a novel polypeptide as defined above, including complementary strands thereof. Polynucleotide as used herein refers to both RNA and DNA, and it may be single stranded or double stranded. The polynucleotide may also be a fragment of said polynucleotides comprising at least 20 nucleotides, e.g. at least 25, 30 or 40 nucleotides. According to one embodiment of the invention it is at least 100, 200 or 300 nucleotides in length. Further the polynucleotide may be degener-

ate as a result of the genetic code to any one of the sequences as defined above. This means that different codons may code for the same amino acid.

According to one embodiment of the invention the polynucleotide is "comprised in" SEQ ID NO: 3, 5, 11, 13, 15, 19, 21, 23 or 25, which means that the sequence has at least part of the sequence mentioned. According to another embodiment of the invention, the polynucleotide comprises a gene similar to that included in a microorganism having accession number DSM 16728, DSM 16729, DSM 17324, DSM 17323, DSM 17729, DSM 16726, DSM 16725, DSM 17325 or DSM 17667.

The novel proteins/polypeptides may be prepared as described above. The novel polynucleotides may be inserted into a vector, which is capable of expressing the polypeptide encoded by the heterologous sequence, and the vector may be inserted into a host cell capable of expressing said polypeptide. The host cell is preferably of the genus *Trichoderma* or *Aspergillus*.

A heterologous gene encoding the novel polypeptides has been introduced on a plasmid into an *Escherichia coli* strain having accession number DSM 16728, DSM 16729, DSM 17324, DSM 17323, DSM 17729, DSM 16726, DSM 16725, DSM 17325 or DSM 17667.

The novel enzymes may be components of an enzyme preparation. The enzyme preparation may comprise one or more of the novel polypeptides, and it may be e.g. in the form of spent culture medium, powder, granules or liquid. According to one embodiment of the invention it comprises cellobiohydrolase, endoglucanase, beta-glucosidase, and optionally xylanase activity and/or other enzyme activities. It may further comprise any conventional additives.

The novel enzymes may be applied in any process involving cellulolytic enzymes, such as in fuel, textile, detergent, pulp and paper, food, feed or beverage industry, and especially in hydrolysing cellulosic material for the production of biofuel comprising ethanol. In the pulp and paper industry they may be used to modify cellulosic fibre for example in treating kraft pulp, mechanical pulp, or recycled paper.

The invention is illustrated by the following non-limiting examples. It should be understood, however, that the embodiments given in the description above and in the examples are for illustrative purposes only, and that various changes and modifications are possible within the scope of the invention.

Examples

Example 1. Screening for strains expressing cellulolytic activity and their cultivation for purification

About 25 fungal strains from the Roal Oy culture collection were tested for cellulolytic activity including beta-glucosidases. After preliminary screening six strains were chosen for further studies. These were *Thermoascus aurantiacus* ALKO4239 and ALKO4242, *Acremonium thermophilum* ALKO4245, *Talaromyces thermophilus* ALKO4246 and *Chaetomium thermophilum* ALKO4261 and ALKO4265.

The strains ALKO4239, ALKO4242 and ALKO4246 were cultivated in shake flasks at 42°C for 7 d in the medium 3 x B, which contains g/litre: Solka Floc cellulose 18, distiller's spent grain 18, oats spelt xylan 9, $CaCO_3$ 2, soybean meal 4.5, $(NH_4)HPO_4$ 4.5, wheat bran 3.0, KH_2PO_4 1.5, $MgSO_4 \cdot H_2O$ 1.5, NaCl 0.5, KNO_3 0.9, locust bean gum 9.0, trace element solution #1 0.5, trace element solution #2 0.5 and Struktol (Stow, OH, USA) antifoam 0.5 ml; the pH was adjusted to 6.5. Trace element solution #1 has g/litre: $MnSO_4$ 1.6, $ZnSO_4 \cdot 7H_2O$ 3.45 and $CoCl_2 \cdot 6H_2O$ 2.0; trace element solution #2 has g/litre: $FeSO_4 \cdot 7H_2O$ 5.0 with two drops of concentrated H_2SO_4.

The strain ALKO4261 was cultivated in shake flasks in the medium 1xB, which has one third of each of the constituents of the 3 x B medium (above) except it has same concentrations for $CaCO_3$, NaCl and the trace elements. The strain was cultivated at 45°C for 7 d.

The strain ALKO4265 was cultivated in shake flasks in the following medium, g/l: Solka Floc cellulose 40, Pharmamedia™ (Traders Protein, Memphis, TN, USA) 10, corn steep powder 5, $(NH_4)_2SO_4$ 5 and KH_2PO_4 15; the pH was adjusted to 6.5. The strain was cultivated at 45°C for 7 d.

After the cultivation the cells and other solids were collected by centrifugation down and the supernatant was recovered. For the shake flask cultivations, protease inhibitors PMSF (phenylmethyl-sulphonylfluoride) and pepstatin A were added to 1 mM and 10 µg/ml, respectively. If not used immediately, the preparations were stored in aliquots at –20°C.

For the estimation of the thermoactivity of the enzymes, assays were performed of the shake flask cultivation preparations at 50°C, 60°C, 65°C, 70°C and 75°C for 1 h, in the presence of 100 µg bovine serum albumin (BSA) /ml as a stabilizer. Preliminary assays were performed at 50°C and

65°C at two different pH values (4.8/5.0 or 6.0) in order to clarify, which pH was more appropriate for the thermoactivity assay.

All shake flask supernatants were assayed for the following activities:

Cellobiohydrolase I –like activity ('CBHI') and the endoglucanase I –like activity ('EGI'):

These were measured in 50 mM Na-acetate buffer with 0.5 mM MUL (4-methylumbelliferyl-beta-D-lactoside) as the substrate. Glucose (100 mM) was added to inhibit any interfering beta-glucosidase activity. The liberated 4-methylumbelliferyl was measured at 370 nm. The 'CBHI' and the 'EGI' activities were distinguished by measuring the activity in the presence and absence of cellobiose (5 mM). The activity that is not inhibited by cellobiose represents the 'EGI' activity and the remaining MUL activity represents the 'CBHI' activity (van Tilbeurgh *et al*, 1988). The assay was performed at pH 5.0 or 6.0 (see below).

The endoglucanase (CMCase) activity:

This was assayed with 2% (w/v) carboxymethylcellulose (CMC) as the substrate in 50 mM citrate buffer essentially as described by Bailey and Nevalainen 1981; Haakana *et al.* 2004. Reducing sugars were measured with the DNS reagent. The assay was performed at pH 4.8 or 6.0 (see below).

Beta-glucosidase (BGU) activity:

This was assayed with 4-nitrophenyl-β-D-glucopyranoside (1 mM) in 50 mM citrate buffer as described by Bailey and Nevalainen 1981. The liberated 4-nitrophenol was measured at 400 nm. The assay was performed at pH 4.8 or 6.0 (see below).

The relative activities of the enzymes are presented in Figure 1. The relative activities were presented by setting the activity at 60°C as 100% (Figure 1). All strains produced enzymes, which had high activity at high temperatures (65°C–75°C).

For protein purifications. ALKO4242 was also grown in a 2 litre bioreactor (Braun Biostat® B, Braun, Melsungen, Germany) in the following medium, g/litre: Solka Floc cellulose 40, soybean meal 10, NH_4NO_3 5, KH_2PO_4 5, $MgSO_4 \cdot 7H_2O$ 0.5, $CaCl_2 \cdot 2H_2O$ 0.05, trace element solution #1 0.5, trace element solution #2 0.5. The aeration was 1 vvm, antifoam control with Struktol, stirring 200–800 rpm and temperature at 47°C. Two batches were run, one at pH 4.7 ± 0.2 (NH_3 / H_2SO_4) and the other with initial pH of pH 4.5. The

cultivation time was 7 d. After the cultivation the cells and other solids were removed by centrifugation.

The strain ALKO4245 was grown in 2 litre bioreactor (Braun Biostat® B, Braun, Melsungen, Germany) in the following medium, g/litre: Solka Floc cellulose 40, corn steep powder 15, distiller's spent grain 5, oats spelt xylan 3, locust bean gum 3, $(NH_4)_2SO_4$ 5 and KH_2PO_4 5. The pH range was 5.2± 0.2 (NH_3 / H_2SO_4), aeration 1 vvm, stirring 300–600 rpm, antifoam control with Struktol and the temperature 42°C. The cultivation time was 4 d. After the cultivation the cells and other solids were removed by centrifugation.

For enzyme purification, ALKO4261 was grown in a 10 litre bioreactor (Braun Biostat® ED, Braun, Melsungen, Germany) in the following medium, g/litre: Solka Floc cellulose 30, distiller's spent grain 10, oats spelt xylan 5, $CaCO_3$ 2, soybean meal 10, wheat bran 3.0, $(NH_4)_2SO_4$ 5, KH_2PO_4 5, $MgSO_4 \cdot 7H_2O$ 0.5, NaCl 0.5, KNO_3 0.3, trace element solution #1 0.5 and trace element solution #2 0.5. The pH range was 5.2± 0.2 (NH_3 / H_2SO_4), aeration 1 vvm, stirring 200–600 rpm, antifoam control with Struktol and the temperature 42°C. The cultivation time was 5 d. A second batch was grown under similar conditions except that Solka Floc was added to 40 g/l and spent grain to 15 g/l. The supernatants were recovered by centrifugation and filtering through Seitz-K 150 and EK filters (Pall SeitzSchenk Filtersystems GmbH, Bad Kreuznach, Germany). The latter supernatant was concentrated about ten fold using the Pellicon mini ultrafiltration system (filter NMWL 10 kDa; Millipore, Billerica, MA, USA).

For enzyme purification, ALKO4265 was also grown in a 10 litre bioreactor (Braun Biostat® ED, Braun, Melsungen, Germany) in the same medium as above, except KH_2PO_4 was added to 2.5 g/l. The pH range was 5.3 ± 0.3 (NH_3 / H_3PO_4), aeration 0.6 vvm, stirring 500 rpm, antifoam control with Struktol and the temperature 43°C. The cultivation time was 7 d. The supernatants were recovered by centrifugation and filtering through Seitz-K 150 and EK filters (Pall SeitzSchenk Filtersystems GmbH, Bad Kreuznach, Germany). The latter supernatant was concentrated about 20 fold using the Pellicon mini ultrafiltration system (filter NMWL 10 kDa; Millipore, Billerica, MA, USA).

Example 2. Purification and characterization of cellobiohydrolases from *Acremonium thermophilum* ALKO4245 and *Chaetomium thermophilum* ALKO4265

Acremonium thermophilum ALKO4245 and *Chaetomium thermophilum* ALKO4265 were grown as described in Example 1. The main cellobiohydrolases were purified using *p*-aminobenzyl 1-thio-β-cellobioside-based affinity column, prepared as described by Tomme *et al.,* 1988.

The culture supernatants were first buffered into 50 mM sodium acetate buffer pH 5.0, containing 1 mM δ-gluconolactone and 0.1 M glucose in order to retard ligand hydrolysis in the presence of β-glucosidases. Cellobiohydrolases were eluted with 0.1 M lactose and finally purified by gel filtration chromatography using Superdex 200 HR 10/30 columns in the ÄKTA system (Amersham Pharmacia Biotech). The buffer used in gel filtration was 50 mM sodium phosphate pH 7.0, containing 0.15 M sodium chloride.

Purified cellobiohydrolases were analysed by SDS-polyacrylamide gel electrophoresis and the molecular mass of both proteins was determined to be approximately 70 kDa evaluated on the basis of the molecular mass standards (Low molecular weight calibration kit, Amersham Biosciences). Purified *Acremonium* and *Chaetomium* cellobiohydrolases were designated as At Cel7A and Ct Cel7A, respectively, following the scheme in Henrissat *et al.* (1998) (Henrissat, 1991; Henrissat and Bairoch, 1993).

The specific activity of the preparations was determined using 4-methylumbelliferyl-β-D-lactoside (MUL), 4-methylumbelliferyl-β-D-cellobioside (MUG2) or 4-methylumbelliferyl-β-D-cellotrioside (MUG3) as substrate (van Tilbeurgh *et al.*, 1988) in 0.05 M sodium citrate buffer pH 5 at 50°C for 10 min. Endoglucanase and xylanase activities were determined by standard procedures (according to IUPAC, 1987) using carboxymethyl cellulose (CMC) and birch glucuronoxylan (Bailey *et al.*, 1992) as substrates. Specific activity against Avicel was calculated on the basis of reducing sugars formed in a 24 h reaction at 50°C, pH 5.0, with 1% substrate and 0.25 µM enzyme dosage. The protein content of the purified enzyme preparations was measured according to Lowry *et al.*, 1951. To characterize the end products of hydrolysis, soluble sugars liberated in 24 h hydrolysis experiment, as described above, were analysed by HPLC (Dionex). Purified cellobiohydrolase I (CBHI/Cel7A) of *Trichoderma reesei* was used as a reference.

The specific activities of the purified enzymes and that of *T. reesei* CBHI/Cel7A are presented in Table 1. The purified At Cel7A and Ct Cel7A cellobiohydrolases possess higher specific activities against small synthetic substrates as compared to *T. reesei* CBHI/Cel7A. The specific activity against Avicel was clearly higher with the herein disclosed enzymes. Low activities of the purified enzyme preparations against xylan and CMC may either be due to the properties of the proteins themselves, or at least partially to the remaining minor amounts of contaminating enzymes. The major end product of cellulose hydrolysis by all purified enzymes was cellobiose which is typical to cellobiohydrolases.

Table 1. Specific activities (nkat/mg) of the purified cellobiohydrolases and the reference enzyme of *T. reesei* (50°C, pH 5.0, 24 h).

Substrate	*A. thermophilum* ALKO4245 Cel7A	*C. thermophilum* ALKO4265 Cel7A	*T. reesei* Cel7A
Xylan	11.3	6.7	1.3
CMC	26.2	5.5	1.0
MUG2	9.2	18.9	4.3
MUG3	1.3	1.5	0.9
MUL	21.5	54.0	21.9
Avicel	1.8	1.4	0.6

Thermal stability of the purified cellobiohydrolases was determined at different temperatures. The reaction was performed in the presence of 0.1% BSA at pH 5.0 for 60 min using 4-methylumbelliferyl-β-D-lactoside as substrate. *C. thermophilum* ALKO4265 CBH/Cel7A and *A. thermophilum* ALKO4245 CBH/Cel7A were stable up to 65° and 60°C, respectively. The *T. reesei* reference enzyme (CBHI/Cel7A) retained 100% of activity up to 55°C.

Example 3. Purification and characterization of an endoglucanase from *Acremonium thermophilum* ALKO4245

Acremonium thermophilum ALKO4245 was grown as described in Example 1. The culture supernatant was incubated at 70°C for 24 hours after which it was concentrated by ultrafiltration. The pure endoglucanase was obtained by sequential purification with hydrophobic interaction and cation exchange chromatography followed by gel filtration. The endoglucanase activity of the fractions collected during purification was determined using carboxymethyl cellulose (CMC) as substrate (procedure of IUPAC 1987). Protein content was measured by BioRad Assay Kit (Bio-Rad Laboratories) using bovine serum albumine as standard.

The concentrated culture supernatant was applied to a HiPrep 16/10 Butyl FF hydrophobic interaction column equilibrated with 20 mM potassium phosphate buffer pH 6.0, containing 1 M $(NH_4)_2SO_4$. Bound proteins were eluted with the linear gradient from the above buffer to 5 mM potassium phosphate, pH 6.0. Fractions were collected and the endoglucanase activity

was determined as described above. The endoglucanase activity was eluted in a broad conductivity area of 120 to 15 mS/cm.

Combined fractions were applied to a HiTrap SP XL cation exchange column equilibrated with 8 mM sodium acetate, pH 4.5. Bound proteins were eluted with a linear gradient from 0 to 0.25 M NaCl in the equilibration buffer. The protein containing endoglucanase activity was eluted at the conductivity area of 3–7 mS/cm. Cation exchange chromatography was repeated and the protein eluate was concentrated by freeze drying.

The dissolved sample was loaded onto a Superdex 75 HR10/30 gel filtration column equilibrated with 20 mM sodium phosphate buffer pH 7.0, containing 0.15 M NaCl. The main protein fraction was eluted from the column with the retention volume of 13.3 ml. The protein eluate was judged to be pure by SDS-polyacryl amide gel electrophoresis and the molecular weight was evaluated to be 40 kDa. The specific activity of the purified protein, designated as At EG_40, at 50°C was determined to be 450 nkat/mg (procedure of IUPAC 1987, using CMC as substrate).

Thermal stability of the purified endoglucanase was determined at different temperatures. The reaction was performed in the presence of 0.1 mg/ml BSA at pH 5.0 for 60 min using carboxymethyl cellulose as substrate. *A. thermophilum* EG_40/Cel45A was stable up to 80°C. The *T. reesei* reference enzymes EGI (Cel7B) and EGII (Cel5A) retained 100% of activity up to 60°C and 65°C, respectively.

Example 4. Purification of an endoglucanase from *Chaetomium thermophilum* ALKO4261

Chaetomium thermophilum ALKO4261 was grown as described in Example 1. The pure endoglucanase was obtained by sequential purification with hydrophobic interaction and cation exchange chromatography followed by gel filtration. The endoglucanase activity of the fractions collected during purification was determined using carboxymethyl cellulose (CMC) as substrate (procedure of IUPAC 1987).

Ammonium sulfate was added to the culture supernatant to reach the same conductivity as 20 mM potassium phosphate pH 6.0, containing 1 M $(NH_4)_2SO_4$. The sample was applied to a HiPrep 16/10 Phenyl FF hydrophobic interaction column equilibrated with 20 mM potassium phosphate pH 6.0, containing 1 M $(NH_4)_2SO_4$. Elution was carried out with a linear gradient of 20 to 0 mM potassium phosphate, pH 6.0, followed by 5 mM potassium

phosphate, pH 6.0 and water. Bound proteins were eluted with a linear gradient of 0 to 6 M Urea. Fractions were collected and the endoglucanase activity was analysed as described above. The protein containing endoglucanase activity was eluted in the beginning of the urea gradient.

The fractions were combined, equilibriated to 16 mM Tris-HCl pH 7.5 (I = 1.4 mS/cm) by 10DG column (Bio-Rad) and applied to a HiTrap DEAE FF anion exchange column equilibrated with 20 mM Tris-HCl, pH 7.5. Bound proteins were eluted with a linear gradient from 0 to 1 M NaCl in the equilibration buffer. Fractions were collected and analyzed for endoglucanase activity as described above. The protein was eluted in the range of 10–20 mS/cm.

The sample was equilibrated to 15 mM sodium acetate, pH 4.5 by 10DG column (Bio-Rad) and applied to a HiTrap SP XL cation exchange column equilibrated with 20 mM sodium acetate pH 4.5. Proteins were eluted with a linear gradient from 0 to 0.4 M sodium acetate, pH 4.5. Endoglucanase activity was eluted in the range of 1–10 mS/cm. The collected sample was lyophilized.

The sample was dissolved in water and applied to a Superdex 75 HR 10/30 gel filtration column equilibrated with 20 mM sodium phosphate pH 6.0, containing 0.15 M NaCl. Fractions were collected and those containing endoglucanase activity were combined. The protein eluate was judged to be pure by SDS-polyacrylamide gel electrophoresis and the molecular mass was evaluated on the basis of molecular mass standards (prestained SDS-PAGE standards, Broad Range, Bio-Rad) to be 54 kDa. The pI of the purified protein, designated as Ct EG_54 was determined with PhastSystem (Pharmacia) to be ca 5.5.

Example 5. Purification of an endoglucanase from *Thermoascus aurantiacus* ALKO4242

Thermoascus aurantiacus ALKO4242 was grown as described in Example 1. The pure endoglucanase was obtained by sequential purification with hydrophobic interaction and anion exchange chromatography followed by gel filtration. The endoglucanase activity of the fractions collected during purification was determined using carboxymethyl cellulose (CMC) as substrate (procedure of IUPAC 1987). Protein content was measured by BioRad Assay Kit (Bio-Rad Laboratories) using bovine serum albumine as standard.

The culture supernatant was applied to a HiPrep 16/10 Butyl hydrophobic interaction column equilibrated with 20 mM potassium phosphate buffer pH 6.0, containing 0.7 M $(NH_4)_2SO_4$. Bound proteins were eluted with 0.2 M $(NH_4)_2SO_4$ (I = 39 mS/cm). Fractions containing endoglucanase activity were combined and concentrated by ultrafiltration.

The sample was desalted in 10DG columns (Bio-Rad) and applied to a HiTrap DEAE FF anion exchange column equilibrated with 15 mM Tris-HCL, pH 7.0. Bound proteins were eluted with a linear gradient from 0 to 0.4 M NaCl in the equilibration buffer. The protein containing endoglucanase activity was eluted at the conductivity area of 15–21 mS/cm. Collected fractions were combined and concentrated as above.

The sample was applied to a Sephacryl S-100 HR 26/60 gel filtration column equilibrated with 50 mM sodium acetate buffer pH 5.0, containing 0.05 M NaCl. The protein fraction containing endoglucanase activity was eluted from the column with a retention volume corresponding to a molecular weight of 16 kDa. Collected fractions were combined, concentrated and gel filtration was repeated. The protein eluate was judged to be pure by SDS-polyacryl amide gel electrophoresis and the molecular weight was evaluated to be 28 kDa. The pI of the purified protein, designated as Ta EG_28, was determined in an IEF gel (PhastSystem, Pharmacia) to be about 3.5. The specific activity of Ta EG_28 at 50°C was determined to be 4290 nkat/mg (procedure of IUPAC 1987, using CMC as substrate).

Example 6. Purification and characterization of a β-glucosidase from *Acremonium thermophilum* ALKO4245

Acremonium thermophilum ALKO4245 was grown as described in Example 1. The pure β-glucosidase was obtained by sequential purification with hydrophobic interaction and anion exchange chromatography followed by gel filtration. The β-glucosidase activity of the fractions collected during purification was determined using 4-nitrophenyl-β-D-glucopyranoside as substrate (Bailey and Linko, 1990). Protein content was measured by BioRad Assay Kit (Bio-Rad Laboratories) using bovine serum albumine as standard.

The culture supernatant was applied to a HiPrep 16/10 Phenyl Sepharose FF hydrophobic interaction column equilibrated with 20 mM potassium phosphate pH 6.0, containing 1 M $(NH_4)_2SO_4$. Bound proteins were eluted with a linear gradient from the equilibration buffer to 5 mM potassium phosphate in

the conductivity area 137–16 mS/cm. Collected fractions were combined and concentrated by ultrafiltration.

The sample was desalted in 10DG columns (Bio-Rad) and applied to a HiTrap DEAE FF anion exchange column equilibrated with 10 mM potassium phosphate pH 7.0. Bound proteins were eluted with a linear gradient from the equilibration buffer to the same buffer containing 0.25 M NaCl in the conductivity area 1.5–12 mS/cm. Anion exchange chromatography was repeated as above, except that 4 mM potassium phosphate buffer pH 7.2 was used. Proteins were eluted at the conductivity area of 6–9 mS/cm. Fractions containing β-glucosidase activity were collected, combined, and concentrated.

The active material from the anion exchange chromatography was applied to a Sephacryl S-300 HR 26/60 column equilibrated with 20 mM sodium phosphate pH 6.5, containing 0.15 M NaCl. The protein with β-glucosidase activity was eluted with a retention volume corresponding to a molecular weight of 243 kDa. The protein was judged to be pure by SDS-polyacrylamide gel electrophoresis and the molecular weight was evaluated to be 101kDa. The pI of the purified protein, designated as At βG_101, was determined in an IEF gel (PhastSystem, Pharmacia) to be in the area of 5.6–4.9. The specific activity of At βG_101 at 50°C was determined to be 1100 nkat/mg (using 4-nitrophenyl-β-D-glucopyranoside as substrate, Bailey and Linko, 1990).

Thermal stability of the purified β-glucosidase was determined at different temperatures. The reaction was performed in the presence of 0.1 mg/ml BSA at pH 5.0 for 60 min using 4-nitrophenyl-β-D-glucopyranoside as substrate. *A. thermophilum* βG_101 was stable up to 70°C. The *Aspergillus* reference enzyme (Novozym 188) retained 100% of activity up to 60°.

Example 7. Purification of a β-glucosidase from *Chaetomium thermophilum* ALKO4261

Chaetomium thermophilum ALKO4261 was grown as described in Example 1. The pure β-glucosidase was obtained by sequential purification with hydrophobic interaction, anion and cation exchange chromatography followed by gel filtration. The β-glucosidase activity of the fractions collected during purification was determined using 4-nitrophenyl-β-D-glucopyranoside as substrate (Bailey and Linko, 1990).

The culture supernatant was applied to a HiPrep 16/10 Phenyl Sepharose FF hydrophobic interaction column equilibrated with 20 mM potassium phosphate pH 6.0, containing 0.8 M $(NH_4)_2SO_4$. The elution was carried out

with a linear gradient from the equilibration buffer to 3 mM potassium phosphate, pH 6.0, followed by elution with water and 6 M urea. The first fractions with β-glucosidase activity were eluted in the conductivity area of 80–30 mS/cm. The second β-glucosidase activity was eluted with 6 M urea. The active fractions eluted by urea were pooled and desalted in 10DG columns (Bio-Rad) equilibrated with 10 mM Tris-HCl pH 7.0.

After desalting, the sample was applied to a HiTrap DEAE FF anion exchange column equilibrated with 15 mM Tris-HCl pH 7.0. The protein did not bind to the column but was eluted during the sample feed. This flow-through fraction was desalted in 10DG columns (Bio-Rad) equilibrated with 7 mM Na acetate, pH 4.5.

The sample from the anion exchange chromatography was applied to a HiTrap SP FF cation exchange column equilibrated with 10 mM sodium acetate pH 4.5. Bound proteins were eluted with a linear gradient from 10 mM to 400 mM sodium acetate, pH 4.5. The fractions with β-glucosidase activity eluting in conductivity area of 6.5–12 mS/cm were collected, desalted in 10DG columns (Bio-Rad) equilibrated with 7 mM sodium acetate, pH 4.5 and lyophilized.

The lyophilized sample was diluted to 100 µl of water and applied to a Superdex 75 HF10/30 gel filtration column equilibrated with 20 mM sodium phosphate pH 4.5, containing 0.15 M NaCl. The β-glucosidase activity was eluted at a retention volume of 13.64 ml. Collected fractions were combined, lyophilized and dissolved in water. The protein was judged to be pure by SDS-polyacryl amide gel electrophoresis and the molecular weight was evaluated to be 76 kDa. The protein was designated as Ct βG_76.

Example 8. Purification and characterization of a β-glucosidase from *Thermoascus aurantiacus* ALKO4242

Thermoascus aurantiacus ALKO4242 was grown as described in Example 1. The pure β-glucosidase was obtained by sequential purification with hydrophobic interaction, anion and cation exchange chromatography followed by gel filtration. The β-glucosidase activity of the fractions collected during purification was determined using 4-nitrophenyl-β-D-glucopyranoside as substrate (Bailey and Linko, 1990). Protein content was measured by BioRad Assay Kit (Bio-Rad Laboratories) using bovine serum albumine as standard.

The culture supernatant was applied to a HiPrep 16/10 Phenyl Sepharose FF hydrophobic interaction column equilibrated with 20 mM potassium

phosphate pH 6.0, containing 0.7 M $(NH_4)_2SO_4$. Bound proteins were eluted with a linear gradient from 0.2 M $(NH_4)_2SO_4$ to 5 mM potassium phosphate, pH 6.0. The β-glucosidase activity was eluted during the gradient in the conductivity area of 28.0–1.1 mS/cm. Fractions were combined and concentrated by ultrafiltration.

The sample was desalted in 10DG columns (Bio-Rad) and applied to a HiTrap DEAE FF anion exchange column equilibrated with 20 mM Tris-HCl pH 7.0. The enzyme was eluted with a linear gradient from 0 to 0.2 M NaCl in the equilibration buffer and with delayed elution by 20 mM Tris-HCl, containing 0.4 M NaCl. The sample eluting in the conductivity area of ca. 10–30 mS/cm was concentrated by ultrafiltration and desalted by 10DG column (Bio-Rad).

The sample was applied to a HiTrap SP XL cation exchange column equilibrated with 9 mM sodium acetate pH 4.5. The enzyme was eluted with a linear gradient from 10 mM to 400 mM NaAc and by delayed elution using 400 mM NaAc pH 4.5 Proteins with β-glucosidase activity were eluted broadly during the linear gradient in the conductivity area of 5.0–11.3 mS/cm.

The active material from the cation exchange chromatography was applied to a Sephacryl S-300 HR 26/60 column equilibrated with 20 mM sodium phosphate pH 7.0, containing 0.15 M NaCl. The protein with β-glucosidase activity was eluted with a retention volume corresponding to a molecular weight of 294 kDa. Collected fractions were combined, lyophilized and dissolved in water. The protein was judged to be pure by SDS-polyacrylamide gel electrophoresis and the molecular weight was evaluated to be 81 kDa, representing most likely the monomeric form of the protein. Isoelectric focusing (IEF) was carried out using a 3–9 pI gel. After silver staining, a broad area above pI 5.85 was stained in addition to a narrow band corresponding to pI 4.55. The specific activity of the purified protein, designated as Ta βG_81, at 50°C was determined to be 600 nkat/mg using 4-nitrophenyl-β-D-glucopyranoside as substrate (Bailey and Linko, 1990).

Thermal stability of the purified β-glucosidase was determined at different temperatures. The reaction was performed in the presence of 0.1 mg/ml BSA at pH 5.0 for 60 min using 4-nitrophenyl-β-D-glucopyranoside as substrate. *T. aurantiacus* βG_81 was stable up to 75°C. The *Aspergillus* reference enzyme (Novozym 188) retained 100% of activity up to 60°C.

Example 9. Purification of a xylanase from *Acremonium thermophilum* ALKO4245

Acremonium thermophilum ALKO4245 was grown as described in Example 1. The culture supernatant was incubated at 70°C for 24 hours after which, it was concentrated by ultrafiltration. The pure xylanase was obtained by sequential purification with hydrophobic interaction and cation exchange chromatography followed by gel filtration. The xylanase activity was determined using birch xylan as substrate (procedure of IUPAC 1987). Protein was assayed by BioRad Protein Assay Kit (Bio-Rad Laboratories) using bovine serum albumin as standard.

The concentrated culture supernatant was applied to a HiPrep 16/10 Butyl FF hydrophobic interaction column equilibrated with 20 mM potassium phosphate buffer pH 6.0, containing 1 M $(NH_4)_2SO_4$. Bound proteins were eluted with the linear gradient from the above buffer to 5 mM potassium phosphate, pH 6.0. The protein fraction was eluted in a broad conductivity area of 120 to 15 mS/cm.

The sample from the hydrophobic interaction column was applied to a HiTrap SP XL cation exchange column equilibrated with 8 mM sodium acetate, pH 4.5. The protein did not bind to this column but was eluted in the flow-through during sample feed. This eluate was concentrated by ultrafiltration. The hydrophobic chromatography was repeated as described above. The unbound proteins were collected and freeze dried.

The dissolved sample was loaded onto the Superdex 75 HR10/30 gel filtration column equilibrated with 20 mM sodium phosphate buffer pH 7.0, containing 0.15 M NaCl. The protein eluted from the column with the retention volume of 11.2 ml was judged to be pure by SDS-polyacryl amide gel electrophoresis. The molecular mass of the purified protein was evaluated on the basis of molecular mass standards (prestained SDS-PAGE standards, Broad Range, Bio-Rad) to be 60 kDa. The specific activity of the protein, designated as At XYN_60, at 50°C was determined to be 1800 nkat/mg (procedure of IUPAC 1987, using birch xylan as substrate). The relative activity was increased about 1.2 fold at 60°C and 1.65 fold at 70°C (10 min, pH 5.0) as compared to 50°C. The specific activity against MUG2 (4-methylumbelliferyl-β-D-cellobioside), MUL (4-methylumbelliferyl-beta-D-lactoside) and MUG3 (4-methylumbelliferyl-β-D-cellotrioside) were 54, 33 and 78 nkat/mg (50°C pH 5.0 10 min), respectively. This is in agreement with the fact that the family 10

xylanases also show activity against the aryl glucopyranosides (Biely *et al.* 1997).

Example 10. Purification of a xylanase from *Thermoascus aurantiacus* ALKO4242

Thermoascus aurantiacus ALKO4242 was grown as described in Example 1. The pure xylanase was obtained by sequential purification with hydrophobic interaction, anion, and cation exchange chromatography followed by gel filtration. The xylanase activity was determined using birch xylan as substrate (procedure of IUPAC 1987). Protein was assayed by BioRad Protein Assay Kit (Bio-Rad Laboratories) using bovine serum albumin as standard.

The culture supernatant was applied to a HiPrep 16/10 Phenyl Sepharose FF hydrophobic interaction column equilibrated with 20 mM potassium phosphate buffer pH 6.0, containing 0.7 M $(NH_4)_2SO_4$. Bound proteins were eluted with a two-step elution protocol. The elution was carried out by dropping the salt concentration first to 0.2 M $(NH_4)_2SO_4$ and after that a linear gradient from 20 mM potassium phosphate pH 6.0, containing 0.2 M $(NH_4)_2SO_4$ to 5 mM potassium phosphate pH 6.0 was applied. The protein was eluted with 0.2 M $(NH_4)_2SO_4$ (I = 39 mS/cm).

The sample was desalted in 10DG columns (Bio-Rad) and applied to a HiTrap DEAE FF anion exchange column equilibrated with 15 mM Tris-HCL, pH 7.0. The protein did not bind to the anion exchange column but was eluted in the flow-through. The conductivity of the sample was adjusted to correspond that of 20 mM sodium acetate, pH 4.5 by adding water and pH was adjusted to 4.5 during concentration by ultrafiltration.

The sample was applied to a HiTrap SP XL cation exchange column equilibrated with 20 mM sodium acetate, pH 4.5. Bound proteins were eluted with a linear gradient from the equilibration buffer to the same buffer containing 1 M NaCl. The enzyme was eluted at the conductivity area of 1–7 mS/cm. The sample was lyophilized and thereafter dissolved in water.

The lyophilised sample was dissolved in water and applied to a Superdex 75 HR 10/30 gel filtration column equilibrated with 20 mM sodium phosphate pH 7.0, containing 0.15 M NaCl. The protein was eluted from the column with a retention volume corresponding to a molecular weight of 26 kDa. The protein was judged to be pure by SDS-polyacrylamide gel electrophoresis. The molecular mass of the pure protein was 30 kDa as evaluated on the basis of molecular mass standards (prestained SDS-PAGE standards, Broad Range,

Bio-Rad). The pI of the purified protein, designated as Ta XYN_30 was determined with PhastSystem (Pharmacia) to be ca. 6.8. The specific activity of Ta XYN_30 at 50°C was determined to be 4800 nkat/mg (procedure of IUPAC 1987, using birch xylan as substrate).

Example 11. Internal amino acid sequencing

The internal peptides were sequenced by electrospray ionization combined to tandem mass spectrometry (ESI-MS/MS) using the Q-TOF1 (Micromass) instrument. The protein was first alkylated and digested into tryptic peptides. Generated peptides were desalted and partially separated by nano liquid chromatography (reverse-phase) before applying to the Q-TOF1 instrument. The internal peptide sequences for *Chaetomium thermophilum* and *Acremonium thermophilum* cellobiohydrolases are shown in Table 2. The peptides from *Chaetomium* CBH were obtained after the corresponding *cbh* gene had been cloned. The peptides determined from *Acremonium* CBH were not utilized in the cloning of the corresponding gene.

Table 2. Internal peptide sequences determined from *Chaetomium thermophilum* ALKO4265 CBH (1_C – 4_C) and *Acremonium thermophilum* ALKO4245 CBH (1_A – 4_A).

Peptide	Sequence
Peptide 1_C	T P S T N D A N A G F G R
Peptide 2_C	V A F S N T D D F N R
Peptide 3_C	F S N T D D F N R K
Peptide 4_C	P G N S L/I T Q E Y C D A Q/K K
Peptide 1_A	V T Q F I/L T G
Peptide 2_A	M G D T S F Y G P G
Peptide 3_A	C D P D G C D F N
Peptide 4_A	S G N S L/I T T D F

I/L = leucine and isoleucine have the same molecular mass and cannot be distinguished in ESI-MS/MS analysis

Q/K = the molecular mass of glutamine and lysine differs only 0.036 Da and cannot be distinguished in ESI-MS/MS analysis

The internal peptide sequences of purified endoglucanases, β-glucosidases, and xylanases of *Acremonium thermophilum* ALKO4245, *Chaetomium thermophilum* ALKO4261 and *Thermoascus aurantiacus* ALKO4242 are listed in Table 3, Table 4 and Table 5.

Table 3. Internal peptide sequences determined from *Acremonium thermophilum* ALKO4245 EG_40, *Chaetomium thermophilum* ALKO4261 EG_54 and *Thermoascus aurantiacus* ALKO4242 EG_28 endoglucanases.

Protein	Peptide	Sequence[a]
At EG_40	Peptide 1	Q S C S S F P A P L K P G C Q W R
	Peptide 2	Y A L T F N S G P V A G K
	Peptide 3	V Q C P S E L T S R
	Peptide 4	N Q P V F S C S A D W Q R
	Peptide 5	Y W D C C K P S C G W P G K
	Peptide 6	P T F T
Ct EG_54	Peptide 1	E P E P E V T Y Y V
	Peptide 2	Y Y L L D Q T E Q Y
	Peptide 3	R Y C A C M D L W E A N S R
	Peptide 4	P G N T P E V H P Q/K
	Peptide 5	S I/L A P H P C N Q/K
	Peptide 6	Q Q Y E M F R
	Peptide 7	A L N D D F C R
	Peptide 8	W G N P P P R
Ta EG_28	Peptide 1	I/L T S A T Q W L R
	Peptide 2	G C A I/L S A T C V S S T I/L G Q E R
	Peptide 3	P F M M E R
	Peptide 4	Q Y A V V D P H N Y G R

(a I/L = leucine and isoleucine have the same molecular mass and cannot be distinguished in ESI-MS/MS analysis, Q/K = the molecular mass of glutamine and lysine differs only 0.036 Da and cannot be distinguished in ESI-MS/MS analysis.

Table 4. Internal peptide sequences determined from *Acremonium thermophilum* ALKO4245 βG_101, *Chaetomium thermophilum* ALKO4261 βG_76 and *Thermoascus aurantiacus* ALKO4242 βG_81 beta-glucosidases.

Protein	Peptide	Sequence[a]
At βG_101	Peptide 1	S P F T W G P T R
	Peptide 2	V V V G D D A G N P C
	Peptide 3	A F V S Q L T L L E K
	Peptide 4	G T D V L/I Y T P N N K
	Peptide 5	Q P N P A G P N A C V L/I R
Ct βG_76	Peptide 1	E G L F I D Y R
	Peptide 2	P G Q S G T A T F R
	Peptide 3	E T M S S N V D D R
	Peptide 4	I A L V G S A A V V
	Peptide 5	M W L C E N D R
	Peptide 6	Y P Q L C L Q D G P L G I R
	Peptide 7	E L N G Q N S G Y P S I
Ta βG_81	Peptide 1	T P F T W G K
	Peptide 2	L C L Q D S L P G V R
	Peptide 3	G V D V Q L G P V A G V A P R
	Peptide 4	V N L T L E
	Peptide 5	F T G V F G E D V V G
	Peptide 6	N D L P L T G Y E K

(a I/L = leucine and isoleucine have the same molecular mass and cannot be distinguished in ESI-MS/MS analysis

Table 5. Internal peptide sequences determined from *Acremonium thermophilum* ALKO4245 XYN_60 and *Thermoascus aurantiacus* ALKO4242 XYN_30 xylanases.

Protein	Peptide	Sequence
At XYN_60	Peptide 1	Y N D Y N L E Y N Q K
	Peptide 2	F G Q V T P E N
	Peptide 3	V D G D A T Y M S Y V N N K
	Peptide 4	K P A W T S V S S V L A A K
	Peptide 5	S Q G D I V P R A K
Ta XYN_30	Peptide 1	V Y F G V A T D Q N R
	Peptide 2	N A A I I Q A D F G Q V T P E N S M K
	Peptide 3	G H T L V W H S Q L P S W V S S I T D K
	Peptide 4	N H I T T L M T R
	Peptide 5	A W D V V N E A F N E D G S L R
	Peptide 6	L Y I N D Y N L D S A S Y P K
	Peptide 7	A S T T P L L F D G N F N P K P A Y N A I V Q D L Q Q
	Peptide 8	Q T V F L N V I G E D Y I P I A F Q T A R

Example 12. Construction of genomic libraries for *Thermoascus aurantiacus*, *Chaetomium thermophilum* and *Acremonium thermophilum*

The genomic library of *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 were made to Lambda DASH®II vector (Stratagene, USA) according to the instructions from the supplier. The chromosomal DNAs, isolated by the method of Raeder and Broda (1985), were partially digested with *Sau*3A. The digested DNAs were size-fractionated and the fragments of the chosen size (≈ 5–23 kb) were dephosphorylated and ligated to the *Bam*HI digested lambda vector arms. The ligation mixtures were packaged using Gigapack III Gold packaging extracts according to the manufacturer's instructions (Stratagene, USA). The titers of the *Chaetomium thermophilum* and *Acremonium thermophilum* genomic libraries were 3.6 x 10^6 pfu/ml and 3.7 x 10^5 pfu/ml and those of the amplified libraries were 6.5 x 10^{10} pfu/ml and 4.2 x 10^8 pfu/ml, respectively.

Lambda FIX® II/Xho I Partial Fill-In Vector Kit (Stratagene, USA) was used in the construction of the genomic libraries for *Thermoascus aurantiacus* ALKO4242 and *Chaetomium thermophilum* ALKO4261 according to the instructions from the supplier. The chromosomal DNAs, isolated by the method of Raeder and Broda (1985), were partially digested with *Sau*3A. The digested DNAs were size-fractionated and the fragments of the chosen size (≈ 6–23 kb) were filled-in and ligated to the *Xho*I digested Lambda FIX II vector arms. The ligation mixtures were packaged using Gigapack III Gold packaging extracts according to the manufacturer´s instructions (Stratagene, USA). The titers of the *Thermoascus aurantiacus* ALKO4242 and *Chaetomium thermophilum* ALKO4261 genomic libraries were 0.2×10^6 and 0.3×10^6 pfu/ml and those of the amplified libraries were 1.8×10^9 and 3.8×10^9 pfu/ml, respectively.

Example 13. Cloning of the cellobiohydrolase (*cbh/cel7*) genes from *Thermoascus aurantiacus*, *Chaetomium thermophilum* and *Acremonium thermophilum*

Standard molecular biology methods were used in the isolation and enzyme treatments of DNA (plasmids, DNA fragments), in *E. coli* transformations, etc. The basic methods used are described in the standard molecular biology handbooks, e.g., Sambrook *et al*. (1989) and Sambrook and Russell (2001).

The probes for screening the genomic libraries which were constructed as described in Example 12 were amplified by PCR using the *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 genomic DNAs as templates in the reactions. Several primers tested in PCR reactions were designed according to the published nucleotide sequence (WO 03/000941, Hong *et al*., 2003b). The PCR reaction mixtures contained 50 mM Tris-HCl, pH 9.0, 15 mM $(NH_4)_2SO_4$, 0.1% Triton X-100™, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 5 µM each primer and 1 units of Dynazyme EXT DNA polymerase (Finnzymes, Finland) and ≈ 0.5–1 µg of the genomic DNA. The conditions for the PCR reactions were the following: 5 min initial denaturation at 95°C, followed by 30 cycles of 1 min at 95°C, either 1 min annealing at 62°C (±8°C gradient) for *Thermoascus* ALKO4242 and *Chaetomium* ALKO4265 templates or 1 min annealing at 58°C (±6°C gradient) for *Acremonium* ALKO4245 template, 2 min extension at 72°C and a final extension at 72°C for 10 min.

Date Reçue/Date Received 2021-02-11

DNA products of the expected sizes (calculated from published *cbh* sequences) were obtained from all genomic templates used. The DNA fragments of the expected sizes were isolated from the most specific PCR reactions and they were cloned to pCR® Blunt-TOPO® vector (Invitrogen, USA). The inserts were characterized by sequencing and by performing Southern blot hybridizations to the genomic DNAs digested with several restriction enzymes. The PCR fragments, which were chosen to be used as probes for screening of the *Thermoascus aurantiacus*, *Chaetomium thermophilum* and *Acremonium thermophilum* genomic libraries are presented in Table 6.

Table 6. The primers used in the PCR reactions and probes chosen for screening of the *cbh/cel7* genes from *Thermoascus aurantiacus*, *Chaetomium thermophilum* and *Acremonium thermophilum* genomic libraries. The genomic template DNA and the name of the plasmid containing the probe fragment are shown.

Gene	Forward primer	Reverse primer	Template DNA	Fragment (kb)	Plasmid
Ta *cbh*	TCEL11 atgcgaactggcgttgggtcc	TCEL12 gaatttggagctagtgtcgacg	*Thermoascus* ALKO4242	0.8 kb	pALK1633
Ct *cbh*	TCEL7 cgatgccaactggcgctggac	TCEL8 ttcttggtggtgtcgacggtc	*Chaetomium* ALKO4265	0.8 kb	pALK1632
At *cbh*	TCEL13 agctcgaccaactgctacacg	TCEL4 accgtgaacttcttgctggtg	*Acremonium* ALKO4245	0.7 kb	pALK1634

The deduced amino acid sequences from all these probes had homology to several published CBH sequences (BLAST program, version 2.2.9 at NCBI, National Center for Biotechnology Information; Altschul *et al.*, 1990) of glycoside hydrolase family 7 (Henrissat, 1991; Henrissat and Bairoch, 1993).

The inserts from the plasmids listed in Table 6 were labeled with digoxigenin according to the supplier´s instructions (Roche, Germany), and the amplified genomic libraries ($2 \times 10^5 - 3 \times 10^5$ plaques) were screened with the labeled probe fragments. The hybridization temperature for the filters was 68°C and the filters were washed 2 x 5 min at RT using 2 x SSC – 0.1% SDS

followed by 2 x 15 min at 68°C using 0.1 x SSC – 0.1% SDS with the homologous probes used. Several positive plaques were obtained from each of the hybridizations. In screening of the *Acremonium* ALKO4245 genomic libraries, some of the positive plaques were strongly hybridizing to the probe in question but, in addition, there was an amount of plaques hybridizing more weakly to the probes. This suggested that other cellobiohydrolase gene(s) might be present in the genome, causing cross-reaction. From four to five strongly hybridizing plaques were purified from *Thermoascus* ALKO4242 and *Chaetomium* ALKO4265 genomic library screenings. In the case of the *Acremonium thermophilum* ALKO4245, four out of six purified plaques hybridized weakly by the probe used. The phage DNAs were isolated and characterized by Southern blot hybridizations. The chosen restriction fragments hybridizing to the probe were subcloned to pBluescript II KS+ vector and the relevant regions of the clones were sequenced.

In total four *cbh/cel7* genes were cloned; one from *Thermoascus aurantiacus* ALKO4242, one from *Chaetomium thermophilum* ALKO4265 and two from *Acremonium thermophilum* ALKO4245 (at the early phase of the work, these had the codes At_cbh_C and At_cbh_A, and were then designated as At *cel7*A and At *cel7*B, respectively). Table 7 summarizes the information on the probes used for screening the genes, the phage clones from which the genes were isolated, the chosen restriction fragments containing the full-length genes with their promoter and terminator regions, the plasmid names, and the DSM deposit numbers for the *E. coli* strains carrying these plasmids.

Table 7. The probes used for cloning of *cbh/cel7* genes, the phage clone and the subclones chosen, the plasmid number and the number of the deposit of the corresponding *E. coli* strain.

Gene	Probe used in screening	Phage clone	The fragment subcloned to pBluescript II	Plasmid no	*E. coli* deposit no
Ta *cel7*A	pALK1633	F12	3.2 kb *Xba*I	pALK1635	DSM 16723
Ct *cel7*A	pALK1632	F36	2.3 kb *Pvu*I - *Hin*dIII	pALK1642	DSM 16727
At *cel7*B	pALK1634	F6	3.1 kb *Eco*RI	pALK1646	DSM 16728
At *cel7*A	pALK1634	F2	3.4 kb *Xho*I	pALK1861	DSM 16729

The relevant information on the genes and the deduced protein sequences (SEQ ID NO: 1–8) are summarized in Table 8 and Table 9, respectively.

The peptide sequences of the purified CBH proteins from *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 (Table 2) were found from the deduced amino acid sequences of the clones containing the Ct *cel7A* and At *cel7A* genes. Thus, it could be concluded that the genes encoding the purified CBH/Cel7 proteins from *Chaetomium thermophilum* and *Acremonium thermophilum* were cloned.

Table 8. Summary on the *cbh/cel7* genes isolated from *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245.

Cbh gene	Length with introns (bp) [a]	Coding region (bp) [b]	No of introns	Lengths of introns (bp)	SEQ ID NO:
Ta *cel7*A	1439	1371	1	65	1
Ct *cel7*A	1663	1596	1	64	7
At *cel7*B	1722	1377	3	134, 122, 87	3
At *cel7*A	1853	1569	4	88, 53, 54, 86	5

(a The STOP codon is included.

(b The STOP codon is not included.

Table 9. Summary of amino acid sequences deduced from the *cbh/cel7* gene sequences from *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALK4265 and *Acremonium thermophilum* ALKO4245. ss, signal sequence.

CBH protein	No of aas	Length of ss NN/HMM[a]	C-terminal CBD[b]	Predicted MW (Da, ss not incl)[c]	Predicted pI (ss not incl)	Putative N-glycosylation sites[d]	SEQ ID NO:
Ta Cel7A	457	17/17	NO	46 873	4.44	2	2
Ct Cel7A	532	18/18	YES, T497 to L532	54 564	5.05	3	8
At Cel7B	459	21/21	NO	47 073	4.83	2	4
At Cel7A	523	17/17	YES, Q488 to L523	53 696	4.67	4	6

(a The prediction on the signal sequence was made using the program SignalP V3.0 (Nielsen *et al.*, 1997; Bendtsen *et al.*, 2004); the NN value was obtained using neural networks and HMM value using hidden Markov models.

(b The cellulose-binding domain (CBD), the amino acids of the C-terminal CBD region are indicated (M1 (Met #1) included in numbering)

(c The predicted signal sequence was not included. The prediction was made using the Compute pI/MW tool at ExPASy server (Gasteiger *et al.*, 2003).

(d The number of sequences N-X-S/T.

The deduced amino acid sequences of *Thermoascus aurantiacus* Cel7A and *Acremonium thermophilum* Cel7A (core, without the CBD) were most homologous to each other (analyzed by Needleman-Wunsch global alignment, EMBOSS 3.0.0 Needle, with Matrix EBLOSUM62, Gap Penalty 10.0 and Extend Penalty 0.5; Needleman and Wunsch, 1970). In addition, the deduced *Acremonium thermophilum* Cel7A had a lower identity to the deduced *Chaetomium thermophilum* Cel7A. The *Acremonium thermophilum* Cel7B was most distinct from the CBH/Cel7 sequences of the invention.

The deduced *Chaetomium* Cel7A sequence possessed the highest identities (analyzed by Needleman-Wunsch global alignment, EMBOSS Needle, see above) to polypeptides of *Chaetomium thermophilum*, *Scytalidium thermophilum* and *Thielavia australiensis* CBHI described in WO 03/000941. Similarly, the deduced *Thermoascus aurantiacus* Cel7A sequence was highly identical to the published CBHI of the *Thermoascus aurantiacus* (WO 03/000941, Hong *et al.*, 2003b). *Acremonium thermophilum* Cel7B had significantly lower identities to the previously published sequences, being more closely related to the CBHI polypeptide from *Oryza sativa*. The highest homologies of the deduced *Acremonium thermophilum* Cel7A sequence were to *Exidia gladulosa* and *Acremonium thermophilum* CBHI polynucleotides (WO 03/000941). The alignment indicates that the cloned *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 sequences encode the CBH proteins having high homology to the polypeptides of the glycoside hydrolase family 7, therefore these were designated as Cel7A or Cel7B (Henrissat *et al.* 1998).

The comparison of the deduced amino acid sequences of the *cbh/cel7* genes from *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245 *Thielavia* to each other, and further to the sequences found from the databases, are shown in Table 10.

Table 10. The highest homology sequences to the deduced amino acid sequences of the *cbh/cel7* genes from *Thermoascus aurantiacus* ALKO4242, *Chaetomium thermophilum* ALKO4265 and *Acremonium thermophilum* ALKO4245. The alignment was made using Needleman-Wunsch global alignment (EMBLOSUM62, Gap penalty 10.0, Extend penalty 0.5). *indicates an amino acid sequence derived from one of the cellobiohydrolase genes cloned in this work. 'Core' indicates alignment without the CBD.

Organism, enzyme and accession number	Identity, (%)
* ***Thermoascus aurantiacus* Cel7A**	100.0
Thermoascus aurantiacus, AY840982	99.6
Thermoascus aurantiacus, AX657575	99.1
Thermoascus aurantiacus, AF421954	97.8
Talaromyces emersonii, AY081766	79.5
Chaetomidium pingtungium, AX657623	76.4
Trichophaea saccata, AX657607	73.4
* *Acremonium thermophilum* Cel7A (core)	70.6
Emericella nidulans, AF420020 (core)	70.4
* *Chaetomium thermophilum* Cel7A (core)	66.4
* ***Chaetomium thermophilum* Cel7A**	100.0
Chaetomium thermophilum, AY861347	91.9
Chaetomium thermophilum, AX657571	91.7
Scytalidium thermophilum, AX657627	74.7
Thielavia australiensis, AX657577	74.6
Acremonium thermophilum, AX657569	72.3
Exidia glandulosa, AX657613	68.0
* *Acremonium thermophilum* Cel7A	66.9
* *Thermoascus aurantiacus* Cel7A (core)	66.4
Exidia glandulosa, AX657615	60.8
Chaetomium pingtungium, AX657623	60.7
* *Acremonium thermophilum* Cel7B (core)	60.2
* ***Acremonium thermophilum* Cel7B**	100.0
Oryza sativa, AK108948	66.1
Exidia glandulosa, AX657615	65.0
Acremonium thermophilum, AX657569 (core)	64.8

Thermoascus aurantiacus, AX657575	64.8
* *Acremonium thermophilum* Cel7A	64.6
* *Thermoascus aurantiacus* Cel7A	64.4
Trichophaea saccata, AX657607	63.6
* *Chaetomium thermophilum* Cel7A (core)	60.2
**** Acremonium thermophilum* Cel7A**	100.0
Exidia glandulosa, AX657613	77.9
Exidia glandulosa, AX657615	77.9
Acremonium thermophilum, AX657569	77.5
Thielavia australiensis, AX657577	71.0
* *Thermoascus aurantiacus* Cel7A (core)	70.6
Scytalidium thermophilum, AX657627	67.5
Chaetomium thermophilum, AX657571	67.5
Chaetomium pingtungium, AX657623	67.3
* *Chaetomium thermophilum* Cel7A	66.9
* *Acremonium thermophilum* Cel7B (core)	64.6

Example 14. Production of recombinant CBH/Cel7 proteins in *Trichoderma reesei*

Expression plasmids were constructed for production of the recombinant CBH /Cel7 proteins from *Thermoascus aurantiacus* (Ta Cel7A), *Chaetomium thermophilum* (Ct Cel7A) and *Acremonium thermophilum* (At Cel7A, At Cel7B; at early phase of the work these proteins had the temporary codes At CBH_C and At CBH_A, respectively). The expression plasmids constructed are listed in Table 11. The recombinant *cbh/cel7* genes, including their own signal sequences, were exactly fused to the *T. reesei cbh1* (*cel7A*) promoter by PCR. The transcription termination was ensured by the *T. reesei cel7A* terminator and the *A. nidulans amdS* marker gene was used for selection of the transformants as described in Paloheimo *et al.* (2003). The linear expression cassettes (Fig. 2), were isolated from the vector backbones after *Eco*RI digestion and were transformed into *T. reesei* A96 and A98 protoplasts (both strains have the genes encoding the four major cellulases CBHI/Cel7A, CBHII/Cel6A, EGI/Cel7B and EGII/Cel5A deleted). The transformations were performed as in Penttilä *et al.* (1987) with the modifications described in Karhunen *et al.* (1993), selecting with acetamide as a sole nitrogen source. The transformants were

purified on selection plates through single conidia prior to sporulating them on PD.

Table 11. The expression cassettes constructed to produce CBH/Cel7 proteins of *Thermoascus aurantiacus* ALKO4242 (Ta Cel7A), *Chaetomium thermophilum* ALKO4265 (Ct Cel7A), and *Acremonium thermophilum* ALKO4245 (At Cel7A, At Cel7B) in *Trichoderma reesei*. The overall structure of the expression cassettes was as described in Fig. 2. The cloned *cbh/cel7* genes were exactly fused to the *T. reesei cbh1/cel7A* promoter.

CBH/Cel7	Expression plasmid	Size of the expr. cassette [a]	*cel7*A terminator [b]
Ta Cel7A	pALK1851	9.0 kb	245 bp (*Xba*I)
Ct Cel7A	pALK1857	9.2 kb	240 bp (*Hin*dIII)
At Cel7B	pALK1860	9.4 kb	361 bp (*Eco*RI)
At Cel7A	pALK1865	9.5 kb	427 bp (*Eco*RV)

(a The expression cassette for T. *reesei* transformation was isolated from the vector backbone by using *Eco*RI digestion.

(b The number of the nucleotides from the genomic *cbh1/cel7A* terminator region after the STOP codon. The restriction site at the 3'-end, used in excising the genomic gene fragment, is included in the parenthesis.

The CBH/Cel7 production of the transformants was analysed from the culture supernatants of the shake flask cultivations (50 ml). The transformants were grown for 7 days at 28°C in a complex lactose-based cellulase-inducing medium (Joutsjoki *et al.* 1993) buffered with 5% KH_2PO_4. The cellobiohydrolase activity was assayed using 4-methylumbelliferyl-β-D-lactoside (MUL) substrate according to van Tilbeurgh *et al.*, 1988. The genotypes of the chosen transformants were confirmed by using Southern blots in which several genomic digests were included and the respective expression cassette was used as a probe. Heterologous expression of the Ta Cel7A, Ct Cel7A, At Cel7A and At Cel7B proteins was analyzed by SDS-PAGE with subsequent Coomassive staining. The findings that no cellobiohydrolase activity or heterologous protein production in SDS-PAGE could be detected for the At Cel7B transformants containing integrated expression cassette, suggest that

At Cel7B is produced below detection levels in *Trichoderma* using the described experimental design.

The recombinant CBH/Cel7 enzyme preparations were characterized in terms of pH optimum and thermal stability. The pH optimum of the recombinant CBH/Cel7 proteins from *Thermoascus aurantiacus*, *Chaetomium thermophilum,* and *Acremonium thermophilum* were determined in the universal McIlvaine buffer within a pH range of 3.0–8.0 using 4-methylumbelliferyl-β-D-lactoside (MUL) as a substrate (Fig 3 A). The pH optimum for Ct Cel7A and At Cel7A enzymes is at 5.5, above which the activity starts to gradually drop. The pH optimum of the recombinant crude Ta Cel7A is at 5.0 (Fig 3 A). Thermal stability of the recombinant Cel7 enzymes was determined by measuring the MUL activity in universal McIlvaine buffer at the optimum pH with reaction time of 1 h. As shown from the results Ta Cel7A and Ct Cel7A retained more than 60% of their activities at 70°C, whereas At Cel7A showed to be clearly less stable at the higher temperatures (≥65°C) (Fig 3 B).

The chosen CBH/Cel7 transformants were cultivated in lab bioreactors at 28°C in the medium indicated above for 3–4 days with pH control 4.4 ± 0.2 (NH_3/H_3PO_4) to obtain material for the application tests. The supernatants were recovered by centrifugation and filtering through Seitz-K 150 and EK filters (Pall SeitzSchenk Filtersystems GmbH, Bad Kreuznach, Germany).

Example 15. Production of the recombinant *Thermoascus aurantiacus* Cel7A+CBD fusion proteins in *T. reesei*

Thermoascus aurantiacus Cel7A (AF478686, Hong *et al.*, 2003b; SEQ ID. NO: 1) was fused to linker and CBD of *Trichoderma reesei* CBHI/Cel7A (AR088330, Srisodsuk *et al.* 1993) (= Tr CBD) followed by the production of the fusion protein (SEQ ID NO: 28 corresponding nucleic acid SEQ ID. NO: 27) in the *T. reesei* as was described in FI20055205/US 11/119,526; filed April 29, 2005. In addition, *Thermoascus aurantiacus* Cel7A was fused to linker and CBD of *Chaetomium thermophilum* Cel7A (SEQ ID. NO: 7) (Ct CBD). For that purpose, the coding sequence of the linker and the CBD of *Chaetomium thermophilum* Cel7A were synthesized by PCR using following primers:

5'-TTAAACATATGTTATCTACTCCAACATCAAGGTCGGACCCATCGGCTC-GACCGTCCCTGGCCTTGAC-3' (forward sequence)

and

5'-TATATGCGGCCGCAAGCTTTACCATCAAGTTACTCCAGCAAATCAGGG-AACTG-3' (reverse sequence).

The PCR reaction mixture contained 1x DyNAzyme™ EXT reaction buffer (Finnzymes, Finland), 15 mM Mg^{2+}, 0.2 mM dNTPs, 2 µM of each primer, 0.6 units of DyNAzyme™ EXT DNA polymerase (Finnzymes, Finland), and approximately 75 ng / 30 µl of template DNA, containing full-length *cel7A* gene from the *Chaetomium thermophilum*. The conditions for the PCR reaction were the following: 2 min initial denaturation at 98°C, followed by 30 cycles of 30 sec at 98°C, 30 sec annealing at 68°C (±4°C gradient), 30 sec extension at 72°C and a final extension at 72°C for 10 min. The specific DNA fragment in PCR reaction was obtained at annealing temperature range from 64°C to 68.5°C. The synthesized CBD fragment of the *Chaetomium thermophilum* was ligated after *Thermoascus aurantiacus cel7A* gene resulting in a junction point of GPIGST between the domains. The PCR amplified fragment in the plasmid was confirmed by sequencing (SEQ ID. NO: 29). The constructed fusion *cel7A* gene was exactly fused to the *T. reesei cbh1 (cel7A)* promoter. The transcription termination was ensured by the *T. reesei cel7*A terminator and the *A. nidulans amdS* marker gene was used for selection of the transformants as described in Paloheimo *et al.* (2003).

The linear expression cassette was isolated from the vector backbone after *Not*I digestion and was transformed to *T. reesei* A96 protoplasts. The transformations were performed as in Penttilä *et al.* (1987) with the modifications described in Karhunen *et al.* (1993), selecting with acetamide as a sole nitrogen source. The transformants were purified on selection plates through single conidia prior to sporulating them on PD.

Thermoascus aurantiacus Cel7A+CBD (SEQ ID. NO: 28 and 30) production of the transformants was analyzed from the culture supernatants of the shake flask cultivations (50 ml). The transformants were grown for 7 days in a complex cellulase-inducing medium (Joutsjoki *et al.* 1993) buffered with 5% KH_2PO_4 at pH 5.5. The cellobiohydrolase activity was assayed using 4-methylumbelliferyl-β-D-lactoside (MUL) substrate according to van Tilbeurgh *et al.*, 1988. The genotypes of the chosen transformants were confirmed by using Southern blots in which several genomic digests were included and the expression cassette was used as a probe. The SDS-PAGE analyses showed that

the recombinant *Thermoascus aurantiacus* Cel7A+CBD enzymes were produced as stable fusion proteins in *T. reesei*.

The chosen transformant producing the Ta Cel7A+Tr CBD fusion protein (SEQ ID. NO: 28) was also cultivated in 2 litre bioreactor at 28°C in the medium indicated above for 3-4 days with pH control 4.4±0.2 (NH_3/H_3PO_4) to obtain material for the application tests. The supernatants were recovered by centrifugation and filtering through Seitz-K 150 and EK filters (Pall SeitzSchenk Filtersystems GmbH, Bad Kreuznach, Germany).

Example 16. Comparison of the Michaelis-Menten and cellobiose inhibition constants of purified recombinant cellobiohydrolases

The Michaelis-Menten and cellobiose inhibition constants were determined from the cellobiohydrolases produced heterologously in *T. reesei* (Examples 14 and 15). The enzymes were purified as described in Example 2. Protein concentrations of purified enzymes were measured by their absorption at 280 nm using a theoretical molar extinction co-efficient, which were calculated from the amino acid sequences (Gill and von Hippel, 1989).

Kinetic constants (Km and kcat values) and cellobiose inhibition constant (Ki) for Tr CBHI/Cel7A, Ta CBH/Cel7A, At CBH/Cel7A and Ct CBH/Cel7A, were measured using CNPLac (2-Chloro-4-nitrophenyl-β-D-lactoside) as substrate at ambient temperature (22°C) in 50 mM sodium phosphate buffer, pH 5.7. For the determination of the inhibition constant (Ki), eight different substrate concentrations (31–4000 µM) in the presence of a range of five inhibitor concentrations (0–100 µM or 0–400 µM), which bracket the Ki value, were used. All experiments were performed in microtiter plates and the total reaction volume was 200 µl. The initial rates were in each case measured by continuous monitoring the release of the chloro-nitrophenolate anion (CNP, 2-Chloro-4-nitrophenolate) through measurements at 405 nm using Varioscan (Thermolabsystems) microtiter plate reader. The results were calculated from CNP standard curve (from 0 to 100 µM). Enzyme concentrations used were: Tr CBHI/Cel7A 2.46 µM, Ta CBH/Cel7A 1.58 µM, Ct CBH/Cel7A 0.79 µM and At CBH/Cel7A 3 µM. The Km and kcat constants were calculated from the fitting of the Michaelis-Menten equation using the programme of Origin. Lineweaver-Burk plots, replots (LWB slope versus [Glc2; cellobiose]) and Hanes plots were used to distinguish between competitive and mixed type inhibition and to determine the inhibition constants (Ki).

The results from the kinetic measurements are shown in Table 12 and Table 13. As can be seen, Ct CBH/Cel7A has clearly the higher turnover number (kcat) on CNPLac and also the specificity constant (kcat/Km) is higher as compared to CBHI/Cel7A of *T. reesei*. Cellobiose (Glc2) is a competitive inhibitor for all the measured cellulases, and the Tr CBHI/Cel7A (used as a control) has the strongest inhibition (i.e. the lowest Ki value) by cellobiose. The At CBH/Cel7A had over 7-fold higher inhibition constant as compared to that of Tr CBHI/Cel7A. These results indicate that all three novel cellobiohydrolases could work better on cellulose hydrolysis due to decreased cellobiose inhibition as compared to *Trichoderma reesei* Cel7A cellobiohydrolase I.

Table 12. Comparison of the cellobiose inhibition constants of four GH family 7 cellobiohydrolases, measured on CNPLac in 50 mM sodium phosphate buffer pH 5.7, at 22°C.

Enzyme	Ki (µM)	Type of inhibition
Ct Cel7A	39	competitive
Ta Cel7A	107	competitive
At Cel7A	141	competitive
Tr Cel7A	19	competitive

Table 13. Comparison of the Michaelis-Menten kinetic constants of *Chaetomium thermophilum* cellobiohydrolase Cel7A to CBHI/Cel7A of *T. reesei*, measured on CNPLac in 50 mM sodium phosphate buffer pH 5.7, at 22°C.

Enzyme	kcat (min^{-1})	Km (µM)	kcat /Km (min^{-1} M^{-1})
Ct Cel7A	18.8	1960	9.5 103
Tr Cel7A	2.6	520	5.0 103

Example 17. Hydrolysis of crystalline cellulose (Avicel) by the recombinant cellobiohydrolases

The purified recombinant cellobiohydrolases Ct Cel7A, Ta Cel7A, Ta Cel7A+Tr CBD, Ta Cel7A+ Ct CBD, At Cel7A as well as the core version of Ct Cel7A (see below) were tested in equimolar amounts in crystalline cellulose hydrolysis at two temperatures, 45°C and 70°C; the purified *T. reesei* Tr Cel7A

and its core version (see below) were used as comparison. The crystalline cellulose (Ph 101, Avicel; Fluka, Bucsh, Switzerland) hydrolysis assays were performed in 1.5 ml tube scale 50 mM sodium acetate, pH 5.0. Avicel was shaken at 45°C or at 70°C, with the enzyme solution (1.4 µM), and the final volume of the reaction mixture was 325 µl. The hydrolysis was followed up to 24 hours taking samples at six different time points and stopping the reaction by adding 163 µl of stop reagent containing 9 vol of 94% ethanol and 1 vol of 1 M glycine (pH 11). The solution was filtered through a Millex GV13 0.22 µm filtration unit (Millipore, Billerica, MA, USA). The formation of soluble reducing sugars in the supernatant was determined by para-hydroxybenzoic-acidhydrazide (PAHBAH) method (Lever, 1972) using a cellobiose standard curve (50 to 1600 µM cellobiose). A freshly made 0.1 M PAHBAH (Sigma-Aldrich, St. Louis, MO, USA) in 0.5 M NaOH (100 µl) solution was added to 150 µl of the filtered sample and boiled for 10 minutes after which the solution was cooled on ice. The absorbance of the samples at 405 nm was measured.

The core versions of the cellobiohydrolases harboring a CBD in their native form were obtained as follows: Ct Cel7A and Tr Cel7A were exposed to proteolytic digestion to remove the cellulose-binding domain. Papain (Papaya Latex, 14 U/mg, Sigma) digestion of the native cellobiohydrolases was performed at 37°C for 24 h in a reaction mixture composed of 10 mM L-cystein and 2 mM EDTA in 50 mM sodium acetate buffer (pH 5.0) with addition of papain (two papain concentrations were tested: of one fifth or one tenth amount of papain of the total amount of the Cel7A in the reaction mixture). The resultant core protein was purified with DEAE Sepharose FF (Pharmacia, Uppsala, Sweden) anion exchange column as described above. The product was analysed in SDS-PAGE.

The hydrolysis results at 45°C and 70°C are shown in Figure 4 and Figure 5, respectively. The results show clearly that all the cellobiohydrolases show faster and more complete hydrolysis at both temperatures as compared to the state-of-art cellobiohydrolase *T. reesei* Cel7A. At 70°C the thermostable cellobiohydrolases from *Thermoascus aurantiacus* ALKO4242 and *Chaetomium thermophilum* ALKO4265 are superior as compared to the *T. reesei* Cel7A, also in the case where the *Thermoascus* Cel7A core is linked to the CBD of *T. reesei* Cel7A (Ta Cel7A + Tr CBD). It was surprising that the cellobiohydrolases isolated and cloned in this work are superior, when harboring a CBD, in the rate and product formation in crystalline cellulose hydrolysis also at the con-

ventional hydrolysis temperature of 45°C when compared to the state-of-art cellobiohydrolase *T. reesei* Cel7A (CBHI) at the same enzyme concentration. The results are also in agreement with those enzyme preparations (At Cel7A and Ct Cel7A), which were purified from the original hosts and tested in Avicel hydrolysis (50°C, 24 h) (Example 2, Table 1).

Example 18. Cloning of *Acremonium thermophilum* ALKO4245, *Chaetomium thermophilum* ALKO4261, and *Thermoascus aurantiacus* ALKO4242 endoglucanase genes

Standard molecular biology methods were used as described in Example 13. The construction of the *Acremonium*, *Chaetomium*, and *Thermoascus* genomic libraries has been described in Example 12.

The peptides derived from the purified *Acremonium* and *Chaetomium* endoglucanases shared homology with several endoglucanases of glycosyl hydrolase family 45 such as *Melanocarpus albomyces* Cel45A endoglucanase (AJ515703) and *Humicola insolens* endoglucanase (A35275), respectively. Peptides derived from the *Thermoascus* endoglucanase shared almost 100% identity with the published *Thermoascus aurantiacus* EG1 endoglucanase sequence (AF487830). To amplify a probe for screening of the *Acremonium* and *Chaetomium* genomic libraries, degenerate primers were designed on the basis of the peptide sequences. The order of the peptides in the protein sequence and the corresponding sense or anti-sense nature of the primers was deduced from the comparison with the homologous published endoglucanases. Primer sequences and the corresponding peptides are listed in Table 14. Due to almost 100% identity of the *Thermoascus* peptides with the published sequence, the endoglucanase gene was amplified by PCR directly from the genomic DNA.

Table 14. Oligonucleotides synthesized and used as PCR primers to amplify a probe for screening of *Acremonium thermophilum cel*45A (EG_40) and *Chaetomium thermophilum cel*7B (EG_54) gene from the corresponding genomic libraries.

Protein	Peptide	Primer location[a]	Primer sequence[b]
At EG_40	Peptide 5	1–6	TAYTGGGAYTGYTGYAARCC
	WFQNADN [c]		RTTRTCNGCRTTYTGRAACCA
Ct EG_54	Peptide 7	3–7	GCAAGCTTCGRCARAARTCRTCRTT [d]
	Peptide 2	5–9	GGAATTCGAYCARACNGARCARTA [e]

(a Amino acids of the peptide used for designing the primer sequence

(b N = A, C, G, or T; R = A or G; Y = C or T

(c Peptide not derived from the purified *Acremonium* EG_40 protein, but originates from the *M. albomyces* Cel45A sequence (AJ515703) homologous to EG_40.

(d A *Hind*III restriction site was added to the 5' end of the oligonucleotide

(e An *Eco*RI restriction site was added to the 5' end of the oligonucleotide

The *Acremonium thermophilum cel*45A gene specific probe to screen the genomic library was amplified with the forward (TAYTGGGAY-TGYTGYAARCC) and reverse (RTTRTCNGCRTTYTGRAACCA) primers using genomic DNA as a template. The PCR reaction mixtures contained 50 mM Tris-HCl, pH 9.0, 15 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.1 mM dNTPs, 0.5 µg each primer, 1 unit of Dynazyme EXT DNA polymerase (Finnzymes, Finland) and approximately 0.5 µg of *Acremonium* genomic DNA. The conditions for PCR reactions were the following: 5 min initial denaturation at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min annealing at 50–60°C, 2 min extension at 72°C and a final extension at 72°C for 10 min. For amplification of the *Chaetomium thermophilum cel*7B gene (coding for Ct EG_54) specific probe, a forward primer (GGAATTCGAYCARACNGARCARTA) and a reverse primer (GCAAGCTTCGRCARAARTCRTCRTT) were used. The PCR reaction mixtures contained 10 mM Tris-HCl, pH 8.8, 50 mM KCl, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 250 pmol each primer, 2 unit of Dynazyme II DNA polymerase (Finnzymes, Finland) and approximately 2 µg of *Chaetomium*

genomic DNA. The conditions for PCR reaction were as described above, except that annealing was performed at 45–50°C.

Two PCR products were obtained from the *Acremonium* PCR reaction. DNA fragments of about 0.6 kb and 0.8 kb were isolated from agarose gel and were cloned into the pCR4-TOPO® TA vector (Invitrogen, USA) resulting in plasmids pALK1710 and pALK1711, respectively. The DNA products were characterized by sequencing and by performing Southern blot hybridizations to the genomic *Acremonium* DNA digested with several restriction enzymes. The hybridization patterns obtained with the two fragments in stringent washing conditions suggest that two putative endoglucanase genes could be screened from the *Acremonium* genomic library. The deduced amino acid sequences of both PCR products have homology to several published endoglucanase sequences of glycosyl hydrolase family 45 (BLAST program, National Center for Biotechnology Information; Altschul *et al.*, 1990).

One PCR product of expected size (estimated from the homologous *Humicola insolens* endoglucanase sequence, A35275) was obtained from the *Chaetomium* PCR reaction. This DNA fragment of about 0.7 kb was cloned into the pCR4-TOPO® TA vector (Invitrogen, USA) resulting in plasmid pALK2005 and analyzed as described above. The deduced amino acid sequence of the PCR product has homology to several published cellulase sequences of glycosyl hydrolase family 7 (BLAST program, version 2.2.9 at NCBI, National Center for Biotechnology Information; Altschul *et al.*, 1990).

The insert from plasmids pALK1710, pALK1711, and pALK2005 was isolated by restriction enzyme digestion and labeled with digoxigenin according to the supplier´s instructions (Roche, Germany). About 1–2 x 10^5 plaques from the amplified *Acremonium* or *Chaetomium* genomic library were screened. The temperature for hybridisation was 68°C and the filters were washed 2 x 5 min at RT using 2 x SSC – 0.1 % SDS followed by 2 x 15 min at 68°C using 0.1 x SSC – 0.1% SDS. Several positive plaques were obtained, of which five to six strongly hybridizing plaques were purified from each screening. Phage DNAs were isolated and analysed by Southern blot hybridization. Restriction fragments hybridizing to the probe were subcloned into the pBluescript II KS+ vector (Stratagene, USA) and the relevant parts were sequenced. In all cases the subcloned phage fragment contains the full-length gene of interest. Table 15 summarises the information of the probes used for screening of the endoglucanase genes, phage clones from which the genes

were isolated, chosen restriction fragments containing the full-length genes with their promoter and terminator regions, names of plasmids containing the subcloned phage fragment, and the deposit numbers in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH culture collection (DSM) for *E. coli* strains carrying these plasmids.

Table 15. Probes used for cloning of endoglucanase gene, phage clone and the subclone chosen, plasmid name and the corresponding deposit number of the *E. coli* strain.

Gene	Genomic library	Probe used in screening	Phage clone	Subcloned fragment	Plasmid	*E. coli* deposit no.
At *cel*45A	*A. thermophilum* ALKO4245	pALK1710	P24	5.5 kb *Sma*I	pALK1908	DSM 17324
At *cel*45B	*A. thermophilum* ALKO4245	pALK1711	P41	6.0 kb *Xho*I	pALK1904	DSM 17323
Ct *cel*7B	*C. thermophilum* ALKO4261	pALK2005	P55	5.1 kb *Bam*HI	pALK2010	DSM 17729

*Thermoascus aurantiacus cel*5A gene (coding for EG_28) (SEQ ID NO: 9) was amplified directly from the isolated genomic DNA by PCR reaction. The forward (ATTAACCGCGGACTGCGCATCATGAAGCTCGGCTCTCTCGTGCTC) and reverse (AACTGAGGCATAGAAACTGACGTCATATT) primers that were used for amplification were designed on the basis of the published *T. aurantiacus eg*1 gene (AF487830). The PCR reaction mixtures contained 1 x Phusion HF buffer, 0.3 mM dNTPs, 0.5 µM of each primer, 2 units of PhusionTM DNA polymerase (Finnzymes, Finland) and approximately 0.25 µg of *Thermoascus* genomic DNA. The conditions for PCR reactions were the following: 5 min initial denaturation at 95°C, followed by 25 cycles of 30 s at 95°C, 30 s annealing at 57–67°C, 2.5 min extension at 72°C and a final extension at 72°C for 5 min. The amplified 1.3 kb product containing the exact gene (from START to STOP codon) was cloned as a *Sac*II-*Pst*I fragment into the pBluescript II KS+ vector. Two independent clones were sequenced and one clone was selected and designated as pALK1926. The deposit number of the *E. coli* strain containing

pALK1926 in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH culture collection is DSM 17326.

Relevant information of the genes and the deduced protein sequences (SEQ ID NO: 9–16) are summarized in Table 16 and Table 17, respectively. Peptide sequences of the purified *Acremonium* EG_40 (gene At *cel*45A), *Chaetomium* EG_54 (gene Ct *cel*7B), and *Thermoascus* EG_28 (gene Ta *cel*5A) endoglucanases were found in the corresponding deduced amino acid sequences of the cloned genes confirming that appropriate genes were cloned.

Table 16. Summary of the endoglucanase genes isolated from *Acremonium thermophilum*, *Chaetomium thermophilum*, and *Thermoascus aurantiacus*.

Endoglucanase gene	Length with introns (bp) [a]	Coding region (bp) [b]	No of introns	Lengths of introns (bp)	SEQ ID NO:
At *cel*45A	1076	891	2	59, 123	11
At *cel*45B	1013	753	2	155, 102	13
Ct *cel*7B	1278	1275	-	-	15
Ta *cel*5A	1317	1005	5	55, 60, 59, 74, 61	9

(a The STOP codon is included.

(b The STOP codon is not included.

Table 17. Summary of the deduced endoglucanase sequences of *Acremonium thermophilum*, *Chaetomium thermophilum*, and *Thermoascus aurantiacus*. ss, signal sequence.

Endoglucanase protein	No of aas	Length of ss NN/HMM[a]	CBD[b]	Predicted MW (Da, ss not incl)[c]	Predicted pI (ss not incl)	Putative N-glycosylation sites[d]	SEQ ID NO:
At EG_40	297	21/21	Yes, K265 to L297	28625	4.79	2	12
At EG_40_like	251	20/20	No	23972	6.11	2	14
Ct EG_54	425	17/17	No	45358	5.44	1	16
Ta EG_28	335	30(e	No	33712	4.30	1	10

(a The prediction of the signal sequence was made using the program SignalP V3.0 (Nielsen *et al.*, 1997; Bendtsen *et al.*, 2004); the NN value was obtained using neural networks and HMM value using hidden Markov models.

(b Presence of a cellulose binding domain in the protein, the amino acids of the C- terminal CBD are indicated (numbering according to the full length polypeptide)

(c The predicted signal sequence is not included. Prediction was made using the Compute pI/MW tool at ExPASy server (Gasteiger *et al.*, 2003).

(d The putative N-glycosylation sites N-X-S/T were predicted using the program NetNGlyc 1.0 (Gupta *et al.*, 2004).

(e According to Hong *et al.* 2003a

The deduced protein sequences of *Acremonium* EG_40 (At Cel45A) and EG_40_like (At Cel45B), *Chaetomium* EG_54 (Ct Cel7B), and *Thermoascus* EG_28 (Ta Cel5A) endoglucanases share homology with cellulases of glycosyl hydrolase family 45 (*Acremonium*), family 7 (*Chaetomium*), and family 5 (*Thermoascus*), thus identifying the isolated genes as members of these gene families. The closest homologies of the *Acremonium* endoglucanases EG_40/Cel45A and EG_40_like/Cel45B are endoglucanases of *Thielavia terrestris* (CQ827970, 77.3% identity) and *Myceliophthora thermophila*

(AR094305, 66.9% identity), respectively (Table 18). The two isolated *Acremonium* family 45 endoglucanases share only an identity of 53.7% with each other. Of these enzymes only EG_40/Cel45A contains a cellulose binding domain (CBD).

The closest homology for the predicted protein sequence of *Chaetomium* EG_54/Cel7B endoglucanase is found in the *Melanocarpus albomyces* Cel7A cellulase sequence (AJ515704). The identity between these two protein sequences is 70.6%.

The protein sequence of the isolated *Thermoascus aurantiacus* endoglucanase is completely identical with that of the published *T. aurantiacus* EGI (AF487830, Table 18). The closest homology was found in a β-glucanase sequence of *Talaromyces emersonii* (AX254752, 71.1% identity).

Table 18. Comparison of the deduced *Acremonium thermophilum* EG_40, EG_40_like/Cel45B, *Chaetomium thermophilum* EG_54/Cel7B, and *Thermoascus aurantiacus* EG_28/Cel5A endoglucanases with their homologous counterparts. The alignment was performed using the Needle programme of the EMBOSS programme package. *indicates an endoglucanase encoded by a gene cloned in this work.

Organism, enzyme, and accession number	Identity (%)
***Acremonium thermophilum* EG_40**	100.0
Thielavia terrestris EG45, CQ827970	77.3
Melanocarpus albomyces Cel45A, AJ515703	75.3
Neurospora crassa, hypothetical XM_324477	68.9
Humicola grisea var *thermoidea*, EGL3, AB003107	67.5
Humicola insolens EG5, A23635	67.3
Myceliophthora thermophila fam 45, AR094305	57.9
* *Acremonium thermophilum* EG_40_like	53.7
***Acremonium thermophilum* EG_40_like**	100.0
Myceliophthora thermophila fam 45, AR094305	66.9
Magnaporthe grisea 70-15 hypothetical, XM_363402	61.9
Thielavia terrestris EG45, CQ827970	
* *Acremonium thermophilum* EG_40	56.8
Melanocarpus albomyces Cel45A, AJ515703	53.7
	52.8
***Chaetomium thermophilum* EG_54**	100.0

Melanocarpus albomyces Cel7A, AJ515704	70.6
Humicola grisea var *thermoidea* EGI, D63516	68.8
Humicola insolens EGI, AR012244	67.7
Myceliophthora thermophila EGI, AR071934	61.7
Fusarium oxysporum var *lycopercisi* EGI, AF29210	53.5
Fusarium oxysporum EGI, AR012243	52.6
***Thermoascus aurantiacus* EG_28**	100.0
Thermoascus aurantiacus EG, AX812161	100.0
Thermoascus aurantiacus EGI, AY055121	99.4
Talaromyces emersonii β-glucanase, AX254752	71.1
Talaromyces emersonii EG, AF440003	70.4
Aspergillus niger EG, A69663	70.1
Aspergillus niger EG, A62441	69.9
Aspergillus niger EG, AF331518	69.6
Aspergillus aculeatus EGV, AF054512	68.5

Example 19. Production of recombinant endoglucanases in *Trichoderma reesei*

Expression plasmids were constructed for production of the recombinant *Acremonium* EG_40/Cel45A, EG_40_like/Cel45B, and *Thermoascus* EG_28/Cel5A proteins as described in Example 14. Linear expression cassettes (Table 19) were isolated from the vector backbone by restriction enzyme digestion, transformed into T. *reesei* A96 and transformants purified as described in Example 14.

Table 19. The expression cassettes constructed for production of *Acremonium thermophilum* EG_40/Cel45A, EG_40_like/Cel45B, and *Thermoascus aurantiacus* EG_28/Cel5A endoglucanases in *Trichoderma reesei*. The schematic structure of the expression cassettes is described in Figure 2.

Endoglucanase	Expression plasmid	Size of the expression cassette[a]	Heterologous terminator[b]
At EG_40	pALK1920	10.9 kb *Not*I	156 bp (*Hin*dIII)
At EG_40_like	pALK1921	8.6 kb *Eco*RI	282 bp (*Ssp*I)
Ta EG_28	pALK1930	8.6 kb *Not*I	none

(a The expression cassette for *T. reesei* transformation was isolated from the vector backbone by *Eco*RI or *Not*I digestion.

(b The number of nucleotides after the STOP codon of the cloned gene that are included in the expression cassette are indicated. The restriction site at the 3'-region of the gene that was used in construction of the expression cassette is indicated in parenthesis.

The endoglucanase production of the transformants was analyzed from the culture supernatants of shake flask cultivations (50 ml). Transformants were grown as in Example 14 and the enzyme activity of the recombinant protein was measured from the culture supernatant as the release of reducing sugars from carboxymethylcellulose (2% (w/v) CMC) at 50°C in 50 mM citrate buffer pH 4.8 essentially as described by Bailey and Nevalainen 1981; Haakana *et al.* 2004. Production of the recombinant proteins was also detected from culture supernatants by SDS-polyacrylamide gel electrophoresis. *Acremonium* EG_40-specific polyclonal antibodies were produced in rabbits (University of Helsinki, Finland). The expression of EG_40 was verified by Western blot analysis with anti-EG_40 antibodies using the ProtoBlot Western blot AP system (Promega). The genotypes of the chosen transformants were analysed by Southern blotting using the expression cassette as a probe.

The pH optimum of the heterologously produced endoglucanases was determined in the universal McIlvaine's buffer within a pH range of 4.0–8.0 using carboxymethylcellulose as substrate. As shown in Figure 6 A the broad-

est pH range (4.5–6.0) is that of the *Acremonium* EG_40/Cel45A protein, the optimum being at pH 5.5. The pH optima for the other heterologously produced endoglucanases are pH 5.0–5.5 and 6.0 for *Acremonium* EG_40_like/Cel45B and *Thermoascus* EG_28/Cel5A, respectively. The optimal temperature for enzymatic activity of these endoglucanases was determined at the temperature range of 50–85°C as described above. The highest activity of the enzymes was determined to be at 75°C, 60°C, and 75°C for the *Acremonium* EG_40/Cel45A, EG_40_like/Cel45B, and *Thermoascus* EG_28/Cel5A, respectively (Figure 6 B).

The chosen transformants were cultivated, as described in Example 14, in a 2 litre bioreactor for four days (28°C, pH 4.2) to obtain material for the application tests.

Example 20. Cloning of *Acremonium thermophilum* ALKO4245, *Chaetomium thermophilum* ALKO4261, and *Thermoascus aurantiacus* ALKO4242 beta-glucosidase genes

Standard molecular biology methods were used as described in Example 13. The construction of the *Acremonium*, *Chaetomium*, and *Thermoascus* genomic libraries has been described in Example 12.

The peptides derived from the purified *Acremonium*, *Chaetomium*, and *Thermoascus* β-glucosidases shared homology with several β-glucosidases of glycosyl hydrolase family 3 such as *Acremonium cellulolyticus* (BD168028), *Trichoderma viride* (AY368687), and *Talaromyces emersonii* (AY072918) β-glucosidases, respectively. To amplify a probe for screening of the *Acremonium*, *Chaetomium*, or *Thermoascus* genomic libraries, degenerate primers were designed on the basis of the peptide sequences. The order of the peptides in the protein sequence and the corresponding sense or anti-sense nature of the primers was deduced from the comparison with the homologous published β-glucosidases. Primer sequences and the corresponding peptides are listed in Table 20.

Table 20. Oligonucleotides synthesized and used as PCR primers to amplify a probe for screening of *Acremonium thermophilum cel*3A (βG_101), *Chaetomium thermophilum cel*3A (βG_76), and *Thermoascus aurantiacus cel*3A (βG_81) gene from the corresponding genomic libraries.

Protein	Peptide	Primer location[a]	Primer Sequence[b]
At βG_101	EKVNLT[c]		GARAARGTNAAYCTNAC
	Peptide 4	6–11	YTTRCCRTTRTTSGGRGTRTA
Ct βG_76	Peptide 6	4–9	TNTGYCTNCARGAYGG
	Peptide 1	3–8	TCRAARTGSCGRTARTCRATRAASAG
Ta βG_81	Peptide 3	1–5	AARGGYGTSGAYGTSCAR
	Peptide 1	2–7	YTTRCCCCASGTRAASGG

(a Amino acids of the peptide used for designing the primer sequence

(b To reduce degeneracy, some codons were chosen according to fungal preference. N = A, C, G, or T; R = A or G; S = C or G; Y = C or T

(c Peptide not derived from the purified *Acremonium* βG_101 protein, but originates from the *A. cellulolyticus* β-glucosidase sequence (BD168028) homologous to βG_101.

The probes for screening genomic libraries constructed were amplified with the listed primer combinations (Table 20) using *Acremonium*, *Chaetomium*, or *Thermoascus* genomic DNA as template. The PCR reaction mixtures contained 50 mM Tris-HCl, pH 9.0, 15 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 1.5 mM $MgCl_2$, 0.1–0.2 mM dNTPs, 0.25 μg each primer, 1 unit of Dynazyme EXT DNA polymerase (Finnzymes, Finland) and approximately 0.5 μg of genomic DNA. The conditions for PCR reactions were the following: 5 min initial denaturation at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min annealing at 40°C (*Acremonium* DNA as a template), at 50°C (*Chaetomium* DNA as a template), or at 63°C (*Thermoascus* DNA as a template), 2–3 min extension at 72°C and a final extension at 72°C for 5–10 min.

Specific PCR products of expected size (estimated from the homologous β-glucosidase sequences BD168028, AY072918, and AY368687) were

isolated from the agarose gel. DNA fragments of about 1.8 kb (*Acremonium*), 1.5 kb (*Chaetomium*), and 1.52 kb (*Thermoascus*) were cloned into the pCR4-TOPO® TA vector (Invitrogen, USA) resulting in plasmids pALK1924, pALK1935, and pALK1713, respectively. The DNA products were characterized by sequencing and by performing Southern blot hybridizations to the genomic DNA digested with several restriction enzymes. The hybridization patterns in stringent washing conditions suggest that one putative β-glucosidase gene could be isolated from the *Acremonium*, *Chaetomium*, and *Thermoascus* genomic library. The deduced amino acid sequences of all three PCR products have homology to several published β-glucosidase sequences of glycosyl hydrolase family 3 (BLAST program, National Center for Biotechnology Information; Altschul *et al.*, 1990).

The insert from plasmids pALK1713, pALK1924, and pALK1935 was isolated by restriction enzyme digestion and labeled with digoxigenin according to the supplier's instructions (Roche, Germany). About 1–2 x 10^5 plaques from the amplified *Acremonium*, *Chaetomium*, or *Thermoascus* genomic library were screened as described in Example 18. Several positive plaques were obtained, of which five to six strongly hybridizing plaques were purified from each screening. Phage DNAs were isolated and analysed by Southern blot hybridization. Restriction fragments hybridizing to the probe were subcloned into the pBluescript II KS+ vector (Stratagene, USA) and the relevant parts were sequenced. In all cases the subcloned phage fragment contains the full-length gene of interest. Table 21 summarises the information of the probes used for screening of the β-glucosidase genes, phage clones from which the genes were isolated, chosen restriction fragments containing the full-length genes with their promoter and terminator regions, names of plasmids containing the subcloned phage fragment, and the deposit numbers in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH culture collection (DSM) for *E. coli* strains carrying these plasmids.

Table 21. Probes used for cloning of β-glucosidase gene, phage clone and the subclone chosen, plasmid name and the corresponding deposit number of the *E. coli* strain.

Gene	Genomic library	Probe used in screening	Phage clone	Sub-cloned fragment	Plasmid	*E. coli* deposit no.
At *cel*3A	A. *thermophilum* ALKO4245	pALK1924	P44	6.0 kb *Hin*dIII	pALK1925	DSM 17325
Ct *cel*3A	C. *thermophilum* ALKO4261	pALK1935	P51	7.0 kb *Xba*I	pALK2001	DSM 17667
Ta *cel*3A	T. *aurantiacus* ALKO4242	pALK1713	P21	5.3 kb *Bam*HI	pALK1723	DSM 16725

Relevant information of the genes and deduced protein sequences (SEQ ID NO: 21–26) are summarized in Table 22 and Table 23, respectively. Peptide sequences of the purified *Acremonium* βG_101 (At Cel3A), *Chaetomium* βG_76 (Ct Cel3A), and *Thermoascus* βG_81 (Ta Cel3A) proteins were found in the corresponding deduced amino acid sequences of the cloned genes confirming that appropriate genes were cloned.

Table 22. Summary of the β-glucosidase genes isolated from *Acremonium thermophilum*, *Chaetomium thermophilum*, and *Thermoascus aurantiacus*.

β-glucosidase gene	Length with introns (bp) [a]	Coding region bp) [b]	No of introns	Lengths of introns (bp)	SEQ ID NO:
At *cel*3A	2821	2583	3	92, 74, 69	23
Ct *cel*3A	2257	2202	1	52	25
Ta *cel*3A	3084	2529	7	134,67,56,64,59,110,62	21

(a The STOP codon is included.

(b The STOP codon is not included.

Table 23. Summary of the deduced β-glucosidase sequences of *Acremonium thermophilum*, *Chaetomium thermophilum*, and *Thermoascus aurantiacus*. ss, signal sequence.

β-glucosidase protein	No of aas	Length of ss NN/HMM[a]	CBD[b]	Predicted MW (Da, ss not incl)[c]	Predicted pI ss not incl)	Putative N-glycosylation sites[d]	SEQ ID NO:
At βG_101	861	19/18	No	91434	5.46	8	24
Ct βG_76	734	20/20	No	76457	6.3	2	26
Ta βG_81	843	19/19	No	89924	4.95	8	22

(a The prediction of the signal sequence was made using the program SignalP V3.0 (Nielsen *et al.*, 1997; Bendtsen *et al*, 2004); the NN value was obtained using neural networks and HMM value using hidden Markov models.

(b Presence of a cellulose binding domain in the protein.

(c The predicted signal sequence is not included. Prediction was made using the Compute pI/MW tool at ExPASy server (Gasteiger *et al.*, 2003).

(d The putative N-glycosylation sites N-X-S/T were predicted using the program NetNGlyc 1.0 (Gupta *et al.*, 2004).

The deduced protein sequences of *Acremonium* βG_101/Cel3A, *Chaetomium* βG_76/Cel3A, and *Thermoascus* βG_81/Cel3A β-glucosidases share homology with enzymes of glycosyl hydrolase family 3, thus identifying that the isolated genes belong to this gene family. The closest counterparts of the *Acremonium*, *Chaetomium*, and *Thermoascus* β-glucosidases are those of *Magnaporthe grisea* (β-glucosidase, AY849670), *Neurospora crassa* (hypothetical, XM_324308), and *Talaromyces* emersonii (β-glucosidase, AY072918), respectively (Table 24). The highest sequence identity (73.2%) found was that of *C. thermophilum* βG_76/Cel3A to *N. crassa* hypothetical protein indicating that novel enzymes genes were cloned.

Table 24. Comparison of the deduced *Acremonium thermophilum* βG_101/Cel3A, *Chaetomium thermophilum* βG_76/Cel3A, and *Thermoascus aurantiacus* βG_81/Cel3A β-glucosidases with their homologous counterparts. The alignment was performed using the Needle programme of the EMBOSS programme package. *indicates a β-glucosidase encoded by a gene cloned in this work.

Organism, enzyme, and accession number	Identity (%)
***Acremonium thermophilum* βG_101**	100.0
Magnaporthe grisea β-glucosidase, AY849670	73.1
Neurospora crassa hypothetical, XM_330871	71.1
Trichoderma reesei Cel3B, AY281374	65.2
* *Thermoascus aurantiacus* βG_81	62.2
Aspergillus aculeatus β-glucosidase, D64088	59.5
Talaromyces emersonii β-glucosidase, AY072918	58.9
Aspergillus oryzae, AX616738	58.2
Acremonium cellulolyticus β-glucosidase, BD168028	57.2
* *Chaetomium thermophilum* βG_76	40.9
***Chaetomium thermophilum* βG_76**	100.0
Neurospora crassa, hypothetical XM_324308	76.9
Magnaporthe grisea, hypothetical XM_364573	70.2
Trichoderma viridae BGI, AY368687	65.8
Acremonium cellulolyticus β-glucosidase, BD168028	41.2
* *Acremonium thermophilum* βG_101	40.9
Trichoderma reesei Cel3B, AY281374	40.0
* *Thermoascus aurantiacus* βG_81	39.9
***Thermoascus aurantiacus* βG_81**	100.0
Talaromyces emersonii β-glucosidase, AY072918	73.2
Aspergillus oryzae, AX616738	69.5
Aspergillus aculeatus β-glucosidase, D64088	68.0
Acremonium cellulolyticus β-glucosidase, BD168028	65.7
* *Acremonium thermophilum* βG_101	62.2
Trichoderma reesei Cel3B, AY281374	57.9
* *Chaetomium thermophilum* βG_76	39.9

Example 21. Production of recombinant beta-glucosidases in *Trichoderma reesei*

Expression plasmids were constructed for production of the recombinant *Acremonium* βG_101/Cel3A, *Chaetomium* βG_76/Cel3A, and *Thermoascus* βG_81/Cel3A proteins as described in Example 14. Linear expression cassettes (Table 25) were isolated from the vector backbone by restriction enzyme digestion, transformed into *T. reesei* A96 or A33 (both strains have the genes encoding the four major cellulases CBHI/Cel7A, CBHII/Cel6A, EGI/Cel7B and EGII/Cel5A deleted) and transformants purified as described in Example 14.

Table 25. The expression cassettes constructed for production of *Acremonium thermophilum* βG_101/Cel3A, *Chaetomium thermophilum* βG_76/Cel3A, and *Thermoascus aurantiacus* βG_81/Cel3A β-glucosidases in *Trichoderma reesei*. The schematic structure of the expression cassettes is described in Figure 2.

β-glucosidase	Expression plasmid	Size of the expression cassette[a]	Heterologous terminator[b]
At βG_101	pALK1933	10.5 kb *Not*I	300 bp (*Hin*dIII)
Ct βG_76	pALK2004	10.1 kb *Eco*RI	528 bp (*Xba*I)
Ta βG_81	pALK1914	10.9 kB *Eco*RI	452 bp (*Apo*I)

(a The expression cassette for *T. reesei* transformation was isolated from the vector backbone by *Eco*RI or *Not*I digestion.

(b The number of nucleotides after the STOP codon of the cloned gene that are included in the expression cassette are indicated. The restriction site at the 3'-region of the gene that was used in construction of the expression cassette is indicated in parenthesis.

The beta-glucosidase production of the transformants was analyzed from the culture supernatants of shake flask cultivations (50 ml). Transformants were grown as in Example 14 and the enzyme activity of the recombinant protein was measured from the culture supernatant using 4-nitrophenyl-β-D-glucopyranoside substrate as described by Bailey and Nevalainen 1981. Production of the recombinant proteins was also detected from culture supernatants by SDS-polyacrylamide gel electrophoresis. In addition, the expression

of *Thermoascus* βG_81 was verified by Western blot analysis with anti-βG_81 antibodies as described in Example 19. The genotypes of the chosen transformants were analysed by Southern blotting using the expression cassette as a probe.

The pH optimum of the heterologously produced β-glucosidases was determined in the universal McIlvaine's buffer within a pH range of 3.0–8.0 using 4-nitrophenyl-β-D-glucopyranoside as substrate. The pH optima for the *Acremonium* βG_101, *Chaetomium* βG_76, and *Thermoascus* βG_81 are pH 4.5, 5.5, and 4.5, respectively (Figure 7 A). The optimal temperature for enzymatic activity of these β-glucosidases was determined at the temperature range of 50–85°C as described above. The highest activity of the enzymes was determined to be at 70°C, 65°C, and 75°C for the *Acremonium* βG_101/Cel3A, *Chaetomium* βG_76/Cel3A, and *Thermoascus* βG_81/Cel3A, respectively (Figure 7 B).

The chosen transformants were cultivated, as described in Example 14, in a 2 litre bioreactor for four days (28°C, pH 4.2) to obtain material for the application tests.

Example 22. Cloning of *Acremonium thermophilum* ALKO4245 and *Thermoascus aurantiacus* ALKO4242 xylanase genes

Standard molecular biology methods were used as described in Example 13. The construction of the *Acremonium* genomic library has been described in Example 12.

The peptides derived from the purified *Acremonium* xylanase shared homology with xylanases of the glycosyl hydrolase family 10 such as *Humicola grisea* XYNI (AB001030). All peptides derived from the *Thermoascus* xylanase were completely identical with the published *Thermoascus aurantiacus* XYNA sequence (AJ132635) thus identifying the purified protein as the same enzyme. Due to this the *Thermoascus* xylanase gene was amplified by PCR from the genomic DNA.

To amplify a probe for screening of the *Acremonium* xylanase gene from the genomic library, degenerate primers were designed on the basis of the peptide sequences (Example 11, Table 5). The order of the peptides in the protein sequence and the corresponding sense or antisense nature of the primers was deduced from the comparison with the homologous *Humicola insolens* XYNI sequence (AB001030). The sense primer sequence (GAYGGYGAYGCSACYTAYATG) is based on Peptide 3 (amino acids 2–8)

and anti-sense primer (YTTYTGRTCRTAYTCSAGRTTRTA) on Peptide 1 (amino acids 4–11).

A PCR product of expected size (estimated from the homologous *Humicola insolens* XYNI sequence AB001030) was obtained from the reaction. This DNA fragment of about 0.7 kb was cloned into the pCR4-TOPO® TA vector (Invitrogen, USA) resulting in plasmid pALK1714, and was characterized by sequencing. The deduced amino acid sequence of the PCR product has homology to several published xylanase sequences of glycosyl hydrolase family 10 (BLAST program, National Center for Biotechnology Information; Altschul *et al.*, 1990).

The insert from plasmid pALK1714 was isolated by restriction enzyme digestion and labeled with digoxigenin according to the supplier's instructions (Roche, Germany). About 1–2 x 10^5 plaques from the amplified *Acremonium* genomic library were screened as described in Example 18. Several positive plaques were obtained, of which five strongly hybridizing plaques were purified. Phage DNAs were isolated and analysed by Southern blot hybridization. A 3.0 kb *Xba*I restriction fragment hybridizing to the probe was subcloned into the pBluescript II KS+ vector (Stratagene, USA) resulting in plasmid pALK1725. Relevant parts of pALK1725 were sequenced and found to contain the full-length *Acremonium thermophilum xyn*10A gene (SEQ ID NO: 19). The deposit number of the *E. coli* strain containing pALK1725 in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH culture collection is DSM 16726.

*Thermoascus aurantiacus xyn*10A gene (SEQ ID NO: 17) was amplified directly from the isolated genomic DNA by PCR reaction. The forward (TTATACCGCGGGAAGCCATGGTTCGACCAACGATCCTAC) and reverse (TTATAGGATCCACCGGTCTATACTCACTGCTGCAGGTCCTG) primers that were used in the amplification of the gene were designed on the basis of the published *T. aurantiacus xyn*A gene (AJ132635). The PCR reaction mixtures contained 50 mM Tris-HCl, pH 9.0, 15 mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 1.5 mM MgCl2, 0.3 mM dNTPs, 1 µM each primer, 1 unit of Dynazyme EXT DNA polymerase (Finnzymes, Finland) and approximately 0.5 µg of *Thermoascus* genomic DNA. The conditions for PCR reactions were the following: 5 min initial denaturation at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min annealing at 60–66°C, 3 min extension at 72°C and a final extension at 72°C for 10 min. The amplified 1.9 kb product containing the exact gene (from START to

STOP codon) was cloned as a *Sac*II-*Bam*HI fragment into the pBluescript II KS+ vector. Three independent clones were sequenced and one clone was selected and designated as pALK1715. The deposit number of the *E. coli* strain containing pALK1715 in the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH culture collection is DSM 16724.

Relevant information of the genes and deduced protein sequences (SEQ ID NO: 17–20) are summarized in Table 26 and Table 27, respectively. Peptide sequences of the purified *Acremonium* XYN_60 and *Thermoascus* XYN_30 proteins were found in the corresponding deduced amino acid sequences of the cloned genes (At *xyn*10A and Ta *xyn*10A, respectively) confirming that appropriate genes were cloned.

Table 26. Summary of the xylanase genes isolated from *Acremonium thermophilum* and *Thermoascus aurantiacus*.

Xylanase gene	Length with introns (bp) [a]	Coding region (bp) [b]	No of introns	Lengths of introns (bp)	SEQ ID NO:
At *xyn*10A	1471	1248	2	135, 85	19
Ta *xyn*10A	1913	987	10	73, 74, 68, 103, 69, 65, 93, 66, 100, 212	17

(a The STOP codon is included.

(b The STOP codon is not included.

Table 27. Summary of the deduced xylanase sequences of *Acremonium thermophilum* and *Thermoascus aurantiacus*. ss, signal sequence.

Xylanase protein	No of aas	Length of ss NN/HMM[a]	CBD[b]	Predicted MW (Da, ss not incl)[c]	Predicted pI (ss not incl)	Putative N-glyco-sylation sites[d]	SEQ ID NO:
At XYN_60	416	19/19	Yes, W385 to L416	42533	6.32	1-2	20
Ta XYN_30	329	26(e	No	32901	5.81	0	18

(a The prediction of the signal sequence was made using the program SignalP V3.0 (Nielsen *et al.*, 1997; Bendtsen *et al*, 2004); the NN value was obtained using neural networks and HMM value using hidden Markov models.

(b Presence of a carbohydrate binding domain CBD, the amino acids of the C- terminal CBD are indicated (numbering according to the full length polypeptide)

(c The predicted signal sequence is not included. Prediction was made using the Compute pI/MW tool at ExPASy server (Gasteiger *et al.*, 2003).

(d The putative N-glycosylation sites N-X-S/T were predicted using the program NetNGlyc 1.0 (Gupta *et al.*, 2004).

(e According to Lo Leggio *et al.*, 1999

The deduced protein sequences of *Acremonium* and *Thermoascus* xylanases share homology with several enzymes of glycosyl hydrolase family 10, identifying the corresponding genes as members of family 10 xylanases. The closest counterpart for the *Acremonium* XYN_60/Xyn10A found is the *Humicola grisea* XYLI (AB001030) showing 67.1% identity with XYN_60 (Table 28). The predicted protein sequence of the isolated *Thermoascus aurantiacus* XYN_30/Xyn10A xylanase is completely identical with that of the published *T. aurantiacus* XYNA (P23360, Table 28). The closest homology was found in a xylanase sequence of *Aspergillus niger* (A62445, 69.7% identity).

Table 28. Comparison of the deduced *Acremonium thermophilum* XYN_60/Xyn10A and *Thermoascus aurantiacus* XYN_30/Xyn10A xylanases with their homologous counterparts. The alignment was performed using the Needle programme of the EMBOSS programme package. *indicates a xylanase encoded by a gene cloned in this work.

Organism, enzyme, and accession number	Identity (%)
***Thermoascus aurantiacus* XYN_30**	100.0
Thermoascus aurantiacus XynA, P23360	100.0
Thermoascus aurantiacus XynA, AF127529	99.4
Aspergillus niger xylanase, A62445	69.7
Aspergillus aculeatus xylanase, AR137844	69.9
Aspergillus terreus fam 10 xyn, DQ087436	65.0
Aspergillus sojae, XynXI AB040414	63.8
Penicillium chrysogenum xylanase, AY583585	62.5
***Acremonium thermophilum* XYN_60**	100.0
Humicola grisea XYL I, AB001030	67.1
Magnaporthe grisea 70-15, hypothetical XM_364947	63.8
Aspergillus aculeatus xylanase, AR149839	53.7
Talaromyces emersonii xylanase, AX403831	51.8
Gibberella zeae xylanase, AY575962	51.4
Magnaporthe grisea XYL5, AY144348	48.5
Talaromyces emersonii, AX172287	46.9

Example 23. Production of recombinant xylanases in *Trichoderma reesei*

Expression plasmids were constructed for production of the recombinant *Acremonium* XYN_60/Xyn10A and *Thermoascus* XYN_30/Xyn10A proteins as described in Example 14. Linear expression cassettes (Table 29) were isolated from the vector backbone by restriction enzyme digestion, transformed into *T. reesei* A96, and transformants purified as described in Example 14.

Table 29. The expression cassettes constructed for production of *Acremonium thermophilum* XYN_60/Xyn10A and *Thermoascus aurantiacus* XYN_30/Xyn10A xylanases in *Trichoderma reesei*. The schematic structure of the expression cassettes is described in Figure 2.

Xylanase	Expression plasmid	Size of the expression cassette[a]	Heterologous terminator[b]
At XYN_60	pALK1912	9.0 kb	150 bp (*Bam*HI)
Ta XYN_30	pALK1913	9.3 kb	none

(a The expression cassette for *T. reesei* transformation was isolated from the vector backbone by *Eco*RI digestion.

(b The number of nucleotides after the STOP codon of the cloned gene that are included in the expression cassette are indicated. The restriction site at the 3'-region of the gene that was used in construction of the expression cassette is indicated in parenthesis.

The xylanase production of the transformants was analyzed from the culture supernatants of shake flask cultivations (50 ml). Transformants were grown as in Example 14 and the enzyme activity of the recombinant protein was measured from the culture supernatant as the release of reducing sugars from birch xylan (1% w/v) at 50°C in 50 mM citrate buffer pH 5.3 as described by Bailey and Poutanen 1989. Production of the recombinant protein was also analyzed from culture supernatant by SDS-polyacrylamide gel electrophoresis. In addition, the expression of both xylanases was determined by Western blot analysis with anti-XYN_30 or anti-XYN_60 antibodies as described in Example 19. The genotypes of the chosen transformants were analysed by Southern blotting using the expression cassette as a probe.

Thermoascus XYN_30/Xyn10A was produced in *T. reesei* and the pH optimum of the heterologously produced protein was determined in the universal McIlvaine's buffer within a pH range of 3.0–8.0 using birch xylan as substrate (Figure 8 A). The optimal pH was determined to be 4.5. The temperature optimum for the enzymatic activity of XYN_30 was determined to be 75°C (Figure 8 B).

The chosen transformants were cultivated, as described in Example 14, in a 2 litre bioreactor for four days (28°C, pH 4.2) to obtain material for the application tests.

Example 24. Performance of the recombinant cellobiohydrolases in the hydrolysis

The performance of the purified recombinant cellobiohydrolases was evaluated in the hydrolysis studies with purified T. *reesei* enzymes. Hydrolysis was carried out with controlled mixtures of purified enzymes on several pre-treated substrates. Culture filtrates of *T. reesei*, containing different cloned CBH/Cel7 enzymes were obtained as described in Examples 14 and 15, and the CBH enzymes were purified by affinity chromatography as described in Example 2. In addition, pure *T. reesei* cellulases (purified as described by Suurnäkki *et al.*, 2000) were used in the enzyme mixtures. The cellobiohydrolases used in the experiment were:

Thermoascus aurantiacus ALKO4242 CBH (Ta Cel7A)

Thermoascus aurantiacus ALKO4242 CBH (Ta Cel7A) with genetically attached CBD of *Trichoderma reesei* (Ta Cel7A +Tr CBD)

Thermoascus aurantiacus ALKO4242 CBH (Ta Cel7A) with genetically attached CBD of *Chaetomium thermophilum* (Ta Cel7A +Ct CBD)

Acremonium thermophilum ALKO4245 CBH (At Cel7A)

Chaetomium thermophilum ALKO4265 CBH (Ct Cel7A).

Each CBH/Cel7 to be tested (dosage 14.5 mg/g dry matter of substrate) was used either together with EGII/Cel5A of *T. reesei* (3.6 mg/g) or with a mixture containing *T. reesei* EGI/Cel7B (1.8 mg/g), EGII/Cel5A (1.8 mg/g), xylanase pI 9 (Tenkanen *et al.* 1992) (5000 nkat/g) and acetyl xylan esterase (AXE) (Sundberg and Poutanen, 1991) (250 nkat/g). All mixtures were supplemented with additional β-glucosidase from a commercial enzyme preparation Novozym 188 (176 nkat/g d.w.). Triplicate tubes containing the enzyme mixture and 10 mg (dry matter)/ml of the substrate suspended in 0.05 M sodium acetate were incubated in mixing by magnetic stirring at 45°C for 48 h. Reference samples with inactivated enzymes and corresponding substrates were also prepared. The release of hydrolysis products was measured as reducing sugars with DNS method using glucose as standard (Table 30).

The following substrates were used in the experiment:

Crystalline cellulose (Avicel)

Washed steam pre-treated spruce fibre (impregnation with 3% w/w SO_2 for 20 min, followed by steam pre-treatment at 215°C for 5 min), dry matter 25.9% (SPRUCE).

Washed wet oxidized corn stover fibre (WOCS).

Washed steam pre-treated willow fibre (pre-treatment for 14 min at 210°C), dry matter 23.0% (WILLOW).

Table 30. Hydrolysis products with CBH enzymes (45°C, pH 5.0). Reaction products after 48 h hydrolysis as reducing sugars (mg/ml), measured glucose as standard. Abbreviations: CBH = cellobiohydrolase; EGI = endoglucanase I (Cel7B) of *T. reesei*, EGII = endoglucanase II (Cel5A) of *T. reesei*; bG = β-glucosidase (from Novozym 188); XYL= xylanase pl 9 (XYN II) of *T. reesei*, AXE = acetyl xylan esterase of *T. reesei*; nd = not done.

Enzymes		**Substrates**			
CBH	Additional enzymes	Avicel	SPRUCE	WOCS	WILLOW
Ta Cel7A	EGII, bG	2.0	2.0	2.8	2.0
Ta Cel7A +Tr CBD	EGII, bG	5.8	4.0	4.4	4.0
Ta Cel7A +Ct CBD	EGII, bG	4.9	3.7	4.6	3.7
At Cel7A	EGII, bG	5.3	3.3	4.5	3.3
Ct Cel7A	EGII, bG	6.0	2.6	3.4	2.6
Cel7A of *T. reesei*	EGII, bG	4.7	2.9	2.9	2.9
Ta Cel7A	EGII, EGI, XYL, AXE, bG	nd	nd	4.3	2.8
Ta Cel7A +Tr CBD	EGII, EGI, XYL, AXE, bG	nd	nd	7.2	5.9
Ta Cel7A +Ct CBD	EGII, EGI, XYL, AXE, bG	nd	nd	7.2	5.6
At Cel7A	EGII, EGI, XYL, AXE, bG	nd	nd	6.4	5.4
Ct Cel7A	EGII, EGI, XYL, AXE, bG	nd	nd	5.6	4.0
Cel7A of *T. reesei*	EGII, EGI, XYL, AXE, bG	nd	nd	6.0	4.1

In Table 30 the different cellobiohydrolases have been compared based on the same protein dosage in the hydrolysis. The results show that on cellulosic substrates (Avicel and spruce fibre) Cel7A of *Thermoascus aurantiacus* with genetically attached CBD showed clearly higher hydrolysis than *T. reesei* CBHI/Cel7A. Without CBD, *T. aurantiacus* Cel7A was less efficient on these substrates. The performance of *Acremonium thermophilum* and *Chaetomium thermophilum* cellobiohydrolases was also better than that of

T. reesei CBHI/Cel7A on several substrates; in particular, *C. thermophilum* Cel7A showed high efficiency on pure cellulose (Avicel).

In the case of substrates containing notable amounts of hemicellulose (willow and corn stover) the CBH/Cel7 enzymes clearly needed additionally both hemicellulases and endoglucanases to perform efficiently. If no additional hemicellulases were present, Cel7A of *T. aurantiacus* with genetically attached CBD showed again clearly highest hydrolysis. With the most important hemicellulose-degrading enzymes (xylanase, acetyl xylan esterase and EGI) Cel7A of T. *aurantiacus* with genetically attached CBD performed again with highest efficiency. *A. thermophilum* Cel7A was more efficient than *T. reesei* enzyme and *C. thermophilum* Cel7A produced hydrolysis products on the same level than *T. reesei* CBHI/Cel7A. The cellulose binding domain of *T. reesei* seemed to give slightly better efficiency than CBD of *C. thermophilum* in the hydrolytic performance of *T. aurantiacus* Cel7A, even though the difference was rather small.

It can be concluded that when CBHI/ Cel7A was replaced in the mixture of *Trichoderma* enzymes by the herein produced cellobiohydrolases, the hydrolysis efficiency as judged by this experimental arrangements was clearly improved in the case of *T. aurantiacus* Cel7A with genetically attached CBD, and also improved in the case of *A. thermophilum* Cel7A and *C. thermophilum* Cel7A. Considering also the better temperature stability of the herein produced cellobiohydrolases, the results indicate that the performance of cellulase enzyme mixtures in higher temperatures than 45°C can be clearly improved by using the herein produced cellobiohydrolases.

Example 25. Performance of the recombinant endoglucanases in the hydrolysis

The preparations containing the endoglucanases were compared in hydrolysis studies mixed with the purified CBH/Cel7 and CBH/Cel6 enzymes on several pre-treated substrates. Culture filtrates of *T. reesei*, containing different cloned endoglucanase enzymes were obtained as described in Example 19. The enzymes were enriched by removing thermolabile proteins from the mixtures by a heat treatment (60°C, 2 h, pH 5) and the supernatant was used for the hydrolysis studies. In addition, pure *T. reesei* cellulases (purified as described by Suurnäkki *et al.*, 2000) were used in the enzyme mixtures. The endoglucanases used in the experiment were:

Acremonium thermophilum ALKO4245 endoglucanase At EG_40/Cel45A (ALKO4245 EG_40)

Acremonium thermophilum ALKO4245 endoglucanase At EG_40_like/Cel45B (ALKO4245 EG_40_like)

Thermoascus aurantiacus ALKO4242 endoglucanase Ta EG_28/Cel5A (ALKO4242 EG_28).

The following substrates were used in the experiment:

Washed steam pre-treated spruce fibre (impregnation with 3% SO_2 for 20 min, followed by steam pre-treatment at 215°C for 5 min), dry matter 25.9% (SPRUCE).

Steam exploded corn stover fibre (steam pre-treatment at 210°C for 5 min), dry matter 31.0% (SECS).

The endoglucanases to be studied (dosage 840 nkat/g dry matter, based on endoglucanase activity against HEC according to IUPAC, 1987) were used either with cellobiohydrolases of *T. reesei* (CBHI/Cel7A, 8.1 mg/g d.m. and CBHII/Cel6A, 2.0 mg/g d.m.) or with *Thermoascus aurantiacus* Cel7A with genetically attached CBD of *T. reesei* (10.1 mg/g d.m.). Purified (Suurnäkki *et al.*, 2000) EGI (Cel7B) and EGII (Cel5A) of *T. reesei* were also included in the experiments for comparison. All mixtures were supplemented with additional β-glucosidase from Novozym 188 (to make the total β-glucosidase dosage 560 nkat/g d.w., the relatively high dosage was used to compensate the differences in the background activities of the different EG preparations). Triplicate tubes were incubated in mixing at 45°C for 48 h and reference samples with inactivated enzymes and corresponding substrates were prepared. The release of hydrolysis products was measured as reducing sugars with DNS method using glucose as standard (Table 31).

Table 31. Hydrolysis products with different endoglucanase preparations when used together with cellobiohydrolases from *T. reesei* or with *T. aurantiacus* Cel7A harbouring CBD of *T. reesei*. Reaction products after 48 h hydrolysis (45°C, pH 5.0) as reducing sugars (mg/ml), measured glucose as standard. Abbreviations: CBHI = cellobiohydrolase I (Cel7A) of *T. reesei*; CBHII = cellobiohydrolase II (Cel6A) of *T. reesei*; EGI = endoglucanase I (Cel7B) of *T. reesei*, EGII = endoglucanase II (Cel5A) of *T. reesei*; bG = β-glucosidase (from Novozym 188); nd. = not done.

Enzymes		**Substrate**	
Endoglucanase	CBH/Cel7	SPRUCE	SECS
no added EG	CBHI and CBHII of *T. reesei*	2.4	3.2
EGI	CBHI and CBHII of *T. reesei*	3.5	4.6
EGII	CBHI and CBHII of *T. reesei*	3.8	3.5
At EG_40	CBHI and CBHII of *T. reesei*	4.9	4.3
At EG_40like	CBHI and CBHII of *T. reesei*	4.5	4.8
Ta EG_28	CBHI and CBHII of *T. reesei*	3.0	3.9
no added EG	*T. aurantiacus* Cel7A + Tr CBD	1.8	2.1
EG I	*T. aurantiacus* Cel7A + Tr CBD	nd.	4.2
EG II	*T. aurantiacus* Cel7A + Tr CBD	3.2	nd.
At EG_40	*T. aurantiacus* Cel7A + Tr CBD	4.8	4.0
Ta EG_28	*T. aurantiacus* Cel7A + Tr CBD	1.5	nd.

In Table 31 the different endoglucanases have been compared based on the same activity dosage in the hydrolysis. This may favour enzymes with low specific activity against the substrate (hydroxyethyl cellulose) used in the assay and underestimate the efficiency of enzymes with high specific activity against hydroxyethyl cellulose. In any case, the results show that *Acremonium thermophilum* endoglucanases perform very well in the hydrolysis when affecting together with both cellobiohydrolases used in the mixture. *A. thermophilum* endoglucanases have similar performance to *T. reesei* EGI/Cel7B which is a very efficient enzyme on hemicellulose-containing corn stover substrate due to its strong xylanase side activity. *T. aurantiacus* endoglucanase Cel5A (ALKO4242 EG_28) showed lower hydrolysis than *T. reesei* enzymes.

It can be concluded that the endoglucanases from *A. thermophilum* perform with comparable or enhanced efficiency when compared to the corresponding *Trichoderma* enzymes in the hydrolysis as judged by this experimental arrangement. Considering also the temperature stability of the herein described endoglucanases, the results indicate that the performance of cellulase enzyme mixtures in higher temperatures than 45°C can be improved by using the herein described endoglucanases.

Example 26. Hydrolysis of steam pre-treated spruce at high temperatures

Washed steam exploded spruce fibre (impregnation with 3% w/w SO_2 for 20 min, followed by steam pre-treatment at 215°C for 5 min), with dry matter of 25.9% was suspended in 5 ml of 0.05 M sodium acetate buffer in the consistency of 10 mg/ml. This substrate was hydrolysed using different enzyme mixtures in test tubes with magnetic stirring in the water bath adjusted in different temperatures for 72 h. For each sample point, a triplicate of test tubes was withdrawn from hydrolysis, boiled for 10 min in order to terminate the enzyme hydrolysis, centrifuged, and the supernatant was analysed for reaction products from hydrolysis. The blanks containing the substrate alone (only buffer added instead of enzymes) were also incubated in the corresponding conditions.

A mixture of thermophilic cellulases was prepared using the following components:

Thermophilic CBH/Cel7 preparation containing *Thermoascus aurantiacus* ALKO4242 Cel7A with genetically attached CBD of *T. reesei* CBHI/Cel7A. The protein preparation was produced as described in Example 15 and purified according to Example 2 resulting in the purified Ta Cel7A + Tr CBD preparation with protein content of 5.6 mg/ml.

Thermophilic endoglucanase preparation containing *Acremonium thermophilum* ALKO4245 endoglucanase At EG_40/Cel45A. The protein was produced in *T. reesei* as described in Example 19. In order to enrich the thermophilic components, the spent culture medium was heat treated (60°C for 2 hours). The preparation obtained contained protein 4.9 mg/ml and endoglucanase activity (according to IUPAC, 1987) 422 nkat/ml.

Thermophilic β-glucosidase preparation prepared as described in Example 21 containing *Thermoascus aurantiacus* ALKO4242 β-glucosidase Ta βG_81/Cel3A. In order to enrich the thermophilic components, the fermentor broth was heat treated (65°C for 2 hours). The preparation obtained con-

tained 4.3 mg/ml protein and β-glucosidase activity of 6270 nkat/ml (according to Bailey and Linko, 1990).

These enzyme preparations were combined as follows (per 10 ml of mixture): CBH/Cel7-preparation 4.51 ml, endoglucanase preparation 5.19 ml and β-glucosidase preparation 0.29 ml. This mixture was used as "MIXTURE 1" of the thermophilic enzymes.

As a comparison and reference, a state-of art mixture of commercial *Trichoderma reesei* enzymes was constructed combining (per 10 ml): 8.05 ml Celluclast 1,5 L FG (from Novozymes A/S) and 1.95 ml Novozym 188 (from Novozymes A/S). This was designated as "*T. REESEI* ENZYMES".

Enzymes were dosed on the basis of the FPU activity of the mixtures: "MIXTURE 1" using the dosage of 5.5 FPU per 1 gram of dry matter in the spruce substrate, and "*T. REESEI* ENZYMES" using 5.8 FPU per 1 gram of dry matter in the spruce substrate.

Samples were taken from the hydrolysis after 24, 48 and 72 h and treated as described above. The hydrolysis products were quantified using the assay for reducing sugars (Bernfeld, 1955), using glucose as standard. The amount of hydrolysis products as reducing sugars is presented in Figure 9.

The results clearly show better performance of the herein described enzymes as compared to the state-of-art *Trichoderma* enzymes in 55°C and 60°C on the spruce substrate. On the basis of HPLC analysis the maximum yield of sugars from the substrate would be 5.67 mg per 10 mg of dry spruce substrate. Because of the relatively low dosage of enzyme the final sugar yields were clearly lower. For thermostable enzymes the sugar yield based on reducing sugar assay was 66% and 57% of theoretical in 55°C and 60°C, respectively. For state-of art *Trichoderma* enzymes it was only 31% and 11% in 55°C and 60°C, respectively.

Example 27. Hydrolysis of steam pre-treated corn stover at high temperatures

Steam exploded corn stover fibre (treatment at 195°C for 5 min), with dry matter of 45.3% was suspended in 5 ml of 0.05 M sodium acetate buffer in the consistency of 10 mg/ml. The treatments and measurements were performed as described in Example 26.

A mixture of herein described thermophilic cellulases was constructed using the following components:

Thermophilic CBH preparation containing *Thermoascus aurantiacus* ALKO4242 Cel7A with genetically attached CBD of *T. reesei* CBHI/Cel7A (Ta Cel7A + Tr CBD, Example 15). The protein content of the preparation was 31 mg/ml.

Thermophilic endoglucanase preparation containing *Acremonium thermophilum* ALKO4245 endoglucanase At EG_40/Cel45A was obtained as described in Example 19. The concentrated enzyme preparation contained endoglucanase activity (according to IUPAC, 1987) of 2057 nkat/ml.

Thermophilic β-glucosidase preparation containing *Thermoascus aurantiacus* ALKO 4242 β-glucosidase Ta βG_81/Cel3A was obtained as described in Example 21 containing β-glucosidase activity (according to Bailey and Linko, 1990) of 11500 nkat/ml.

Thermophilic xylanase product containing an AM24 xylanase originating from *Nonomuraea flexuosa* DSM43186. The product was prepared by using a recombinant *Trichoderma reesei* strain that had been transformed with the expression cassette pALK1502, as described in WO2005/100557. The solid product was dissolved in water to make a 10% solution and an enzyme preparation with xylanase activity (assayed according to Bailey *et al.*, 1992) of 208000 nkat/ml was obtained.

These enzyme preparations were combined as follows (per 10 ml of mixture): CBH/Cel7 preparation 7.79 ml, endoglucanase preparation 0.96 ml, β-glucosidase preparation 1.14 ml and xylanase preparation 0.31 ml. This mixture was used as "MIXTURE 2" of the thermophilic enzymes.

As a comparison and reference, a state-of art mixture of commercial *Trichoderma reesei* enzymes was constructed by combining (per 10 ml) 8.05 ml Celluclast 1,5 L FG (from Novozymes A/S) and 1.95 ml Novozym 188 (from Novozymes A/S). This was designated as "*T. REESEI* ENZYMES".

Samples were taken from the hydrolysis after 24, 48 and 72 h and treated as described above. The hydrolysis products were quantified using the assay for reducing sugars (Bernfeld, 1955), using glucose as standard. The results from the substrate blanks were subtracted from the samples with enzymes, and the concentration of hydrolysis products as reducing sugars is presented in Figure 10.

The results clearly show better performance of the herein described enzymes as compared to the state-of-art *Trichoderma* enzymes. In 45°C the mixture of thermophilic enzymes showed more efficient hydrolysis as com-

pared to *T. reesei* enzymes: The hydrolysis was faster and higher sugar yields were also obtained. On the basis of HPLC analysis the maximum yield of sugars (including free soluble sugars in the unwashed substrate that was used) from the substrate would be 5.73 mg per 10 mg of dry substrate. Thus, the hydrolysis by the MIXTURE 2 enzymes was nearly complete within 48 hours. In 55°C and 57.5°C the herein described thermophilic enzymes showed also clearly better performance in the hydrolysis as compared to the state-of art *Trichoderma* enzymes.

Example 28. Hydrolysis of pre-treated corn stover at high temperatures using mixture with a thermostable xylanase

The procedure explained in Example 27 was repeated except that the xylanase product XT 02026A3 was replaced by thermophilic xylanase preparation containing *Thermoascus aurantiacus* ALKO4242 xylanase Ta XYN_30/Xyn10A produced in *T. reesei*. The fermentor broth, produced as described in Example 23 contained xylanase activity of 132 000 nkat/ml (assayed according to Bailey *et al.*, 1992).

These enzyme preparations were combined as follows (per 10 ml of mixture): CBH/Cel7-preparation 7.64 ml, endoglucanase preparation 0.96 ml, β-glucosidase preparation 1.15 ml and xylanase preparation 0.25 ml. This mixture was used as "MIXTURE 3" of the thermophilic enzymes.

As a comparison and reference, a state-of-art mixture of commercial *Trichoderma reesei* enzymes was constructed by combining (per 10 ml) 8.05 ml Celluclast 1,5 L FG (from Novozymes A/S) and 1.95 ml Novozym 188 (from Novozymes A/S). This was designated as "*T. REESEI* ENZYMES".

Samples were taken from the hydrolysis after 24, 48 and 72 h and treated as described above. The hydrolysis products were quantified using the assay for reducing sugars (Bernfeld, 1955), using glucose as standard. The results from the substrate blanks were subtracted from the samples with enzymes, and the concentration of hydrolysis products as reducing sugars is presented in Figure 11.

The results clearly show better performance of the mixture of the herein described enzymes as compared to the state-of-art *Trichoderma* enzymes. In 45°C the mixture of thermophilic enzymes showed more efficient hydrolysis as compared to *T. reesei* enzymes. In 55°C and 60°C the herein described thermophilic enzymes showed clearly better performance in the hydrolysis as compared to the state-of art *Trichoderma* enzymes. The perfor-

mance of the new enzyme mixture at 60°C was at the same level than the performance of state-of-art enzymes at 45°C.

Example 29. Hydrolysis of pre-treated spruce at high temperatures using mixture with a thermostable xylanase

Procedure as described in Example 28 was repeated with washed steam exploded spruce fibre (impregnation with 3% w/w SO_2 for 20 min, followed by steam pre-treatment at 215°C for 5 min, with dry matter of 25,9%) as substrate using hydrolysis temperatures 45°C, 55°C and 60°C. Samples were taken from the hydrolysis after 24, 48 and 72 h and treated as described above. The hydrolysis products were quantified using the assay for reducing sugars (Bernfeld, 1955), using glucose as standard. The results from the substrate blanks were subtracted from the samples with enzymes, and the concentration of hydrolysis products as reducing sugars is presented in Figure 12.

The results clearly show better performance of the mixture of herein described enzymes as compared to the state-of-art *Trichoderma* enzymes in all the temperatures studied. At 45°C the mixture of thermophilic enzymes showed more efficient hydrolysis as compared to *T. reesei* enzymes, evidently due to the better stability in long term hydrolysis. At 55°C the efficiency of the mixture of herein described enzymes was still on the same level than at 45°C, whereas the state-of-art mixture was inefficient with the substrate used in this temperature. At 60°C the herein described thermophilic enzymes showed decreased hydrolysis although the hydrolysis was nearly at the same level as the performance of the state-of-art enzymes at 45°C.

Example 30. Evaluation of glucose inhibition of β-glucosidases from *Acremonium thermophilium* ALKO4245, *Chaetomium thermophilum* ALKO4261 and *Thermoascus aurantiacus* ALKO4242

The culture filtrates produced by *Acremonium thermophilium* ALKO4245, *Chaetomium thermophilum* ALKO4261 and *Thermoascus aurantiacus* ALKO4242 strains are described in Example 1. The β-glucosidase activities (measured according to Bailey and Linko, 1990) of these preparations were 21.4 nkat/ml, 5.6 nkat/ml and 18.6 nkat/ml, respectively. For comparison, commercial enzymes Celluclast 1,5L (β-glucosidase 534 nkat/ml) and Novozym 188 (β-glucosidase 5840 nkat/ml) were also included in the experiment.

In order to evaluate the sensitivity of the different β-glucosidases towards glucose inhibition, the standard activity assay procedure was performed in the presence of different concentrations of glucose. The substrate (p-nitrophenyl-β-D-glucopyranoside) solutions for β-glucosidase activity assay were supplemented by glucose so that the glucose concentration in the assay mixture was adjusted to the values from 0 to 0.5 M. Except this glucose addition the assay was performed using the standard procedure (Bailey and Linko, 1990). The activities in the presence of varying glucose concentrations as a percentage of the activity without glucose are presented in Figure 13.

The results show that β-glucosidases from *C. thermophilum* and *T. aurantiacus* were affected less by glucose inhibition than the β-glucosidases present in the commercial enzymes: *Aspergillus*-derived β-glucosidase in Novozym 188 or *Trichoderma*-derived β-glucosidase in Celluclast 1,5L. *A. thermophilum* enzyme showed behaviour comparable to *T. reesei* enzyme of Celluclast. Especially *C. thermophilum* enzyme was clearly less affected by high glucose concentration. Thus, these results indicate that considering glucose inhibition the use of the new β-glucosidases, especially from strains *Acremonium thermophilium* ALKO4242 and *Chaetomium thermophilum* ALKO4261, would give clear advantages in hydrolysis in industrial conditions with high glucose concentration.

Example 31. Filter paper activity of enzyme mixtures in high temperatures

Filter paper activity of enzyme preparations was measured according to the method of IUPAC (1987) as described in the procedure except enzyme reaction was performed at temperatures from 50°C to 70°C. The calculated FPU activity is based on the amount of enzyme required to hydrolyse 4% of filter paper substrate in 1 h under the experimental conditions. The FPU activity is considered to represent the total overall cellulase activity of an enzyme preparation.

The enzyme mixtures were MIXTURE 2 prepared as described in Example 27, MIXTURE 3 prepared as described in Example 28, and MIXTURE 4. MIXTURE 4 was prepared by combining enzyme preparations described in Example 27 as follows (per 10 ml of mixture): CBH/Cel7-preparation 7.84 ml, endoglucanase preparation 0.99 ml and β-glucosidase preparation 1.17 ml.

The enzyme mixtures used as reference, representing the state-of art-mixtures, were:

"T. REESEI ENZYMES A" prepared as preparation "T. REESEI ENZYMES" described in Example 26.

"T. REESEI ENZYMES B" was constructed combining (per 10 ml) 8.05 ml Econase CE (a commercial *T. reesei* cellulase preparation from AB Enzymes Oy, Rajamäki, Finland) and 1.95 ml Novozym 188 (from Novozymes A/S).

The FPU activities measured for the enzyme preparations at different temperatures are presented in Figure 14 as percentages of the activity under standard (IUPAC, 1987) conditions (at 50°C).

Results clearly show that the mixtures of the invention show higher overall cellulase activity in elevated (60–70°) temperatures as compared to the state-of art mixtures based on enzymes from *Trichoderma* and *Aspergillus*.

Example 32. Use of the novel beta-glucosidases in preparation of sophorose

A high concentration starch hydrolysate mixture (Nutriose 74/968, Roquette) was treated with *Thermoascus aurantiacus* βG_81/Cel3A enriched enzyme preparation produced as described in Example 21 to produce a sugar mixture containing appreciable amounts of cellulase inducer (sophorose) to overcome the glucose repression.

The Ta βG_81/Cel3A enriched enzyme preparation was added to a 70% (w/w) Nutriose solution to a final concentration of 1 g total protein /litre. The container of the mixture was incubated in a water bath at 65°C for 3 days with constant stirring and used as a carbon source in a shake flask medium for two different *Trichoderma*-strains (A47 and Rut-C30). The effect of the enzyme treatment was measured as an endoglucanase activity formed during a 7 days shake flask cultivation. As a reference cultivations were performed under the same conditions with untreated Nutriose as a carbon source. More than two-fold increase in the activities was obtained in the shake flask cultivations performed on Ta βG_81/Cel3A pretreated Nutriose media with the strains tested. Results are shown in Figure 15.

List of deposited organisms

Strain	Plasmid contained	Deposition authority	Deposition date	Deposition number
Acremonium thermophilum ALKO4245	-	CBS[(1)]	20 Sep 2004	CBS 116240
Thermoascus aurantiacus ALKO4242	-	CBS[(1)]	20 Sep 2004	CBS 116239
Chaetomium thermophilum ALKO4265	-	CBS[(2)]	Nov 8, 1995	CBS 730.95[(4)]
Escherichia coli	pALK1635	DSMZ[(3)]	16 Sep 2004	DSM 16723
Escherichia coli	pALK1642	DSMZ	16 Sep 2004	DSM 16727
Escherichia coli	pALK1646	DSMZ	16 Sep 2004	DSM 16728
Escherichia coli	pALK1861	DSMZ	16 Sep 2004	DSM 16729
Escherichia coli	pALK1715	DSMZ	16 Sep 2004	DSM 16724
Escherichia coli	pALK1723	DSMZ	16 Sep 2004	DSM 16725
Escherichia coli	pALK1725	DSMZ	16 Sep2004	DSM 16726
Escherichia coli	pALK1904	DSMZ	13 May 2005	DSM 17323
Escherichia coli	pALK1908	DSMZ	13 May 2005	DSM 17324
Escherichia coli	pALK1925	DSMZ	13 May 2005	DSM 17325
Escherichia coli	pALK1926	DSMZ	13 May 2005	DSM 17326
Escherichia coli	pALK2001	DSMZ	18 Oct 2005	DSM 17667
Escherichia coli	pALK2010	DSMZ	18 Nov 2005	DSM 17729

[(1)] the Centralbureau Voor Schimmelcultures at Uppsalalaan 8, 3584 CT, Utrecht, the Netherlands

[(2)] the Centralbureau Voor Schimmelcultures at Oosterstraat 1, 3742 SK BAARN, The Netherlands

[(3)]Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ), Mascheroder Weg 1 b, D-38124 Braunschweig, Germany

[(4)][After termination of the current deposit period, samples will be stored under agreements as to make the strain available beyond the enforceable time of the patent.]

References

Altschul S., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990) Basic local alignment search tool. J. Mol. Biol. 215: 403-410.

Badger, P.C. (2002) Ethanol from cellulose: a general review. In Trends in new crops and new uses. J. Janick and A. Whipkey (eds.). ASHS Press, Alexandria, VA, USA, pp. 17-21.

Bailey M. J. and K. M. H. Nevalainen (1981) Induction, isolation and testing of stable *Trichoderma reesei* mutants with improved production of solubilizing cellulase. Enz Microbiol Technol. 3: 153-157.

Bailey, M.J., Biely, P. and Poutanen, K. (1992) Interlaboratory testing for assay of xylanase activity. J Biotechnol. 23:257-270.

Bailey, M. J. and Linko, M. (1990) Production of β-galactosidase by *Aspergillus oryzae* in submerged bioreactor cultivation. J. Biotechnol. 16:57-66.

Bailey M. J. and Poutanen K. (1989) Production of xylanases by strains of *Aspergillus*. Appl. Microbiol. Biotechnol. 30:5-10.

Bailey M., Siika-aho M., Valkeajärvi A. and Penttilä M. (1993) Hydrolytic properties of two cellulases of *Trichoderma reesei* expressed in yeast. Biotehnol. Appl. Biochem 17: 65-76.

Bendtsen J.D., Nielsen H., von Heijne G. and Brunak S. (2004) Improved prediction of signal peptides: SignalP 3.0. J. Mol. Biol. 340: 783-795.

Bernfeld, B. (1955) Amylases, α and β. In: Methods in Enzymology, vol. 1, Eds. Colowick, S.P. and Kaplan, N.O. Academic Press, New York, pp. 149-158.

Biely P., Vrsanska M., Tenkanen M., Kluepfel D. (1997) Endo-beta-1,4-xylanase families: differences in catalytic properties. Journal of Biotechnology 57: 151-166.

Coen, D.M. (2001) The polymerase chain reaction. In: Ausubel, F.M., Brent, R., Kingston, R.E., More, D.D., Seidman, J.G., Smith, K. and Struhl, K. (eds.) Current protocols in molecular biology. John Wiley & Sons. Inc., Hoboken, USA.

Gasteiger, E., Gattiker A., Hoogland C., Ivanyi I., Appel R. D. and Bairoch A. (2003) ExPASy: the proteiomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31: 3784-3788.

Gellissen, G. (ed.) (2005) Production of recombinant proteins. Novel microbial and eukaryotic expression systems. Wiley-VCH Verlag Gmbh&Co. Weinheim, Germany.

Gill, S.C, and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal. Biochem. 182: 319-326.

Gupta, R., E. Jung and S. Brunak. (2004) Prediction of N-glycosylation sites in human proteins, In preparation.

Haakana H., Miettinen-Oinonen A., Joutsjoki V., Mäntylä A., Suominen P, and Vehmaanperä J. (2004) Cloning of cellulase genes from Melanocarpus albomyces and their efficient expression in *Trichoderma reesei*. Enz. Microbiol. Technol. 34: 159-167.

Henrissat B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 280: 309-316.

Henrissat B. and Bairoch A. (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 293: 781-788.

Henrissat B. and Bairoch A. (1996).Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316: 695-696

Henrissat B., Teeri T. T. and Warren R. A. J. (1998) A scheme for designating enzymes that hydrolyse the polysaccharides in the cell wall of plants. FEBS Letters 425: 352-354.

Hong J., H. Tamaki, K. Yamamoto, and Kumagai H. (2003a) Cloning of a gene encoding a thermo-stabile endo-β-1,4-glucanase from *Thermoascus aurantiacus* and its expression in yeast. Biotech. Letters 25: 657-661.

Hong J., Tamaki H., Yamamoto K. and Kumagai H. (2003b) Cloning of a gene encoding thermostable cellobiohydrolase from *Thermoascus aurantiacus* and its expression in yeast. Appl. Microbiol. Biotechnol. 63: 42-50.

IUPAC (International Union of Pure and Applied Chemistry) (1987). Measurement of cellulase activities. Pure and Appl. Chem. 59: 257-268.

Joutsjoki, V. V., Torkkeli T.K. and Nevalainen K. M. H. (1993) Transformation of Trichoderma reesei with the Hormoconis resinae glucoamylase P (gamP) gene: production of a heterologous glucoamylase by *Trichoderma reesei*. Curr. Genet. 24: 223-228.

Karhunen T., Mäntylä A., Nevalainen K. M. H. and Suominen P. L. (1993) High frequency one-step gene replacement in *Trichoderma reesei*. I. Endoglucanase I overproduction. Mol. Gen. Genet. 241: 515-522.

Kurabi A., Berlin A, Gilkes N., Kilburn D., Markov A., Skomarovsky A., Gusakov A., Okunev O., Sinitsyn A., Gregg D. Xie D. and Saddler J. (2005) Enzymatic hydrolysis of steam-exploded and ethanol organosolv-pretreated Douglas-Fir by novel and commercial fungal cellulases. Appl. Biochem and Biotechn. Vol 121-124: 219-229.

Lever, M. (1972) A new reaction for colorimetric determination of carbohydrates. Anal. Biochem., 47: 276-279.

Lo Leggio, L., Kalogiannis S., Bhat M.K., and Pickersgill R.W. (1999) High resolution structure and sequence of the *T. aurantiacus* xylanase I: implications for evolution of thermostability in family 10 xylanases and enzymes with (beta) alpha-barrel architecture. Proteins 36(3): 295-306.

Lowry, O., Rosenbrough, N., Farr, A. and Randall, R. (1951) Protein measuremen with the Folin phenol reagent. J. Biol. Chem. 193: 265-275.

Needleman S. and Wunsch C. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. Journal of Molecular Biology 48, 443-453.

Nielsen H., Engelbrecht J., Brunak S. and von Heijne G. (1997) Identification of prokaryotic and eykaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10: 1-6.

Paloheimo M., Mäntylä A., Kallio J., and Suominen P. (2003) High-yield production of a bacterial xylanase in the filamentous fungus *Trichoderma reesei* requires a carrier polypeptide with an intact domain structure. Appl. Env. Microbiol. 69: 7073-7082.

Parry N., Beever D., Owen E., Nerinckx W. Claeyssens M, Van Beeumen J. and Bhat M. (2002) Biochemical characterization and mode of action of a thermostable endoglucanase purified from *Thermoascus aurantiacus*, Arch. of Biochem. and Biophys. 404: 243-253.

Penttilä M., Nevalainen H., Rättö M., Salminen E. and Knowles J. (1987) A versatile transformation system for the cellulolytic filamentous fungus *Trichoderma reesei*. Gene 61:155-164.

Raeder U. and Broda P. (1985) Rapid preparation of DNA from filamentous fungi. Lett. Appl. Microbiol. 1: 17-20.

Rice P, Longden I and Bleasby A. (2000). EMBOSS: The European Molecular Biology Open Software Suite. Trends in Genetics 16:276-277.

Sambrook J., Fritsch E. F. and Maniatis T. (1989) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, New York, US.

Sambrook J. and Russell D. W. (2001) Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, New York, US.

Srisodsuk M, Reinikainen T, Penttilä M and Teeri T. (1993) Role of the interdomain linker peptide of *Trichoderma reesei* cellobiohydrolase I in its interaction with crystalline cellulose. J. Biol. Chem. Oct 5;268(28): 20756-61.

Sundberg, M., and Poutanen, K. (1991) Purification and properties of two acetylxylan esterases of *Trichoderma reesei*. Biotechnol. Appl. Biochem. 13: 1-11.

Suurnäkki, A., Tenkanen M., Siika-aho, M., Niku-Paavola, M.-L., Viikari, L. and Buchert, J. (2000) *Trichoderma reesei* cellulases and their core domains in the hydrolysis and modification of chemical pulp. Cellulose 7: 189-209.

Tenkanen, M., Puls, J. and Poutanen, K (1992) Two major xylanases of *Trichoderma reesei*. Enzyme Microbiol. Technol. 14: 566-574.

Tomme, P. McRae, S., Wood, T. and Claeyssens, M. (1988) Chromatographic separation of cellulolytic enzymes. Methods in Enzymol. 160: 187-192.

Tuohy M., Walsh J., Murray P., Claeyssens M., Cuffe M., Savage A. and Coughan M. (2002) Kinetic parameters and mode of action of cellobiohydrolases produced by *Talaromyces emersonii*. Biochem. Biophys. Acta 1596: 366–380 (abstract).

Van Petegem *et al* (2002) Atomic resolution structure of major endoglucanase from *Thermoascus aurantiacus*, Biochem. and Biophys. Res. Comm. 296, pp. 161-166.

Van Tilbeurgh, H., Loonties, F., de Bruyne, C. and Claeyssens, M. (1988) Fluorogenic and chromogenic glycosides as substrates and ligands of carbohydrases. Methods Enzymol. 160:45-59.

Wyman, C.E. (2001) Twenty years of trials, tribulations, and research progress in bioethanol technology. Applied Biochemistry and Biotechnology 91-93: 5-21.

Claims

1. A polypeptide having cellulolytic activity and comprising an amino acid sequence having at least 80% identity over the full length to SEQ ID NO: 22.

2. The polypeptide of claim 1 comprising
an amino acid sequence having at least 85% identity over the full length to SEQ ID NO: 22.

3. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
a) the nucleotide sequence of SEQ ID NO: 21;
b) a sequence encoding the polypeptide of claim 1; and
c) the full length complement of the nucleotide sequence of a).

4. A vector, which comprises as a heterologous sequence the polynucleotide of claim 3.

5. The vector of claim 4, which expresses the polypeptide of claim 1.

6. A host cell comprising the vector of claim 4.

7. The host cell of claim 6, which expresses the polypeptide encoded by the heterologous polynucleotide sequence.

8. The host cell of claim 7, which is a strain from the genus *Trichoderma* or *Aspergillus*.

9. An *Escherichia coli* strain having accession number DSM 16725.

10. An enzyme preparation comprising the polypeptide of claim 1 in an admixture with a biologically compatible carrier, in a form of spent culture medium, powder, granules, or liquid.

Date Reçue/Date Received 2022-03-03

11. The enzyme preparation of claim 10, which comprises cellobiohydrolase, endoglucanase, and beta-glucosidase.

12. The enzyme preparation of claim 10 or 11, further comprising at least one of xylanase activity and other enzyme activities selected from the group consisting of proteases, amylases, laccases, lipases, pectinases, esterases and peroxidases.

13. The enzyme preparation of any one of claims 10–12, which further comprises conventional additives.

14. Use of the polypeptide according to claim 1 or 2, or the enzyme preparation according to any one of claims 10-13 in the fuel, textile, detergent, pulp and paper, food, feed or beverage industry.

15. The use according to claim 14, wherein the enzyme preparation is used in treatment of kraft pulp, mechanical pulp, or recycled paper.

16. The use according to claim 14, wherein the enzyme preparation is spent culture medium.

17. A method for preparing a polypeptide having cellulolytic activity and comprising an amino acid sequence having at least 80% identity over the full length to SEQ ID NO: 22,

said method comprising transforming a host cell with a vector encoding said polypeptide, and culturing said host cell under conditions enabling expression of said polypeptide, and recovering and purifying the polypeptide produced.

Date Reçue/Date Received 2022-03-03

Fig. 1A.

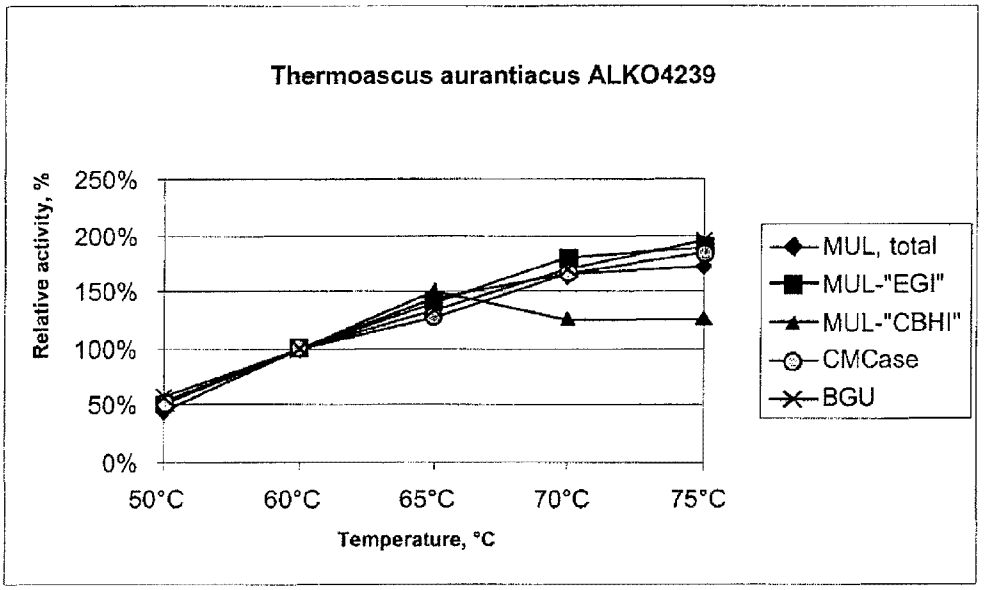

Thermoascus aurantiacus ALKO4239
Relative activity, %
250%
200%
150%
100%
50%
0%
50°C
60°C
65°C
70°C
75°C
Temperature, °C
MUL, total
MUL-"EGI"
MUL-"CBHI"
CMCase
BGU

Fig. 1B.

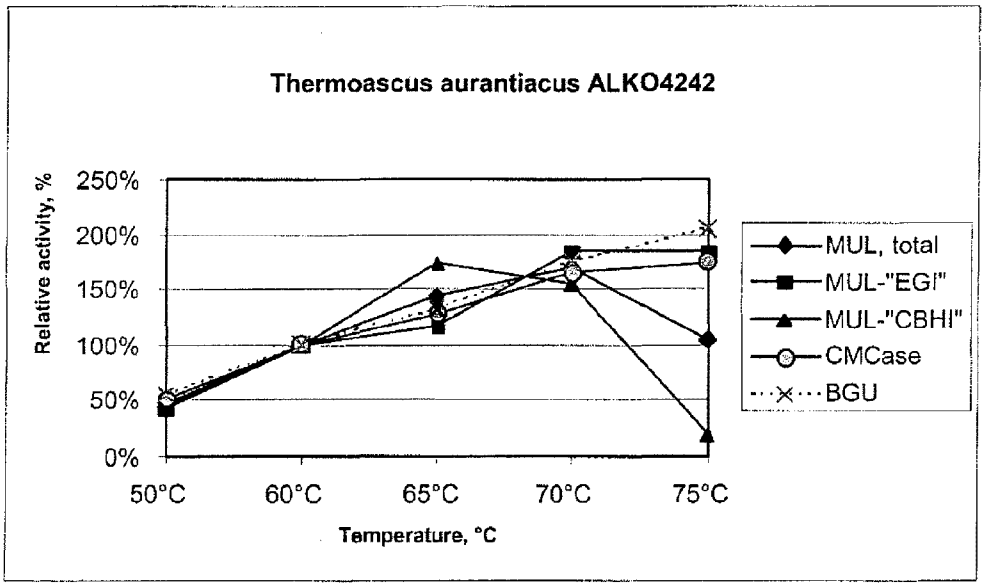

Thermoascus aurantiacus ALKO4242
Relative activity, %
250%
200%
150%
100%
50%
0%
50°C
60°C
65°C
70°C
75°C
Temperature, °C
MUL, total
MUL-"EGI"
MUL-"CBHI"
CMCase
BGU

Fig. 1C.

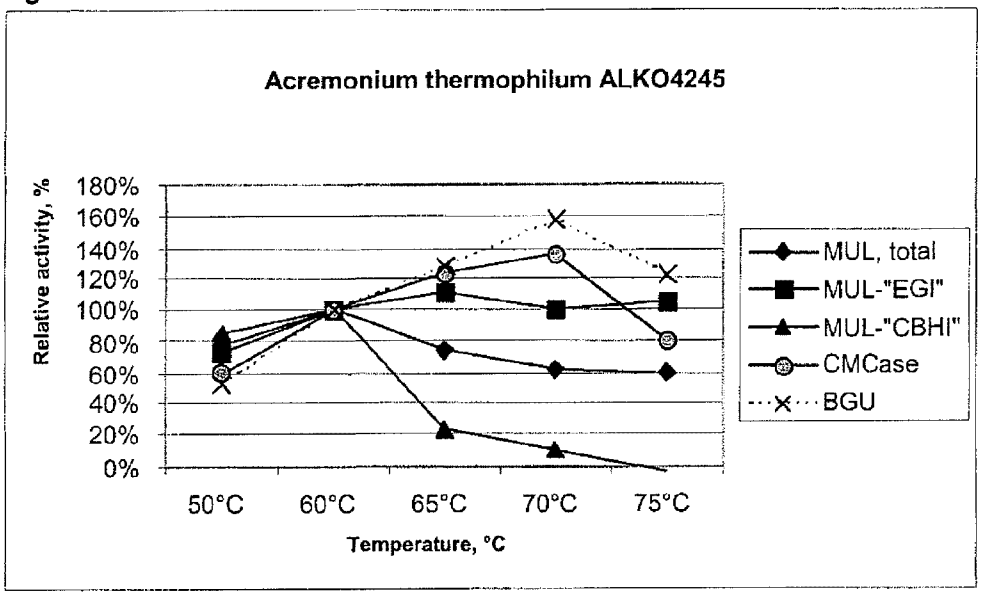

Acremonium thermophilum ALKO4245
Relative activity, %
180%
160%
140%
120%
100%
80%
60%
40%
20%
0%
50°C
60°C
65°C
70°C
75°C
Temperature, °C
MUL, total
MUL-"EGI"
MUL-"CBHI"
CMCase
BGU

Fig. 1D.

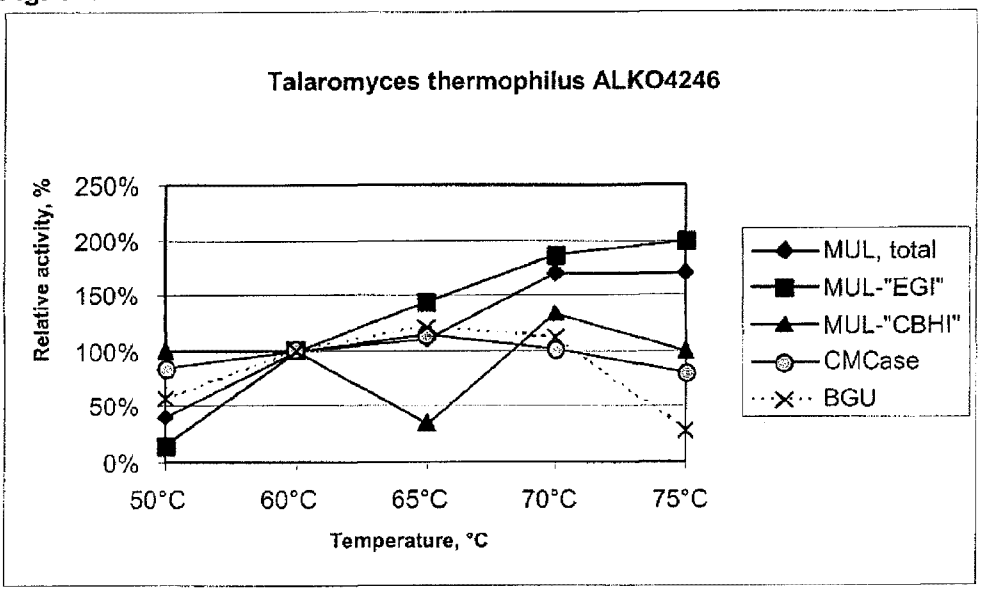

Talaromyces thermophilus ALKO4246
Relative activity, %
250%
200%
150%
100%
50%
0%
50°C
60°C
65°C
70°C
75°C
Temperature, °C
MUL, total
MUL-"EGI"
MUL-"CBHI"
CMCase
BGU

Fig. 1E.

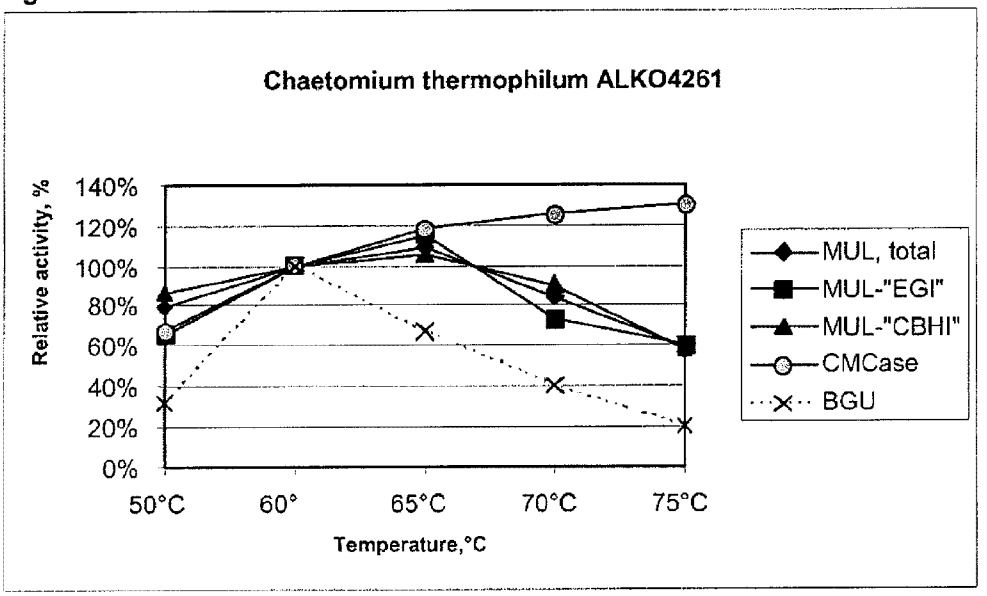

Chaetomium thermophilum ALKO4261
Relative activity, %
140%
120%
100%
80%
60%
40%
20%
0%
50°C
60°
65°C
70°C
75°C
Temperature,°C
MUL, total
MUL-"EGI"
MUL-"CBHI"
CMCase
BGU

Fig. 1F.

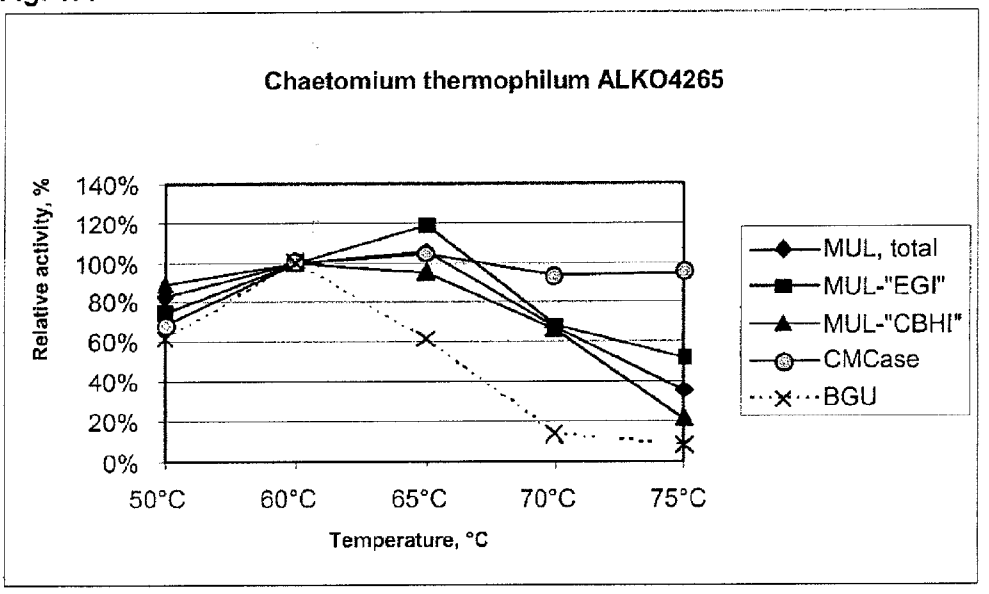

Chaetomium thermophilum ALKO4265
Relative activity, %
140%
120%
100%
80%
60%
40%
20%
0%
50°C
60°C
65°C
70°C
75°C
Temperature, °C
MUL, total
MUL-"EGI"
MUL-"CBHI"
CMCase
BGU

Fig. 2.

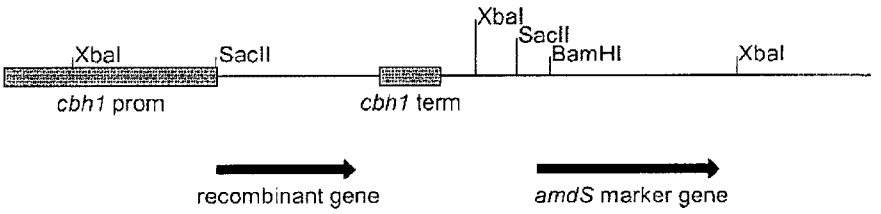

XbaI
SacII
XbaI
SacII
BamHI
XbaI
cbh1 prom
cbh1 term
recombinant gene
amdS marker gene

Fig. 3A.

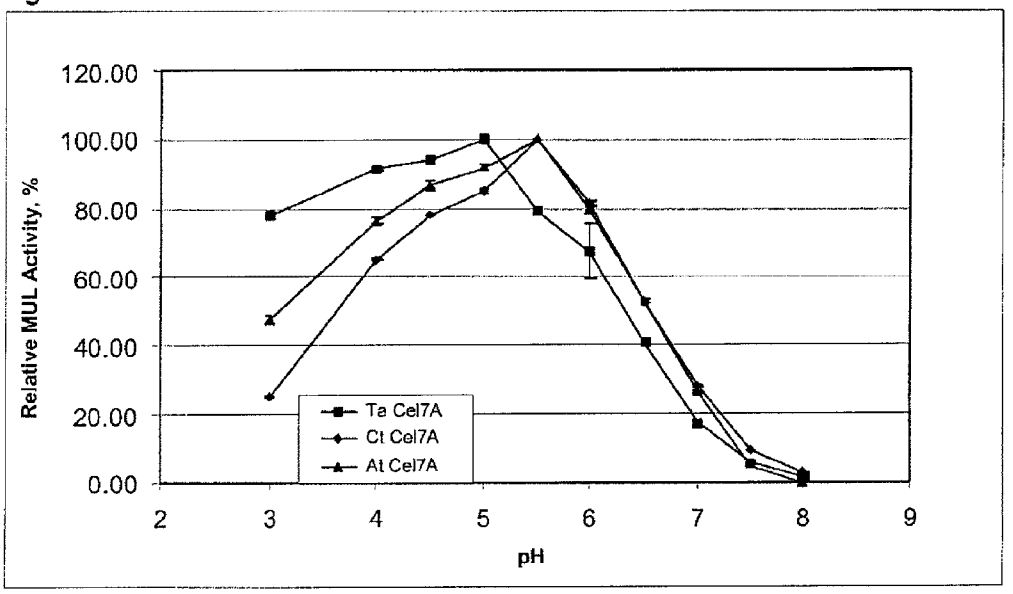

120.00
100.00
80.00
60.00
40.00
20.00
0.00
Relative MUL Activity, %
2
3
4
5
6
7
8
9
pH
Ta Cel7A
Ct Cel7A
At Cel7A

Fig. 3B.

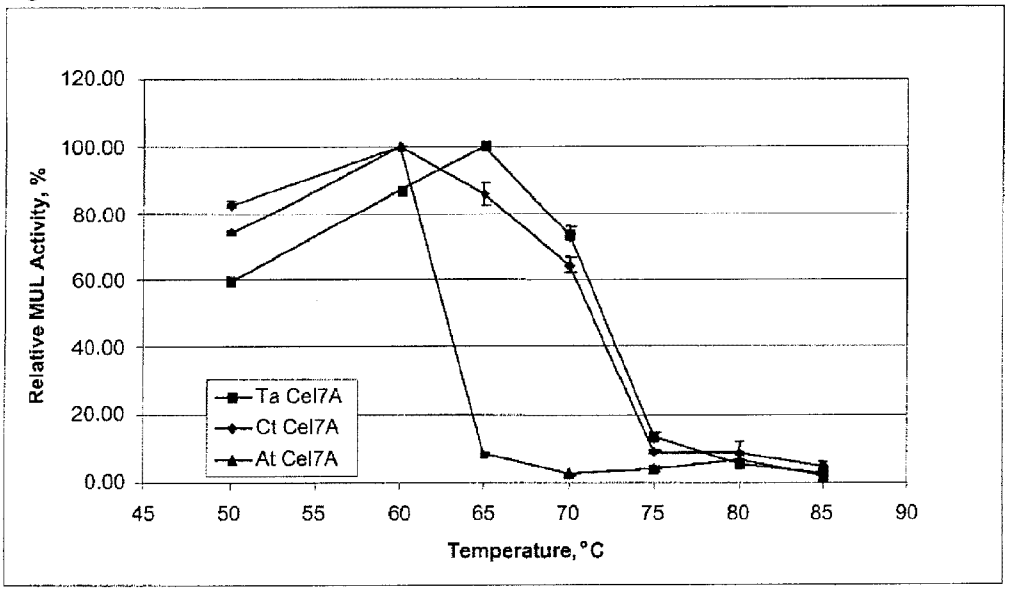

120.00
100.00
80.00
60.00
40.00
20.00
0.00
Relative MUL Activity, %
45
50
55
60
65
70
75
80
85
90
Temperature, °C
Ta Cel7A
Ct Cel7A
At Cel7A

Fig. 4A.

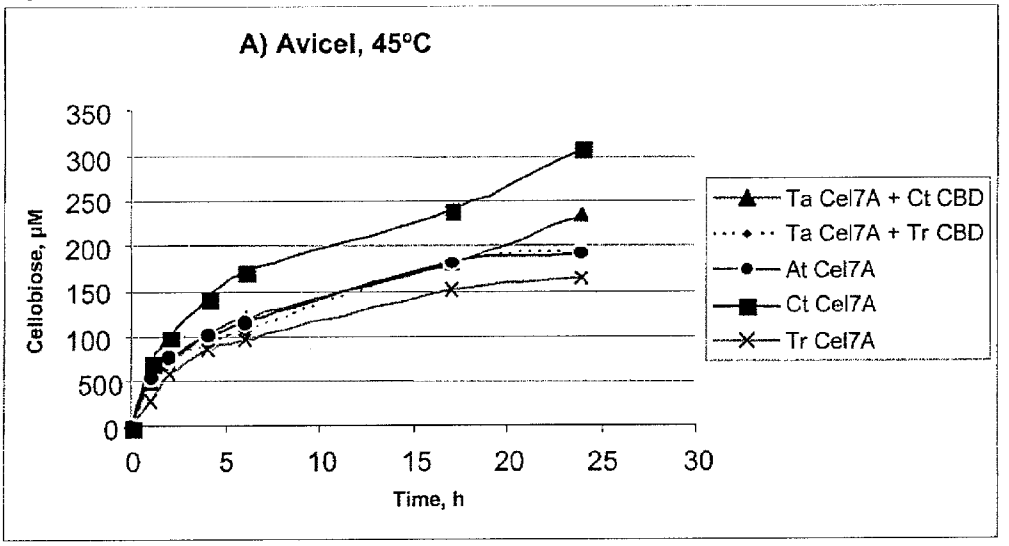

A) Avicel, 45°C
Cellobiose, µM
350
300
250
200
150
100
500
0
0
5
10
15
20
25
30
Time, h
Ta Cel7A + Ct CBD
Ta Cel7A + Tr CBD
At Cel7A
Ct Cel7A
Tr Cel7A

Fig. 4B.

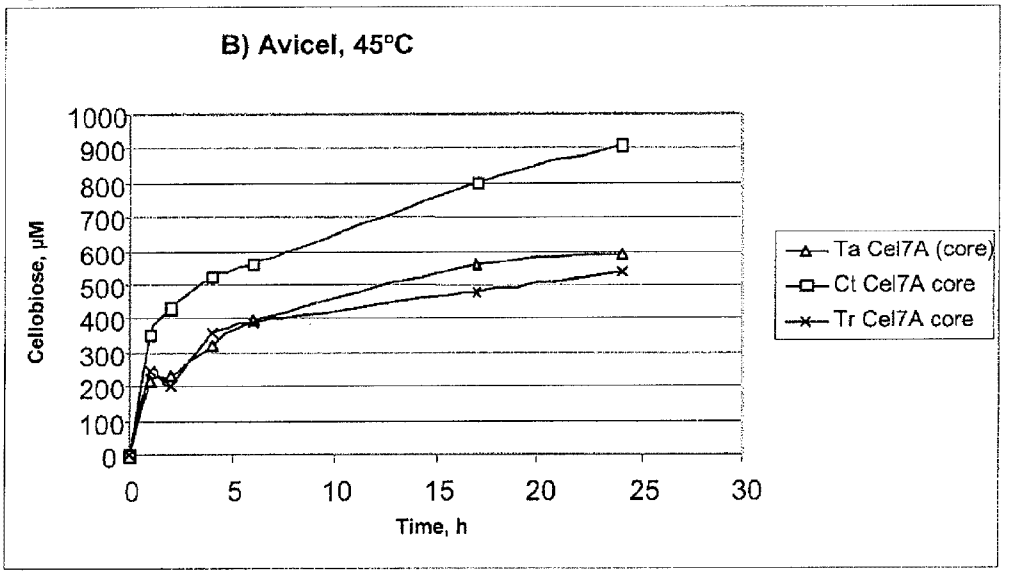

B) Avicel, 45°C
Cellobiose, µM
1000
900
800
700
600
500
400
300
200
100
0
0
5
10
15
20
25
30
Time, h
Ta Cel7A (core)
Ct Cel7A core
Tr Cel7A core

Fig. 5A.

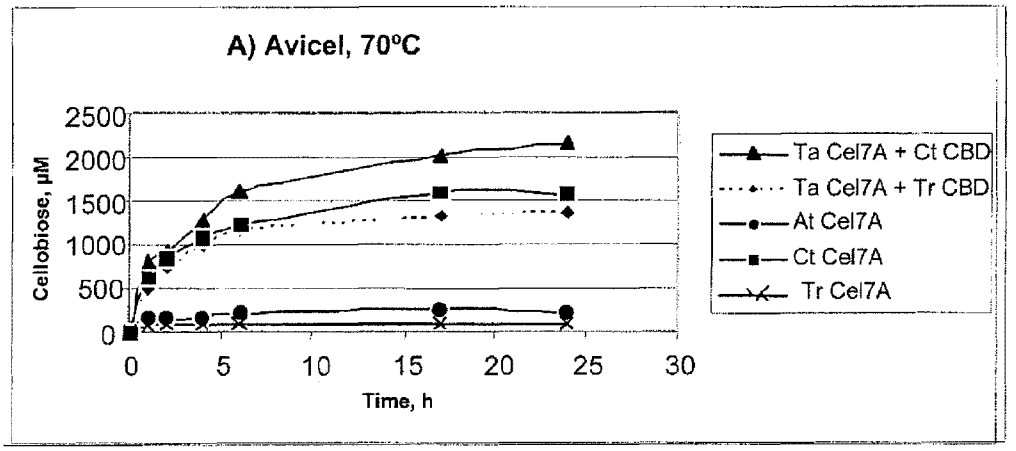

A) Avicel, 70°C
Cellobiose, µM
2500
2000
1500
1000
500
0
0
5
10
15
20
25
30
Time, h
Ta Cel7A + Ct CBD
Ta Cel7A + Tr CBD
At Cel7A
Ct Cel7A
Tr Cel7A

Fig. 5B.

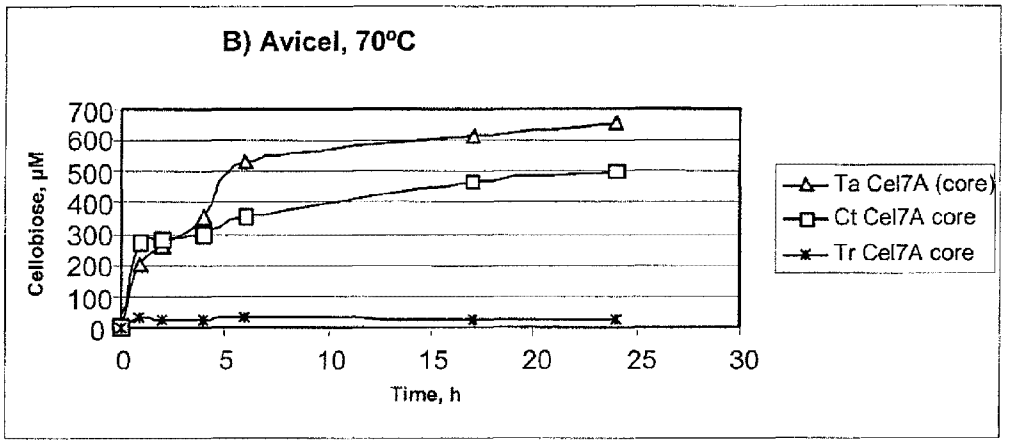

B) Avicel, 70°C
Cellobiose, µM
700
600
500
400
300
200
100
0
0
5
10
15
20
25
30
Time, h
Ta Cel7A (core)
Ct Cel7A core
Tr Cel7A core

Fig. 6A.

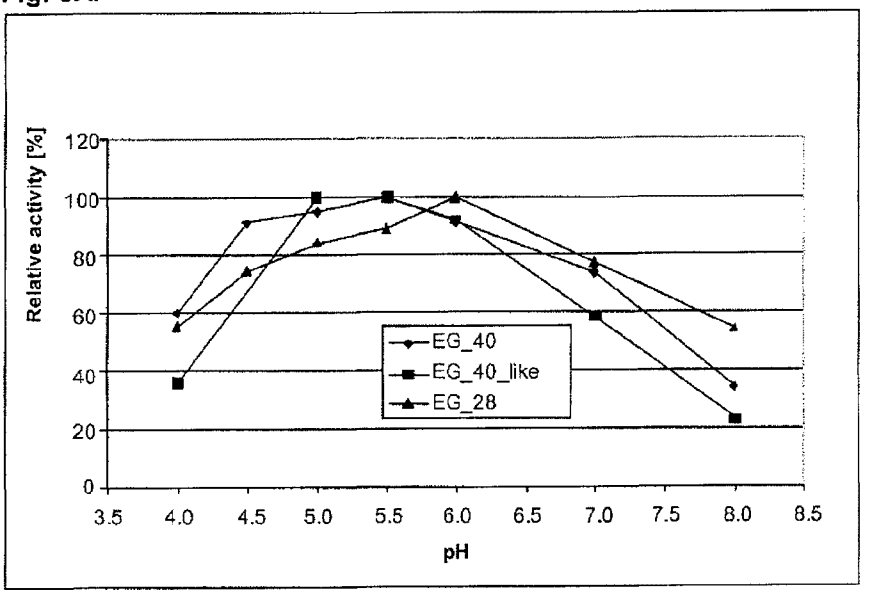

Relative activity [%]
120
100
80
60
40
20
0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
pH
EG_40
EG_40_like
EG_28

Fig. 6B.

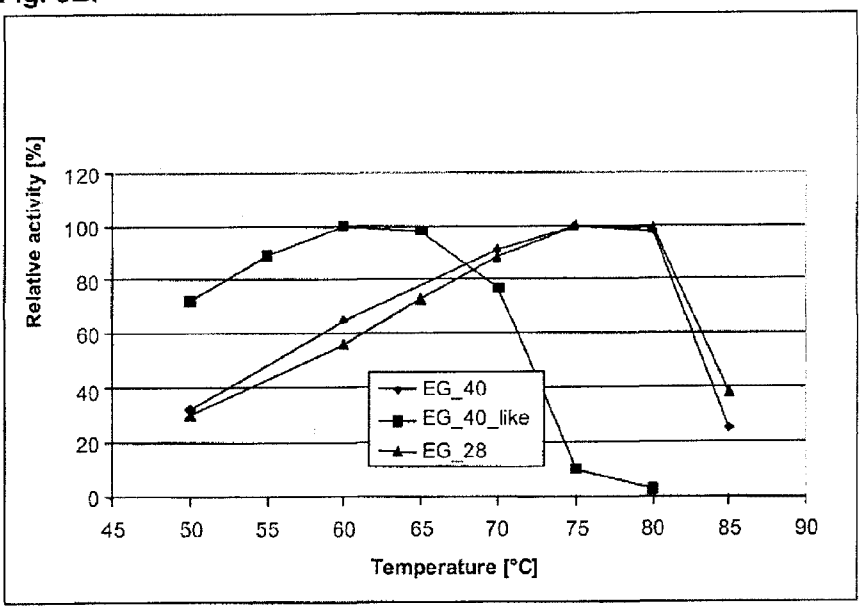

Relative activity [%]
120
100
80
60
40
20
0
45
50
55
60
65
70
75
80
85
90
Temperature [°C]
EG_40
EG_40_like
EG_28

Fig. 7A.

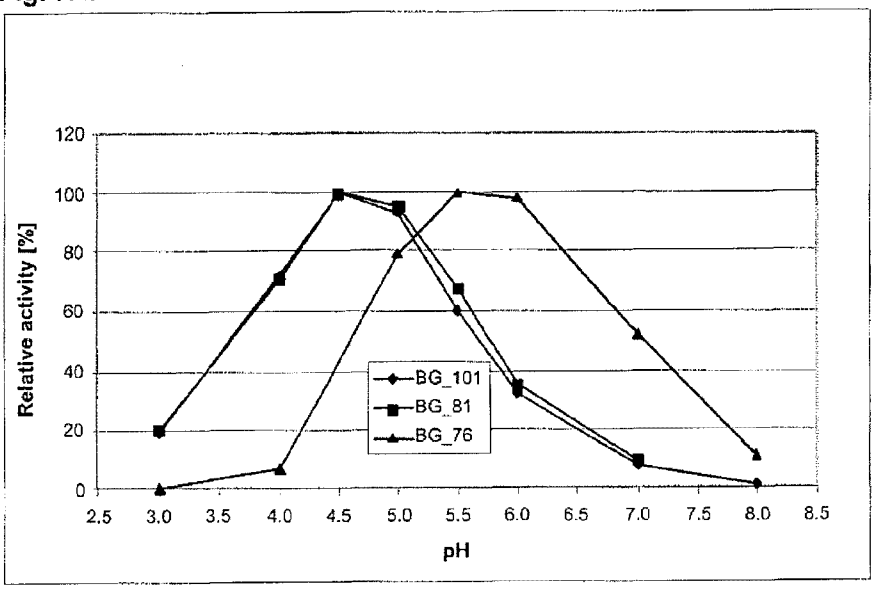

120
100
80
60
40
20
0
Relative activity [%]
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
pH
BG_101
BG_81
BG_76

Fig. 7B.

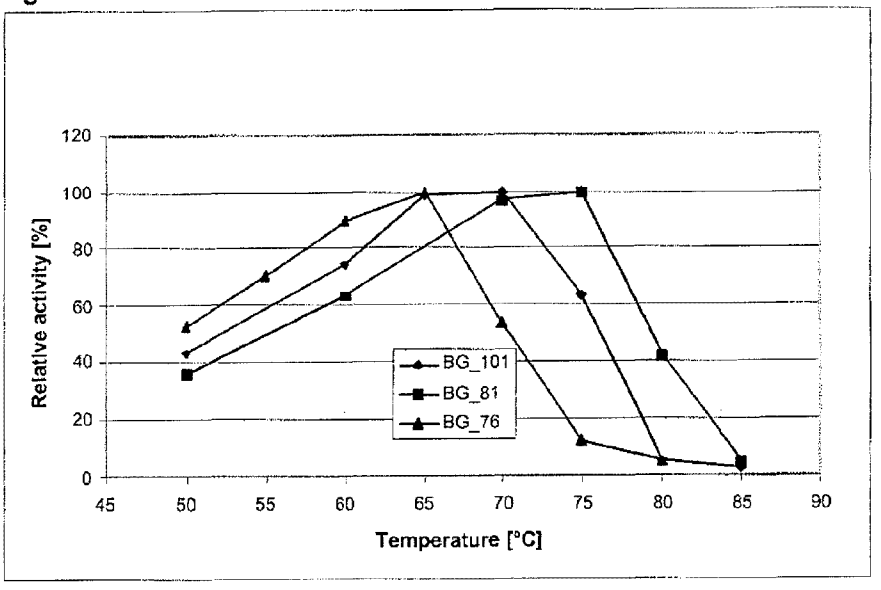

120
100
80
60
40
20
0
Relative activity [%]
45
50
55
60
65
70
75
80
85
90
Temperature [°C]
BG_101
BG_81
BG_76

Fig. 8A.

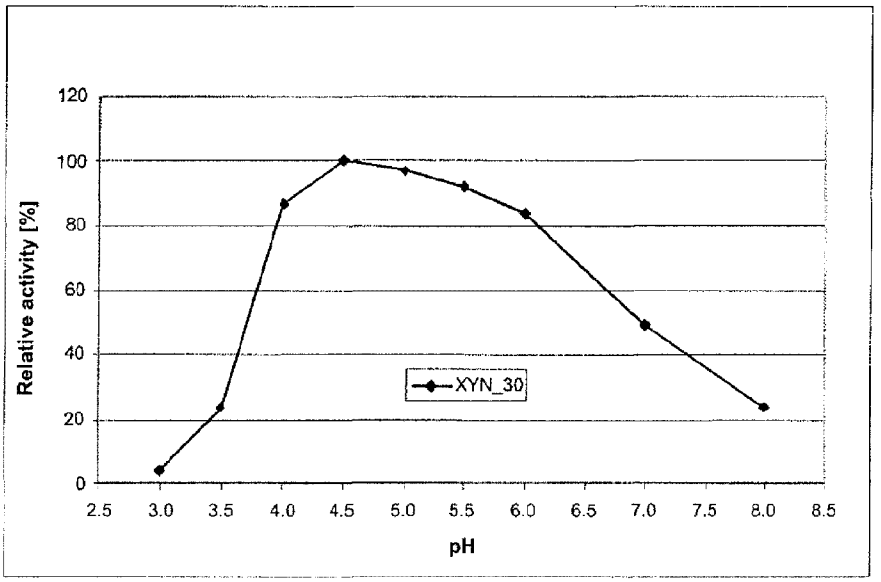

120
100
80
60
40
20
0
Relative activity [%]
XYN_30
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5
8.0
8.5
pH

Fig 8B.

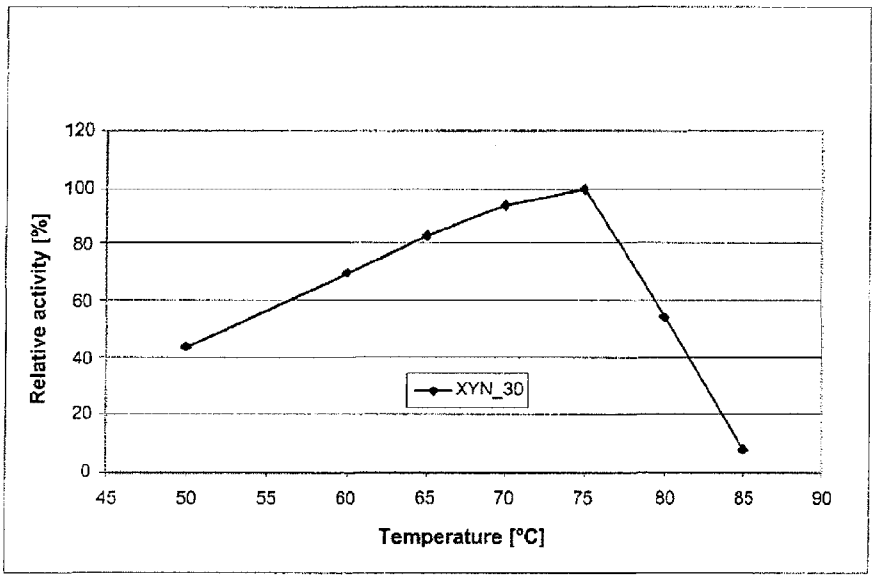

120
100
80
60
40
20
0
Relative activity [%]
XYN_30
45
50
55
60
65
70
75
80
85
90
Temperature [°C]

Fig. 9.

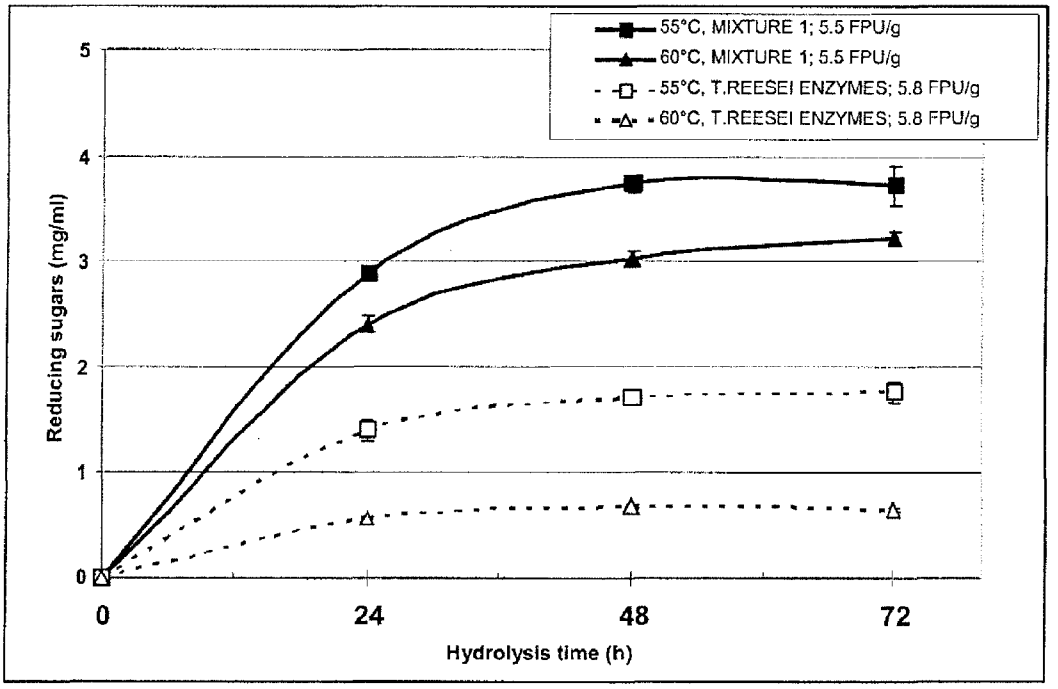

55°C, MIXTURE 1; 5.5 FPU/g
60°C, MIXTURE 1; 5.5 FPU/g
55°C, T.REESEI ENZYMES; 5.8 FPU/g
60°C, T.REESEI ENZYMES; 5.8 FPU/g
Reducing sugars (mg/ml)
0
1
2
3
4
5
0
24
48
72
Hydrolysis time (h)

Fig. 10.

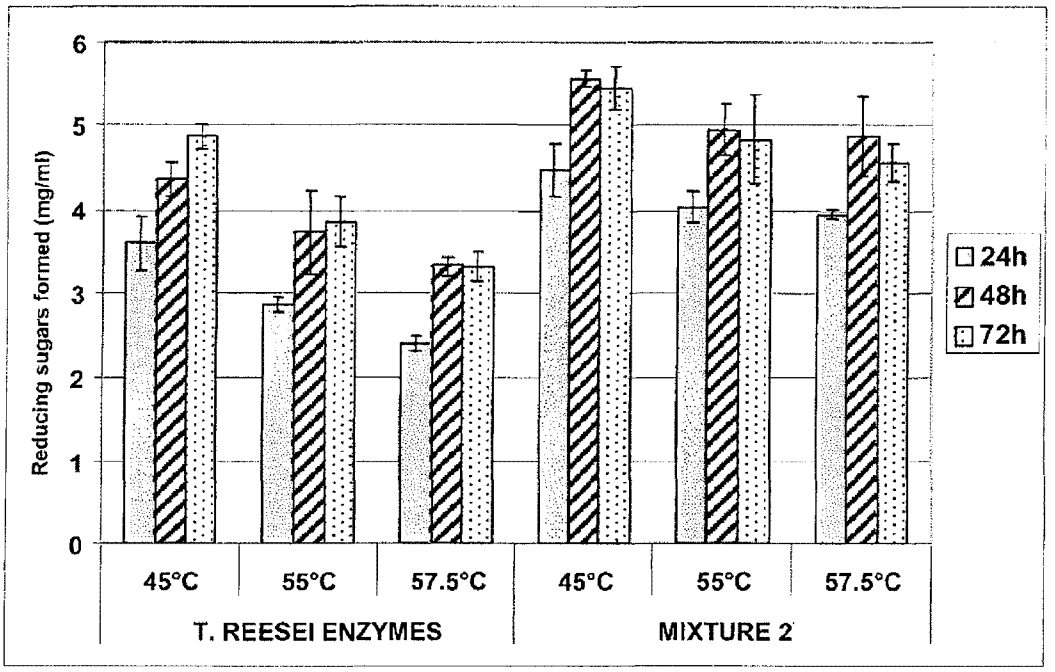

Reducing sugars formed (mg/ml)
0
1
2
3
4
5
6
45°C
55°C
57.5°C
45°C
55°C
57.5°C
T. REESEI ENZYMES
MIXTURE 2
24h
48h
72h

Fig. 11.

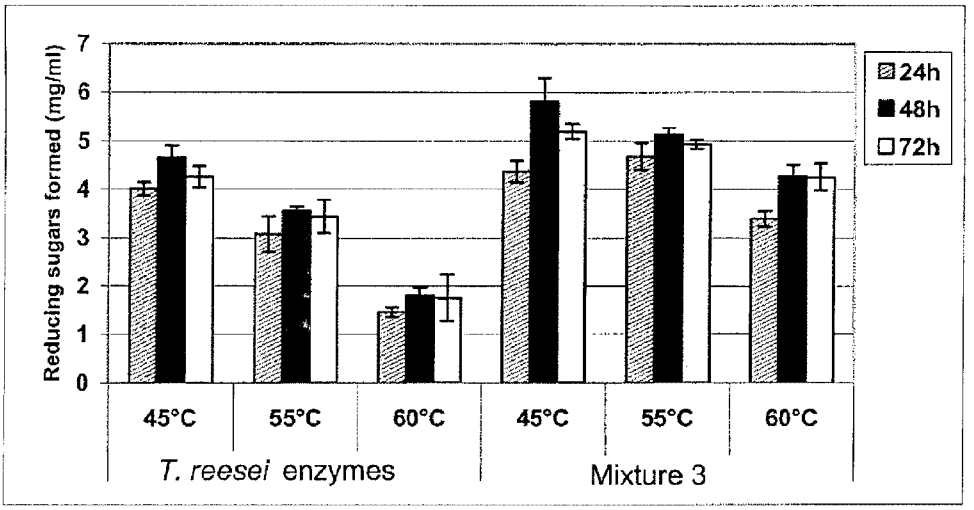

Reducing sugars formed (mg/ml)
0
1
2
3
4
5
6
7
24h
48h
72h
45°C
55°C
60°C
45°C
55°C
60°C
T. reesei enzymes
Mixture 3

Fig. 12.

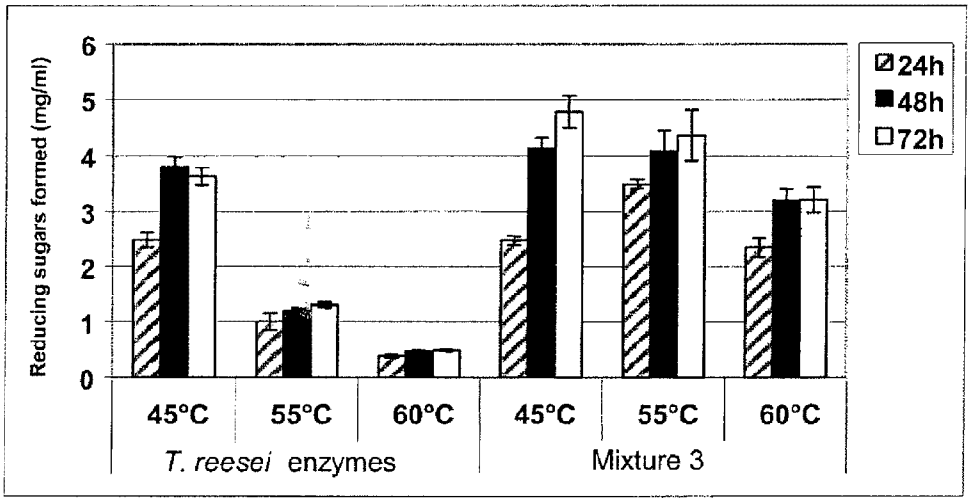

Reducing sugars formed (mg/ml)
0
1
2
3
4
5
6
24h
48h
72h
45°C
55°C
60°C
45°C
55°C
60°C
T. reesei enzymes
Mixture 3

Fig. 13.

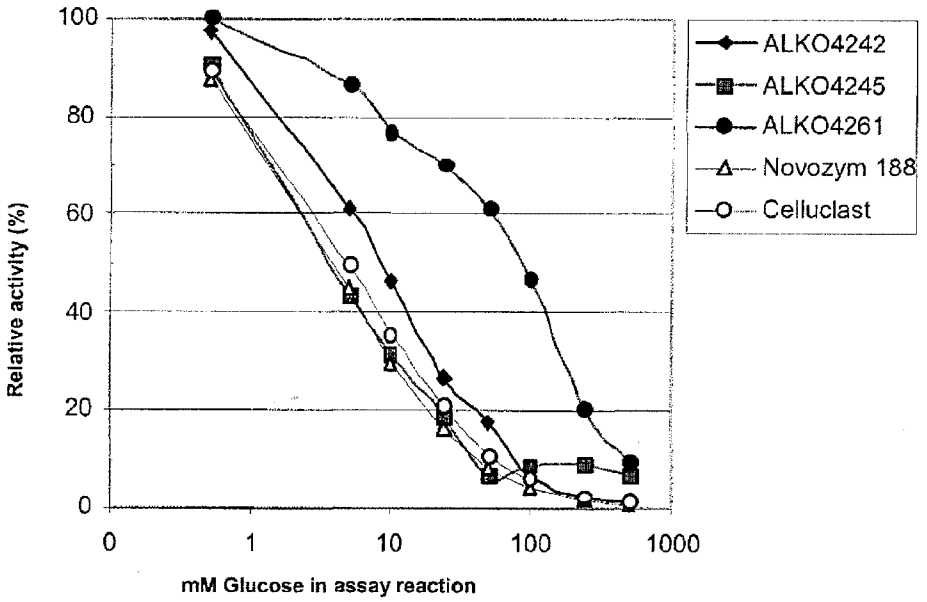

100
80
60
40
20
0
Relative activity (%)
0
1
10
100
1000
mM Glucose in assay reaction
ALKO4242
ALKO4245
ALKO4261
Novozym 188
Celluclast

Fig. 14.

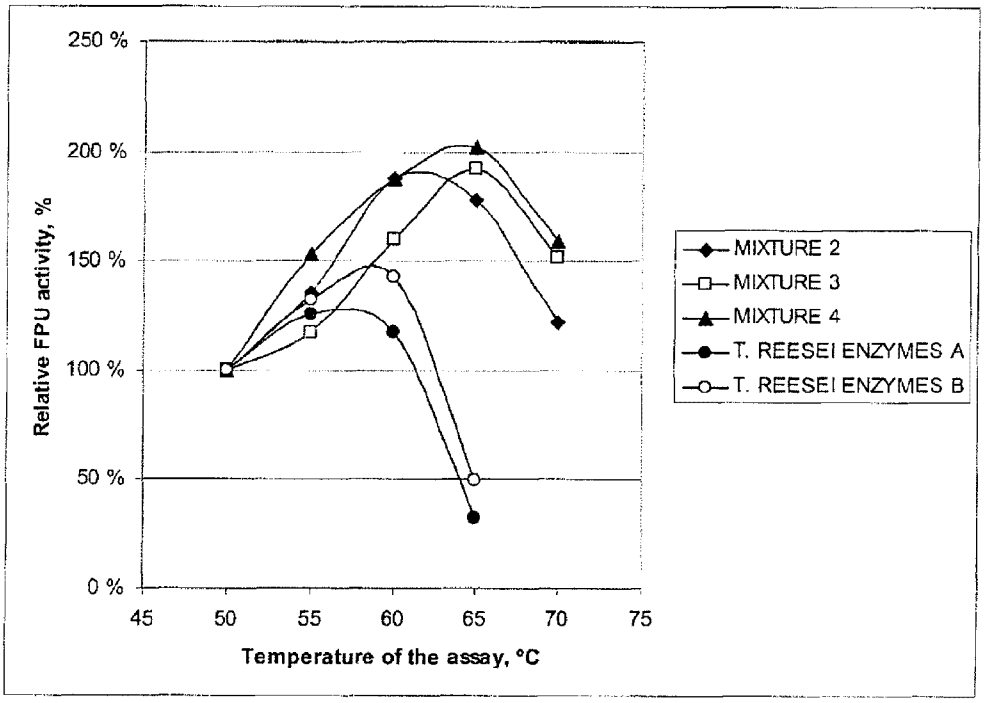
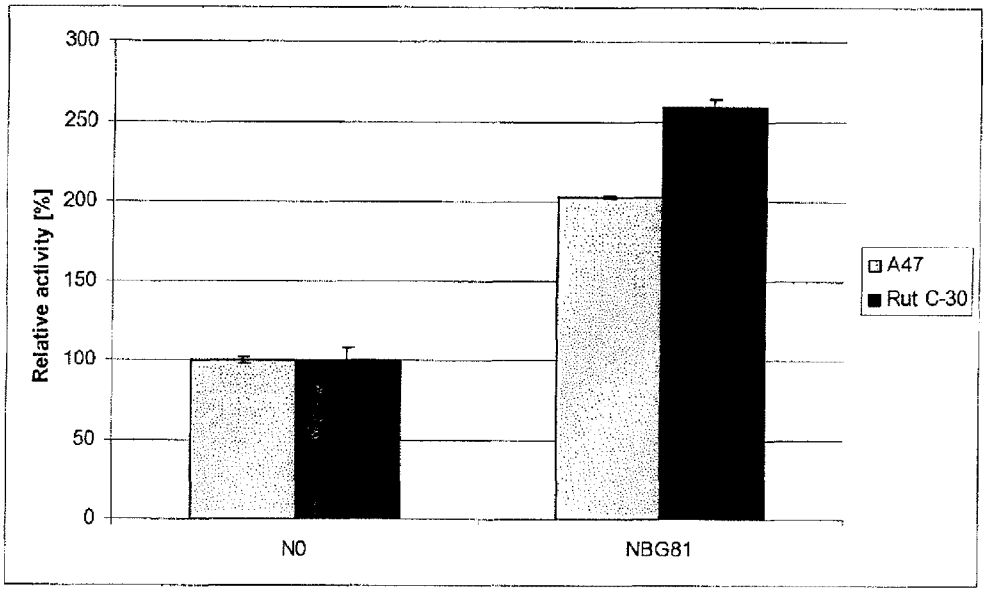

250 %
200 %
150 %
100 %
50 %
0 %
Relative FPU activity, %
45
50
55
60
65
70
75
Temperature of the assay, °C
MIXTURE 2
MIXTURE 3
MIXTURE 4
T. REESEI ENZYMES A
T. REESEI ENZYMES B

Fig. 15.


300
250
200
150
100
50
0
Relative activity [%]
N0
NBG81
A47
Rut C-30