(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 October 2012 (04.10.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/129699 A1

(51) International Patent Classification:
C12N 15/56 (2006.01) *C12N 9/42* (2006.01)
C11D 3/386 (2006.01) *C12P 19/00* (2006.01)
C11D 7/42 (2006.01) *C12P 7/10* (2006.01)
C12N 15/53 (2006.01) *D06M 16/00* (2006.01)
C12N 15/55 (2006.01) *D21C 5/00* (2006.01)
C12N 9/02 (2006.01) *D21C 9/10* (2006.01)
C12N 9/14 (2006.01) *C08H 7/00* (2011.01)
C12N 9/24 (2006.01)

(21) International Application Number:
PCT/CA2012/050200

(22) International Filing Date:
29 March 2012 (29.03.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/470,836 1 April 2011 (01.04.2011) US

(72) Inventors; and
(71) Applicants : **TSANG, Adrian** [CA/CA]; 3816 Draper Avenue, Montréal, Québec H4A 2P1 (CA). **POWLOWSKI, Justin** [CA/CA]; 4555 Montclair Avenue, Montréal, Québec H2B 2J8 (CA). **BUTLER, Gregory** [AU/CA]; Apt 301, 1700 blvd René Lévesque Ouest, Montréal, Québec H3H 2V1 (CA).

(74) Agent: **NORTON ROSE CANADA LLP/S.E.N.C.R.L., S.R.L.**; Suite 2500, 1 place Ville-Marie, Montreal, Québec H3B 1R1 (CA).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

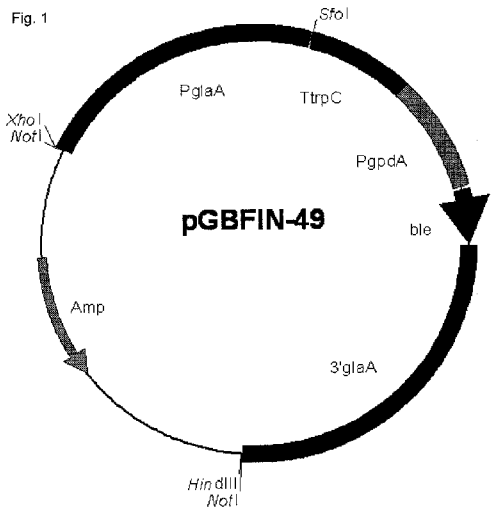
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2012/129699 A1

(54) **Title**: CELL WALL DECONSTRUCTION ENZYMES OF *THERMOMYCES LANUGINOSUS* AND USES THEREOF

Fig. 1



(57) **Abstract**: The present invention relates to novel Thermomyces lanuginosus enzymes or proteins for cell wall deconstruction, polynucleotide sequences encoding the polypeptides according to the invention, a production process for the enzymes according to the invention and the use of the enzymes according to the invention in various industrial processes.

CELL WALL DECONSTRUCTION ENZYMES OF THERMOMYCES LANUGINOSUS AND USES THEREOF

Field of the invention

The invention relates to newly identified polynucleotide sequences comprising genes that encode novel cell wall deconstruction enzymes. The enzymes may be isolated from the fungus, *Thermomyces lanuginosus* ATCC strain 200065. The invention features the full length coding sequences of the novel genes, the genomic sequences of each gene, as well as the amino acid sequences of the full-length functional proteins and functional equivalents of the genes or the amino acid sequences. The invention also relates to methods of using these proteins in industrial processes. Also included in the invention are cells transformed with a polynucleotide according to the invention suitable for producing these proteins and cells wherein a protein according to the invention is genetically modified to enhance or reduce its activity and/or level of expression.

Background of the invention

The present invention relates to novel cell wall deconstruction enzymes, suitable for use in several industrial applications, for example in food applications, such as for example cereal-based food products, in the textile industry for the treatment of cellulose-based fabrics; in the feed-enzyme industry such as for example increasing the digestibility of nutrients; in the pulp and paper industry such as enhancing bleachability of the pulp; in the waste treatment industry for the decolorization of synthetic dyes; and in the bioethanol industry such as for example improving the ethanol yield and increasing the efficiency and economy of ethanol production.

The conversion of biomass into second-generation biofuels, driven by the limited availability of fossil fuels, is heavily dependent on inexpensive and effective enzymes for the conversion of lignocellulose to ethanol. Cellulase enzyme cocktails require the concerted action of endoglucanases, cellobiohydrolases, and beta-glucosidases. The current cost of cellulose enzymes is too high for bioethanol to compete economically with fossil fuels: lowered cellulose costs may

result from the discovery of cellulase enzymes with higher specific activity, lower production costs, or greater compatibility with processing conditions including temperature, pH and the presence of inhibitors in the biomass, or produced as the result of biomass pre-treatment.

Conversion of plant biomass to glucose may also be enhanced by supplementing cellulose cocktails with enzymes that degrade the other components of biomass, including hemicelluloses, pectins and lignins, and their linkages, to improve the accessibility of cellulose to the cellulase enzymes. These enzymes include: xylanases, mannanases, arabinanases, esterases, glucuronidases, xyloglucanases and arabinofuranosidases for hemicelluloses; lignin peroxidases, manganese-dependent peroxidases, versatile peroxidases, and laccases for lignin; and pectate lyase, pectin lyase, polygalacturonase, pectin acetyl esterase, alpha-arabinofuranosidase, beta-galactosidase, galactanase, arabinanase, rhamnogalacturonase, rhamnogalacturonan lyase, and rhamnogalacturonan acetyl esterase, xylogalacturonosidase, xylogalacturonase, and rhamnogalacturonan lyase. Additionally, glycoside hydrolase family 61 (GH61) proteins have been shown to stimulate the activity of cellulase preparations.

The enzymes described may also be useful for other purposes in processing biomass. The lignin modifying enzymes may be used to alter the structure of lignin to produce novel materials, and hemicelluases may be employed to produce 5-carbon sugars from hemicelluloses, which may then be further converted to chemical products.

There is also a need for improved enzymes for feed and food processing applications. Cereal-based food products such as pasta, noodles and bread can be prepared from dough which is usually made from the basic ingredients (cereal) flour, water and optionally salt. As a result of a consumer-driven need to replace the chemical additives by more natural products, several enzymes have been developed with dough and/or cereal-based food product improving properties and which are used in all possible combinations depending on the specific application conditions. Suitable enzymes include xylanase, starch degrading enzymes, oxidizing enzymes, fatty material splitting enzymes, protein degrading, and modifying or crosslinking enzymes. Many of these enzymes are also used for treating animal feed or animal feed additives, to make them more digestible or to improve their nutritional quality. Amylases are used for the conversion of plant

starches to glucose. Pectin-active enzymes are used in fruit processing, for example to increase the yield of juices, and in fruit juice clarification, as well as in other food processing steps.

Object of the invention

It is an object of the invention to provide novel polynucleotides encoding novel cell wall deconstruction enzymes. A further object is to provide naturally and recombinantly produced cell wall deconstruction enzymes as well as recombinant strains producing these. Also fusion polypeptides are part of the invention as well as methods of making and using the polynucleotides and polypeptides according to the invention.

Summary of the invention

The invention provides for novel polynucleotides encoding novel cell wall deconstruction enzymes.

In particular, the invention provides for polynucleotides having a nucleotide sequences that hybridize preferably under high stringent conditions to the complement of any one sequence according to SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, or a functional equivalent thereof. Consequently, the invention provides nucleic acids that are at least 30%, preferably at least 40%, 50%, 60%, or 70%, more preferably at least 80% or 90%, even more preferably at least 95%, 96%, 97%, 98% or 99% homologous to the sequences listed above. Actual percentage will differ per gene/protein depending on closest known homologues in prior art.

In one embodiment the invention provides for an isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising nucleic acid sequence as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92,

94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, or a functional equivalent thereof;

b) a polynucleotide encoding a polypeptide comprising a polypeptide sequence as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, or a functional equivalent thereof;

c) a polynucleotide hybridizing under stringent conditions to the polynucleotide of a) or b); and

d) a polynucleotide hybridizing under stringent conditions to the reverse complement of the polynucleotide of a) or b), said polynucleotide.

In another embodiment the invention provides for such an isolated polynucleotide obtainable from a fungus, in particular *Thermomyces* is preferred and even more preferred *Thermomyces lanuginosus*.

Still in another embodiment, the invention provides for a vector comprising the polynucleotide of the present invention. The vector may further comprise a regulatory sequence operatively linked to the polynucleotide for expression of same in a suitable host cell, preferably a filamentous fungus.

In a further embodiment, the invention provides a host cell comprising the polynucleotide or the vector of the present invention.

In another embodiment the invention provides an isolated polypeptide as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, or a functional equivalent thereof. Such polypeptide is preferably obtainable from *Thermomyces lanuginosus*. In an alternate embodiment, the polypeptide is obtainable by expressing the polynucleotide or the vector of the present invention in an appropriate host cell.

Still in another embodiment, the invention provides a method for manufacturing the polypeptide comprising the steps of transforming a suitable host cell with the isolated polynucleotide or the vector of the present invention, culturing said cell under conditions allowing expression of said polynucleotide and optionally purifying the encoded polypeptide from said cell or culture medium.

The polypeptide of the present invention can be used in the preparation of a food product, in the preparation of a detergent, in the preparation of an animal feed, for prebleaching kraft pulp, for processing lignin, for producing ethanol, for treating textiles or dyed textiles or for degrading biomass or pretreated biomass.

The invention provides a process for degrading biomass or pretreated biomass to sugars wherein an enzyme is used comprising a polypeptide having

a. a polypeptide sequence as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132;

b. a polypeptide that is at least 60%, preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, 95%, 96%, 97%, 98% or 99% homologous to the any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132;

c. a polypeptide sequence encoded by nucleic acid sequence as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, or nucleic acids that are at least 60%, preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, 95%, 96%, 97%, 98% or 99% homologous to any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130;

d. a polypeptide sequence encoded by a nucleic acid sequence

hybridizing under stringent conditions to the polynucleotide as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130; or

e. a polypeptide sequence encoded by a nucleic acid sequence hybridizing under stringent conditions to the reverse complement of a polynucleotide as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

Preferably the polypeptide is obtainable from *Thermomyces lanuginosus*. The enzyme is preferably a cellulase-enhancing protein, a glycoside hydrolase, more preferably a glycoside of the GH61 family, a glycosidase (for example a beta-glucosidase), an endoglucanase (for example endoglucanase-1), a beta-hexosaminidase, a xylanase (for example an endo-1,4-beta-xylanase) a laccase, a polygalacturonase or a xyloglucan:xyloglucosyltransferase.

The invention provides for novel polynucleotides encoding novel cell wall deconstruction enzymes.

In particular, the invention provides for polynucleotides having a nucleotide sequences that hybridize preferably under high stringent conditions to the complement of any one sequence according to SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130. Consequently, the invention provides nucleic acids that are at least 30%, preferably at least 40%, 50%, 60%, or 70%, more preferably at least 80% or 90%,

even more preferably at least 95%, 96%, 97%, 98% or 99% homologous to the sequences listed above. Actual percentage will differ per gene/protein depending on closest known homologues in prior art.

In one embodiment the invention provides for an isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising nucleic acid sequence as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130;

b) a polynucleotide encoding a polypeptide comprising a polypeptide sequence as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132;

c) a polynucleotide hybridizing under stringent conditions to the polynucleotide of a) or b); and

d) a polynucleotide hybridizing under stringent conditions to the reverse complement of the polynucleotide of a) or b), said polynucleotide.

In another embodiment the invention provides for such an isolated polynucleotide obtainable from a fungus, in particular *Thermomyces* is preferred and even more preferred *Thermomyces lanuginosus*.

Still in another embodiment, the invention provides for a vector comprising the polynucleotide of the present invention. The vector may further comprise a regulatory sequence operatively linked to the polynucleotide for expression of same in a suitable host cell, preferably a filamentous fungus.

In a further embodiment, the invention provides a host cell comprising the polynucleotide or the vector of the present invention.

In another embodiment the invention provides an isolated polypeptide as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102,

105, 108, 111, 114, 117, 120, 123, 126, 129, and 132. Such polypeptide is preferably obtainable from *Thermomyces lanuginosus*. In an alternate embodiment, the polypeptide is obtainable by expressing the polynucleotide or the vector of the present invention in an appropriate host cell.

Still in another embodiment, the invention provides a method for manufacturing the polypeptide comprising the steps of transforming a suitable host cell with the isolated polynucleotide or the vector of the present invention, culturing said cell under conditions allowing expression of said polynucleotide and optionally purifying the encoded polypeptide from said cell or culture medium.

The polypeptide of the present invention can be used in the preparation of a food product, in the preparation of a detergent, in the preparation of an animal feed, for prebleaching kraft pulp, for processing lignin, for producing ethanol, for treating textiles or dyed textiles or for degrading biomass or pretreated biomass.

The examples of activities of enzymes according to the invention are herein intended to at least cover any of the following:

- Enzymes that hydrolyze cellulose, including ***endoglucanases* (EG)** ((E.C. 3.2.1.4) hydrolyze the beta-1,4-linkages between glucose units); **exoglucanases**, also known as ***cellobiohydrolases*** 1 and 2 **(CBH1 or CBHI and CBH2 or CBHII)** ((E.C. 3.2.1.91) hydrolyze cellobiose, a glucose disaccharide, from the reducing and non-reducing ends of cellulose); and ***beta-glucosidases* (BG)** ((E.C. 3.2.1.21) hydrolyze the beta-1,4 glycoside bond of cellobiose to glucose)
- **Glycoside hydrolase family 61 (GH61) proteins**, which enhance the action of cellulose enzymes on lignocellulose substrates. Glycoside hydrolase family 61 (GH61 or sometimes referred to EGIV or EG4) proteins are proteins which enhance the action of cellulases on lignocellulose substrates. GH61 was originally classified as endogluconase based on measurement of very weak endo-1,4-β-d-glucanase activity in one family member.
- Enzymes that degrade or modify xylan and/or xylan-lignin complexes including **xylanase** ((E.C. 3.2.1.8) catalyzes random cleavage of beta-1,4 bonds in xylan or xyloglucan), **xylan 1,4-beta-xylosidase** (EC 3.2.1.37) catalyzes hydrolysis of 1,4-beta-D-xylans, to remove successive D-xylose residues from the non-reducing terminals, and also cleaves xylobiose), **alpha-arabinofuranosidase** ((EC 3.2.1.55) hydrolyzes terminal non-

reducing alpha-L-arabinofuranoside residues in alpha-L-arabinosides including arabinoxylans and arabinogalactans), **alpha-glucuronidase** ((EC 3.2.1.139) hydrolyzes an alpha-D-glucuronoside to the corresponding alcohol and D-glucuronate), **feruloyl esterase** ((EC 3.1.1.73) catalyzes hydrolysis of the 4-hydroxy-3-methoxycinnamoyl (feruloyl) group from an esterified sugar, which is usually arabinose in natural substrates), and **acetyl xylan esterase** ((EC 3.1.1.72) catalyzes deacetylation of xylans and xylo-oligosaccharides)

- Enzymes that degrade or modify mannan including **mannanase** ((EC 3.2.1.78) catalyzes random hydrolysis of 1,4-beta-D-mannosidic linkages in mannans, galactomannans and glucomannans), **mannosidase** ((EC 3.2.1.25) hydrolyzes terminal, non-reducing beta-D-mannose residues in beta-D-mannosides), **alpha-galactosidase** ((EC 3.2.1.22) hydrolyzes terminal, non-reducing alpha-D-galactose residues in alpha-D-galactosides, including galactose oligosaccharides, galactomannans and galactohydrolase), and **mannan acetyl esterase** .
- Enzymes that degrade xyloglucans including **xyloglucanase** ((EC 3.2.1.151) involves endohydrolysis of 1,4-beta-D-glucosidic linkages in xyloglucan while (EC 3.2.1.155) catalyzes exohydrolysis of 1,4-beta-D-glucosidic linkages in xyloglucan), **endoglucanase**, and **cellulase**.
- Enzymes that degrade beta-1,4-glucan including **endoglucanase**, **cellobiohydrolase**, and **beta-glucosidase**.
- Enzymes that degrade beta-1,3-1,4-glucan including **endo-beta-1,3(4)-glucanase** ((EC 3.2.1.6) catalyzes endohydrolysis of 1,3- or 1,4-linkages in beta-D-glucans when the glucose residue whose reducing group is involved in the linkage to be hydrolysed is itself substituted at C-3), **endoglucanase** (beta-glucanase, cellulase), and **beta-glucosidase**.
- Enzymes that degrade galactan include **galactanases** ((EC 3.2.1.23) hydrolyzes terminal non-reducing beta-D-galactose residues in beta-D-galactosides).
- Enzymes that degrade arabinan include **arabinanases** ((EC 3.2.1.99) catalyze endohydrolysis of 1,5-alpha-arabinofuranosidic linkages in 1,5-arabinans).
- Enzymes that degrade starch, including **alpha-amylase** ((EC 3.2.1.1 catalyzes endohydrolysis of 1,4-alpha-D-glucosidic linkages in

polysaccharides containing three or more 1,4-alpha-linked D-glucose units) and **alpha-glucosidase** ((EC 3.2.1.20) hydrolyzes terminal, non-reducing 1,4-linked alpha-D-glucose residues with release of alpha-D-glucose)

- Enzymes that degrade or modify pectin, including **pectate lyase** ((EC 4.2.2.2) carries out eliminative cleavage of pectate to give oligosaccharides with 4-deoxy-alpha-D-gluc-4-enuronosyl groups at their non-reducing ends), **pectin lyase** ((EC 4.2.2.10) catalyzes eliminative cleavage of (1-4)-alpha-D-galacturonan methyl ester to give oligosaccharides with 4-deoxy-6-O-methyl-alpha-D-galact-4-enuronosyl groups at their non-reducing ends), **polygalacturonase** ((EC 3.2.1.15) carries out random hydrolysis of 1,4-alpha-D-galactosiduronic linkages in pectate and other galacturonans), **pectin acetyl esterase** ((EC 3.1.1.11) hydrolyzes acetate from pectin acetyl esters), **alpha-arabinofuranosidase**, **beta-galactosidase**, **galactanase**, **arabinanase**, **rhamnogalacturonase** (EC 3.2.1.-) hydrolyzes alpha-D-galacturonopyranosyl-(1,2)-alpha-L-rhamnopyranosyl linkages in the backbone of the hairy regions of pectins), **rhamnogalacturonan lyase** (EC 4.2.2.-) degrades type I rhamnogalacturonan from plant cell walls and releases disaccharide products), **rhamnogalacturonan acetyl esterase** ((EC 3.1.1.-) hydrolyzes acetate from rhamnogalacturonan), **xylogalacturonosidase**, and **xylogalacturonase** ((EC 3.2.1.-) hydrolyzes xylogalacturonan (xga), a galacturonan backbone heavily substituted with xylose, and which is one important component of the hairy regions of pectin)
- Enzymes that degrade or modify lignin, including **lignin peroxidases** ((EC 1.11.1.14) oxidize lignin and lignin model compounds using hydrogen peroxide) , **manganese-dependent peroxidases** ((EC 1.11.1.13) oxidizes lignin and lignin model compounds using Mn^{2+} and hydrogen peroxide), **versatile peroxidases** ((EC 1.11.1.16) oxidizes lignin and lignin model compounds using an electron donor and hydrogen peroxide and combines the substrate-specificity characteristics of the two other ligninolytic peroxidases, EC 1.11.1.13, manganese peroxidase and EC 1.11.1.14, lignin peroxidase), and **laccases** ((EC 1.10.3.2) a group of multi-copper proteins of low specificity acting on both o- and p-quinols, and often acting

also on lignin)

- Enzymes acting on chitin, including **chitinase** ((EC 3.2.1.14) which catalyzes random hydrolysis of N-acetyl-beta-D-glucosaminide 1,4-beta-linkages in chitin and chitodextrins) and **beta-N-acetylhexosaminidase** ((EC 3.2.1.52) which hydrolyzes terminal non-reducing N-acetyl-D-hexosamine residues in N-acetyl-beta-D-hexosaminides).

The invention also relates to vectors comprising a polynucleotide sequence according to the invention, as well as primers, probes and fragments that may be used to amplify or detect the DNA according to the invention.

In a further preferred embodiment, a vector is provided wherein the polynucleotide sequence according to the invention is functionally linked with at least one regulatory sequence suitable for expression of the encoded amino acid sequence in a suitable host cell, such as a filamentous fungus, for example *Aspergillus*. The invention also provides methods for preparing polynucleotides and vectors according to the invention.

The invention also relates to recombinantly produced host cells that contain heterologous or homologous polynucleotides according to the invention.

In another embodiment, the invention provides recombinant host cells wherein the expression of an enzyme according to the invention is significantly increased or wherein the activity of the enzyme is increased.

In another embodiment the invention provides for a recombinantly produced host cell that contains heterologous or homologous DNA according to the invention and wherein the cell is capable of producing a functional enzyme according to the invention, preferably a cell capable of over-expressing the enzyme according to the invention, for example an *Aspergillus niger* strain comprising an increased copy number of a gene according to the invention.

In yet another aspect of the invention, a purified polypeptide is provided. The polypeptides according to the invention include the polypeptides encoded by the polynucleotides according to the invention. Especially preferred are polypeptides according to any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132.

Fusion proteins comprising a polypeptide according to the invention are also within the scope of the invention. The invention also provides methods of making the polypeptides according to the invention.

The invention also relates to the use of the enzyme according to the invention in any industrial process as described herein.

Brief description of the Figures

Fig. 1 illustrates the structure of pGBFIN-49;

Fig. 2 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on TEC-210 using hWS substrate;

Fig. 3 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on TEC-210 using hWS substrate;

Fig. 4 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on 4E mix using hWS substrate;

Fig. 5 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on TEC-210 using aCS substrate;

Fig. 6 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on 4E mix using aCS substrate;

Fig. 7 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on TEC-210 using aCS substrate;

Fig. 8 illustrates the effect of a set of Thermomyces lanuginosa proteins spiked on 4E mix using aCS substrate;

Fig. 9 illustrates the effect of a Thermomyces lanuginosa GH61 protein spiked on TEC-210 mix using aCS substrate; and

Fig. 10 illustrates the effect of a Thermomyces lanuginosa GH61 protein spiked on a 3E mix using aCS substrate.

Detailed description of the invention

Polynucleotides

The present invention provides polynucleotides encoding enzymes, having amino acid sequences according to any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132. The sequences of the genes were determined by sequencing cDNA clones, mRNA transcripts, or genomic DNA obtained from *Thermomyces lanuginosus* ATCC 200065. The invention provides polynucleotide sequences comprising the genes encoding the enzymes listed in Table 12 as well as their coding sequences.

Accordingly, the invention relates to an isolated polynucleotide comprising the nucleotide sequences according to any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

In particular, the invention relates to an isolated polynucleotide hybridizable under stringent conditions, preferably under high stringent conditions, to the complement of the polynucleotide listed above. Advantageously, such isolated polynucleotide may be obtained from fungi, in particular from *Thermomyces*, preferably from *Thermomyces lanuginosus*. More specifically, the invention relates to isolated polynucleotides having nucleotide sequences according to any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

As used herein, the terms “gene” and “recombinant gene” refer to nucleic acid molecules which may be isolated from chromosomal DNA, which include an open reading frame encoding a protein, e.g. *Thermomyces lanuginosus* enzymes according to the present invention. A gene may include coding sequences, non-coding sequences, introns and regulatory sequences. Moreover, a gene refers to an isolated nucleic acid molecule as defined herein.

A nucleic acid molecule of the present invention, such as a nucleic acid molecule having the nucleotide sequences listed above can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or a portion of these nucleic acid sequences as hybridization probes, nucleic acid molecules according to the invention can be isolated using standard hybridization and cloning techniques (e. g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. Molecular Cloning: A Laboratory Manual.2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26,

28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence information contained in these sequences.

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to or hybridizable to nucleotide sequences according to the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In a preferred embodiment, isolated nucleic acid molecules of the invention comprise the nucleotide sequences shown in any one of SEQ ID NOs: 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 65, 68, 71, 74, 77, 80, 83, 86, 89, 92, 95, 98, 101, 104, 107, 110, 113, 116, 119, 122, 125, 128, and 131,

These sequences correspond to the coding regions of the *Thermomyces lanuginosus* genes shown in Table 12. These DNA sequences encode the *Thermomyces lanuginosus* polypeptides according to any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a reverse complement of the nucleotide sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

A nucleic acid molecule which is complementary to another nucleotide sequence is one which is sufficiently complementary to the other nucleotide sequence such that it can hybridize to the other nucleotide sequence thereby

forming a stable duplex.

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a functional equivalent thereof such as a biologically active fragment or domain, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules.

An "isolated polynucleotide" or "isolated nucleic acid" is a DNA or RNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment, an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promoter) sequences that are immediately contiguous to the coding sequence. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a hybrid gene encoding an additional polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "isolated nucleic acid fragment" is a nucleic acid fragment that is not naturally occurring as a fragment and would not be found in the natural state.

As used herein, the terms "polynucleotide" or "nucleic acid molecule" are intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. The nucleic acid may be *synthesized using oligonucleotide analogs or derivatives (e.g., inosine or* phosphorothioate nucleotides). Such oligonucleotides can be used, for example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases.

Another embodiment of the invention provides an isolated nucleic acid

molecule which is antisense to nucleic acid molecules shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, e.g., the coding strand of these nucleic acid molecules. In a further embodiment, it is also provided an antisense molecule which hybridizes with at least 10 contiguous, 20 contiguous, 40 contiguous, more preferably 50 contiguous, 60 contiguous, at least 80 contiguous and more preferably 100 contiguous nucleotides to any sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130 and 29, e.g., the coding strands of a these molecules. Also included within the scope of the invention are the complement strands of the nucleic acid molecules described herein.

Sequencing errors

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein can be readily used to isolate the complete gene from filamentous fungi, in particular from *Thermomyces lanuginosus* which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other

approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct for such errors.

Nucleic acid fragments, probes and primers

Nucleic acid molecules according to the invention may comprise only a portion or a fragment of the nucleic acid sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, for example a fragment which can be used as a probe or primer or a fragment encoding a portion of these genes. The nucleotide sequence determined from the cloning of the genes shown in Table 12 allows for the generation of probes and primers designed for use in identifying and/or cloning other family members, as well as homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide which typically comprises a region of nucleotide sequence that hybridizes preferably under highly stringent conditions to at least about 12 or 15, preferably about 18 or 20, preferably about 22 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 or more consecutive nucleotides of a nucleotide sequence shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, or a functional equivalent thereof.

Probes based on these nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins for instance in other organisms. In preferred embodiments, the probe further

comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme cofactor. Such probes can also be used as part of a diagnostic test kit for identifying cells which express a protein encoded by the genes shown in Table 12.

Identity & homology

The terms "homology" or "percent identity" are used interchangeably herein. For the purpose of this invention, it is defined here that in order to determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = number of identical positions/total number of positions (i.e. overlapping positions) x 100). Preferably, the two sequences are the same length.

The skilled person will be aware of the fact that several different computer programs are available to determine the homology between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48): 444-453 (1970)) algorithm which has been incorporated into the GAP program in the Accelrys GCG software package (available at http://www.accelrys.com/products/gcg/), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

In yet another embodiment, the percent identity between two nucleotide

sequences is determined using the GAP program in the Accelrys GCG software package (available at http://www.accelrys.com/products/gcg/), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity two amino acid or nucleotide sequence is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989) which has been incorporated into the ALIGN program (version 2.0) (available at the ALIGN Query using sequence data of the Genestream server IGH Montpellier France http://vega.igh.cnrs.fr/bin/align-guess.cgi) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403—10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, word length = 12 to obtain nucleotide sequences homologous to nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, word length = 3 to obtain amino acid sequences homologous to protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17): 3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See the homepage of the National Center for Biotechnology Information at http://www.ncbi.nlm.nih.gov/.

Hybridization

As used herein, the term "hybridizing" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 60%, at least about 70%, at least about 80%, more preferably at least about 85%, even more preferably at least about 90%, more preferably at least 95%, more preferably at least 98% or more preferably at least 99% homologous to each other typically remain hybridized to each other.

A preferred, non-limiting example of such hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45 °C, followed

by one or more washes in 1 X SSC, 0.1% SDS at 50 °C, preferably at 55 °C, preferably at 60 °C and even more preferably at 65 °C.

Highly stringent conditions include, for example, hybridizing at 68 °C in 5x SSC/5x Denhardt's solution / 1.0% SDS and washing in 0.2x SSC/0.1% SDS at room temperature. Alternatively, washing may be performed at 42 °C.

The skilled artisan will know which conditions to apply for stringent and highly stringent hybridization conditions. Additional guidance regarding such conditions is readily available in the art, for example, in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3' terminal poly(A) tract of mRNAs), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to specifically hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

Obtaining full length DNA from other organisms

In a typical approach, cDNA libraries constructed from other organisms, e.g. brown rot fungi, in particular from the micro-organism family *Thermomyces* can be screened.

For example, *Thermomyces* strains can be screened for homologous polynucleotides shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, by Northern blot analysis. Upon detection of transcripts homologous to polynucleotides according to the invention, cDNA libraries can be constructed from RNA isolated from the appropriate strain, utilizing standard techniques well known to those of skill in the art. Alternatively, a total genomic DNA library can be screened using a probe hybridizable to a polynucleotide shown above.

Homologous gene sequences can be isolated, for example, by

performing PCR using two degenerate oligonucleotide primer pools designed on the basis of nucleotide sequences as taught herein.

The template for the reaction can be cDNA obtained by reverse transcription of mRNA prepared from strains known or suspected to express a polynucleotide according to the invention. The PCR product can be subcloned and sequenced to ensure that the amplified sequences represent the sequences of a new nucleic acid sequence corresponding to those shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

The PCR fragment can then be used to isolate a full-length cDNA clone by a variety of known methods. For example, the amplified fragment can be labelled and used to screen a bacteriophage or cosmid cDNA library. Alternatively, the labelled fragment can be used to screen a genomic library.

PCR technology also can be used to isolate full-length cDNA sequences from other organisms. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction can be performed on the RNA using an oligonucleotide primer specific for the most 5' end of the amplified fragment for the priming of first strand synthesis.

The resulting RNA/DNA hybrid can then be "tailed" (e.g., with guanines) using a standard terminal transferase reaction, the hybrid can be digested with RNase H, and second strand synthesis can then be primed (e.g., with a poly-C primer). Thus, cDNA sequences upstream of the amplified fragment can easily be isolated. For a review of useful cloning strategies, see e.g. Sambrook et al., supra; and Ausubel et al., supra.

Vectors

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a protein shown in Table 12 and whose sequence may be found in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132.

As used herein, the term “vector” refers to a nucleic acid molecule

capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. The terms "plasmid" and "vector" can be used interchangeably herein as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vector includes one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operatively linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signal). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in a certain host cell (e.g. tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on

such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, encoded by nucleic acids as described herein (e.g. proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, mutant forms of these proteins, fragments, variants or functional equivalents thereof, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, in prokaryotic or eukaryotic cells. For example, these proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors e.g., vectors derived from bacterial plasmids, bacteriophage, yeast episome, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled person. In a specific embodiment, promoters are preferred that are capable of directing a high expression level of lignocellulose active proteins from fungi. Such promoters are known in the art. The expression constructs may contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the mature transcripts expressed by the

constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, transduction, infection, lipofection, cationic lipid-mediated transfection or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., *Basic Methods in Molecular Biology* (1986) and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methatrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as those encoding proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g. cells that have incorporated the selectable marker gene will survive, while the other cells die).

Expression of proteins in prokaryotes is often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, e.g. to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a

ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein.

As indicated, the expression vectors will preferably contain selectable markers. Such markers include dihydrofolate reductase or neomycin resistance for eukaryotic cell culture and tetracyline or ampicillin resistance for culturing in *E. coli* and other bacteria. Representative examples of appropriate host include bacterial cells, such as *E. coli*, *Streptomyces Salmonella typhimurium* and certain *Bacillus* species; fungal cells such as *Aspergillus* species, for example *A. niger, A. oryzae* and *A. nidulans*, yeast cells such as *Kluyveromyces*, for example *K. lactis* and/or *Pichia,* for example *P. pastoris*; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS and *Bowes melanoma*; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Vectors preferred for use in bacteria are for example disclosed in WO-A1-2004/074468, which are hereby enclosed by reference. Other suitable vectors will be readily apparent to the skilled artisan.

Known bacterial promoters suitable for use in the present invention include the promoters disclosed in WO-A1-2004/074468, which are hereby enclosed by reference.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signal may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide whose sequences are shown in any one of SEQ ID Nos:

3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification.

Polypeptides according to the invention

The invention provides isolated polypeptides having the amino acid sequences shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 alone or in an appropriate host. Also, a peptide or polypeptide comprising a functional equivalent of the above polypeptides is comprised within the present invention. The above polypeptides are collectively comprised in the term "polypeptides according to the invention"

The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context required to indicate a chain of at least two amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than seven amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxyl terminus. The one-letter code of amino acids used herein is commonly known in the art and can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual, 2nd,ed. Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Sequence Listings programs can convert easily this one-letter code of amino acid sequence into a three-letter code.

By "isolated" polypeptide or protein is intended a polypeptide or protein removed from its native environment. For example, recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the invention as are native or recombinant polypeptides which have

been substantially purified by any suitable technique such as, for example, the single-step purification method disclosed in Smith and Johnson, Gene 67:31-40 (1988).

The proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 according to the invention can be recovered and purified from recombinant cell cultures by methods known in the art. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Protein fragments

The invention also features biologically active fragments of the polypeptides according to the invention.

Biologically active fragments of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequences shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, which include fewer amino acids than the full length protein but which exhibit at least one biological activity of the corresponding full-length protein. Typically, biologically active fragments comprise a domain or motif with at least one activity of the full-length protein. A biologically active fragment of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the biological activities of the native form of a polypeptide of the invention.

The invention also features nucleic acid fragments which encode the above biologically active fragments of the proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132.

Fusion proteins

The proteins of the present invention or functional equivalents thereof, e.g., biologically active portions thereof, can be operatively linked to unrelated polypeptides (e.g., heterologous amino acid sequences) to form fusion proteins. "Unrelated polypeptides" refer to polypeptides having amino acid sequences corresponding to proteins which are not substantially homologous to the proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132. Such "unrelated polypeptides" can be derived from the same or a different organism. Within a fusion protein the polypeptide derived from the sequences shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 can correspond to all or a biologically active fragment of these proteins. In a preferred embodiment, a fusion protein comprises at least two biologically active portions of proteins whose sequences are shown above. Within the fusion protein, the term "operatively linked" is intended to indicate that the polypeptide whose sequence is shown above, and the unrelated polypeptide are fused in-frame to each other. The unrelated polypeptide can be fused to the N-terminus or C-terminus of the polypeptide whose sequence is one of those shown above.

For example, in one embodiment, the fusion protein is a fusion protein in which the protein whose sequence as shown above is fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant protein from *Thermomyces lanuginosus*. In another embodiment, the fusion protein is a protein whose sequence is one of those shown above, containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian and yeast host cells), expression and/or secretion of proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21,

24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 can be increased through use of a heterologous signal sequence.

In another example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

A signal sequence can be used to facilitate secretion and isolation of a protein or polypeptide of the invention. Signal sequences are typically characterized by a core of hydrophobic amino acids, which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by known methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence, which facilitates purification, such as with a GST domain. Thus, for instance, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide, which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in *Gentz et al, Proc. Natl. Acad. Sci. USA 86*:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. The HA tag is another peptide useful for purification which corresponds to an epitope derived of influenza hemaglutinin protein, which has been described by Wilson *et al., Cell 37*:767 (1984), for instance.

Preferably, a fusion protein of the invention (corresponding to one of

those whose sequences shown above) is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers, which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding one of the proteins shown in Table 12 can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the this protein.

Functional equivalents

The terms "functional equivalents" and "functional variants" are used interchangeably herein. Functional equivalents of DNA whose sequences are shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, are isolated DNA fragments that encode a polypeptide that exhibits a particular function of the corresponding *Thermomyces lanuginosus* enzyme or protein as shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132. A functional equivalent of a polypeptide according to the invention is a polypeptide that exhibits at least one function of a *Thermomyces lanuginosus* enzyme or protein as defined herein. Functional equivalents therefore also encompass biologically active fragments.

Functional protein or polypeptide equivalents may contain only

conservative substitutions of one or more amino acids of proteins whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 or substitutions, insertions or deletions of non-essential amino acids. Accordingly, a non-essential amino acid is a residue that can be altered in proteins whose sequences are shown above, without substantially altering the biological function. For example, amino acid residues that are conserved among the proteins of the present invention are predicted to be particularly unamenable to alteration. Furthermore, amino acids conserved among the proteins according to the present invention (shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132) and other enzymes are not likely to be amenable to alteration.

The term "conservative substitution" is intended to indicate a substitution in which the amino acid residue is replaced with an amino acid residue having a similar side chain. These families are known in the art and include amino acids with basic side chains (e.g. lysine, arginine and hystidine), acidic side chains (e.g. aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagines, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine tryptophan, histidine).

Functional nucleic acid equivalents may typically contain silent mutations or mutations that do not alter the biological function of encoded polypeptide. Accordingly, the invention provides nucleic acid molecules encoding the proteins whose sequences are shown above, that contain changes in amino acid residues that are not essential for a particular biological activity. Such proteins differ in amino acid sequence from those shown yet retain at least one biological activity thereof. In one embodiment the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises a substantially homologous amino acid sequence of at least about 72%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequences shown above.

For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J.U. et al., Science 247:1306-1310 (1990) and the references cited therein. As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which changes are likely to be permissive at a certain position of the protein.

An isolated nucleic acid molecule encoding a protein homologous to a protein whose sequence is shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, can be created by introducing one or more nucleotide substitutions, additions or deletions into the coding nucleotide sequences above such that one or more amino acid substitutions, deletions or insertions are introduced into the encoded protein. Such mutations may be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis.

The term "functional equivalents" also encompasses orthologues of the *Thermomyces lanuginosus* proteins. Orthologues of the *Thermomyces lanuginosus* proteins are proteins that can be isolated from other strains or species and possess a similar or identical biological activity. Such orthologues can readily be identified as comprising an amino acid sequence that is substantially homologous to one of the sequences shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132.

As defined herein, the term "substantially homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with similar side chain) amino acids or nucleotides to a second amino acid or nucleotide sequence such that the first and the second amino acid or nucleotide sequences have a common domain. For example, amino acid or nucleotide sequences which contain a common domain having about 72%, preferably 75%, more preferably 80%, even more preferably 85%, 90%, 95%, 96%, 97%, 98% or 99% identity or more are defined herein as sufficiently identical.

Also, nucleic acids encoding other family members related to those proteins whose sequences are shown above, which thus have a nucleotide

sequences that differ from sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, are within the scope of the invention. Moreover, nucleic acids encoding proteins corresponding to those whose sequences are shown above from different species which can have a nucleotide sequences which differ from those shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, are within the scope of the invention.

Nucleic acid molecules corresponding to variants (e.g. natural allelic variants) and homologues of the DNA of the invention (shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130) can be isolated based on their homology to the nucleic acids disclosed herein using the cDNAs disclosed herein or a suitable fragment thereof, as a hybridization probe according to standard hybridization techniques preferably under highly stringent hybridization conditions.

In addition to naturally occurring allelic variants of the sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, the skilled person will recognise that changes can be introduced by mutation into those nucleotide sequences thereby leading to changes in the amino acid sequences of the corresponding proteins (whose sequences are shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132,

respectively without substantially altering the functions of the corresponding proteins.

In another aspect of the invention, improved proteins derived from the sequences shown above are provided. Improved proteins are proteins wherein at least one biological activity is improved. Such proteins may be obtained by randomly introducing mutations along all or part of the coding sequences of the polypeptides shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 such as by saturation mutagenesis, and the resulting mutants can be expressed recombinantly and screened for biological activity. For instance, the art provides for standard assays for measuring the enzymatic activity of the resulting protein and thus improved proteins may easily be selected.

In a preferred embodiment the protein has an amino acid sequence according to a sequence shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132. In another embodiment, the polypeptide is substantially homologous to the amino acid sequence according to a sequence shown above and retains at least one biological activity of a polypeptide according to the sequence shown above, yet differs in amino acid sequence due to natural variation or mutagenesis as described above.

In a further preferred embodiment, the protein has an amino acid sequence encoded by an isolated nucleic acid fragment capable of hybridizing to a nucleic acid according to the sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, preferably under highly stringent hybridization conditions.

Accordingly, the protein is preferably a protein which comprises an amino acid sequence at least about 72%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to an amino acid sequence shown in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117,

120, 123, 126, 129, and 132 and retains at least one functional activity of the polypeptide according to the sequences shown above.

Functional equivalents of a protein according to the invention can also be identified e.g. by screening combinatorial libraries of mutants, e.g. truncation mutants, of the protein of the invention for activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g. for phage display). There are a variety of methods that can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening a subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations of truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which

detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering 6(3): 327-331).

In addition to the gene sequences shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, it will be apparent for the person skilled in the art that DNA sequence polymorphisms may exist within a given population, which may lead to changes in the amino acid sequence of the protein sequences as shown herein. Such genetic polymorphisms may exist in cells from different populations or within a population due to natural allelic variation. Allelic variants may also include functional equivalents.

Fragments of a polynucleotide according to the invention may also comprise polynucleotides not encoding functional polypeptides. Such polynucleotides may function as probes or primers for a PCR reaction.

Nucleic acids according to the invention irrespective of whether they encode functional or non-functional polypeptides can be used as hybridization probes or polymerase chain reaction (PCR) primers. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having an activity shown in Table 12, inter alias, (1) isolating the gene encoding the protein, or allelic variants thereof from a cDNA library e.g. from an organism other than *Thermomyces lanuginosus*; (2) in situ hybridization (e.g. FISH) to metaphase chromosomal spreads to provide precise chromosomal location of the gene as described in Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988); (3) Northern blot analysis for detecting expression of mRNA corresponding to one of those shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122,

124, 125, 127, 128, and 130, in specific tissues and/or cells and (4) probes and primers that can be used as a diagnostic tool to analyse the presence of a nucleic acid hybridizable to the a sequence shown above, probe in a given biological (e.g. tissue) sample.

Also encompassed by the invention is a method of obtaining a functional equivalent of a gene corresponding to one of those shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130. Such a method entails obtaining a labelled probe that includes an isolated nucleic acid which encodes all or a portion of the protein sequence according to any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132 and to Table 12 or a variant thereof; screening a nucleic acid fragment library with the labelled probe under conditions that allow hybridization of the probe to nucleic acid fragments in the library, thereby forming nucleic acid duplexes, and preparing a full-length gene sequence from the nucleic acid fragments in any labelled duplex to obtain a gene related to the gene shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

In one embodiment, a nucleic acid of the invention is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to a nucleic acid sequence shown herein or to the reverse complement thereof.

Host cells

In another embodiment, the invention features cells, e.g., transformed host cells or recombinant host cells that contain a nucleic acid or vector encompassed by the invention. A "transformed cell" or "recombinant cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a nucleic acid or vector according to the invention.

Both prokaryotic and eukaryotic cells are included, e.g., bacteria, fungi, yeast, and the like, especially preferred are cells from filamentous fungi, in particular *Thermomyces lanuginosus*. A cell of the invention is typically not a wild-type *Thermomyces lanuginosus* or a naturally-occurring cell. They include, but are not limited to: fungi such as *Aspergillus niger*, *Trichoderma reesii*, *Myceliophthora thermophila* or *Talaromyces emersonii*; yeasts such as *Saccharomyces cerevisiae*, *Yarrowia lipolytica* and *Pichia pastoris*; bacteria such as *Escherichia coli* and *Bacillus sp.*; and plants such as *Nicotiana benthamiana*, *Nicotiana tabacum* and *Medicago sativa*.

A nucleic acid molecule (or a nucleic acid molecule which is comprised within a vector) may be homologous or heterologous with respect to the cell into which it is introduced. In this context, a nucleic acid molecule is homologous to a cell if the nucleic acid molecule naturally occurs in that cell. A nucleic acid molecule is heterologous to a cell if the nucleic acid molecule does not naturally occur in that cell. Accordingly, the invention provides a cell which comprises a heterologous or a homologous sequence corresponding to one of those shown in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

A host cell can be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific, desired fashion. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may facilitate optimal functioning of the protein.

Various host cells have characteristic and specific mechanisms for post-translational processing and modification of proteins and gene products. Appropriate cell lines or host systems familiar to those of skill in the art of molecular biology and/or microbiology can be chosen to ensure the desired and correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product can be used. Such host cells are well known in the art.

Host cells also include, but are not limited to, mammalian cell lines such as CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and choroid plexus

cell lines.

If desired, a stably transfected cell line can produce the polypeptides according to the invention. A number of vectors suitable for stable transfection of mammalian cells are available to the public, methods for constructing such cell lines are also publicly known, e.g., in Ausubel *et al.* (supra).

Use of the *Thermomyces lanuginosus* enzymes in industrial processes

The invention also relates to the use of the enzymes according to the invention in a selected number of industrial processes. Despite the long term experience obtained with these processes, the enzymes according to the invention feature a number of significant advantages over the enzymes currently used. Depending on the specific application, these advantages can include aspects such as lower production costs, higher specificity towards the substrate, greater synergies with existing enzymes, less antigenic effect, less undesirable side activities, higher yields when produced in a suitable microorganism, more suitable pH and temperature ranges, better properties of the final product, and food grade or kosher aspects.

The present invention also relates to methods for preparing a food product comprising incorporating into the food product an effective amount of an enzyme of the present invention. This improves one or more properties of the food product relative to a food product in which the polypeptide is not incorporated.

The phrase "incorporated into the food product" is defined herein as adding the enzyme according to the invention to the food product, any ingredient from which the food product is to be made, and/or any mixture of food ingredients from which the food product is to be made. In other words, the enzyme according to the invention may be added in any step of the food product preparation and may be added in one, two or more steps. The enzyme according to the invention is added to the ingredients of a food product which can then be treated by methods including cooking, boiling, drying, frying, steaming or baking as is known in the art.

The term "effective amount" is defined herein as an amount of the enzyme according to the invention that is sufficient for providing a measurable effect on at least one property of interest of the food product.

The term "improved property" is defined herein as any property of a food

product which is improved by the action of an enzyme according to the invention relative to a food product in which the enzyme according to the invention is not incorporated. The improved property may be determined by comparison of a food product prepared with and without addition of a polypeptide of the present invention. Organoleptic qualities may be evaluated using procedures well established in the food industry, and may include, for example, the use of a panel of trained taste-testers.

The enzymes of the present invention may be in any form suitable for the use in question, e.g., in the form of a dry powder, agglomerated powder, or granulate, in particular a non-dusting granulate, liquid, in particular a stabilized liquid, or protected enzyme such as described in WO01/11974 and WO02/26044. Granulates and agglomerated powders may be prepared by conventional methods, e.g., by spraying the enzyme according to the invention onto a carrier in a fluid-bed granulator. The carrier may consist of particulate cores having a suitable particle size. The carrier may be soluble or insoluble, e.g., a salt (such as NaCl or sodium sulphate), sugar (such as sucrose or lactose), sugar alcohol (such as sorbitol), starch, rice, corn grits, or soy. The enzyme according to the invention and/or additional enzymes may be contained in slow-release formulations. Methods for preparing slow-release formulations are well known in the art. Adding nutritionally acceptable stabilizers such as sugar, sugar alcohol, or another polyol, and/or lactic acid or another organic acid according to established methods may for instance, stabilize liquid enzyme preparations.

The enzyme according to the invention may also be incorporated in yeast comprising compositions such as disclosed in EP-A-0619947, EP-A-0659344 and WO02/49441.

One or more additional enzymes may also be incorporated into the food product. The additional enzyme may be of any origin, including mammalian and plant, and preferably of microbial (bacterial, yeast or fungal) origin and may be obtained by techniques conventionally used in the art. Enzymes may conveniently be produced in microorganisms. Microbial enzymes are available from a variety of sources; *Bacillus* species are a common source of bacterial enzymes, whereas fungal enzymes are commonly produced in *Aspergillus* species.

Suitable additional enzymes include other starch degrading enzymes, xylanases, oxidizing enzymes, fatty material splitting enzymes, or protein-degrading, modifying or crosslinking enzymes.

Starch degrading enzymes are for instance endo-acting enzymes such as alpha-amylase, maltogenic amylase, pullulanase or other debranching enzymes and exo-acting enzymes that cleave off glucose (amyloglucosidase), maltose (beta-amylase), maltotriose, maltotetraose and higher oligosaccharides.

Suitable xylanases are for instance xylanases, pentosanases, hemicellulase, arabinofuranosidase, glucanase, cellulase, cellobiohydrolase, beta-glucosidase, and others.

Oxidizing enzymes are for instance glucose oxidase, hexose oxidase, pyranose oxidase, sulfhydryl oxidase, lipoxygenase, laccase, polyphenol oxidases and others.

Fatty material splitting enzymes are for instance triacylglycerol lipases, phospholipases (such as A_1, A_2, B, C and D) and galactolipases.

Protein degrading, modifying or crosslinking enzymes are for instance endo-acting proteases (serine proteases, metalloproteases, aspartyl proteases, thiol proteases), exo-acting peptidases that cleave off one amino acid, or dipeptide, tripeptide etceteras from the N-terminal (aminopeptidases) or C-terminal (carboxypeptidases) ends of the polypeptide chain, asparagines or glutamine deamidating enzymes such as deamidase and peptidoglutaminase or crosslinking enzymes such as transglutaminase.

In a preferred embodiment, the additional enzyme may be an amylase, such as an alpha-amylase (can be useful for providing sugars fermentable by yeast) or beta-amylase, cyclodextrin glucanotransferase, peptidase, in particular, an exopeptidase (can be useful in flavour enhancement), transglutaminase, lipase (can be useful for the modification of lipids present in the food or food constituents), phospholipase, cellulase, hemicellulase, protein disulfide isomerase, peroxidase, laccase, or oxidase, e.g., an glucose oxidase, hexose oxidase, aldose oxidase, pyranose oxidase, lipoxygenase or L-amino acid oxidase..

When one or more additional enzyme activities are to be added in accordance with the methods of the present invention, these activities may be added separately or together with the polypeptide according to the invention.

In addition to the use of the enzymes according to the present invention in food applications, the present invention also relates to the use of the enzymes according to the present invention in other industrial applications.

The enzymes of the current invention may be used in new or improved

methods for enzymatically degrading or converting plant cell wall polysaccharides from biomass into various useful products. In addition to cellulose and hemicellulose, plant cell walls contain associated pectins and lignins, the removal of which by enzymes of the current invention can improve accessibility to cellulases and hemicellulases, or which can themselves be converted to useful products.

Usually, biomass must be subjected to pre-treatment in order to make the cellulose more accessible and the enzymes of the current invention may also be used in improved methods for the processing of pretreated biomass. Pretreatment technologies may involve chemical, physical, or biological treatments. Examples of pre-treatment technologies include but are not limited to: steam explosion; ammonia; acid hydrolysis; alkaline hydrolysis; solvent extraction; crushing; milling; etc.

One example of a product produced from biomass is bioethanol. Bioethanol is usually produced by the fermentation of glucose to ethanol by yeasts such as *Saccharomyces cerevisiae*: in addition to ethanol, other chemicals may be synthesized starting from glucose. Ethanol, today, is produced mostly from sugars or starches, obtained from sugar cane, fruits and grains. In contrast, cellulosic ethanol is obtained from cellulose, the main component of wood, straw and much of the plants. Sources of biomass for cellulosic ethanol production comprise agricultural residues (such as leftover crop materials from stalks, leaves, and husks of corn plants), forestry wastes (such as chips and sawdust from lumber mills, dead trees, and tree branches), energy crops (such as dedicated fast-growing trees and grasses such as switch grass), municipal solid waste (such as household garbage and paper products), food processing and other industrial wastes (such as black liquor, paper manufacturing by-product, etc.).

Plant biomass is a mixture of plant polysaccharides, including cellulose, hemicelluloses, and pectin, together with the structural polymer, lignin. Glucose is released from cellulose by the action of mixtures of enzymes, including: endoglucanases, exoglucanases (cellobiohydrolases 1 and 2) and beta-glucosidases. Efficient large-scale conversion of cellulosic materials by such mixtures requires the full complement of enzymes, and can be enhanced by the addition of enzymes that attack the other plant cell wall components (hemicelluloses, pectins, and lignins), as well as chemical linkages between these components. Hence enzymes of the current invention that are highly expressed,

or have high specific activity, stability, or resistance to inhibitors would improve the efficiency of the process, and lower enzyme costs. It would be an advantage to the art to improve the degradation and conversion of plant cell wall polysaccharides by composing cellulase mixtures using cellulase enzymes with such properties. Furthermore, enzymes of the current invention that are able to function at extremes of pH and temperature are desirable, both since improved enzyme robustness decreases costs, and because enzymes that function at high temperature will allow high processing temperatures under high substrate consistency conditions that decrease viscosity and thus improve yields.

Glycoside hydrolases from family GH61 are known to stimulate the activity of cellulose cocktails on lignocellulosic substrates and are thus considered to exhibit cellulose-enhancing activity (P. V. Harris et al., Biochemistry 49, 3305 (2010)). They have no known enzymatic activities of their own. Enhancement of cellulase cocktail efficiency by GH61 proteins of the current invention would contribute to lowering the costs of cellulase enzymes used for the production of glucose from plant cell biomass, as described above.

Enzymatic hydrolysis of plant hemicellulose yields 5-carbon sugars that either may be fermented to ethanol by some species of yeast, or converted to other types of chemical products. Enzymatic deconstruction of hemicellulose is also known to improve the accessibility of plant cell wall cellulose to cellulase enzymes for the production of glucose from lignocellulosic materials. Hemicellulase enzymes of the current invention that enhance glucose production from lignocellulose would find utility in the bioethanol industry and in other process that rely on glucose or pentose streams from lignocellulose.

Lignin is composed of methoxylated phenyl-propane units linked by ether linkages and carbon-carbon bonds. The chemical composition of lignin may, depending on species, include guaiacyl, 4-hydroxyphenyl, and syringyl groups. Enzymatic modification of lignin by the enzymes of the current invention can be used for the production of structural materials from plant biomass, or alternatively improve the accessibility of plant cellulose and hemicelluloses to cellulase enzymes for the release of glucose from biomass as described above. Enzymes that degrade the lignin component of lignocellulose include lignin peroxidases, manganese-dependent peroxidases, versatile peroxidases, and laccases (Vicuna, 2000, Molecular Biotechnology 14: 173-176; Broda et al., 1996, Molecular Microbiology 19: 923-932). These enzymes of the current invention may also in

certain instances be active in the decolourization of industrial dyes, and thus useful for the treatment and detoxification of chemical wastes.

Pectin degrading enzymes of the current invention can also enhance the action of cellulases on plant biomass by improving the accessibility of cellulase to the cellulose component of lignocellulose.

The enzymes of the present invention may also be used in other applications for hydrolyzing non-starch polysaccharide (NSP).

One application is in the detergent industry for removal from laundry of carbohydrate-based stains. The textile industry uses various enzymes to improve the properties of its products. Such improvement relates to softness, quality of the finish, "stone-wash look" of denim, etc. Enzymes are used in detergents in order to improve its efficacy to remove most types of dirt. Enzymes have been used in textile processing since the early part of this century to remove starch-based sizing, but only in the past decade has serious attention been given to using enzymes for a wide range of textile applications. Enzymes are expected to have an even greater impact on effluent quality as more fibre preparation, pre-treatment and value-added finishing processes convert to biotreatment. In addition, enzymes are very effective catalysts even under mild conditions and do not require the high energy input often associated with chemical processes. The use of enzymes of the present invention finds utility in the detergent industry for removal from laundry of carbohydrate-based stains.

Feed enzymes have an important role to play in current farming systems. They can increase the digestibility of nutrients, leading to greater efficiency in the production of animal products such as meat and eggs. At the same time they can play a role in minimizing the environmental impact of increased animal production. Non-starch polysaccharides (NSP) can increase the viscosity of the digesta which can, in turn, decrease nutrient availability and animal performance.

Endoxylanases and phytases are the best-known feed-enzyme products. Phytase enzymes hydrolyse phytic acid and release inorganic phosphate, thereby avoiding the need to add inorganic phosphates to the diet and reducing phosphorus excretion. Addition of xylanases to feed has also been shown to have positive effects on animal growth. Adding specific nutrients to feed improves animal digestion and thereby reduces feed costs. A lot of feed additives are being currently used and new concepts are continuously developed. Use of specific enzymes like non-starch carbohydrate degrading enzymes could breakdown the

fibre releasing energy as well as increasing the protein digestibility due to better accessibility of the protein when the fibre gets broken down. In this way the feed cost could come down as well as the protein levels in the feed also could be reduced.

Non-starch polysaccharides (NSPs) are also present in virtually all feed ingredients of plant origin. NSPs are poorly utilized and can, when solubilized, exert adverse effects on digestion. Exogenous enzymes can contribute to a better utilization of these NSPs and as a consequence reduce any antinutritional effects. The hemicellulases and other polysaccharide-active enzymes of the present invention can be used for this purpose in cereal-based diets for poultry and, to a lesser extent, for pigs and other species.

The xylanases of the present invention can be used for prebleaching of kraft pulp. Xylanases have been found to be most effective for that purpose.

Xylanases attract increasing scientific and commercial attention due to applications in the pulp and paper industry for removal of hemicellulose from dissolving pulps or for enhancement of the bleachability of pulp and, thus, reduction of the use of environmentally harmful bleaching chemicals. A similar application of xylanases for pulp prebleaching is an already well-established technology and has greatly stimulated research on hemicellulases in the past decade. Although lignin-active peroxidases of the present invention may also be active in modification of lignin and hence have bleaching properties, such enzymes are generally less attractive for bleaching dues to the need to use and recycle expensive redox mediators.

Xylanases of the present invention can be used to pre-bleach pulp to reduce the amount of bleaching chemicals to obtain a given brightness. It is suggested that xylanase depolymerises xylan blocks and increases accessibility or helps liberation of residual lignin by releasing xylan-chromophore fragments. In addition to brownstock prior to bleaching, xylanases of the present invention can save on bleaching chemicals. The enzymes hydrolyze surface xylans and are able to break linkages between hemicellulose and lignin. Other hemicellulase active enzymes of the present invention which can break these linkages can function effectively in bleaching or pre-bleaching of pulp.

In addition, xylanases of the present invention can also be used in antibacterial formulation as well as in pharmaceutical products such as throat lozenges, toothpastes, and mouthwash.

Chitin is a β-(1,4)-linked polymer of N-acetyl D-glucosamine (GlcNAc), found as a structural polysaccharide in fungal cell walls as well as in the exoskeleton of arthropods and the outer shell of crustaceans. Approximately 75% the total weight of shellfish is considered waste and a large proportion of the material making up the waste is chitin. Chitin degrading enzymes of the current invention are useful in the modification and degradation of chitin, allowing the production of chitin-derived material, such as chitooligosaccharides and N-acetyl D-glucosamine, from chitin waste; another use of chitinase enzymes as antifungal agents.

For the nomenclature of microorganisms generally applied and accepted nomenclature in the art is used herein. Renaming of a species, strain, genus etc. may lead to a new name of the species, strain, genus etc. In case a species, strain, genus etc. is or will be renamed, this will not have an effect on the present invention or the scope of its protection. The name of the species, strain, genus etc. used herein may be the former, the present or the future name of a microorganism. *Rasamsonia* is a new genus comprising thermotolerant and thermophilic *Talaromyces* and *Geosmithia* species (J. Houbraken et al., Antonie van Leeuwenhoek 2012 Feb; 101(2): 403-21). Based on phenotypic, physiological and molecular data, Houbraken et al proposed to transfer the species *T. emersonii, T. byssochlamydoides, T. eburneus, G. argillacea* and *G. cylindrospora* to *Rasamsonia* gen. nov.

EXAMPLES

Fermentation of the organism

Materials & Methods

In general, for each species, starter mycelium was grown in rich medium (either mycological broth or yeast malt broth (the latter being indicated with YM in the growth conditions table)) and then washed with water. The starter was then used to inoculate different liquid media or solid substrate and the resulting mycelium was used for RNA extraction and library construction.

Following are the medium recipes and the solid substrates with a referenced source (if available) as well as a table listing the media variations,

since in some cases the basic recipes of the referenced source have been altered depending on the species grown. This is then followed by a summary of the specific species as grown in the examples.

A. **Mycological broth**

Per liter: 10g soytone, 40g D-glucose, 1ml Trace Element solution, Double-distilled water;

Adjust pH to 5.0 with hydrochloric acid (HCl) and bring volume to 1L with double-distilled water.

Trace Element Solution contains 2mM Iron(II) sulphate heptahydrate ($FeSO_4 \cdot 7H_2O$), 1 mM Copper (II) sulphate pentahydrate ($CuSO_4 \cdot 5H_2O$), 5 mM Zinc sulphate heptahydrate ($ZnSO_4 \cdot 7H_2O$), 10 mM Manganese sulphate monohydrate ($MnSO_4 \cdot H_2O$), 5 mM Cobalt(II) chloride hexahydrate ($CoCl_2 \cdot 6H_2O$), 0.5 mM Ammonium molybdate tetrahydrate ($(NH_4)_6Mo_7O_{24} \cdot 4H_2O$), and 95 mM Hydrochloric acid (HCl)dissolved in double-distilled water.

B. **Yeast-Malt broth (YM)**

(Reference: ATCC medium No. 200)

Per liter: 3g yeast extract, 3g malt extract, 5g peptone, 10g D-glucose, Double-distilled water to 1L.

C. ***Thermomyces lanuginosus* Defined Medium (TDM)**

(Reference: I. D. Reid and M. G. Piace. *Effect of Residual lignin type and amount on biological bleaching of kraft pulp by Thermomyces lanuginosus.* Applied Environmental Microbiology **60**: 1395-1400, 1994.)

Per liter: 10 g D-glucose, 0.75 g L-Asparagine monohydrate, 0.68 g Potassium phosphate monobasic (KH_2PO_4), 0.25 g Magnesium sulphate heptahydrate ($Mg\ SO_4 \cdot 7H_2O$), 15 mg Calcium chloride dihydrate ($CaCl_2 \cdot 2H_2O$), 100 µg Thiamine hydrochloride, 1 ml Trace Element solution, 0.5 g Tween 80, Double distilled water;

Adjust pH to 5.5 with 3M potassium hydroxide and bring volume to 1 L with double-distilled water.

Table 1. Variations of TDM media used for library construction

Variation	Description
TDM-1	Medium was prepared as in basic recipe described above.
TDM-2	Quantity of asparagine monohydrate was reduced to 0.15g.
TDM-3	Manganese sulphate monohydrate was omitted from the medium.
TDM-4	The quantity of manganese sulphate monohydrate was raised to 0.2mM final concentration in the medium.
TDM-5	The quantity of copper (II) sulphate pentahydrate was raised to 20µM.
TDM-6	Glucose was replaced with 10g per liter of cellulose (**Solka-Floc, 200FCC**)
TDM-7	Glucose was replaced with 10g per liter of xylan from birchwood (Sigma Cat. **# X-0502**)
TDM-8	Glucose was replaced with 10g per liter of wheat bran[1].
TDM-9	Glucose was replaced with 10g per liter of citrus pectin (Sigma Cat. **# P-9135).**
TDM-10	Tween80 was omitted from the medium.
TDM-11	The double-distilled water was replaced with whitewater[2] collected from peroxide bleaching (which occurs during the manufacture of fine paper).
TDM-12	The double-distilled water was replaced with whitewater[2] collected from newsprint manufacture.
TDM-13	Glucose was replaced with 5g per liter of ground hardwood kraft pulp[3].
TDM-14	The medium's pH was raised to 7.5.
TDM-15	The strain was incubated at 5°C above its optimum growth temperature.
TDM-16	The strain was incubated at 10°C below its optimum growth temperature.
TDM-17	One half of the double-distilled water was replaced with whitewater from newsprint manufacture. Glucose was omitted.
TDM-18	Potassium phosphate monobasic was replaced with 5mM phytic acid from rice (Sigma Cat. **# P3168).**
TDM-19	Asparagine monohydrate was increased to 4g per liter.
TDM-20	Asparagine monohydrate was increased to 4g per liter and glucose was replaced with 2% fructose.
TDM-21	Asparagine monohydrate was increased to 4g per liter; 100ml of double-distilled water was replaced with 100ml kerosene[4]. Glucose was omitted.
TDM-22	Asparagine monohydrate was increased to 4g per liter; 100ml of double-distilled water was replaced with 100ml hexadecane (Sigma cat. # H0255). Glucose was omitted.
TDM-23	Asparagine monohydrate was increased to 4g per liter; one half of the double-distilled water was replaced with 25% whitewater from newsprint manufacture plus 25% white water from peroxide bleaching. Glucose was omitted.

TDM-24	Asparagine monohydrate was increased to 4g per liter and the quantity of manganese sulphate monohydrate was raised to 0.2mM final concentration in the medium.
TDM-25	Asparagine monohydrate was increased to 4g per liter and manganese sulphate monohydrate was omitted from the medium.
TDM-26	Asparagine monohydrate was increased to 4g per liter; and potassium phosphate monobasic was replaced with 5mM phytic acid from rice (Sigma Cat. **# P3168).**
TDM-27	Glucose was replaced with 10g per liter of olive oil (Sigma cat. # **O1514**)
TDM-28	One half of the double-distilled water was replaced with whitewater from peroxide bleaching. Glucose was omitted.
TDM-29	Glucose was replaced with 10g per liter of tallow.
TDM-30	Glucose was replaced with 10g per liter of yellow grease.
TDM-31	Glucose was replaced with 10g per liter of defined lipid (Sigma cat. # L0288).
TDM-32	Glucose was replaced with 50g per liter of D-xylose.
TDM-33	Glucose was replaced with 20g per liter of glycerol and 20ml per liter of ethanol.
TDM-34	Glucose was reduced to 1g per liter and 10g per liter of bran was added.
TDM-35	Glucose was reduced to 1g per liter and 10g per liter of pectin (Sigma Cat. **# P-9135)** was added.
TDM-36	Glucose was replaced with 10g per liter of biodiesel.
TDM-37	Glucose was replaced with 10g per liter of soy feedstock.
TDM-38	Glucose was replaced with 10g per liter of locust bean gum (Sigma cat **# G0753**).
TDM-39	One half of double-distilled water was replaced with a 1:1 ratio of whitewater from newsprint manufacture and white water from peroxide bleaching. Glucose was omitted.
TDM-40	The medium's pH was raised to 8.5.
TDM-41	One half of double-distilled water was replaced with whitewater from peroxide bleaching; plus yeast extract was added to 1g per liter. Glucose was omitted.
TDM-42	Glucose was replaced with 5g per liter of yellow grease and 5g per liter of soy feedstock
TDM-43	Glucose was replaced with 20g per liter of fructose.
TDM-44	Glucose was replaced with 10g per liter of cellulose (**Solka-Floc, 200FCC**) plus 1g per liter of sophorose.
TDM-45	The medium's pH was raised to 8.84.

[1] Food grade wheat bran sourced from the supermarket was used.

[2] All whitewaters were sourced from Quebec paper mills by PAPRICAN on the Applicant's behalf.

[3] Hardwood kraft pulp was sourced from Quebec paper mills by PAPRICAN on the Applicant's behalf.

[4] Kerosene was sourced from a general hardware store.

D. **Asparagine Salts Medium (AS):**

(Reference: R. Ikeda, T. Sugita, E. Jacobson, and T. Shinoda. *Laccase and Melanization in Clinically Important Cryptococcus Species Other Than Cryptococcus neoformans* Journal of Clinical Microbiology 40: 1214-1218, 2002)

Per liter: 3.0 g D-glucose, 1.0 g L-Asparagine monohydrate, 3.0 g KH_2PO_4, 0.5 g Mg $SO_4 \cdot 7H_2O$, 1 mg Thiamine.

Table 2. Variations of AS media used for library construction

Variation	Description
AS-1	Medium was prepared as in basic recipe described above.
AS-2	Glucose was replaced with 10g per liter of pectin.
AS-3	One half of double-distilled water was replaced with a 1:1 ratio of whitewater from newsprint manufacture and white water from peroxide bleaching. Glucose was omitted.

E. **Solid substrates used:**

SS-1 5 g Wheat Bran.

SS-2 5g Wheat bran plus 5ml defined lipid.

SS-3 5g Oat bran (food grade, sourced from supermarket).

The *Thermomyces lanuginosus* strain was grown according to the methods described above under the following growth conditions: TDM-1, -2, -3, -4, -5, -6, -7, -8, 9, -10, -13, -14, -15, -39; YM, whereby the following optimal growth temperature was used: 25°C.

The strains carrying the recombinant genes were grown according to the methods described above under the following growth conditions: minimal medium as described in Käfer (1977, Adv Genet. 19:33-131) except that the salt concentrations were raised ten-fold and the glucose concentration was 150 grams per litre, at 30°C

Genome sequencing and assembly

Genomic DNA was isolated from mycelium when the growth culture had reached the mid log phase. Genomic DNA was sequenced using the Roche 454 Titanium technology (http://www.454.com) to a genome coverage of over 20-fold according to the instruction of the manufacturer. The sequences were assembled

using the Newbler and Celera assemblers.
(http://sourceforge.net/apps/mediawiki/wgs-assembler).

Building the cDNA library

Total RNA was isolated from fungal cells or mycelia when the growth cultures had reached the late log phase. The mycelia were collected by filtration through Miracloth and washed with water by filtration. The mycelia were padded dry using paper towels, and frozen in liquid nitrogen and stored at -80^{0}C. To extract total RNA, the frozen mycelia or cells were ground to a fine powder in liquid nitrogen using pestle and mortar. Approximately 1-1.5 gram of frozen fungal powder was dissolved in 10 ml of TRIzol® reagent and RNA was extracted according to the manufacturer's protocol (Invitrogen Life Sciences, Catalog #15596-018). Following extraction, the RNA was dissolved at 1-1.5 mg/ml of DEPC-treated water.

The PolyATtract® mRNA Isolation Systems (Promega, Catalog #Z5300) was used to isolate poly(A)+RNA. In general, equal amounts of total RNA extracted from up to ten culture conditions were pooled. One milligram of total RNA was used for isolation of poly(A)+RNA according to the protocol provided by the manufacturer. The purified poly(A)+RNA was dissolved at 200-500 µg/ml of DEPC-treated water.

Five micrograms of poly(A)+RNA were used for the construction of cDNA library. Double-stranded cDNA was synthesized using the ZAP-cDNA® Synthesis Kit (Stratagene, Catalog #200400) according to the manufacturer's protocol with the following modifications. An anchored oligo(dT) linker-primer was used in the first-strand synthesis reaction to force the primer to anneal to the beginning of the poly(A) tail of the mRNA. The anchored oligo(dT) linker-primer has the sequence: 5'-GAGAGAGAGAGAGAGAGAGAACTAGTCTCGAGTTTTTTTTTTTTTTTTTTVN-3' (SEQ ID NO: 133) where V is A, C, or G and N is A, C, G, or T. A second modification was made by adding trehalose at a final concentration of 0.6M and betaine at a final concentration of 2M in the buffer of the first-strand synthesis reaction to promote full-length synthesis. Following synthesis and size fractionation, fractions of double-stranded cDNA with sizes longer than 600 bp were pooled. The pooled cDNA was cloned directionally into the plasmid vector BlueScript KS+® (Stratagene) or a modified BlueScript KS+ vector that contained Gateway® (Invitrogen) recombination sites. The cDNA library was transformed into

E. coli strain XL10-Gold ultracompetent cells (Stratagene, Catalog #Z00315) for propagation.

Bacterial cells carrying cDNA clones were grown on LB agar containing the antibiotic Ampicillin for selection of plasmid-borne bacteria and X-gal and IPTG to use the blue/white system to screen for the presence cDNA inserts. The white bacterial colonies, those carrying cDNA inserts, were transferred by a colony-picking robot to 384-well MTP for replication and storage. Clones that were to be analyzed by sequencing were transferred to 96-well deep blocks using liquid-handling robots. The bacteria were cultured at 37^{0}C with shaking at 150 rpm. After 24 hours of growth, plasmid DNA from the cDNA clones was prepared by alkaline lysis and sequenced from the 5' end using ABI 3730xl DNA analyzers (Applied Biosystems). The chromatograms obtained following single-pass sequencing of the cDNA clones were processed using Phred (available at http://www.phrap.org) to assign sequence quality values, Lucy as described in Chou and Holmes (2001, *Bioinformatics,* 17(12) 1093-1104) to remove vector and low quality sequences, and Phrap (available at http://www.phrap.org/) to assemble overlapping sequences derived from the same gene into contigs.

Annotation

An in-house automated annotation pipeline was used to predict genes in the assembled genome sequence. The analysis pipeline used in part the *ab initio* tool Genemark (http://exon.biology.gatech.edu/) for prediction. It also used the predictor Augustus (http://augustus.gobics.de/) trained on *de novo* assembled sequences and orthologus sequences for gene finding. Sequence similarity searches against the mycoCLAP (http://cubique.fungalgenomics.ca/mycoCLAP/) and NCBI non-redundant databases were performed with BLASTX as described in Altschul *et al.*, (1997) (Nucleic Acids Res. 25(17): 3389-3402). Proteins encoding biomass-degrading enzymes possess conserved domains. We used the domains available at the European Bioinformatics Institute (www.ebi.ac.uk/Tools/InterProScan/) to assist in the identification of target enzymes.

Proteins targeted to the extracellular space by the classical secretory pathway possess an N-terminal signal peptide, composed of a central hydrophobic core surrounded by N- and C- terminal hydrophilic regions. We used Phobius (available at http://phobius.cgb.ki.se) and SignalP version 3 (available at

http://www.cbs.dtu.dk/services/SignalP) to recognize the presence of signal peptides encoded by the cDNA clones. The tools TargetP (available at http://www.cbs.dtu.dk/services/TargetP) and Big-PI Fungal Predictor (available at http://mendel.imp.ac.at/gpi/fungi_server.html) were used to remove sequences that encode proteins which are targeted to the mitochondria or bound to the cell wall. Finally, sequences predicted to encode soluble secreted protein by these automated tools were analyzed manually. Clones that comprise full-length cDNAs which are predicted to encode soluble secreted proteins were sequenced completely. For genes identified from the genome sequence, oligonucleotide primers specific to the target genes were designed and used to PCR amplified the target genes from double-stranded cDNA or genomic DNA. The PCR amplified products were cloned into an appropriate expression vector for protein production in host cells.

General Molecular Biology Procedures:

Standard molecular cloning techniques such as DNA isolation, gel electrophoresis, enzymatic restriction modifications of nucleic acids, *E. coli* transformation etc. were performed as described by Sambrook *et al.*, 1989, (*Molecular cloning: a laboratory manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York and Innes et al. (1990) *PCR protocols, a guide to methods and applications*, Academic Press, San Diego, edited by Michael A. Innis et al). Primers were prepared by IDT (Integrated DNA Technologies). Sanger DNA sequencing was performed using an Applied Biosystem's 3730xl DNA Analyzer technology at the Innovation Centre (Génome Québec), McGill University in Montréal.

TEC-210 was fermented according to the inoculation and fermentation procedures described in WO2011/000949.

The 4E mix (4 enzymes mixture or 4 enzyme mix) containing CBHI, CBHII, GH61 and BG (30%, 25%, 36% and 9%, respectively as described in WO2011/098577) was used.

Construction of pGBFIN49 expression plasmids

Genes of interest were cloned into the expression vector pGBFIN-49. This vector is a derivative of pGBFIN-41 that contains the *A. niger glaA* promoter, *A. niger* TrpC terminator, *A. nidulans gpdA* promoter, gene encoding the

pheomycin resistance gene, *A. niger gla*A terminator and an *E. coli* backbone. Figure 1 represents a schematic map of pGBFIN-49 and the complete nucleotide sequence is presented as SEQ ID NO: 134.
Details of the construction of pGBFIN-49 are as follows:

1. TtrpC terminator PCR amplification (0.7kb):

TtrpC terminator was PCR amplified using purified pGBFIN33 plasmid as a template. The following primers and PCR program were used:
Primer-3: 5'-GTCCGTCGCCGTCCTTCAccgccggtccgacg-3' (SEQ ID NO:135)
Primer-4: 5'-GCGGCCGGCGTATTGGGTGttacggagc-3' (SEQ ID NO:136)
Primer-4 is entirely specific to TtrpC 3' end. Primer-3 was designed to suit the LIC cloning strategy but also to keep TtrpC sequence as close as the original sequence. To do so, five adenines were replaced by thymines (underlined).

PCR master mix:

pGBFIN33	1 µl (5-10ng)
Primer-3 (10mM)	1 µl
Primer-4 (10mM)	1 µl
dNTPs (2mM)	5 µl
HF Buffer (5x)	10 µl
Phusion DNA pol	0.5 µl
Nuclease-free water	31.5 µl
Total	50 µl

PCR program: 1x98°C - 2 min; 25x (98°C - 30 sec, 68°C - 30 sec, 72°C - 1 min); 72°C - 7 min.

Reaction conditions: 5 µl of the PCR reaction was ran on 1.0% agarose gel and remaining was purified using QIAEX II gel Extraction kit (QIAGEN) and resuspended in nuclease-free water.

2. pGBFIN41 vector PCR amplification (8.3kb):

Vector backbone was PCR amplified using pGBFIN41 as a template. Primers were designed outside of the ccdA region (not included in pGBFIN49). The

following primers and PCR program were used:

Primer-2: 5'-CACCCAATACGCCGGCCGCgcttccagacagctc-3' (SEQ ID NO:137)
Primer-1C: 5'-GGTGTTTTGTTGCTGGGGAtgaagctcaggctctcagttgcgtc-3' (SEQ ID NO:138)

Primer-2 contains a pgpdA-specific region and an extra sequence specific to TtrpC 3' end (also included in Primer-4). Primer-1C was designed to suit the LIC cloning strategy but also to keep PgalA region as close as the original sequence. To do so, three thymines were replaced by adenines (underlined).

PCR master mix:

pGBFIN41	1 µl (50 ng)
Primer-2 (10mM)	1 µl
Primer-1C (10mM)	1 µl
dNTPs (2mM)	5 µl
HF Buffer (5x)	10 µl
Phusion DNA pol	0.5 µl
DMSO	1 µl
Nuclease-free water	30.5 µl
Total	50 µl

PCR program:1x98°C - 3 min; 10x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min); 20 x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min+10 sec/cycle); 72°C - 10 min.

Reaction conditions: 5 µl of the PCR reaction was ran on 0.5% agarose gel and remaining was purified using QIAEX II gel Extraction kit (QIAGEN) and resuspended in nuclease-free water.

3. pGBFIN41 + TtrpC overlap-extension PCR:

Overlap-extension / Long range PCR was performed to a) fused the two PCR pieces together; b) add SfoI restriction site to re-circulate the vector. No

primers were used in the overlap-extension stage. Primer-11 and Primer-12 were used for the long range PCR reaction.

Primer-11: 5'-CACCGGCGCCGTCCGTCGCCGTCCTTC -3' (SEQ ID NO:139)

Primer-12: 5'-ACGGCGCCGGTGTTTTGTTGCTGGGGATG -3' (SEQ ID NO:140)

Primers-11 is specific to the LIC tag located on the TtrpC terminator while Primer-12 is specific to the LIC tag located on the PglaA region. SfoI restriction site sequence is underlined.

A standard PCR master mix was prepared to perform overlap-extension PCR using pGBFIN41 and TtrpC purified PCR products as templates. No primers were added.

Overlap-extension master mix:

TtrpC	1 µl
pGBFIN41	9 µl
Buffer GC (5x)	10 µl
dNTPs (2mM)	5 µl
Phusion DNA pol	0.5 µl
Nuclase-free water	24.5 µl
Total	50 µl

PCR program – overlap (no primers):1x 98°C - 2 min; 5x (98°C - 15 sec, 58° - C30 sec, 72°C - 5 min), 5x (98°C - 15 sec, 63°C - 30 sec, 72°C - 5 min), 5x (98°C - 15 sec, 68°C - 30 sec, 72°C - 5 min); 72°C - 10 min).

The overlap-extension PCR product was then, purified on QIAEX II column and 5 µl of the purified reaction was used as template DNA for Long range PCR step with primers-11 and -12.

PCR master mix:

Overlap product	5ul
Primer-11 (10mM)	1 µl
Primer-12 (10mM)	1 µl
dNTPs (2mM)	5 µl
HF Buffer (5x)	10 µl
Phusion DNA pol	0.5 µl
DMSO	1 µl

Nuclease-free water	26.5 µl
Total	50 µl

PCR program – Long range: 1x 98°C - 3 min; 10x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min); 20 x (98°C - 30 sec, 68°C - 30 sec, 72°C - 5 min+10 sec/cycle); 72°C - 10 min.

Reaction conditions: 5 µl of the PCR reaction was ran on 0.5% agarose gel and remaining was purified using QIAEX II gel Extraction kit and resuspended in nuclease-free water. Then, SfoI digestion was performed and digested product was purified using QIAEX II gel extraction kit follow the procedure as described by the manufacture.

4. Ligation:

100 ng of the purified digested fragment was ligated to itself using 1 µl of T4 DNA Ligase (New England Biolabs, M0202), and incubated at 16C overnight. Enzyme inactivation was performed at 65°C for 10 minutes.

Then, 10 µl of ligation product were transformed in DH5α E. coli competent cells and plated on 2xYT agar containing 100 ug/ml ampicillin. DNA extraction was performed on single colonies the next day. Restriction analysis and sequencing were done to confirm the structure.

Cloning of *Thermomyces lanuginosus* genes in *E. coli*

Cloning genes of interest in the pGBFIN-49 expression vector was performed using the Ligation-independent cloning (LIC) method according to Aslanidis, C., de Jong, P. (1990) Nucleic Acids Research Vol. 18 No. 20, 6069-6074.

Coding sequences from genes of interest were amplified by PCR using primers containing LIC tags which are homologous to P*gla* and TrpC sequences in the pGBFIN-49 cloning vector fused to sequences homologous to the coding sequences of the gene of interest, and either genomic DNA or cDNA as template. Primers have following sequences:

Forward primer: 5'-**CCCCAGCAACAAAACACCT**CAGCAATG...15-20 nucleotides specific to each gene to be cloned (SEQ ID NO: 141)

Reverse primer: 5'- **GAAGGACGGCGACGGACT**TCA...15-20 nucleotides

specific to each gene to be cloned (SEQ ID NO:142)

PCR mix consists of following components:

Template (gDNA or cDNA) 1-10 ng/µl	1 µl
5X Phusion HF Buffer (Finnzymes)	10 µl
2 mM dNTPs	5 µl
LIC primer (F+R) mix 10 mM	0.5 µl
Phusion DNA Polymerase (Finnzymes)	0.5 µl
DMSO	1.5 µl
H2O	31.5ul
TOTAL	50 µl

PCR amplification was carried out with following conditions:

Cycle step	**3-step protocol**		**Cycles**
	Temp.	**Time**	
Initial denaturation	98°C	30 s	1
Denaturation Annealing Extension	98°C 58°C 72°C	10 s 30 s 30 s	10
Denaturation Annealing Extension	98°C 68°C 72°C	10 s 30 s 30 s	20
Final extension	70°C	10 min	1
End of PCR storage	4°C	hold	1

Following PCR, 90 µl milliQ water was added to each sample and the mix was purified using a MultiScreen PCR_{96} Filter Plate (Millipore) according to manufacturer's instructions. The PCR product was eluted from the filter in 25 µl 10 mM Tris-HCl pH8.0.

Expression vector pGBFIN-49 was PCR amplified using primers with following sequences:

Forward primer: 5'- **GTCCGTCGCCGTCCTTCA**CCG -3' (SEQ ID NO:143)
Reverse primer: 5'- **GGTGTTTTGTTGCTGGGGA**TGAAGC -3' (SEQ ID NO:144)
(Primers are located at either site of the SfoI restriction site.)

PCR mix consists of following components:

pGBFIN-49 plasmid DNA (10ng/ µl)	2 µl
5X Phusion HF Buffer (Finnzymes)	20 µl
2 mM dNTPs	10 µl
LIC Primer mix (F+R) 10 mM	2 µl
Phusion DNA Polymerase (Finnzymes)	1.5 µl
DMSO	3 µl
H2O	61.5 µl
TOTAL	100 µl

PCR amplification was carried out with following conditions:

Cycle step	**2-step PCR protocol**		**Cycles**
	Temp.	**Time**	
Initial denaturation	98°C	2 min	1
Denaturation Annealing+Extension	98°C 68°C	10 s 4 min +10s/cycle	35
Final extension	70°C	10 min	1
End of PCR storage	4°C	Hold	1

Following PCR, 1 µl DpnI was added to the PCR mix and digestion was allowed overnight at 37°C. Digested PCR product was purified using the Qiaquick PCR purification kit (Qiagen) according to manufacturer's instructions.

Obtained PCR fragments were treated with T4 DNA polymerase in the presence of dTTP to create single stranded tails at the ends of the PCR fragments. The single stranded tails of the PCR fragment are complementary to those at the vector, thus permitting non-covalent bi-molecular associations e.g. circularization between molecules.

Reaction mix of T4 DNA polymerase treatment pGBFIN-49 PCR fragment consists of following components:

Purified pGBFIN-49 PCR fragment	600 ng
10X Neb Buffer 2	2 µl
25 mM dTTP	2 µl
DTT 100 µM	0.8 µl
T4 DNA Polymerase 3U/ µl	1 µl
H_2O	Up to 20 µl
TOTAL	20 µl

Reaction mix of T4 DNA polymerase treatment of Gene of Interest (GOI) PCR fragment consists of following components:

Purified GOI PCR	5 µl
10X NEB Buffer 2	2 µl
25 mM dATP	2 µl
DTT 100 µM	0.8 µl
T4 DNA Polymerase 3U/ µl	1 µl
H_2O	9.2 µl
TOTAL	20 µl

Reaction conditions were as follows:

Step	Temperature °C	duration
Annealing	22	30 min
Enzyme inactivation	75	20 min
End	4	Hold

Following T4 DNA polymerase treatment, 2 µl pGBFIN-49 vector and 4 µl of the GOI were mixed and incubated at room temperature allowing annealing of GOI fragment with pGBFIN-49 vector fragment. The bi-molecular forms are used to transform *E. coli*. Plasmid DNA of resulting transformants was isolated and verified by sequence analyses for correct amplification and cloning of the gene of interest.

Transformation of *Thermomyces lanuginosus* gene expression cassettes into *A. niger*.

As host strain for enzyme production, *A. niger* GBA307 was used. Construction of *A. niger* GBA307 is described in WO2011009700.

Transformation of *A. niger* was performed essentially according to the method described by Tilburn, J. et. al. (1983) *Gene* 26, 205-221 and Kelly, J & Hynes, M. (1985) *EMBO* J., 4, 475-479 with the following modifications:

- Spores were grown for 16-24 hours at 30°C in a rotary shaker at 250 rpm in *Aspergillus* minimal medium. *Aspergillus* minimal medium contains per liter: 6 g $NaNO_3$; 0.52 g KCl; 1.52 g KH_2PO_4; 1.12 ml 4 M KOH; 0.52 g $MgSO_4.7H_2O$; 10 g glucose; 1 g casamino acids; 22 mg $ZnSO_4.7H_2O$; 11 mg H_3BO_3; 5 mg $FeSO_4.7H_2O$; 1.7 mg $CoCl_2.6H_2O$; 1.6 mg $CuSO_4.5H_2O$; 5 mg $MnCl_2.2H_2O$; 1.5 mg $Na_2MoO_4.2\ H_2O$; 50 mg EDTA; 2 mg riboflavin; 2 mg thiamine-HCl; 2 mg nicotinamide; 1 mg pyridoxine-HCl; 0.2 mg panthotenic acid; 4 µg biotin; 10 ml Penicillin (5000IU/ml/Streptomycin (5000 UG/ml) solution (Invitrogen);
- Glucanex 200G (Novozymes) was used for the preparation of protoplasts;
- After protoplast formation (2-3 hours) 10 ml TB layer (per liter: 109.32 g Sorbitol; 100 ml 1 M Tris-HCl pH 7.5) was pipetted gently on top of the protoplast suspension. After centrifugation for 10 min at 4330 x g at 4°C in a swinging bucket rotor, the protoplasts on the interface were transferred to a fresh tube and washed with STC buffer (1.2 M Sorbitol, 10 mM Tris-HCl pH7.5, 50 mM $CaCl_2$). The protoplast suspension was centrifuged for 10 min at 1560 x g in a swinging bucket rotor and resuspended in STC-buffer at a concentration of 10^8 protoplasts/ml;
- To 200 µl of the protoplast suspension, 20 µl ATA (0.4 M Aurintricarboxylic acid), the DNA dissolved in 10 µl in TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA), 100 µl of a PEG solution (20% PEG 4000 (Merck), 0.8M sorbitol, 10 mM Tris-HCl pH 7.5, 50 mM $CaCl_2$) was added;
- After incubation of the DNA-protoplast suspension for 10 min at room temperature, 1.5 ml PEG solution (60% PEG 4000 (Merck), 10 mM Tris-HCl pH7.5, 50 mM $CaCl_2$) was added slowly, with repeated mixing of the tubes. After incubation for 20 min at room temperature, suspensions were diluted with 5 ml 1.2 M sorbitol, mixed by inversion and centrifuged for 10 min at 2770 x g at room temperature.

- The protoplasts were resuspended gently in 1 ml 1.2 M sorbitol and plated onto selective regeneration medium consisting of *Aspergillus* minimal medium without riboflavin, thiamine.HCl, nicotinamide, pyridoxine, panthotenic acid, biotin, casamino acids and glucose, supplemented with 150 µg/ml Phleomycin (Invitrogen), 0.07 M $NaNO_3$, 1 M sucrose, solidified with 2 % bacteriological agar #1 (Oxoid, England). After incubation for 5-10 days at 30°C, single transformants were isolated on PDA (Potato Dextrose Agar (Difco) supplemented with 150 µg/ml Phleomycin in 96 wells MTP. After 5-7 days growth at 30°C single transformants were used for MTP fermentation.

<u>*Aspergillus niger* shake MTP fermentation</u>

96 wells microtiter plates (MTP) with sporulated *Aspergillus niger* strains were used to harvest spores for MTP fermentations. To do this, 200 µl of 10% glycerol was added to each well and after resuspending the mixture, 40 µl of spore suspension was used to inoculate 2 ml *A. niger* medium (70 g/l glucose.H_20, 10 g/l yeast extract, 10 g/l $(NH4)_2SO_4$, 2 g/l K_2SO_4, 0.5 g/l $MgSO_4.7H_2O$, 0.5 g/l $ZnSO_4.7H_2O$, 0.2 g/l $CaCl_2$, 0.01 g/l $MnSO_4.7H_2O$, 0.05 g/l $FeSO_4.7H_2O$, 0.002 $Na_2MoO_4.2H_2O$, 0.25g/l Tween-80, 10 g/l citric acid, 30 g/l MES; pH 5.5 adjusted with 4 M NaOH) in a 24 well MTP. The MTP's were incubated in a humidity shaker (Infors) at 34°C at 550 rpm, and 80% humidity for 6 days. Plates were centrifuged 10 minutes 2750 rpm at 4°C and the supernatants were harvested.

TEC-210 was fermented according to the inoculation and fermentation procedures described in WO2011/000949.

The 4E mix (4 enzymes mixture or 4 enzyme mix) containing CBHI, CBHII, GH61 and BG (30%, 25%, 36% and 9%, respectively as described in WO2011/098577) was used.

<u>Protein concentration determination with TCA-biuret method</u>

Concentrated protein samples (supernatants) were diluted with water to a concentration between 2 and 8 mg/ml. Bovine serum albumin (BSA) dilutions (0, 1, 2, 5, 8 and 10 mg/ml) were made and included as samples to generate a calibration curve. 1 ml of each diluted protein sample was transferred into a 10 ml tube containing 1 ml of a 20% (w/v) trichloro acetic acid solution in water and mixed thoroughly. Subsequently, the tubes were incubated on ice water for one hour and centrifuged for 30 minutes, at 4°C and 6000 rpm. The supernatant was

discarded and pellets were dried by inverting the tubes on a tissue and letting them stand for 30 minutes at room temperature. Next, 4 ml BioQuant Biuret reagent mix was added to the pellet in the tube and the pellet was solubilised upon mixing. Next, 1 ml water was added to the tube, the tube was mixed thoroughly and incubated at room temperature for 30 minutes. The absorption of the mixture was measured at 546 nm with a water sample used as a blank measurement and the protein concentration was calculated via the BSA calibration line.

Microtiter plate (MTP) sugar-release activity assay

For each (hemi-)cellulase assay, the stored samples were analyzed twice; 100 µL sample and 100 µl of a (hemi-)cellulase base mix [3.5 mg/g DM TEC-210 or a 4 enzyme mix at a total dosage of 3.5 mg/g DM consisting of 0,3 mg/g DM BG (9% of total protein 4E mix), 1 mg/g DM CBHI (30% of total protein 4E mix), 0,9 mg/g DM CBHII (25% of total protein 4E mix) and 1.3 mg/g DM GH61 (36% of total protein 4E mix)] was transferred to two suitable vials: one vial containing 800 µL 2.5 % (w/ w) dry matter of the acid pre treated corn stover, hot water treated washed wheat straw, or hot water treated washed corn fiber substrate in a 50 mM citrate buffer, buffered at pH 4.5. The other vial consist of a blank, where the 800 µL 2.5 % (w/ w) dry matter, acid pre treated corn stover, hot water treated wheat straw, or hot water treated corn fiber substrate suspension was replaced by 800 µL 50 mM citrate buffer, buffered at pH 4.5. The assay samples were incubated for 72 hrs at 65 °C. After incubation of the assay samples, a fixed volume of D2O (with 0.5 g/L DSS) containing an internal standard, maleic acid (20 g/L) and EDTA (40 g/L), was added. The amount of sugar released, is based on the signal between 4.65 – 4.61 ppm, relative to DSS, and is determined by means of 1D 1H NMR operating at a proton frequency of 500 MHz, using a pulseprogram with water suppression, at a temperature of 27oC.

The (hemi)-cellulase enzyme solution may contain residual sugars. Therefore, the results of the assay are corrected for the sugar content measured after incubation of the enzyme solution.

Sugar-release activity assay

An *A. niger* strain expressing *Thermomyces lanuginosa* THELA_2_00078

was fermented in shake flask cultures, as described above, in order to obtain more material for further testing. The fermentation supernatant (volume between 75 and 100 ml) was concentrated using a 10 kDa spin filter to a volume of approximately 5 ml. The cellulase activity of this protein sample was tested in an assay where 50 µl supernatant was spiked on top of an enzyme base mix (TEC-210 2.5 mg/g DM) in the presence of 2% (w/w) acid pretreated corn stover in a total reaction volume of 1 ml in a 2ml tube. 'To spike' or 'spiking of' a supernatant or an enzyme indicates in this context the addition of a supernatant or an enzyme to a (hemi)-cellulase base mix. For example a 3E mix (3 enzymes mixture or 3 enzyme mix) is spiked with a fourth enzyme to form the 4E mix. All experiments were performed in duplicate and were incubated for 72 hours at 65°C and 600rpm in a shaking eppendorf incubator.

A blank sample was included. For this blank sample, 50 µl supernatant was added to 1 ml water in the absence of feedstock or base enzyme mix. This allowed a determination of the background sugar present in the concentrated supernatant. Furthermore, base enzyme mix (TEC-210 2.5 mg/g DM) alone was incubated with 2% aCS (w/w) feedstock in a total volume of 1 ml in a 2 ml tube. The blank and base enzyme mix-only samples were also performed in duplicate.

After incubation, the samples were centrifuged and 500µl of the supernatant was transferred to a deep-well microtiter plate (MTP). Subsequently, soluble sugars were analysed by proton NMR as described below.

Soluble sugar analysis by proton NMR

Hydrolysis samples were analyzed for the presence of soluble sugars by proton NMR. 500 µl of each sample was collected in MTP (as described above) and subsequently to each sample 100 µL internal standard solution, containing maleic acid (20 g.L-1) as internal standard, EDTA (40 g.L-1) and trace amounts of DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid), was added.

After lyophilizing, the samples were re-dissolved in D2O. Subsequently the samples were transferred to NMR tubes. 1D 1H NMR spectra were recorded on a Bruker Avance III 700 MHz, equipped with a 5 mm cryoprobe operating at 300K using a pulse program without water suppression and relaxation delay of 30 s.

The amount of sugars present is calculated based on the following signals (δ in ppm relative to DSS):

α-galactose peak at 5.25 ppm (d, 0.29 H, J = 4 Hz)
α-xylose peak at 5.18 ppm (d, 0.37 H, J = 4 Hz)
α-mannose peak at 5.17 ppm (d, 0,665 H, J = 2 Hz)
β-glucose peak at 4.63 ppm (d, 0.62 H, J = 8 Hz)
α-arabinose peak at 4.50 ppm (d, 0.60 H, J = 8 Hz)

In some cases cellobiose is present in the samples. The amount present is calculated on the following signal (δ in ppm relative to DSS):

β-cellobiose peak at 4.65 ppm (d, 0.62 H, J = 8Hz)

The signal used for the standard;
maleic acid peak at 6.11 ± 0.2 ppm (s, 2H).

Example 1. Identification of THELA genes that encode a secreted protein

Genes were identified that based on curation (described above) encoded a secreted protein. A list of these genes is shown in Table 12.

Example 2. Improvement of thermophilic cellulase mixture by *Thermomyces lanuginosa* proteins in the MTP activity assay

Thermomyces lanuginosa (hemi)cellulosic proteins were cloned and expressed in *A. niger* as described above. Supernatants from MTP fermentations were used in MTP sugar release activity assays as described above, using two different feedstocks.

In one set of experiments, hot-water pretreated wheat straw (hWS) was used as the substrate. Supernatants of protein fermentations were added to a base mix of TEC-210, as described above. Two *Thermomyces lanuginosa* proteins showed increased sugar release as shown below in Table 3 / Figure 2:

Protein ID	Glucose (AU)
THELA_2_0003	30.6
THELA_2_0003	29
TEC-210 only	25.7

Table 3: Effect of *Thermomyces lanuginosa* proteins spiked on TEC-210 using hWS substrate

In a second set of experiments with hot-water pretreated wheat straw (hWS) as the substrate, supernatants of a different set of protein fermentations were added to two different cellulase enzyme base mixes, TEC-210 and 4E mix, as described above. A few different proteins from *Thermomyces lanuginosa* showed increased sugar release as shown below in Table 4 / Figure 3 and Table 5 / Figure 4:

Protein ID	Glucose (AU)
THELA_00008	36.6
THELA_00460	36.5
THELA_2_00078	35.1
THELA_00236	34.4
TEC-210 only	28.2

Table 4: Effect of a different set of *Thermomyces lanuginosa* proteins spiked on TEC-210 using hWS substrate

Protein ID	Glucose (AU)
THELA_00008	12.4
THELA_00306	11.3
THELA_2_00078	11.2
THELA_00236	10.8
THELA_00460	9.5
THELA_00077	7.8
4E only	7.0

Table 5: Effect of a different set of *Thermomyces lanuginosa* proteins spiked on 4E mix using hWS substrate

In another set of experiments, acid pretreated corn stover (aCS) was used as the substrate. Supernatants of *Thermomyces lanuginosa* protein fermentations were added to two different cellulase enzyme base mixes, TEC-210 and 4E mix, as described above. Different proteins from *Thermomyces lanuginosa* showed increased sugar release as shown below in Table 6 / Figure 5 and Table 7 / Figure 6:

Protein ID	Glucose (AU)
THELA_2_00027	32.9
THELA_2_00032	29.1
TEC-210 only	25.1

Table 6: Effect of a set of *Thermomyces lanuginosa* proteins spiked on TEC-210 using aCS substrate

Protein ID	Glucose (AU)
THELA_2_00032	27.1
THELA_2_00027	25.8
4E only	22.8

Table 7: Effect of a set of *Thermomyces lanuginosa* proteins spiked on 4E mix using aCS substrate

In second set of experiments, acid pretreated corn stover (aCS) was again used as the substrate with supernatants of a different set of *Thermomyces lanuginosa* protein fermentations. As before these protein supernatants were added to two different cellulase enzyme base mixes, TEC-210 and 4E mix, as described above. Several different proteins from *Thermomyces lanuginosa* showed increased sugar release as shown below in Table 8 / Figure 7 and Table 9 / Figure 8:

Protein ID	Glucose (AU)
THELA_2_00078	39.2
THELA_00008	36.5
THELA_00236	36.4
THELA_00460	35.4
THELA_00478	34.8
THELA_00306	32.3
TEC-210 only	28.5

Table 8: Effect of a different set of *Thermomyces lanuginosa* proteins spiked on TEC-210 using aCS substrate.

Protein ID	Glucose (AU)
THELA_00008	20.4
THELA_00306	20.0
THELA_00460	19.1
THELA_2_00078	19.1
THELA_00077	18.8
THELA_00236	18.2
4E only	16.1

Table 9: Effect of a different set of *Thermomyces lanuginosa* proteins spiked on 4E mix using aCS substrate.

Example 3. Improvement of thermophilic cellulase mixture by a *Thermomyces lanuginosa* GH61 protein in the lab scale activity assay

Thermomyces lanuginosa THELA_2_00078 GH61 was cloned and expressed in *A. niger* as described above. Concentrated supernatant from a shake flask fermentation culture was used in sugar release activity assay as described above, using aCS NREL as feedstock and adding protein based on volume to the cellulase enzyme base mix TEC-210, as described above. Spiking of *Thermomyces lanuginosa* THELA_2_00078 showed an increased sugar release as shown below in Table 10 / Figure 9.

Protein ID	Average Glucose (g/l)	st dev
THELA_2_00078	5.4	0.2
TEC-210 alone	4.4	0.02

Table 10: Effect of spiking *Thermomyces lanuginosa* THELA_2_00078 GH61 on TEC-210 using aCS substrate; st dev, standard deviation.

Example 4. Substitution of *Rasamsonia emersonii* GH61 with *Thermomyces lanuginosa* THELA_2_00078 in the 4 enzyme mixture activity assay

The cellulase enhancing activity of THELA_2_00078 was further analysed. *Thermomyces lanuginosa* THELA_2_00078 GH61 protein was cloned and expressed in *A. niger* as described above. The supernatant of an *A. niger* expressing THELA_2_00078 GH61 shake flask fermentation was concentrated

and spiked in a volume of 125 µl on top of a base activity of a three enzyme base mix (3.2 mg/g DM composed of: BG at 0.45 g/g DM, CBHI at 1.5 mg/g DM and CBHII at 1.25 mg/g DM) at a feedstock concentration of 2% (w/w) aCS. As a control, the 4 enzyme base mix was taken into account, which effectively consisted of the 3 E mix as mentioned here, spiked with 1.8 mg/g DM GH61. The four enzymes were added in a combined volume of 0.5 ml (125µl of each enzyme) to a 0.5 ml solution of 4% (w/w) aCS in 50mM acetate buffer, pH 4.5, in a 2 ml tube. The 3 enzyme base mix with addition of 125 µl 50mM acetate buffer, pH 4.5, instead of supernatant was used as the control blank. Furthermore the background sugars in the supernatant sample of THELA_2_00078 were determined by adding 125 µl THELA_2_00078 supernatant into 875µl 50mM acetate buffer, pH 4.5. All experiments were performed at least in duplicate and were incubated for 72 hours at 65°C and 600rpm in a shaking eppendorf incubator. After incubation, the samples were centrifuged and 500µl of the supernatant was transferred to a deep-well MTP plate. Subsequently, soluble sugars were analysed by proton NMR as described below.

Addition of *Thermomyces lanuginosa* THELA_2_00078 showed an increased sugar release up to similar levels as the 3E mix spiked with GH61, which is present in a standard 4E base mix, shown below in Table 11/ Figure 10.

Protein ID	Average Glucose (g/l)	st dev
THELA_2_00078	5.3	0.6
3E mix + GH61	5.3	0.02
3 enzyme mix	4.2	0.1

Table 11: Effect of GH61 THELA_2_00078 protein spiked on top of a 3E mix using aCS substrate; st dev, standard deviation.

Table 12. List of target genes and reference to gene, transcript and protein sequences

Gene ID		Enzyme function	Enzyme family	Genomic sequence SEQ ID NO:	Coding sequence SEQ ID NO:	Amino acid sequence SEQ ID NO:
In priority application	In description					
Thela2_000136		cellulose oxidase	GH61	1	2	3
Thela2_000245		transglycosidase	GH72	4	5	6
Thela2_000274	THELA_00236	Laccase		7	8	9
Thela2_000290	THELA_00306	undefined biomass-degrading enzyme		10	11	12
Thela2_000447	THELA_2_00032	endo-polygalacturonase	GH28	13	14	15
Thela2_000491		Glucan 1,3-beta-glucosidase	GH17	16	17	18
Thela2_000729	THELA_2_00077	cellulose oxidase	GH61	19	20	21
Thela2_000775		Glucoamylase	GH15	22	23	24
Thela2_000777		Alpha-amylase	GH13	25	26	27
Thela2_000968		undefined biomass-degrading enzyme		28	29	30
Thela2_001825		cellulose oxidase	GH61	31	32	33
Thela2_001888		mannosidase	GH125	34	35	36
Thela2_002219	THELA_00008	xylanase	GH11	37	38	39
Thela2_002431		transglycosidase	GH72	40	41	42
Thela2_002535		chitinase	GH18	43	44	45
Thela2_002614	THELA_00460	Endo-1,3(4)-beta-glucanase	GH16	46	47	48
Thela2_002920		chitinase	GH18	49	50	51
Thela2_002936		beta-glucosidase	GH3	52	53	54

Gene ID		Enzyme function	Enzyme family	Genomic sequence SEQ ID NO:	Coding sequence SEQ ID NO:	Amino acid sequence SEQ ID NO:
In priority application	In description					
Thela2_003059	THELA_2_00078	cellulose oxidase	GH61	55	56	57
Thela2_003195		mannosidase	GH47	58	59	60
Thela2_003334		alpha-glucosidase	GH31	61	62	63
Thela2_003335		Endo-1,3(4)-beta-glucanase	GH16	64	65	66
Thela2_003338		Glucan 1,3-beta-glucosidase	GH55	67	68	69
Thela2_003681		glucopyranoside deacetylase	CE14	70	71	72
Thela2_003707		Glucan 1,3-beta-glucosidase	GH17	73	74	75
Thela2_003744		beta-hexosaminidase	GH20	76	77	78
Thela2_004478	THELA_2_00027	beta-hexosaminidase	GH20	79	80	81
Thela2_004736		cellulase	GH5	82	83	84
Thela2_004868		Endo-1,3(4)-beta-glucanase	GH16	85	86	87
Thela2_004870	THELA_00478	Endo-1,3(4)-beta-glucanase	GH16	88	89	90
Thela2_004934		alpha-galactosidase	GH36	91	92	93
Thela2_004974		endo-1,3(4)-beta-glucanase	GH81	94	95	96
Thela2_005373		alpha-glucosidase	GH31	97	98	99
Thela2_005411		chitinase	GH18	100	101	102
Thela2_005602		chitinase	GH18	103	104	105
Thela2_005694		Endo-1,3(4)-beta-glucanase	GH16	106	107	108
Thela2_006091		Alpha-N-acetylglucosaminidase	GH89	109	110	111
Thela2_006257		arylesteras	CE10	112	113	114

Gene ID		Enzyme function	Enzyme family	Genomic sequence SEQ ID NO:	Coding sequence SEQ ID NO:	Amino acid sequence SEQ ID NO:
In priority application	*In description*					
Thela2_006767		beta-glucosidase	GH3	115	116	117
Thela2_006893		beta-galactosidase	GH35	118	119	120
Thela2_007038		Glucan 1,3-beta-glucosidase	GH17	121	122	123
Thela2_007084		mannosidase	GH47	124	125	126
Thela2_007302	THELA_00448	beta-glucosidase	GH3	127	128	129
	Thela2p4_007157	expansin		130	131	132

CLAIMS

1. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising nucleic acid sequence as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, or a functional equivalent thereof;

b) a polynucleotide encoding a polypeptide comprising a polypeptide sequence as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, or a functional equivalent thereof;

c) a polynucleotide hybridizing under stringent conditions to the polynucleotide of a) or b); and

d) a polynucleotide hybridizing under stringent conditions to the reverse complement of the polynucleotide of a) or b).

2. An isolated polynucleotide selected from the group consisting of:

a) a polynucleotide comprising nucleic acid sequence as set forth in any one of SEQ ID NOs: 7, 8, 10, 11, 13, 14, 19, 20, 37, 38, 46, 47, 55, 56, 79, 80, 88, 89, 127, and 128 , or a functional equivalent thereof;

b) a polynucleotide encoding a polypeptide comprising a polypeptide sequence as set forth in any one of SEQ ID Nos: 9, 12, 15, 21, 39, 48, 57, 81, 90, and 129, or a functional equivalent thereof;

c) a polynucleotide hybridizing under stringent conditions to the polynucleotide of a) or b); and

d) a polynucleotide hybridizing under stringent conditions to the reverse complement of the polynucleotide of a) or b).

3. The isolated polynucleotide of claim 1 or 2, obtainable from a fungus.

4. An isolated polynucleotide according to claim 1, 2 or 3, obtainable from *Thermomyces lanuginosus*.

5. A vector comprising a polynucleotide as defined in any one of claims 1 to 4.

6. A vector according to claim 5 further comprising a regulatory sequence operatively linked to the polynucleotide for expression of same in a suitable host cell.

7. A vector according to claim 6, wherein said suitable host cell is a filamentous fungus.

8. A host cell comprising a polynucleotide as defined in any one of claims 1 to 4 or a vector as defined in any one of claims 5 to 7.

9. An isolated polypeptide:

a) as set forth in any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132, or a functional equivalent thereof.

b) that is at least 60%, preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, 95%, 96%, 97%, 98% or 99% homologous to the any one of SEQ ID Nos: 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75, 78, 81, 84, 87, 90, 93, 96, 99, 102, 105, 108, 111, 114, 117, 120, 123, 126, 129, and 132;

c) a polypeptide sequence encoded by nucleic acid sequence as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97,

98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130, or nucleic acids that are at least 60%, preferably at least 70%, more preferably at least 80%, even more preferably at least 90%, 95%, 96%, 97%, 98% or 99% homologous to any one of said nucleic acid sequence;

d) a polypeptide sequence encoded by a nucleic acid sequence hybridizing under stringent conditions to the polynucleotide as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130 ; or

e) a polypeptide sequence encoded by a nucleic acid sequence hybridizing under stringent conditions to the reverse complement of a polynucleotide as set forth in any one of SEQ ID NOs: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25, 26, 28, 29, 31, 32, 34, 35, 37, 38, 40, 41, 43, 44, 46, 47, 49, 50, 52, 53, 55, 56, 58, 59, 61, 62, 64, 65, 67, 68, 70, 71, 73, 74, 76, 77, 79, 80, 82, 83, 85, 86, 88, 89, 91, 92, 94, 95, 97, 98, 100, 101, 103, 104, 106, 107, 109, 110, 112, 113, 115, 116, 118, 119, 121, 122, 124, 125, 127, 128, and 130.

10. An isolated polypeptide as set forth in any one of SEQ ID Nos: 30, 39, 48, 78, 231, 267, 312, 489, 522, and 840, or a functional equivalent thereof.

11. An isolated polypeptide according to claim 9 or 10, obtainable from *Thermomyces lanuginosus*.

12. A polypeptide obtainable by expressing a polynucleotide according to any one of claims 1 to 4 or a vector according to any one of claims 5 to 7 in an appropriate host cell.

13. A method for manufacturing a polypeptide according to any one of claims 9 to 12 comprising the steps of transforming a suitable host cell

with an isolated polynucleotide according to any one of claims 1 to 4 or a vector according to any one of claims 5 to 7, culturing said cell under conditions allowing expression of said polynucleotide and optionally purifying the encoded polypeptide from said cell or culture medium.

14. Use of a polypeptide according to any one of claims 9 to 12 for prebleaching kraft pulp.

15. Use of a polypeptide according to any one of claims 9 to 12 for processing lignin.

16. Use of a polypeptide according to any one of claims 9 to 12 for producing ethanol.

17. The use of any one of claims 14 to 16 further comprising adding a cellulose or cellulases.

18. Use of a polypeptide according to any one of the claims 9 to 12 treating textiles or dyed textiles.

19. Use of a polypeptide according to any one of the claims 9 to 12 for degrading biomass or pretreated biomass.

20. Use of a polypeptide according to any one of the claims 9 to 12 in the manufacture of a detergent.

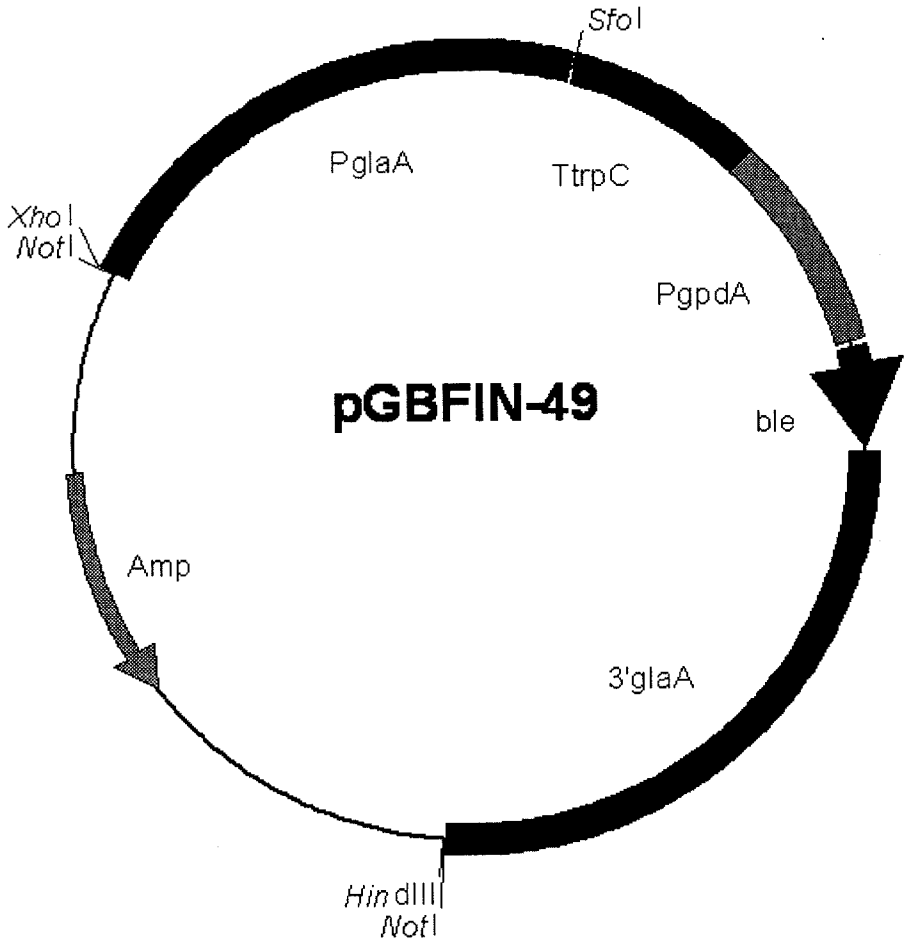

SfoI
PglaA
TtrpC
XhoI
NotI
PgpdA
pGBFIN-49
ble
Amp
3'glaA
HindIII
NotI

Fig. 1

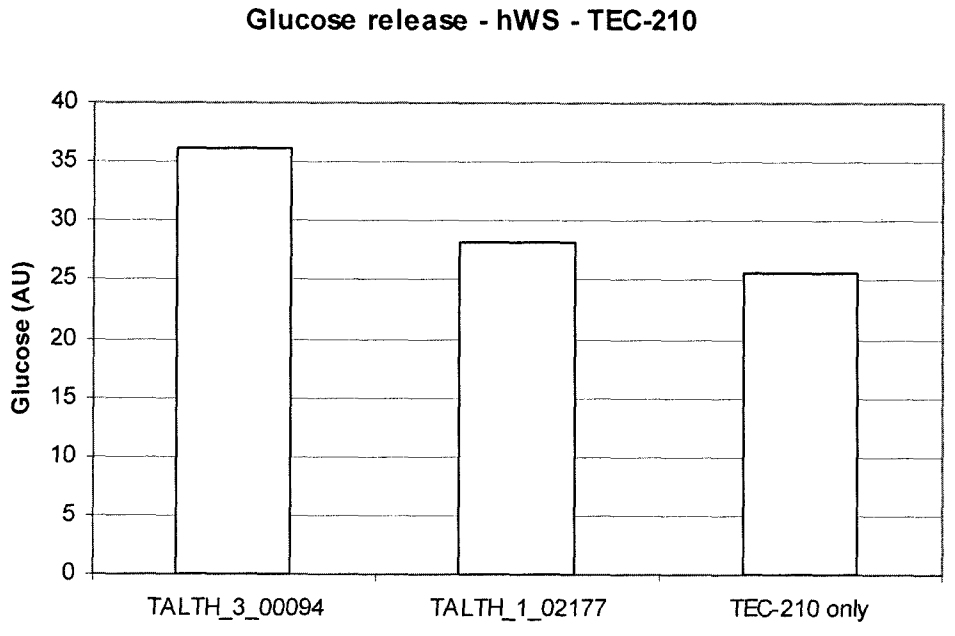

Glucose release - hWS - TEC-210
Glucose (AU)
40
35
30
25
20
15
10
5
0
TALTH_3_00094
TALTH_1_02177
TEC-210 only

Fig. 2

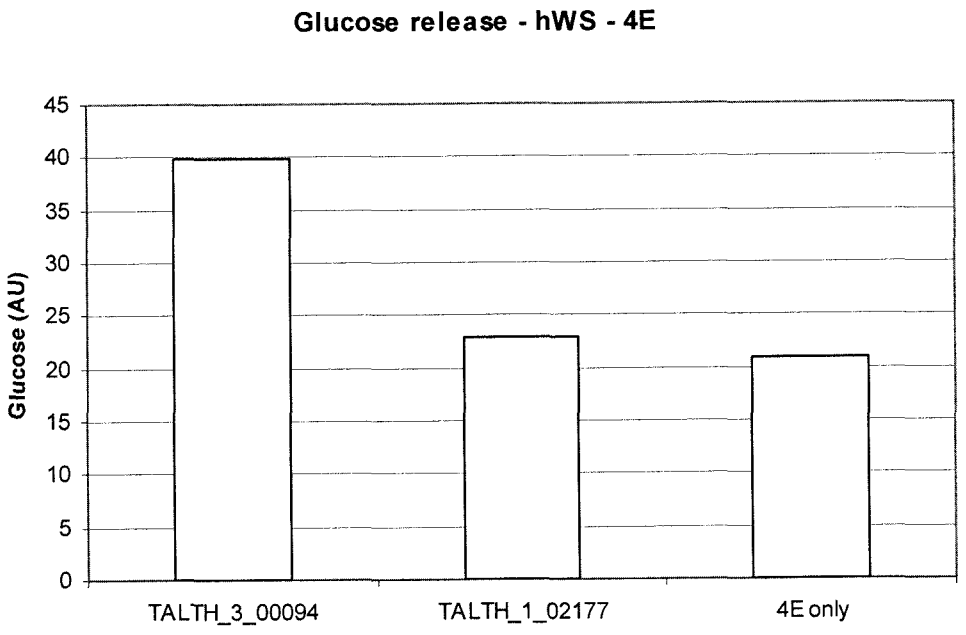

Glucose release - hWS - 4E
Glucose (AU)
45
40
35
30
25
20
15
10
5
0
TALTH_3_00094
TALTH_1_02177
4E only

Fig. 3

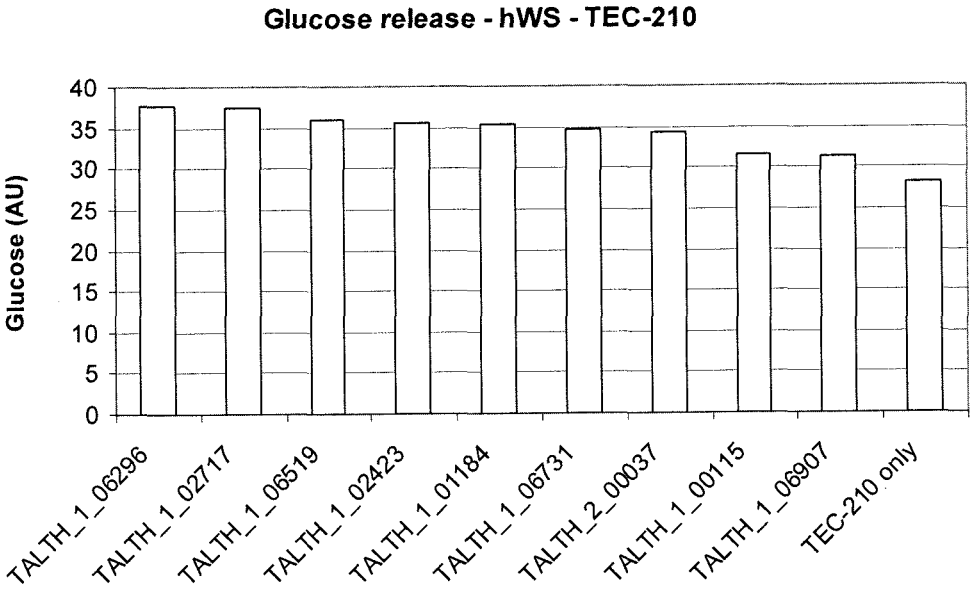

Glucose release - hWS - TEC-210
Glucose (AU)
40
35
30
25
20
15
10
5
0
TALTH_1_06296
TALTH_1_02717
TALTH_1_06519
TALTH_1_02423
TALTH_1_01184
TALTH_1_06731
TALTH_2_00037
TALTH_1_00115
TALTH_1_06907
TEC-210 only

Fig. 4

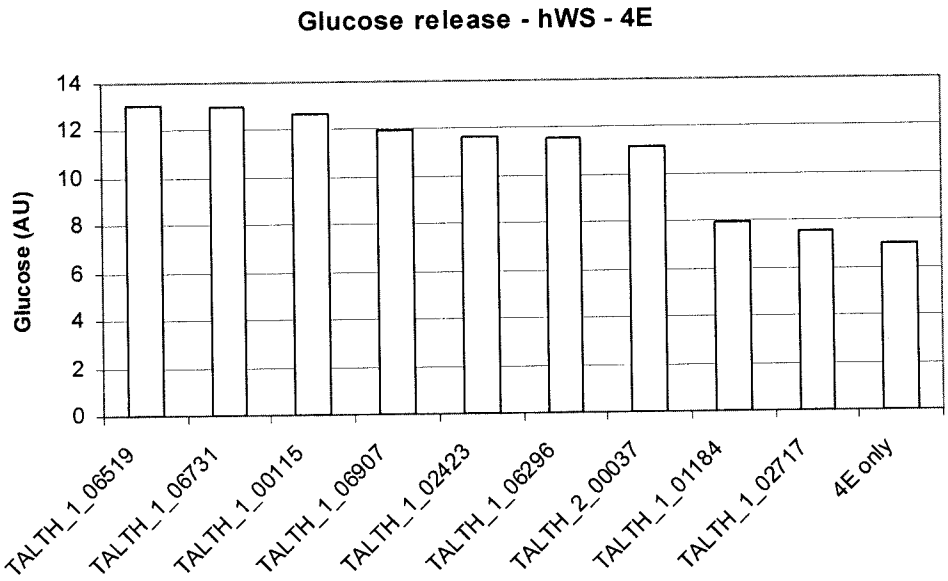

Glucose release - hWS - 4E
Glucose (AU)
14
12
10
8
6
4
2
0
TALTH_1_06519
TALTH_1_06731
TALTH_1_00115
TALTH_1_06907
TALTH_1_02423
TALTH_1_06296
TALTH_2_00037
TALTH_1_01184
TALTH_1_02717
4E only

Fig. 5

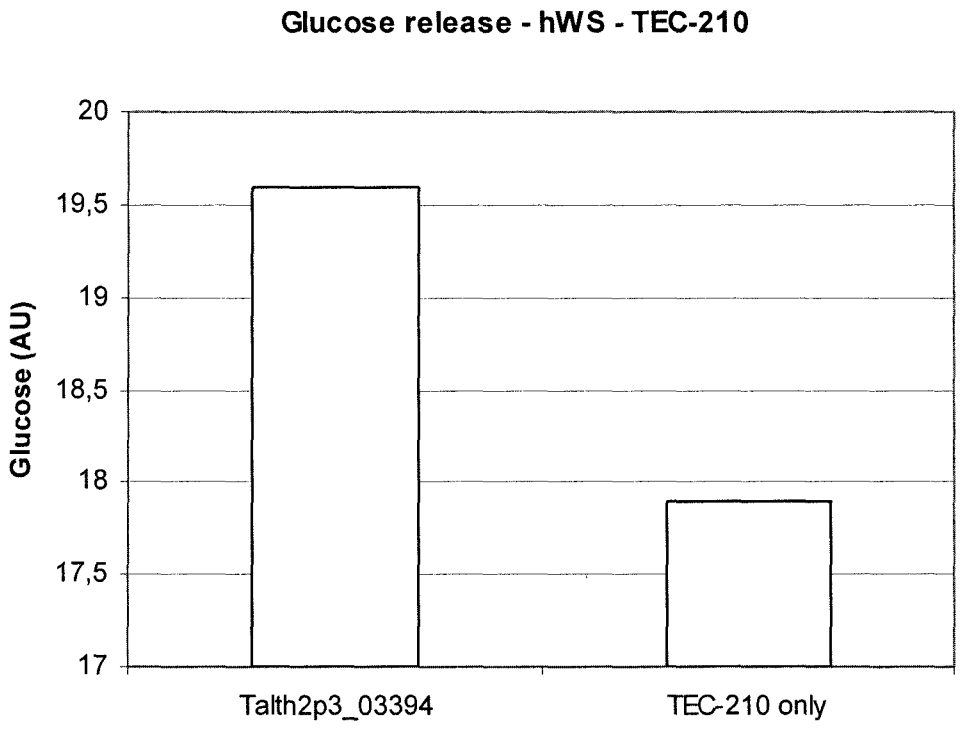

Glucose release - hWS - TEC-210
Glucose (AU)
20
19,5
19
18,5
18
17,5
17
Talth2p3_03394
TEC-210 only

Fig. 6

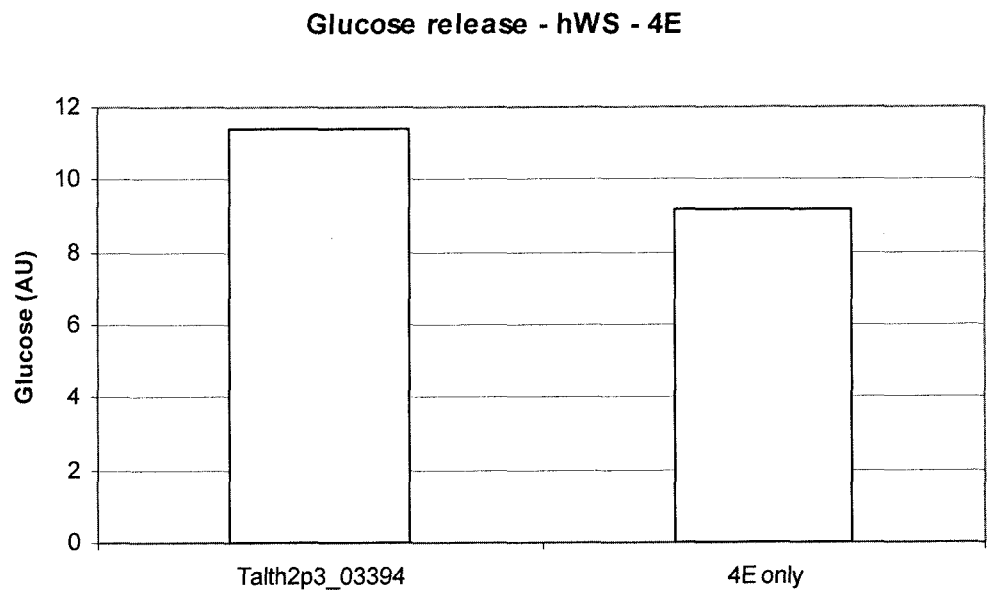

Glucose release - hWS - 4E
Glucose (AU)
12
10
8
6
4
2
0
Talth2p3_03394
4E only

Fig. 7

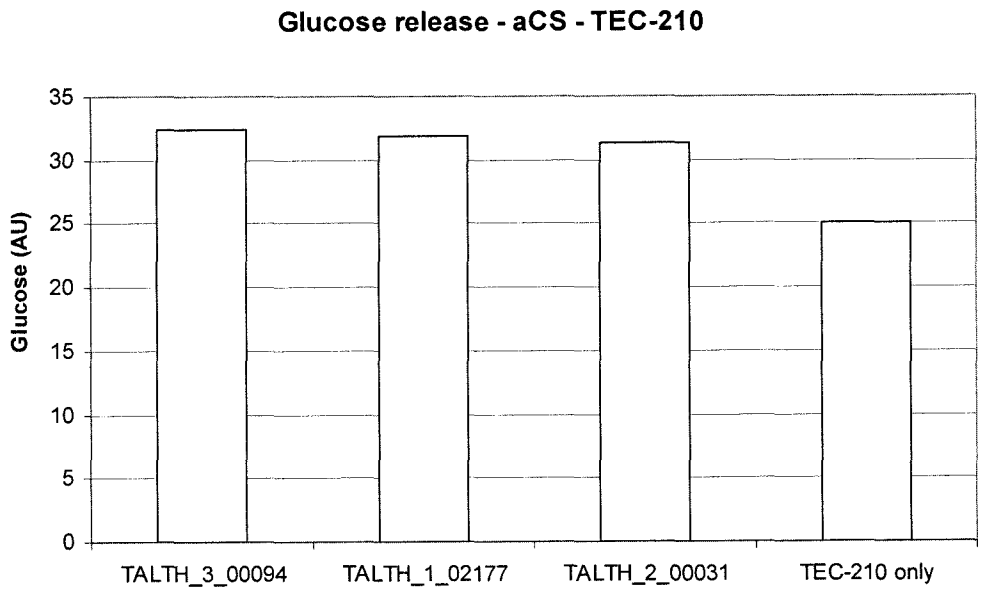

Glucose release - aCS - TEC-210
Glucose (AU)
35
30
25
20
15
10
5
0
TALTH_3_00094
TALTH_1_02177
TALTH_2_00031
TEC-210 only

Fig. 8

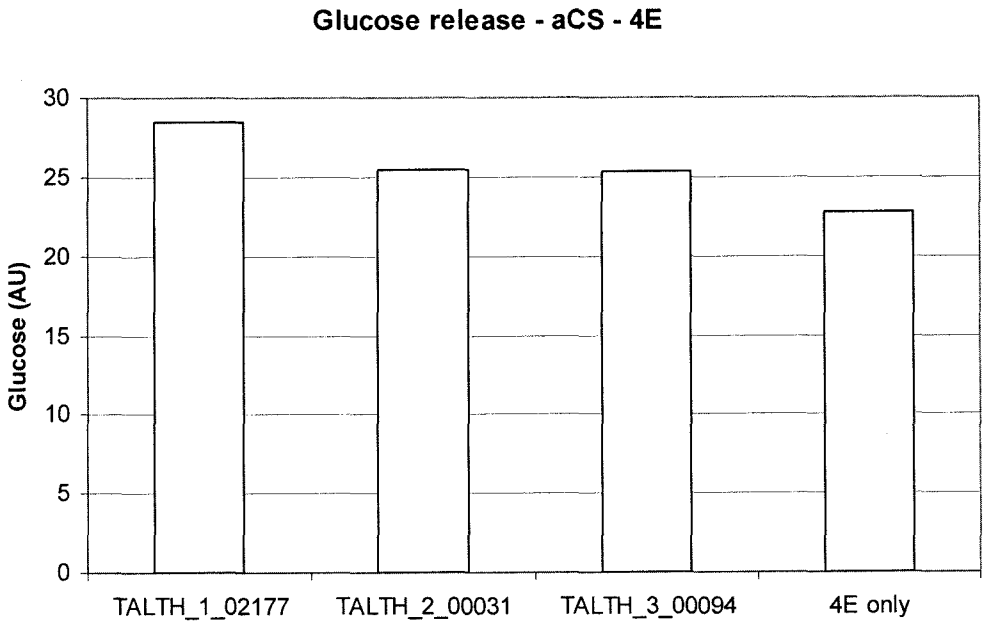

Glucose release - aCS - 4E
Glucose (AU)
30
25
20
15
10
5
0
TALTH_1_02177
TALTH_2_00031
TALTH_3_00094
4E only

Fig. 9

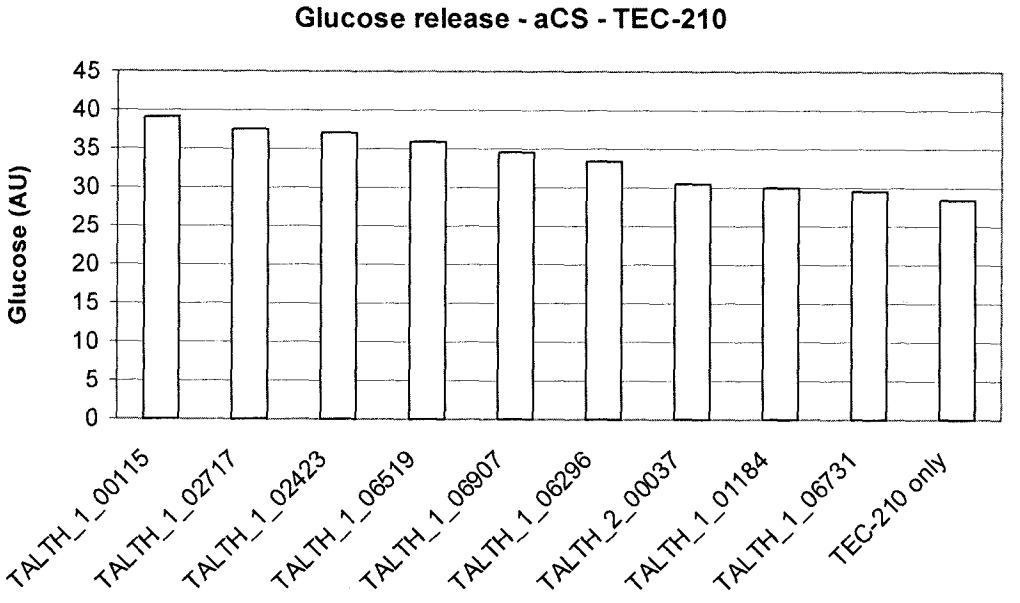

Glucose release - aCS - TEC-210
Glucose (AU)
45
40
35
30
25
20
15
10
5
0
TALTH_1_00115
TALTH_1_02717
TALTH_1_02423
TALTH_1_06519
TALTH_1_06907
TALTH_1_06296
TALTH_2_00037
TALTH_1_01184
TALTH_1_06731
TEC-210 only

Fig. 10

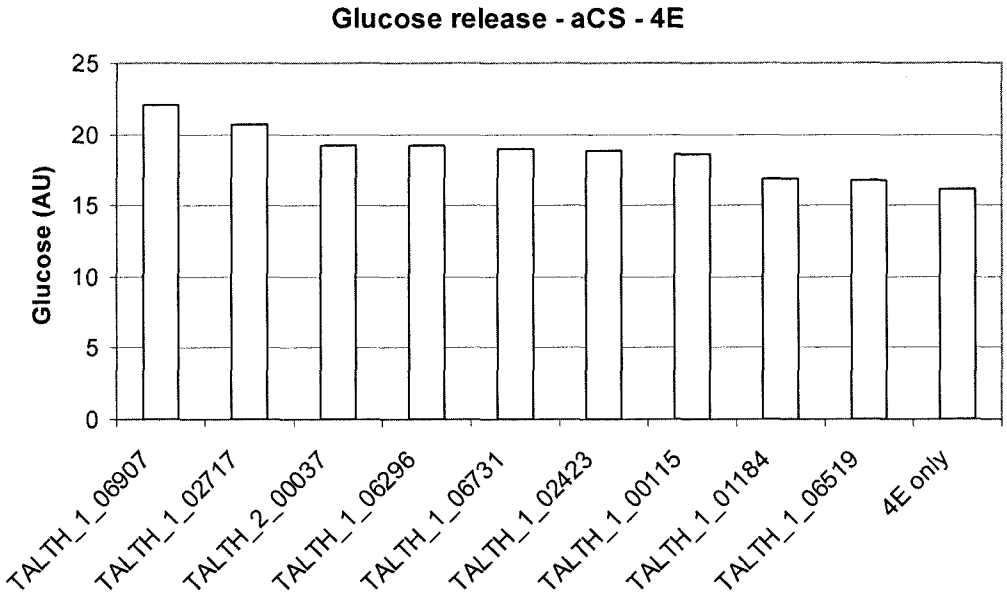

Glucose release - aCS - 4E
Glucose (AU)
25
20
15
10
5
0
TALTH_1_06907
TALTH_1_02717
TALTH_2_00037
TALTH_1_06296
TALTH_1_06731
TALTH_1_02423
TALTH_1_00115
TALTH_1_01184
TALTH_1_06519
4E only

Fig. 11

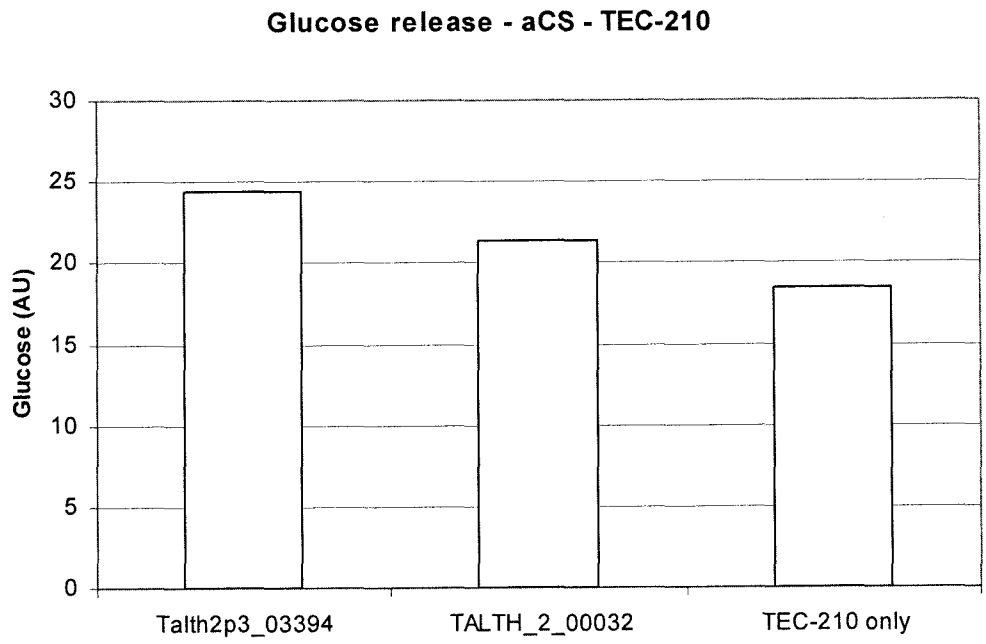

Glucose release - aCS - TEC-210
Glucose (AU)
30
25
20
15
10
5
0
Talth2p3_03394
TALTH_2_00032
TEC-210 only

Fig. 12


Glucose release - aCS - 4E
Glucose (AU)
21
20,5
20
19,5
19
18,5
18
17,5
17
16,5
16
15,5
Talth2p3_03394
4E only

Fig. 13

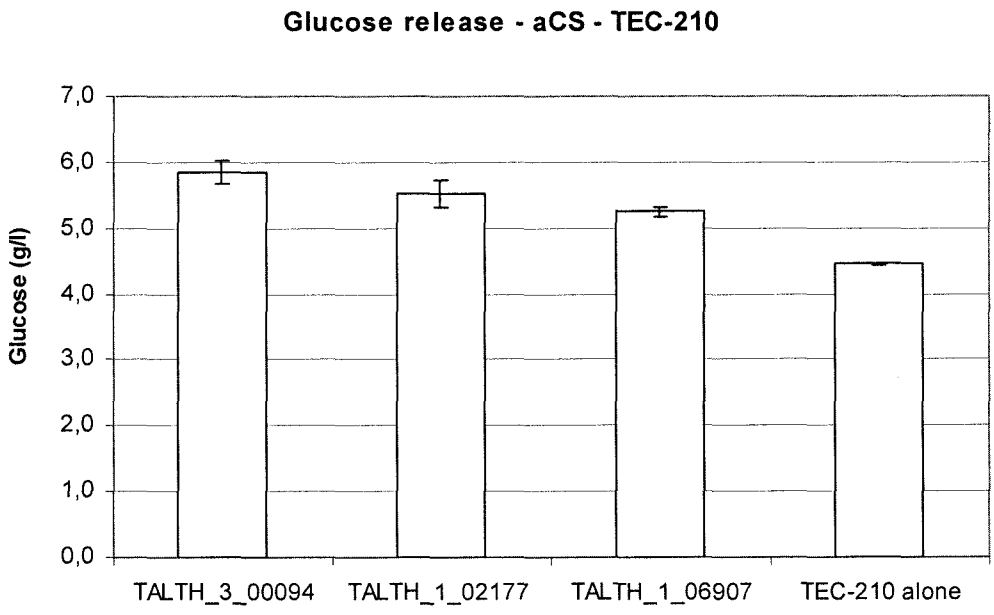

Glucose release - aCS - TEC-210
Glucose (g/l)
7,0
6,0
5,0
4,0
3,0
2,0
1,0
0,0
TALTH_3_00094
TALTH_1_02177
TALTH_1_06907
TEC-210 alone

Fig. 14

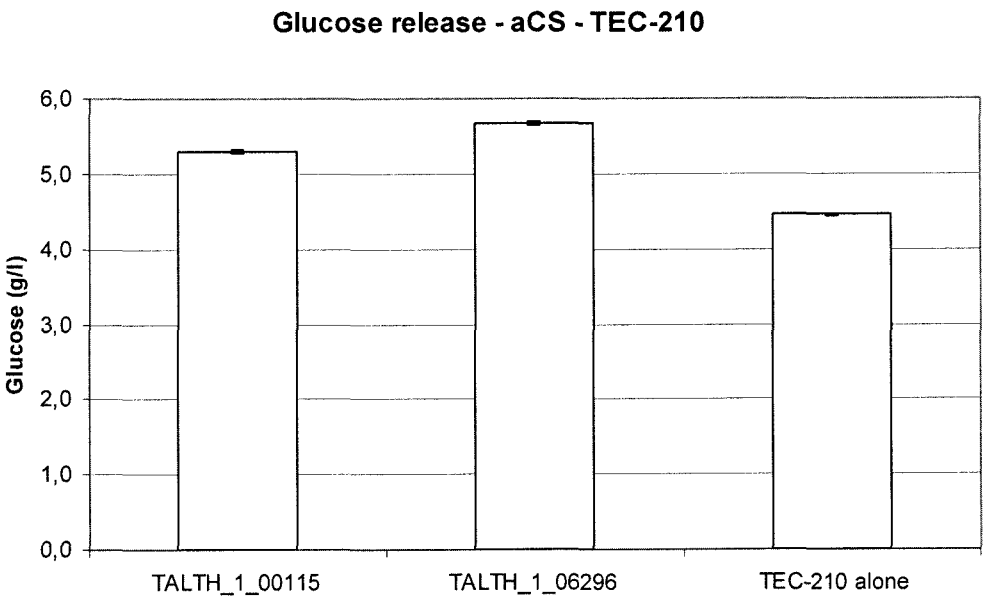

Glucose release - aCS - TEC-210
Glucose (g/l)
6,0
5,0
4,0
3,0
2,0
1,0
0,0
TALTH_1_00115
TALTH_1_06296
TEC-210 alone

Fig. 15

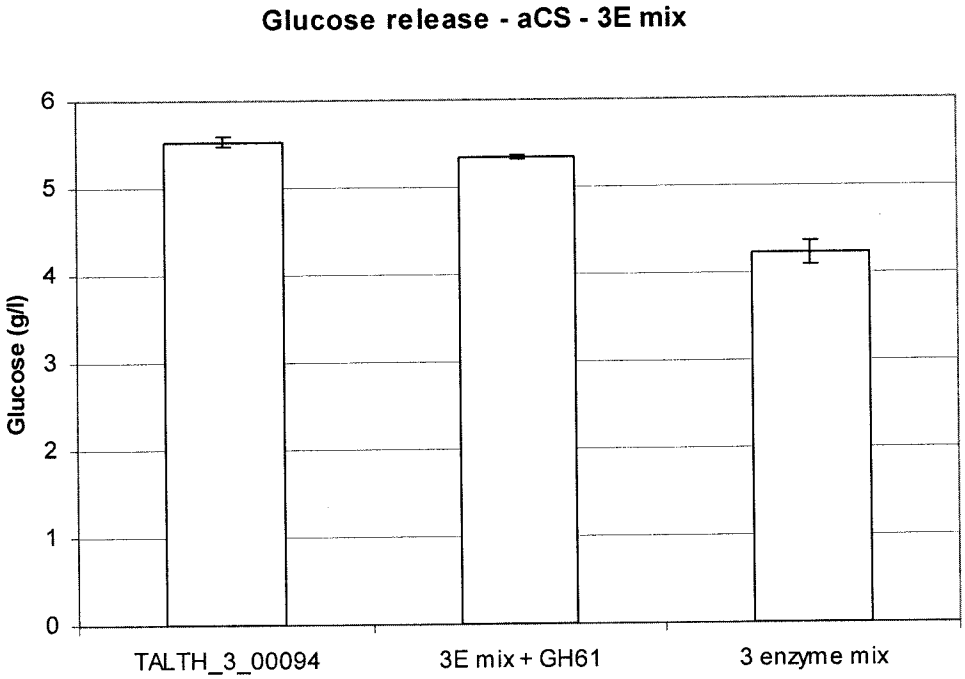

Glucose release - aCS - 3E mix
Glucose (g/l)
0
1
2
3
4
5
6
TALTH_3_00094
3E mix + GH61
3 enzyme mix

Fig 16.

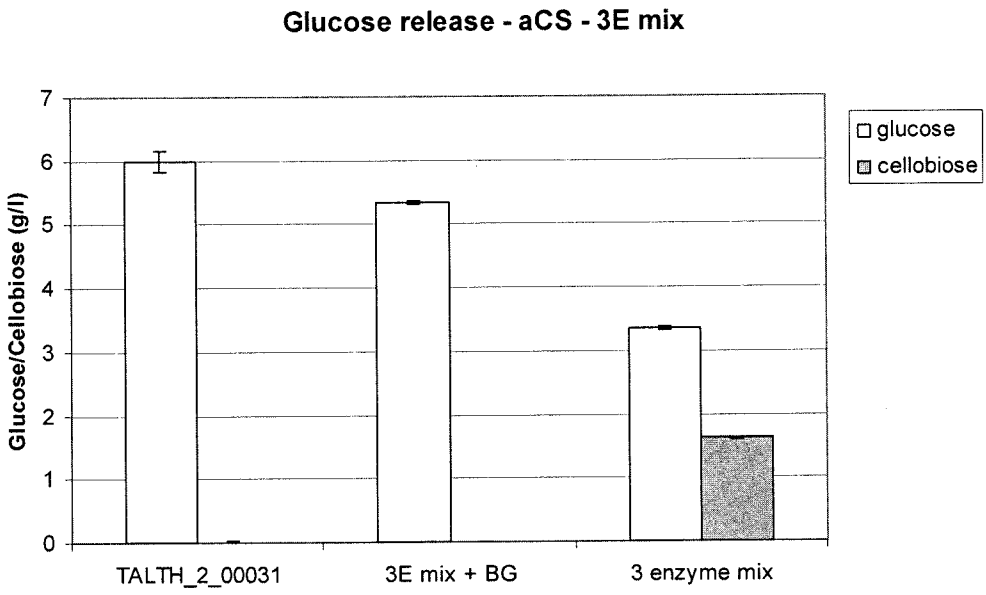

Glucose release - aCS - 3E mix
Glucose/Cellobiose (g/l)
0
1
2
3
4
5
6
7
TALTH_2_00031
3E mix + BG
3 enzyme mix
glucose
cellobiose

Fig. 17

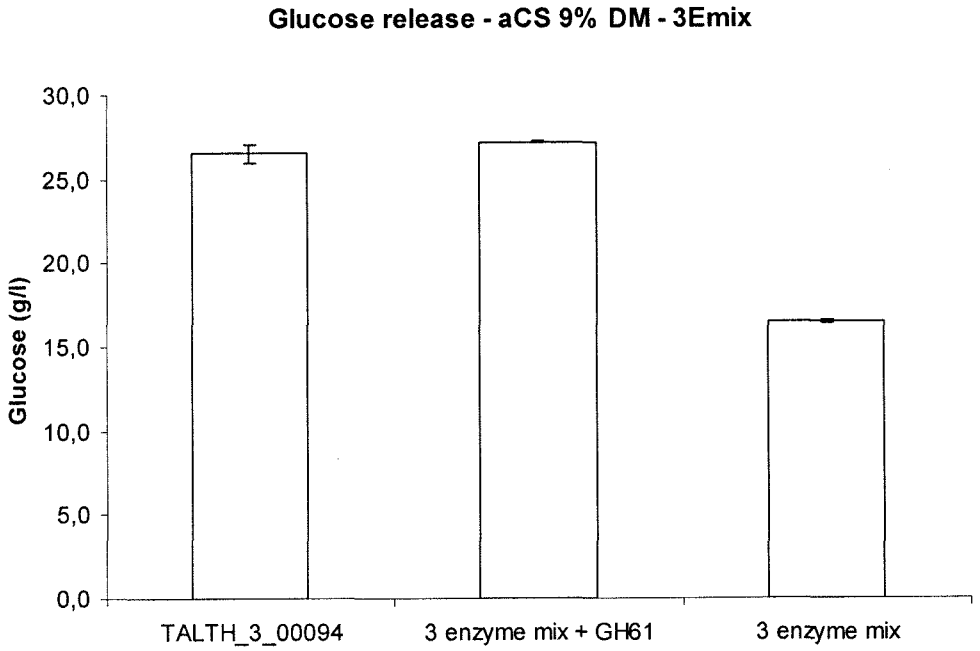

Glucose release - aCS 9% DM - 3Emix
Glucose (g/l)
30,0
25,0
20,0
15,0
10,0
5,0
0,0
TALTH_3_00094
3 enzyme mix + GH61
3 enzyme mix

Fig. 18

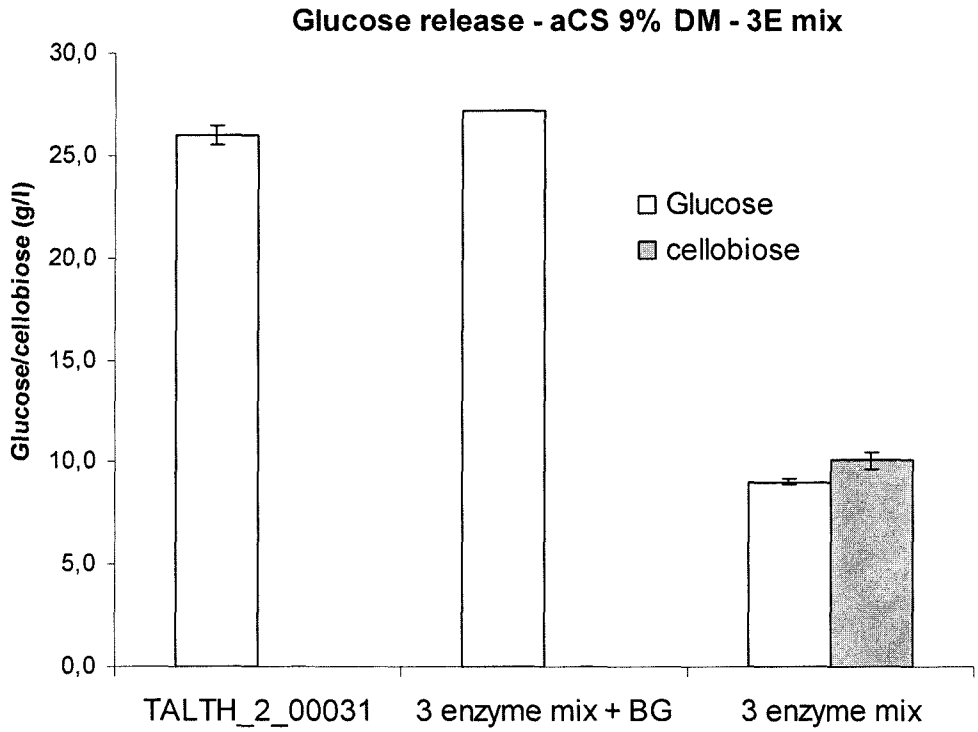

Glucose release - aCS 9% DM - 3E mix
Glucose/cellobiose (g/l)
30,0
25,0
20,0
15,0
10,0
5,0
0,0
Glucose
cellobiose
TALTH_2_00031
3 enzyme mix + BG
3 enzyme mix

Fig. 19

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2012/050200

A. CLASSIFICATION OF SUBJECT MATTER
IPC: ***C12N 15/56*** (2006.01) , ***C11D 3/386*** (2006.01) , ***C11D 7/42*** (2006.01) , ***C12N 15/53*** (2006.01) , ***C12N 15/55*** (2006.01) , ***C12N 9/02*** (2006.01) (more IPCs on the last page)
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC: ***C12N 15/56*** (2006.01) , ***C11D 3/386*** (2006.01) , ***C11D 7/42*** (2006.01) , ***C12N 15/53*** (2006.01) , ***C12N 15/55*** (2006.01) , ***C12N 9/02*** (2006.01) (more IPCs on the last page)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic database(s) consulted during the international search (name of database(s) and, where practicable, search terms used)
Canadian Patent Database; GenomeQuest; TotalPatent; Scopus; Pubmed
Keywords: SEQ ID NOs: 1-3, 19-21, 31-33, and 55-57; cellulose oxidase, *Thermomyces lanuginosus*, Humicola

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	Galagan JE et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae* Nature, 22 December 2005 (22-12-2005), Vol. 438(7071), p. 1105-1115 ISSN: 0028-0836 (Print); 1476-4687 (Electronic) see whole document See also the submitted sequences in the EMBL database under accession no. CBF76006	1-3, and 9 5-8 and 12-20

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 June 2012 (13-06-2012)	05 July 2012 (05-07-2012)

Name and mailing address of the ISA/CA	Authorized officer
Canadian Intellectual Property Office Place du Portage I, C114 - 1st Floor, Box PCT 50 Victoria Street Gatineau, Quebec K1A 0C9 Facsimile No.: 001-819-953-2476	Julie Deschênes-Furry (819) 953-4354

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	Wortman JR et al. The 2008 update of the *Aspergillus nidulans* genome annotation: a community effort Fungal Genet Biol., March 2009 (03-2009), Vol. 46(Supple 1), p. S2-S13 ISSN: 1087-1845 (Print); 1096-0937 (Electronic) see whole document See also the submitted sequences in the GenBank database under accession no. CBF76006 (amino acid sequence) and the EMBL database under accession no. CBF76006	1-3, and 9 5-8 and 12-20
X Y	US 2009/0119022 A1 (Timberlake WE et al.) 7 May 2009 (07-05-2009) see whole document; specifically SEQ ID NO:19074	1-3, and 9 5-8 and 12-20
Y	WO 2010/075402 A1 (Ashton W et al.) 1 July 2010 (01-07-2010) see whole document, specifically page 1, paragraph [03], page 28, paragraph [92] to page 33, paragraph [115]	5-8, and 12-20
A	Guo R-F et al. Purification and Characterization of a Novel Thermostable Chitinase from *Thermomyces lanuginosus* SY2 and Cloning of Its Encoding Gene Agric Science in China, December 2008 (12-2008), Vol. 7(12), p. 1458-1465 ISSN: 1671-2927 see whole document	
A	Thorsen TS et al. Identification and characterization of glucoamylase from the fungus *Thermomyces lanuginosus* Biochim Biophys Acta., April 2006 (04-2006), published online 30 January 2006 (30-01-2006), Vol. 1764(4), p. 671-676 ISSN: 0006-3002 (Print) see whole document	
A	Schlacher A et al. Cloning and characterization of the gene for the thermostable xylanase XynA from *Thermomyces lanuginosus* J Biotechnol., 20 August 1996 (20-08-1996), Vol. 49(1-3), p. 211-218 ISSN: 0168-1656 (Print); 1873-4863 (Electronic) see whole document	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA2012/050200

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of the first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons :

1. [ ] Claim Nos. :
because they relate to subject matter not required to be searched by this Authority, namely :

2. [X] Claim Nos. : 10 (partially) as it relates to SEQ ID NOs: 231, 267, 312, 489, 522, and 840
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically :

Notably, the sequences for SEQ ID NOs: 231, 267, 312, 489, 522, and 840 are not in the sequence listing as filed in electronic form and thus, cannot be adequately searched. Further, as no specific enzymatic activity was provided for said sequences in the description, a limited keyword search could not be performed.

3. [ ] Claim Nos. :
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows :

See Extra sheets

1. [ ] As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claim Nos. :

4. [X] No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim Nos. :

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

[ ] The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

[ ] No protest accompanied the payment of additional search fees.

Continuation of Box No. III:

The claims are directed to a plurality of inventive concepts as follows:

Group A - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 1, 2, 19, 20, 31, 32, 55, and 56) encode and polypeptides (SEQ ID NOs: 3, 21, 33 and 57) correspond to enzymes termed cellulose oxidase.

Group B - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 4, 5, 40, and 41) encode and polypeptides (SEQ ID NOs: 6 and 42) correspond to enzymes termed transglycosidase.

Group C - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 7 and 8) encode and the polypeptide (SEQ ID NO: 9) corresponds to an enzyme termed laccase.

Group D - Claims 1-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 10, 11, 28, and 29) encode and polypeptides (SEQ ID NOs: 12 and 30) correspond to enzymes termed undefined biomass-degrading enzyme.

Group E - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 13 and 14) encode and the polypeptide (SEQ ID NO: 15) corresponds to an enzyme termed endo-polygalacturonase.

Group F - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 16, 17, 67, 68, 73, 74,121, and 122) encode and polypeptides (SEQ ID NOs: 18, 69, 75 and 123) correspond to enzymes termed glucan 1,3-beta-glucosidase.

Group G - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 22 and 23) encode and the polypeptide (SEQ ID NO: 24) corresponds to an enzyme termed glucoamylase.

Group H - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 25 and 26) encode and the polypeptide (SEQ ID NO: 27) corresponds to an enzyme termed alpha-amylase.

Group I - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 34, 35, 58, 59, 124, and 125) encode and polypeptides (SEQ ID NOs: 36, 60, and 126) correspond to enzymes termed mannosidase.

Group J - Claims 1-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 37 and 38) encode and the polypeptide (SEQ ID NOs: 39) corresponds to an enzyme termed xylanase.

Group K - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 43, 44, 49, 50, 100, 101, 103, and 104) encode and polypeptides (SEQ ID NOs: 45, 51, 102, and 105) correspond to enzymes termed chitinase.

Group L - Claims 1-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 46, 47, 64, 65, 85, 86, 88, 89, 94, 95, 106, and 107) encode and polypeptides (SEQ ID NOs: 48, 66, 87, 90, 96 and 108) correspond to enzymes termed endo-1,3(4)-beta-glucanase.

Group M - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 52, 53, 115, 116, 127, and 128) encode and polypeptides (SEQ ID NOs: 54, 117, and 129) correspond to enzymes termed beta-glucosidase.

Group N - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 61, 62, 97, and 98) encode and polypeptides (SEQ ID NOs: 63 and 99) correspond to enzymes termed alpha-glucosidase.

Group O - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 70 and 71) encode and the polypeptide (SEQ ID NO: 72) corresponds to an enzyme termed glucopyranoside deacetylase.

Group P - Claims 1-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 76, 77, 79, and 80) encode and polypeptides (SEQ ID NOs: 78 and 81) correspond to enzymes termed beta-hexosaminidase.

Group Q - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 82 and 83) encode and the polypeptide (SEQ ID NO: 84) corresponds to an enzyme termed cellulase.

Group R - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 91 and 92) encode and the polypeptide (SEQ ID NO: 93) corresponds to an enzyme termed alpha-galactosidase.

Group S - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 109 and 110) encode and the polypeptide (SEQ ID NO: 111) corresponds to an enzyme termed alpha-N-acetylglucosaminidase.

Group T - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 112 and 113) encode and the polypeptide (SEQ ID NO: 114) corresponds to an enzyme termed arylesterase.

Group U - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 118 and 119) encode and the polypeptide (SEQ ID NO: 120) corresponds to an enzyme termed beta-galactosidase.

Group V - Claims 1-9 and 11-20 (all partially) are directed to isolated polynucleotides and isolated polypeptides, vectors comprising said polynucleotides, a host cell comprising said polynucleotides or vectors, methods of manufacturing said polypeptides, and uses thereof, wherein the polynucleotides (SEQ ID NOs: 130 and 131) encode and the polypeptide (SEQ ID NO: 132) corresponds to an enzyme termed expansin.

An *a priori* analysis has concluded that cell wall deconstruction enzymes isolated from the fungus *Thermomyces lanuginosus* ATCC strain 200065 as claimed (claims 1-20, all partially) and listed in Groups A to V do not share a common inventive link, as each Group is directed towards a different family of enzymes that has a different enzymatic activity each one from the other. Further, an *a posteriori* analysis (see Guo R et al., *Agric Sci China* (2008), 7(12): 1458-1465; Thorsen TS et al., *Biochem Biophys Acta.*, (2006), 1764(4): 671-676; and Schlacher A et al., *J Biotechnol.* (1996), 49(1-3): 211-218) has concluded that the cell wall deconstruction enzymes isolated from the fungus, *Thermomyces lanuginosus*, as claimed (claims 1-20, all partially) and listed in Groups A to V are not new and therefore there is no common inventive link uniting each family of enzymes.

Continuation of : CLASSIFICATION OF SUBJECT MATTER

C12N 9/14 (2006.01) , ***C12N 9/24*** (2006.01) , ***C12N 9/42*** (2006.01) , ***C12P 19/00*** (2006.01) , ***C12P 7/10*** (2006.01) , ***D06M 16/00*** (2006.01) , ***D21C 5/00*** (2006.01) , ***D21C 9/10*** (2006.01) , *C08H 7/00* (2011.01)

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CA2012/050200

Patent Document Cited in Search Report	Publication Date	Patent Family Member(s)	Publication Date
US2009119022A1	07 May 2009 (07-05-2009)	None	
WO2010075402A1	01 July 2010 (01-07-2010)	CA2747813A1	01 July 2010 (01-07-2010)
		CN102264892A	30 November 2011 (30-11-2011)
		EP2376629A1	19 October 2011 (19-10-2011)
		KR20110117064A	26 October 2011 (26-10-2011)
		MX2011006779A	03 August 2011 (03-08-2011)
		US2011302722A1	15 December 2011 (15-12-2011)